Likelihood-based brain activation detection from functional Magnetic Resonance Imaging data disturbed by coloured noise

by

Willem Brouwer

A thesis presented to Delft University of Technology in fulfilment of the thesis requirement for the degree of Master of Science in Applied Physics

Delft, The Netherlands, June 2006

Supervisors:
prof.dr.ir. P.M.J. Van den Hof
dr.ir. A.J. den Dekker
ir. R. Bos
Preface

‘And now, the end is near.’ This is also the case of my Master of Science research project. At the end of this period of research, I am glad to present this thesis. In this thesis, I give an overview of all the work I have done in the last year. In this period, I have learned a lot of functional Magnetic Resonance Imaging (fMRI), especially with respect to the possibilities and difficulties of data modelling.

I want to present this thesis to the Delft University of Technology and its research group Delft Center for Systems and Control (DCSC), from which I got the possibility to do my MSc. project and to graduate in Applied Physics. I will give thanks to my supervisors prof.dr.ir. P.M.J. Van den Hof, dr.ir. A.J. den Dekker and ir. R. Bos from DCSC for their support, comments and very useful feedback, each with their own agenda and knowledge. Furthermore, I appreciate the remarks of prof.dr. J. Sijbers and my friend ir. D.H.J. Poot of the University of Antwerp on the subject of fMRI techniques and signal estimation.

Besides the support of those people with respect to the contents of this thesis, I have got very much support from my wife Mieneke to fulfil the research and from my (grand)parents, former fellow occupants and friends. They were all very interested in the process and the results.

I have enjoyed this study. In this thesis, I have tried to explain the basics of fMRI in an accessible way, so that also persons with very few experience in data modelling (for example parents or friends) can understand the subject of my study. I apologise if they cannot follow me to the last chapters, but I guess that is inevitable if one graduate in Applied Physics. Nevertheless, I hope you all enjoy in reading this thesis and that you learn a lot. The aim of science is to get knowledge and to share it, so you may use this thesis for all personal and scientific purposes.

Willem Brouwer
Delft / Den Haag, June 2006
Content

Preface 1
Symbols and abbreviations 4
Summary 8
Samenvatting (Summary in Dutch) 10

1. Introduction 12

2. FMRI: physical background and common mathematical approach 16
   2.1 Use of nuclear magnetic resonance for imaging 16
      2.1.1 Magnetic properties of hydrogen 16
      2.1.2 Imaging 21
   2.2 Neurology and functional MRI 26
      2.2.1 Physiological aspects of brain activation 26
      2.2.2 Use of the haemodynamic response function for fMRI 27
   2.3. The General Linear Model for fMRI data 29
      2.3.1 Model structure 29
      2.3.2 The haemodynamic response function 30
      2.3.3 Trends 31
      2.3.4 Noise 31
      2.3.5 Parameter estimation 33
      2.3.6 Hypothesis tests for activation detection 36
   Concluding remarks Chapter 2 38

3. New mathematical approach: Generalized Likelihood Ratio Test 40
   3.1. Likelihood-based parameter estimation 40
      3.1.1 Maximum likelihood estimation in general 40
      3.1.2 Parameter estimation with likelihood function 42
   3.2. Generalized Likelihood Ratio Test for activation detection 44
      3.2.1 Hypothesis test and test statistic 44
      3.2.2 The false alarm rate and the detection rate 46
   3.3. Noise model order selection criteria 46
      3.3.1 Akaike's Information Criterion 47
      3.3.2 Corrected Kullback Information Criterion 48
      3.3.3 Minimum Description Length 49
   Concluding remarks Chapter 3 53

4. Likelihood-based detection in simulation 54
   4.1 Creation of simulated data 54
      4.1.1 Haemodynamic response function 54
      4.1.2 Trend model 57
      4.1.3 Autoregressive noise 58
4.2. Model estimation and activation detection  
   4.2.1 Maximum likelihood estimation 60  
   4.2.2 Order selection 61  
   4.2.3 Test statistics, hypothesis test and detection 61  
   4.2.4 Multiple runs 62  
Concluding remarks Chapter 4 63

5. Results, discussion and conclusions 64
   5.1 Performances of Generalized Likelihood Ratio Test 64  
      5.1.1 False alarm rate 64  
      5.1.2 Uncorrected detection rate 68  
      5.1.3 Detection rate with false alarm rate correction 69  
      5.1.4 If everything is known 69  
      5.1.5 Conclusions 72  
   5.2 Performances of Generalized Likelihood Ratio Test with order selection 73  
      5.2.1 Loss in comparison with right order 73  
      5.2.2 Profit of order selection with respect to assuming false order 74  
      5.2.3 Conclusions 76  
   5.3 A comparison of the GLRT with the common \( t \)-test 76  
      5.3.1 False alarm rate 76  
      5.3.2 Detection rate 77  
      5.3.3 Conclusions 78

6. Conclusions and recommendations 80

Appendices 84

A. Spectra of activation signal and noise models used 85  
B. Graphical illustrations of the performance analysis of the GLRT 88  
   a. False alarm rate 88  
   b. GLRT detection rate 91  
   c. Corrected GLRT detection rates 99  
   d. Theoretical GLRT detecting rate 107  
   e. Loss of detection from order selection with respect to right order estimation 110  
   f. Profit of detection from order selection with respect to false order estimation 118  
   g. Comparison between the GLRT and the \( t \)-test 126  
C. CD-ROM with programs, data files and digital version of thesis 130

Bibliography 132
Symbols and abbreviations

In this thesis, a ^-mark (circumflex) above a symbol means that the quantity is an estimation of the concerning quantity or parameter. The subscripts x, y and z denote orthogonal directions. Boldface characters denote vectors or matrices, all other characters are scalars.

**Symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>strength of local magnetic field</td>
</tr>
<tr>
<td>B₀</td>
<td>magnetic field, constant magnetic field</td>
</tr>
<tr>
<td>$\mathbb{E}[\xi]$</td>
<td>expected value of variable $\xi$</td>
</tr>
<tr>
<td>G</td>
<td>gradient of magnetic field</td>
</tr>
<tr>
<td>H₀</td>
<td>null hypothesis (no activation)</td>
</tr>
<tr>
<td>H₁</td>
<td>alternative hypothesis (activation)</td>
</tr>
<tr>
<td>I</td>
<td>identity matrix</td>
</tr>
<tr>
<td>K</td>
<td>temporal smoothing matrix</td>
</tr>
<tr>
<td>L</td>
<td>length of MRI scanner</td>
</tr>
<tr>
<td>L</td>
<td>natural logarithm of the likelihood function, $\log_e(l)$</td>
</tr>
<tr>
<td>M</td>
<td>magnetization</td>
</tr>
<tr>
<td>N⁻</td>
<td>number of spins in same direction of magnetic field</td>
</tr>
<tr>
<td>N⁺</td>
<td>number of spins in opposite direction of magnetic field</td>
</tr>
<tr>
<td>$\mathcal{N}(\mu,\sigma^2)$</td>
<td>normal distribution function with mean $\mu$ and variance $\sigma^2$</td>
</tr>
<tr>
<td>P_f</td>
<td>probability of false alarm, false alarm rate</td>
</tr>
<tr>
<td>P_d</td>
<td>probability of detection, detection rate</td>
</tr>
<tr>
<td>Pr(x</td>
<td>y)</td>
</tr>
<tr>
<td>Rₘₐₜ</td>
<td>residual-forming matrix with smoothed data</td>
</tr>
<tr>
<td>T</td>
<td>transpose of a matrix (when denoted right-above a matrix)</td>
</tr>
<tr>
<td>T</td>
<td>test statistic, with subscripts: F: F-test; sm: smoothed data</td>
</tr>
<tr>
<td>T₁</td>
<td>time constant related to the 63% increase of $M_z$</td>
</tr>
<tr>
<td>T₂</td>
<td>time constant related to the 34% decrease of $M_{xy}$</td>
</tr>
<tr>
<td>T₂</td>
<td>time constant related to the decrease of $M_{xy}$, related to extra dephasing due to inhomogeneous magnetic field</td>
</tr>
<tr>
<td>T F</td>
<td>test statistic of F-test</td>
</tr>
<tr>
<td>T sm</td>
<td>test statistic for smoothed data</td>
</tr>
<tr>
<td>T₁</td>
<td>test statistic of Student-t test</td>
</tr>
<tr>
<td>V</td>
<td>covariance matrix of coloured noise</td>
</tr>
<tr>
<td>W</td>
<td>prewhitening weighting matrix for error variance</td>
</tr>
<tr>
<td>X</td>
<td>matrix of deterministic contribution, named ‘design matrix’</td>
</tr>
<tr>
<td>a_i</td>
<td>autoregressive noise model parameter</td>
</tr>
<tr>
<td>e, e_k</td>
<td>white noise error in y (vector), $y_k$ (value at time $k$)</td>
</tr>
<tr>
<td>h(t)</td>
<td>haemodynamic response function</td>
</tr>
<tr>
<td>k</td>
<td>discrete time index</td>
</tr>
<tr>
<td>m</td>
<td>magnetic strength vector of one hydrogen spin</td>
</tr>
<tr>
<td>m</td>
<td>number of deterministic contributions to measurement $y_i$</td>
</tr>
<tr>
<td>l</td>
<td>likelihood function</td>
</tr>
<tr>
<td>n</td>
<td>number of measurements $y_k$</td>
</tr>
<tr>
<td>p</td>
<td>order of AR model of coloured noise</td>
</tr>
<tr>
<td>p₁, p₂</td>
<td>power coefficient used in standard HRF</td>
</tr>
<tr>
<td>q₁, q₂</td>
<td>denominator coefficients used in standard HRF</td>
</tr>
<tr>
<td>r</td>
<td>weight coefficient for gamma functions used in HRF</td>
</tr>
</tbody>
</table>
Symbols and abbreviations

$d_1, d_2$  
time to peak of gamma function in standard HRF

$t$  
analogous time in seconds

$\text{tr}(\xi)$  
trace of matrix $\xi$

$u(t)$  
external stimulus function

$\nu, \nu_k$  
coloured noise error in $y$ (vector), $y_k$ (value)

$x_{\text{hrf}}(t)$  
continuous function for the contribution of the HRF to $y$

$x_{k,\text{hrf}}$  
sampled function for the contribution of the HRF to $y_k$

$x_{k,\text{tr}}$  
sampled function for the contribution of trends to $y_k$

$y$  
time series of measured fMRI data per voxel

$y_k$  
single measurement of $y$ at time $k$

$z$  
position along the centre line of the MRI scanner

$\alpha$  
one-sided user specific threshold of test statistic

$\alpha_a$  
two-sided user specific lower threshold of test statistic

$\alpha_b$  
two-sided user specific upper threshold of test statistic

$\gamma$  
gyromagnetic ratio

$\zeta$  
non-centrality parameter of $\chi^2$ distribution

$\theta$  
parameter related to deterministic contributions to $y$

$\nu_R$  
Larmor precession frequency of a spin

$\lambda$  
ratio of likelihood function under $H_1$ and $H_0$

$\nu_{R0}$  
constant Larmor precession frequency of a spin

$\rho(k)$  
correlation over interval $k$

$\sigma^2_w$  
variance of white noise

$\sigma^2_v$  
variance of coloured noise

$\Delta E$  
energy difference between two magnetic states

*  
Self programmed Matlab function (used after function name)

**Abbreviations**

When an abbreviation is used as subscript, it denotes that the concerning parameter or quantity belongs to that specific method, function or expression.

AIC(c)  
Akaike Information Criterion (corrected)

AR(p)  
autoregressive model of order $p$

BIC  
Bayes Information Criterion

BLUE  
best linear unbiased estimator

BOLD  
Blood Oxygenation Level Dependent

CBF  
cerebral blood flow

CBV  
cerebral blood volume

CT  
Computerized Tomography

EEG  
Electroencephalography

e.m.f.  
electro magnetic field

Eq.  
equation

FAR  
false alarm rate

FIC  
Finite sample Information Criterion

FID  
Free Induction Decay

FM  
Frequency Modulated

fMRI  
functional Magnetic Resonance Imaging

GLM  
General Linear Model

GLRT  
generalized likelihood ratio test

GLS  
generalized least squares

HRF, hrf  
haemodynamic response function

IDD  
independent and identically distributed
(A)KIC(c) (Approximated) Kullback's Information Criterion (corrected)
KL Kullback-Leibler
LR likelihood ratio
LS least squares
(A)MDL (Approximated) Minimum Description Length
MEG Magnetoencephalography
ML Maximum likelihood
MLE maximum likelihood estimator
MRI Magnetic Resonance Imaging
NMR Nuclear Magnetic Resonance
OLS ordinary least squares
OSC order selection criterion
OSM order selection method
PD proton density
PDF probability density function
PET Positron Emission Tomography
RC reflection coefficient
RF Radio Frequency
sm smoothed
SNR signal-to-noise ratio
SPM Statistical Parameter Mapping
tr trend (as subscript)
Functional MRI (fMRI) is a technique used to investigate the function of different parts of the brain. MRI uses the magnetic properties of the hydrogen found in tissues to create images of those tissues. Because different tissues give different contrast in the MRI signal, an image of the structure of tissues can be made using this technique. To investigate the function of the brain, in addition to its structure, fMRI detects and localizes brain activation using the changing properties of blood in the brain during task executing. In this study, we are mainly interested in modelling the fMRI signal.

The observed fMRI signal can be modelled as a time series, that is to say, a succeeding series of values. The signal consists of three contributions: (1) the haemodynamic response to activation, (2) trends coming from small changes in the scanners and (3) noise. In the common approach, the first and second contributions are estimated separately from the third. In this thesis, another approach is elaborated, based on a maximum likelihood estimation of all the contributions at once.

Because beforehand it is not clear how to choose the structure that is suitable for modelling the fMRI signal, some assumptions have to be made concerning the model structures used. In this thesis, we assume that the structures of the haemodynamic response and the trend are known. The noise is modelled as an autoregressive process. In the estimation of the parameters of the noise model, the model order is either fixed or determined using order selection criteria. Five order selection criteria are considered. Three order selection methods are based on Akaike’s information criterion, one on the Minimum Description Length and one on Kullback’s symmetric divergence.

Every contribution in the fMRI signal has its parameter, which denotes the extent of that contribution to the total signal. Because noise has great influence on the observations, the estimated parameters have to be considered as stochastic variables. On the basis of hypothesis tests a decision is made as to whether the value of the estimated activation-related parameter gives one grounds to decide whether there is activation or not. Depending on the method used, a specific test statistic is available. The probability density function (or ‘distribution’) of the test statistic is known, under the assumption that no activation is present in the signal. If the test statistic exceeds a chosen specific threshold, we ‘detect’ activation.

When investigating the likelihood-based approach, we examined the performance of a likelihood-based hypothesis test for brain activation detection using fMRI data. The performance can be expressed in terms of detection rate (right decision that activation is present) and false alarm rate (false decision that activation is present). We will compare the performance of the likelihood-based method to the conventional $t$-test-based method.
In short, our conclusions are as follow. For both methods, we conclude that the false alarm rate for practical fMRI time series length is higher than intended, while measuring for a longer period of time will give better results, as is shown for the likelihood-based method. When the false alarm rate is higher than intended, too often a decision is made that activation is present while no activation was present at all. It also turns out that the performances of tests can differ drastically when one compares different types of noise. When the spectrum of the noise signal resembles the spectrum of the activation signal, activation is difficult to detect. Unfortunately, this is often the case for fMRI signals. Thus more research into fMRI data modelling has to be done, because a combination of too high false alarm rate and a low detection rate gives rise to be very careful with respect to the conclusions from fMRI images. Furthermore, when comparing different order selection criteria, order selection methods perform almost as well as models with right order, while choosing a false order can give bad results. It also turns out that no specific order selection criterion is the best overall. Considering these results, any method of order selection should be used in fMRI modelling.
Samenvatting

Functionele MRI (fMRI) is een techniek om de functionaliteit van delen in de hersenen te onderzoeken. MRI is een techniek die gebruik maakt van de magnetische eigenschappen van waterstofatomen in weefsels voor het maken van afbeeldingen van die weefsels. Omdat verschillende weefsels verschillende niveaus in signaal geven, kan met MRI de structuur van de weefsels weergegeven worden. Om naast de structuur ook de functionaliteit van de hersenen aan te geven, gebruikt fMRI de veranderende eigenschappen van het bloed in de hersenen tijdens het uitvoeren van een bepaalde taak. In deze studie kijken we vooral naar de modelering van het fMRI signaal.


Omdat het niet op voorhand duidelijk is welke model-structuur er gekozen moet worden voor de modelering van het fMRI signaal, moeten enkele aannames gedaan worden met betrekking tot de structuur van het te gebruiken model. In deze studie wordt er van uitgegaan dat de structuur van de respons door de hersenen en van de trend bekend is. De autoregressieve ruis wordt geschat met verschillende vaststaande ordes, en met orde selectie methoden, te weten, Akaike’s informatie criterion, Minimum Description Length en Kullback’s symmetrische divergente.

Aan elke bijdrage in het fMRI wordt een parameter gekoppeld, die aangeeft in hoeverre het betreffende fenomeen bijdraagt in het totale signaal. Omdat de ruis een grote invloed heeft op de metingen, zijn de schattingen van deze parameters te beschouwen als stochastische variabelen. Afhankelijk van de gevolgde methode is er een bepaalde test statistiek beschikbaar die aan de activiteit-parameter gerelateerd is. Deze test-statistiek kan getoetst worden. Van deze test statistiek is de statistische verdeling namelijk bekend, onder de aanname dat er geen activiteit is. Op basis van een hypothese-toets wordt bepaald of de waarde van de schatting van de aan activiteit gerelateerde parameter aanleiding geeft om te besluiten dat er een activiteit-bijdrage aanwezig is in het signaal. Wanneer de test statistiek een bepaalde grenswaarde overschrijdt, ‘detecteren’ we activiteit in het signaal.

Bij het onderzoek naar de methode die gebaseerd is op de meest waarschijnlijke schatting van alle parameters bekijken we de resultaten van de hypothesetoets. De resultaten kunnen worden uitgedrukt in termen van detectie-percentage (een juiste beslissing tot activiteit) en vals-alarm-percentage (een foutieve beslissing tot activiteit). We zullen deze resultaten ook
Samenvatting

vergelijken met die van de conventionele t-test gebaseerde methode.

De volgende conclusies worden getrokken. Voor beide methoden moeten we concluderen dat het vals-alarm-percentage voor praktische fMRI metingen hoger is dan bedoeld. Langer meten zou betere resultaten geven, zoals is aangetoond voor de methode die gebaseerd is op de meest waarschijnlijke schatting. Wanneer het vals-alarm-percentage hoger is dan bedoeld, wordt er te vaak een beslissing gemaakt dat er activiteit is, terwijl het betreffende gebied in de hersenen niet actief is. Het blijkt ook dat de prestaties van testen drastisch kunnen verschillen voor verschillende ruissignalen. Wanneer het spectrum van het ruissignaal overeenkomt met het spectrum van het activiteit-singaal is de activiteit moeilijk te detecteren. Dit blijkt vaak het geval te zijn met fMRI signalen. Daarom moet er meer onderzoek gedaan worden naar de modelering van fMRI metingen, omdat een combinatie van hoge vals-alarm-percentages en lage detectie-percentages reden geven om erg voorzichtig te zijn om conclusies uit fMRI data te trekken. Op het gebied van model-orde kunnen we zeggen dat het niet duidelijk is welk model-orde-selectie-criterium het best scoort. Op voorhand is het niet mogelijk te zeggen of orde selectie beter zal presteren dan het kiezen van een vaste modelorde. Echter, orde selectie methoden geven bijna dezelfde resultaten als wanneer de juiste modelorde gekozen zou zijn, terwijl het kiezen van een verkeerde modelorde slechte resultaten kan opleveren. Uit deze resultaten wordt afgeleid dat er model-orde-selectie-criteria gebruikt moeten worden in fMRI modelering.
Chapter 1

Introduction

In this thesis, we present a new method which described ways to deal with fMRI signals. Functional Magnetic Resonance Imaging (fMRI) is a technique used in medical world to get knowledge about the functioning of the (human) brain that is based on common MRI techniques. Magnetic Resonance Imaging is a relatively new technique for brain imaging. It has several advantages over other scan techniques such as Computerized Tomography (CT), Positron Emission Tomography (PET), Electro-encephalography (EEG) and Magnetoencephalography (MEG). Besides some of the advantages with respect to image resolution and scan time, the main advantage of MRI is that it is a non-invasive technique, so no radiation or contrast fluids are needed. The MRI signal is purely based on magnetic fields, and there is no proof of dangerous short-term or long-term effects. MRI scans can therefore be repeated several times, without causing damage to the brain or the body. Using this property, it is possible to safely track (magnetic) changes in the human brain over a period of time. If the subject executes a certain task, from these magnetic changes, images of brain functionality can be created, presuming that specific brain activity results in specific magnetic signal changes.

Research into the structure and function of the brain using MRI started in the middle of the previous century. In 1946, Felix Bloch and Edward Purcell independently found the nuclear magnetic resonance (NMR) phenomenon of matter. In 1971, Raymond Damadian used NMR for human body research, showing that the relaxing times of hydrogen atoms in tissues and tumours differ, which meant that NMR could be used as a method to examine the brain for disorders. Because of the negative associations with the word ‘nuclear’, that word is dropped in the medical world. In 1990, Belliveau used the coupling between neural activity and blood flow for fast MRI scans of the brain, starting the usage of MRI techniques for brain activation detection. After him, Ogawa found that the oxygenation level of blood can be used to obtain more contrast in the fMRI images. Currently, fMRI is the method most widely used for brain mapping. Much research is still being done into the interpretation of the measured signals, because for several reasons, these are very noisy.

In fMRI research, the main question is whether the measured signal contains an activation signal that responds to a stimulus. Because of the noisy signal, research into activation in the

2. Clare [1997], ch.2; Hornak [1996], ch.1; Logothesis [2002], p.1005
3. Damadian [1977]; Clare [1997], ch.2; Hornak [1996], ch.1
signal has to be done in a statistical way. A two-step approach is generally followed. First, the parameters of the deterministic part and stochastic part in the signal are identified separately, after which a total model of the fMRI signal is constructed with these estimated parameters. This can be done iteratively, using the found structure of one part to estimate the structure of the other part. With the constructed model, the null hypothesis ($H_0$) that no activation signal is present in the data is tested against the alternative hypothesis ($H_1$) that activation is present. This is done by testing whether or not the estimate of the activation-related parameter in the model differs significantly from zero. Usually, this is done using a $t$-test.\(^7\)

In this thesis, a relatively new approach is worked out, which is based on a likelihood-based hypothesis test. In this approach, all the parameters of the fMRI model are estimated simultaneously. It is investigated if estimating all the parameters at once will provide better detection results than the above-described two-step or iterative approaches. The aim of this study is to answer the following question:

**What is the performance of a likelihood-based hypothesis test for brain activation detection from fMRI data in terms of detection rate and false alarm rate; and how does this performance hold in comparison with that of the conventional $t$-test based method for activation detection?**

In the evaluation of the answer to this question, we have taken purely simulated fMRI data, because this data tells us everything, such as the level and shape of activation and trends in the signal and the statistical properties of the noise. With a full knowledge of the properties of the signal, we are able to check the results from the estimations.

Because the underlying theories that justify both the use of the likelihood and $t$-test-based activation detection are based on asymptotic assumptions, in addition to our above defined problem, we question

**how both methods perform for practical finite fMRI sample lengths instead of using asymptotic theory.**

Since the real noise data generating system for fMRI signals is unknown, we have to take into account different possible noise structures and parameter values for simulation. Note that the noise generating process is not the same throughout the whole brain.\(^8\) From these different noise generating models, we want to obtain the behaviour of the likelihood-based method and compare the results with the results from the $t$-test based approach. In view of the fact that the true noise generation system is not known, we will also investigate

**the effects of noise model order selection methods on the performance of the likelihood-based test.**

Our hypothesis about order selection is that order selection will give lower performance than selecting the true order, but will give better results than choosing a false order.

Not only the general properties of the noise are unknown, but also the level and (to a certain extent) the shape of the activation signal. In this thesis, we take into account different levels of activation, but to restrict our scope, we assume that the other properties of the fMRI signal, such as the shape of the activation signal and the structure of trends in the signal, are known. As will be slightly pointed out in this thesis, these properties should also be investigated in more detail, but we leave it to further research. Finally, we want to remark

---

7. E.g. in Clare [1997]; Davidson [2003]; Friston [1995b];
8. Purdon [2001]; Worsley [2002]
that in this thesis we will focus on voxel-wise fMRI data analysis, which means that we focus on data from small brain volumes which data is assumed to be independent from data coming from other volumes.

The following of this thesis is divided into five chapters. Chapter 2 deals with fMRI backgrounds from a physical and more or less neural-biological point of view. Also the common mathematical approach to fMRI observations will be reviewed, explaining the linear model regularly used with three different contributions to the fMRI signal. In Chapter 3, the theoretical elaboration of our new likelihood-based approach is given. The computation of the likelihood function for fMRI signals is worked out, as well as methods to test for activity by the Generalized Likelihood Ratio Test. Finally, a method for selecting the order (complexity) of the model of the noise contribution will be elaborated. The theory given in the second part is elaborated in a practical way in Chapter 4, where the method programming the construction of simulated fMRI data and the estimation of the parameters from these data is given. From these estimated parameters and the related test statistic, a decision is made whether the signal contains a contribution from activation or not. Chapter 5 gives the results of our investigation from different points of view, where some of them need more discussion. Comparisons are made between the obtained likelihood-based results and the results obtained on the basis of the common $t$-test. Based on these results, we can come to conclusions and produce further recommendations in the last chapter (Chapter 6).
Chapter 2

**FMRI: physical background and common mathematical approach**

This chapter gives an introduction into the function of the brain and the possibility of using the magnetic properties of blood and brain matter for imaging the brain. When the physical and physiological backgrounds are explained, the chapter provides a method for modelling the data and the commonly approach used to estimate the parameters of the model. From the parameters a decision is made based on hypothesis test whether activation was present or not.

### 2.1 Use of nuclear magnetic resonance for imaging

The magnetic resonance properties of nuclei in matter are used in Magnetic Resonance Imaging (MRI). ‘Magnetic resonance’ means that when a certain nucleus is placed in a fluctuating magnetic field, it absorbs energy in the radio frequency range of the electromagnetic spectrum ($10^6$-$10^8$ Hz), and re-emits this energy when it relaxes to its original state if the magnetic field is turned off. This phenomenon is called Nuclear Magnetic Resonance (NMR); ‘Nuclear’ because only the nuclei of certain atoms reacts in that way; ‘Magnetic’ as a nuclear magnetic moment has to be available and ‘Resonance’ because of the direct frequency dependence of the measurable phenomenon with the magnetic field, which will be explained further in Section 2.2 of this thesis. The measurable signal can be used to make images of structures in human body or brain.

#### 2.1.1 Magnetic properties of hydrogen

To understand the so-called ‘magnetic resonance’ of matter, we have to zoom in at the matter to atomic scale. Atoms exist of protons, neutrons and electrons, where the protons and neutrons together form the nucleus (Figure 2.1). From quantum mechanics it is known that the electrons turn around the nucleus and the nucleus performs spins around its axis. Since the protons and electrons have a charge, every proton and electron causes a magnetic moment in the same (for positive charged protons) or opposite (for negative charged electrons) direction as its impulse moment. These magnetic moments act as magnetic dipoles. Within an isolated atom the magnetic moments of protons and electrons cancel out.
This is also the case in molecules with so-called ‘covalent bonds’, where the electrons are in the middle between two bonded atoms.

Contrary, the bond between e.g. oxygen and hydrogen (O-H) is a ‘polar bond’, which means that the electron has been dragged slightly away from one (here: hydrogen) nucleus towards the other (here: oxygen) nucleus. Therefore, the originally hydrogen related electron has less influence on the reaction of the hydrogen proton to the external magnetic field. Together with the fact that the amount of hydrogen in the human body is very high, it is very attractive to form the basis of Nuclear Magnetic Resonance with hydrogen. Further on, we will consider the hydrogen atom acting as a single proton, assuming that the effect of the hydrogen electron is negligible. In Figure 2.2 the magnetic dipole of a proton is visualized.

As mentioned above, all protons in a volume react as small magnetic dipoles. However, in general we will not experience any magnetic field from the volume, since the protons’ magnetic dipoles are oriented randomly. Therefore the magnetic field caused by the

Figure 2.2 | Visualization of spinning proton with positive charge (left) and representation of corresponding magnetic field (right).
magnetic dipoles is equal to zero in average.\footnote{11}

However, when an external magnetic field is applied, all dipoles will align with this external field, just as everyday magnets near an other magnet. This alignment could be in two possible states: parallel or anti-parallel with the external field (Figure 2.3). If one wonders about these two alignments and why not all spins align in the same direction as the magnetic field, imagine a dart which one wants to hold in vertical direction (Figure 2.4). This can be done by pointing the dart upwards, against the force of gravity $F_g$ which is an unstable (or high energy) direction. But one can also point the dart downwards, in same direction as the force of gravity. This direction is very stable (the dart will not point upwards by itself) or a so-called ‘low energy state’. Redefining ‘stable’ and ‘unstable’ in terms of energy, an alignment parallel with the external field has a lower energy level than an alignment anti-parallel with the external field.

11. Remark: The Earth magnetic field functions as an external magnetic field, so there will be a small magnetization. The Boltzmann equation given in Eq. (2.1) gives the number of spin-ups and spin-downs.
The Boltzmann equation\(^2\) tells how many spins are in spin-up state, with low energy, pointing into the same direction as \(B_0\) (denoted as \(N^-\)), and spin-down state, with high energy, pointing into opposite direction as \(B_0\) (denoted as \(N^+\)). The ratio between these two states is given by
\[
\frac{N^-}{N^+} = e^{-\frac{\Delta E}{kT}} ,
\]
where \(\Delta E\) stands for the absolute value of the energy difference between the two states, \(k\) is the Boltzmann factor and \(T\) temperature in Kelvin. Because \(\Delta E\) and \(k\) are constants, the ratio depends only on the temperature, which could also be taken as a constant during NMR measurements. The ratio (2.1) will be 99.9\% as a result for an external magnetic field \(B\) of 4 Tesla.

In practice, the magnetic moment is not totally aligned with the magnetic field. In fact, by little magnetic field disturbances, caused by atomic interactions, the spins will precess around their axis just as disturbed spinning tops (Figure 2.5). The precession of the magnetic moment has a certain frequency, depending on the strength of the magnetic field and their spin frequency.\(^3\) This frequency is called the Larmor frequency and is denotes as
\[
\nu_L = \gamma B ,
\]
with \(\gamma\) the gyromagnetic ratio (for hydrogen \(\gamma = 42.58 \times 10^{-6} \text{ Hz/T}\)) and \(B\) the strength of the external magnetic field.\(^4\)

We assume the aligned protons in the external magnetic field precessing with frequency \(\nu\) as starting point from now. The direction of the magnetic field is taken as the positive \(z\)-axis. Applying a circular oscillating electromagnetic wave with the same frequency \(\nu\) to the protons (orthogonal to the magnetic field), their orientations will flip from the \(z\)-axis into the \(xy\)-plane. This can be regarded as giving the precessing proton every cycle a little push and pull, resulting in the falling down of the magnetic moment from the \(z\)-axis to the \(xy\)-plane. This is shown in Figure 2.6. The energy of the photons needed for this action can be
computed by the following formula:

\[ \Delta E = h \nu_L \]  

where \( h \) is the Plank constant \((6.626 \times 10^{-34} \text{ Js})\).\(^{15}\) Because of the specific frequency \( \nu_L \) that is needed, this phenomenon gives rise to the term ‘Resonance’ in NMR. From Eq. (2.3), the specific frequency of that electromagnetic wave can be computed. For typical magnetic fields of strength between 1 and 10 Tesla, this frequency is of order \( 10^{8} \) Hz (100 MHz). This value lays in the radio frequency spectrum (compare to FM-radio frequencies from 88 to 108 MHz). Therefore the electromagnetic wave is called a Radio Frequency (RF) pulse. In practice a little range of frequencies is used as a result of an amplitude sinc-shaped pulse in time domain (see Figure 2.10). The proton will rotate in the xy-plane as long as the RF-pulse is active.

When the RF-pulse is turned off, the magnetic dipoles will align again with the still existing external magnetic field \( B_0 \). This relaxation occurs under emitting a photon, because the energy of a dipole orientated with the magnetic field is lower than the energy of that dipole orientated orthogonal to the magnetic field.

Remember that the proton is still spinning, causing a magnetic moment. This magnetic moment causes a measurable electromagnetic effect outside the proton. When the proton turns around in the xy-plane, the magnetic moment is measured as a circular electromagnetic field (e.m.f.) in the xy-direction, while an alignment with the z-axis causes a measurable e.m.f. signal in the z-direction. The above described relaxation from the xy-plane to the z-axis leads to a decay of the measured signal in the xy-direction during the relaxation. This is called ‘free induction decay’ (FID). The resulting magnetization during time is shown in the Figures 2.7 and 2.8. The measurable e.m.f. signals from the magnetic dipole orientation are used for imaging techniques.

---

15. Hornak [1996], Ch.3
2.1.2 Imaging

As explained in the previous section, each of the hydrogen atoms in water (H-O-H bindings) has a tiny magnetic dipole. This magnetic property is used in the following way to produce images.\(^\text{16}\) The subject of research is placed in a strong magnetic field \(B_0\) of typically 1.5 - 3 Tesla (Earth magnetic field is equal to \(0.5 \times 10^{-4}\) Tesla) by superconductive magnets in a coil.

The magnetic field \(B_0\), generated by the main or primary coil, is orientated along the centre line of the scanner, the (arbitrary) \(z\)-axis. In this way all (initial randomly orientated) hydrogen spins inside the subject turn their orientation in alignment with the \(z\)-axis. All the individual magnetic moments sum up to a net magnetization \(M_z\) as \((N^- - N^+)\mu\), where \(\mu\) is the magnetic strength vector of one hydrogen magnetic moment (now orientated along \(z\)-axis) and \(N^-\) and \(N^+\) as in Eq. (2.1).

By applying an RF-pulse of frequency \(\nu_R\) (the Larmor or resonance frequency), orthogonal to \(B_0\) (or \(B_z\)), the orientation of the hydrogen spins can be turned with a 90° switch from the \(z\)-direction into the \(xy\)-plane, resulting in an oscillating magnetization \(M_{xy}\).

When the RF-signal is turned off, the magnetization of the tissue relaxes from \(M_{xy}\) back to \(M_z\), resulting in a decaying circular e.m.f. in the two pick-up antennas orthogonal to the scanner centre line and an increasing e.m.f. in the \(z\)-direction. The time constant for the decay of \(M_{xy}\) is called \(T_2\), while the time constant for the increase of \(M_z\) is denoted as \(T_1\). The reason of the usage of two different time constants \(T_1\) and \(T_2\) will be given below. The time constants \(T_1\) and \(T_2\) are defined as follows:\(^\text{17}\)

\[
\begin{align*}
M_z(t) &= M_{z,0}(1 - e^{-\nu T_1}) \\
M_{xy}(t) &= M_{xy,0}e^{-\nu T_2},
\end{align*}
\]

(2.4)

with \(M_{z,0}\) the original magnetisation before the applied RF-pulse, while \(M_{xy,0}\) the original \(xy\)-magnetisation just after the applied RF-pulse. As one can derive from these equations, \(T_1\) is the time that the \(z\)-magnetisation is risen to about 63% of the original magnetisation, while \(T_2\) is the moment the \(xy\)-magnetisation is decreased 63% related to the maximum \(xy\)-magnetisation.

Table 2.1 | Water content of several human tissues of interest for MRI. As one can see, bone consists of much less water, so will give no response to MRI. The cortex consists of grey matter, surrounding the inside white matter. Grey matter is that part of the brain which contains the neurons or nerve cells. White matter is the part of the brain containing the nerve tracks.

From: Clare [1997], p.35

<table>
<thead>
<tr>
<th>Tissue</th>
<th>% Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grey matter</td>
<td>70</td>
</tr>
<tr>
<td>White matter</td>
<td>84</td>
</tr>
<tr>
<td>Blood</td>
<td>93</td>
</tr>
<tr>
<td>Bone</td>
<td>12</td>
</tr>
</tbody>
</table>

16. Clare [1997], Sec.2.3; Hornak [1996]; MRI on WebCT [n.d.]; Jan Sijbers, personal communication
17. E.g. Hornak [1996], Ch.3
The measured signal amplitude is related to the number of hydrogen atoms. Measuring per voxel (from volume pixel, the smallest three dimensional imaging scale, about 1 mm³), the number of hydrogen atoms is related to the hydrogen density in that particular voxel. Like other tissues in the body, brain is composed of water for about 70% in average. However, different parts of the brain differ slightly in density of water. Nerve cells are rich of water, while the fatty coating around the nerves contains less. With knowledge about the amount of emitted energy during the relaxation, the density of the hydrogen atoms at a certain voxel can be derived. Combining knowledge about the relation between hydrogen density and tissue material and the actual measured signal gives insight into the structure of human tissue. Even without knowledge of typical proton density (PD), the structure itself can be derived from relative differences.

In the analysis so far, signals from different voxels cannot be distinguished, because all the hydrogen atoms in the subject react in the same way to the magnetization and RF-pulse. To look at the signal of one particular voxel, the next three steps are made during the above indicated procedure. First, a magnetic field $G_z$ with gradient $g_z$ in z-direction is added to the external magnetic field $B_0$. Therefore, only spins in a certain slice react to the RF-pulse, because the varying magnetic field causes different Larmor frequencies, as can be seen from

---

18. Hornak [1996], Ch.1; Parry [2002], p.4
19. Parry [2002], p
Equation (2.2). The thickness of the slice depends on the gradient $g_z$ and the bandwidth of the RF-pulse (Figure 2.9).

The addition of the linear magnetic field $G_z$ causes that the spins in a certain slice will precess at frequency $\nu(z)$:

$$\nu(z) = \nu'(z, \gamma(B_0 + G_z z) = \nu_{z,0} + \gamma G_z z .$$  \hspace{1cm} (2.5)

Spins at one side of the scanner, where $z = -\frac{L}{2}$ (with $L$ the length of the scanner and $z = 0$ in the middle), feel a lower magnetic field than spins at the other side ($z = \frac{L}{2}$) and therefore will precess slower (according to Eq. (2.5)). Depending on the frequency of the RF-pulse, a specific slice of the subject is selected, where the spins will rotate in the $xy$-plane with frequency $\nu(z)$, while in other slices nothing changes and all spins are still orientated in the $z$-direction. So only an e.m.f. of the hydrogen atoms in the selected slice is measured in the $x$ and $y$ antennas. After applying the RF-pulse, the gradient can be switched off. As explained before, when the RF-pulse stops, the spins in the excited slice will relax from the $xy$-plane to the $z$-direction.

When a certain slice is selected using the $z$-gradient, a second gradient is applied in the (arbitrary) $x$-direction during the measurement, causing a second frequency differentiation between the rotating spins in the slice. Spins in lower magnetic field again rotate slower than spins in higher magnetic field, and also cause a lower electromagnetic signal. Remember that we are looking at a certain slice orthogonal to the main scanner axis. Simplified, in this particular slice all the nuclei turn around like a wheel, but due to the $x$-gradient the nuclei
at one side spin faster than nuclei at the other side, or with a higher frequency. A rotation with high frequency results in a received e.m.f. signal with high frequency, and opposite with low frequency. From the different received signal frequencies, the x-position of the voxels can be distinguished. Therefore, this method is called ‘frequency encoding’. Now z and x position are determined.

Last, between two above described measurements of the e.m.f. in the x-direction, a momentary magnetic gradient $G_y$ in y-direction is applied. The short existing gradient effects the spins precessing slightly different to each other for a moment, causing a dephasing in y-direction. When the momentary gradient is turned off, all spins will precess at the same frequency again. The dephasing results in different phases of the received e.m.f. for different y-positions. This action is called ‘phase encoding’.

These three steps give three dimensional information of the voxel-position. From the combined frequency and phase information an image can be made of the selected slice by inverse Fourier transform. Together with the intensity of the signal from the hydrogen density in blood at a certain voxel, we can make an image. The intensity of the signal can be given more contrast using the $T_1$ or $T_2$ properties of tissues. As mentioned above these two time constants are not the same. The time constant $T_1$ (named ‘spin-lattice relaxation time’ is related to the increase of the magnetization $M_z$, while $T_2$ (named ‘spin-spin relaxation time’ is related to the decreasing of $M_{xy}$). The decreasing of $M_{xy}$ has a lower time constant than the increasing of $M_z$, because of dephasing of the magnetic moments in xy-direction. This dephasing exists due to the fact that a specific proton experiences not only the external magnetic field $B_0$, but also the generated magnetic fields of his neighbouring protons. Since the precessing of a proton is proportional to the local magnetic field, the protons will precess at different frequencies, influenced by very small differences in the magnetic field. The precessing at different frequencies results in a net decrease of magnetic moment in the xy-plane during measurement time. This dephasing depends on the mobility of the protons, which is again related to the matter of tissue.

Actually, due to inhomogeneous magnetic field $B_0$, the dephasing of xy-magnetization reacts faster than $T_2$ by a smaller time constant $T_2^*$. This dephasing can be cancelled out for a while by a so-called ‘180° RF-pulse’ or ‘Spin Echo’. This is illustrated in Figure 2.10. Both $T_2^*$ and $T_2$ can be measured and used for imaging.

It is possible to make images of which the contrast is based on the measured values of $T_1$, $T_2$ or $T_2^*$ or with knowledge about the hydrogen density in a certain tissue. All these three
measurable characteristics are related to the properties of tissue (see Table 2.2 for $T_1$ and $T_2$).\textsuperscript{21} As the ratios of $T_1$ and $T_2$ of different tissues differ, the contrast between the tissues in these images will also differ. A $T_1$ weighted image gives other contrasts than a $T_2$ weighted image (see Figure 2.11).

Table 2.2 | Different $T_1$ and $T_2$ for some tissues of interest in fMRI

<table>
<thead>
<tr>
<th>Tissue</th>
<th>$T_1$ ((10^{-3} \text{s}))</th>
<th>$T_2$ ((10^{-3} \text{s}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gray matter</td>
<td>950</td>
<td>100</td>
</tr>
<tr>
<td>White matter</td>
<td>600</td>
<td>80</td>
</tr>
<tr>
<td>Blood</td>
<td>1200</td>
<td>100-200</td>
</tr>
<tr>
<td>Fat</td>
<td>250</td>
<td>60</td>
</tr>
<tr>
<td>Muscle</td>
<td>900</td>
<td>50</td>
</tr>
</tbody>
</table>

Figure 2.11 | Examples of different contrast imaging techniques. Left: a $T_1$ weighted image, based on the relaxation time $T_1$; in the middle: a $T_2$ weighted image, based on the dephasing time $T_2$; right: a PD weighted image, based on the differences in proton density in brain.

From: Lewis Center for NeuroImaging, http://lcni.uoregon.edu/downloads.html

2.2 Neurology and functional MRI

With functional MRI, we want to get insight into the question which brain regions are related to which task. Therefore we first want to know how the brain processes tasks and second how we can recognize these processes in MRI data. This section handles about these two aspects.

2.2.1 Physiological aspects of brain activation

To get a better insight into the work of our brains, we have to know a little about the physical and chemical processes that occur in our brains when we think or act. Therefore it is important to bring up the function of the basic structures of the brains.22

The brains are constructed of many basic unit cells, called neurons. In each volume element of 1 mm³ about $10^6$ neurons are present.23 As showed in Figure 2.12, a typical neuron exists of a cell body (called ‘soma’) and some forks (one ‘axon’ and one or more ‘dendrites’).24 The axon of one neuron makes a connection to a dendrite of a neighbouring neuron. The region of this connection is called ‘synapse’. The cell body functions as a kind of command centre, while the axon is designed to conduct electrical impulses to dendrites of other neurons, operating as receivers. Electrical signals from different dendrites add at the cell body and are redirected from the cell body to the axon. When the power of an electrical impulse travelling through the axon exceeds a certain threshold, it causes the release of neurotransmitters from the end of the axon into the synapse. Neurotransmitters are chemicals that are used to relay, amplify and modulate electrical signals between two neurons. These neurotransmitters transfer the electrical impulse to the receptors on the dendrites of the neighbouring neuron. In such a way a signal is sent through the brain until a large number of neurons are activated in a network.25

For functional MRI, it is not really relevant how the cell body processes the electrical signals exactly, but we look at the consequences of the transfer of neurotransmitters within the synapses, the region of contact between the axon and dendrites. After the neurotransmitters have been released into the synapse, they have to be recycled or re-uptaken to the axon.26 This process requires energy, among other things supplied by oxygen in oxygenated

---

22. Parry [2002]
23. Lange [1997], p.2
24. Structure of the neuron [n.d.]
25. Parry [2002], p.9
26. Logothesis [2002], p.1004
2.2.2 Use of the haemodynamic response function for fMRI

At the base of fMRI stands the need for oxygenated blood in regions of neural activation. Activation of a region in the brain results in an increasing blood supply. This subsection deals with the relation between brain activation and magnetization properties.

As explained in Section 2.1, the use of the different magnetization properties of the tissues is sufficient to distinguish structures and matters in the brain. As explained in the previous subsection, an activation of a brain region as a result of executing a certain task effects a supply of extra blood. In 1890, Roy and Sherrington were the first to postulate the idea of a relation between activation and blood flow. After that, much evidence is obtained. The underlying process of the increasing cerebral blood flow (CBF) was long not completely known, and just recently proven. One can say that neural activity accompanies with need of energy, that can only be obtained from blood, as for all processes in the human body. Unfortunately, the increasing CBF can not be used to recognize activation, because an increase in blood supply does not significantly change the density of hydrogen atoms in the veins, neither other magnetization properties. So merely the extra supply of blood gives not enough information to detect the active regions by magnetization. To overcome this problem, the diamagnetic and paramagnetic properties of oxygenated and deoxygenated haemoglobin, respectively, can be used. In the active area, blood undergoes a process from an oxygenated to a deoxygenated state by metabolism. This is called the haemodynamic response to neural activity. The haemodynamic response function (HRF) gives the relative concentration of oxygen in blood during time after executing a task. In literature, there are many examples or approximations of haemodynamic response functions, where two sketches of them are given in Figure 2.13. It should be noticed that the HRF differs from different subject, different brain region and different task.

It is found out that the blood oxygenation level can help to give a better contrast to the NMR signal from blood. The presence of deoxyhaemoglobin in blood changes the NMR-signal, because deoxygenated blood is paramagnetic, while oxygenated blood is diamagnetic. This gives rise to more magnetization signal when oxygen in blood is delivered to the tissue. This raise in NMR-contrast is called Blood Oxygenation Level Dependent (BOLD) contrast. Further analysis turns out that BOLD contrast not only depends on oxygenation level, but on several parameters, such as cerebral blood flow (CBF) and cerebral blood

31. Mukamel [2005]
32. Ogawa [1990], p.9868
33. Friston [1994]; Lange [1997]; Parry [2002]
34. Lange [1997], p.10; Worsley [2002]; Woolrich [2001]; – [2004], p.217
35. Friston [2004], p.25; Turner [2000], p.46; Woolrich [2004]
37. Lange [1997]; Logothesis [2002], p.1009; Thulborn [1982]
38. About fMRI (General) [n.d.]; Cohen [2001], p202; Ogawa [1990], Parry [2002], p.9
An example of the changes in amounts (and concentrations) of oxygenated and deoxygenated blood is given in Figure 2.14. The total haemodynamic response function starts with an initial dip due to instantaneous oxygen metabolism, while the cerebral blood flow does not react fast enough to supply fresh oxygenated blood. This phenomenon can be recognized in Figure 2.14, subtracting the amounts of deoxygenated and oxygenated blood.

When the subject under study executes a task, there is a haemodynamic response anywhere in the brain. In fMRI, the brain is examined per region (usually voxelwise) during time, so multiple MRI scans of the brain are made during measurements. The brain activation should be recognizable in one or more regions by following the NMR signal changes due to deoxygenation. In fact, BOLD contrast is used to measure the presence of the haemodynamic response as an indication of activation in a brain area, in addition to the usage of the magnetic properties of hydrogen to investigate the structure of brain matter.
2.3 The General Linear Model for fMRI data

To decide whether a brain region (one or more voxels) is active or not, we need a method to distinguish between these two states. The true properties of the signal are unknown. Moreover, the activation-related signal changes are of the order of 1% to 10% of the signal without activation. Also the signal-to-noise ratio (SNR) is usually low. Therefore, brain activation detection has to be done by statistical methods.

2.3.1 Model structure

Measured data is the result of some (unknown) system in the background. Some physical and physiological contributions come together and lead to the measured data. But we do not know the whole physical background. And if we know, we can not compute it because of its complexity. So we need a model which describes what we see in a more or less simple way.

In general, a General Linear Model (GLM) is used to describe fMRI time series data. The GLM relates observations to their expectations by expressing the observations as a linear combination of explanatory variables and a noise term. Considering one voxel, the response fMRI data can be seen as a time series \( y = (y_1, ..., y_n)^T \) of observations, with \( n \) the number of observations per voxel, typically \( n = 80 - 120 \). The superscript \( T \) denotes matrix transposition. The general linear model of time series data \( y_k \) at a certain time \( k \) in a certain voxel describes the measurements as:

\[
y_k = z_{k, \text{hrf}} \theta_{\text{hrf}} + x_{k, \text{tr}} \theta_{\text{tr}} + e_k,
\]

where \( k = 1, ..., n \) indexes the number of the scan, the subscript ‘hrf’ denotes the contribution from the haemodynamic response function (HRF) to \( y_k \), while the subscript ‘tr’ denotes the contribution from trends or drifts to \( y_k \). The trend can be extended to include multiple terms, e.g. polynomial terms or a combination of sines and cosines. The contributions of the HRF, the trends and the noise will be elaborated in the next subsections. The parameters \( \theta_{\text{hrf}} \) and \( \theta_{\text{tr}} \) are unknown real valued parameters related to the corresponding signal contributions. These parameters have to be estimated. After estimation, the parameter \( \theta_{\text{hrf}} \), which corresponds to the contribution \( x_{k, \text{hrf}} \) (describing the HRF), can be used to detect brain activity.

Equation (2.6) can also be written in matrix notation as

\[
y = X\theta + e,
\]

in which \( X \) is an \( n \times m \) matrix, called ‘design matrix’. In \( X \), the \( n \) rows represent the \( n \) measurements, while the \( m \) columns of the design matrix represent the haemodynamic response and trend effects. The vector \( \theta \) is an \( m \times 1 \) matrix of unknown parameters that describe the level of the contributions of the effects modelled by each column of the design matrix. In the next subsections the three contributions to the fMRI signal, as given in Eq. (2.6) will be considered in detail.

42. Lukic [2002]
43. Friston [1995a]; Worsley [1995]
44. Sijbers [2005], p.539
45. Personal communication Sijbers, Den Dekker, March 2006
46. Gautama [2005],p.1212; Hu [2005], p.747; Rowe [2002b]
2.3.2 The haemodynamic response function

The most interesting contribution to the fMRI time series is the contribution of the haemodynamic response. Let \( u(t) \) denote the external stimulus (mostly assumed to be a box-car ‘on-off’ signal) of a certain task at certain time \( t \), whereas \( h(t) \) denotes the real haemodynamic blood oxygen level response. Then \( x_{hrf}(t) \) denotes the voxel-wise measured signal of the haemodynamic response at that time. The response \( x_{hrf}(t) \) reacts not instantaneously on \( u(t) \) due to the reaction time of the cerebral blood flow (CBF) on activation of a brain region, resulting in a delay of the maximum response by about 4 - 6 seconds. A method to describe the relation between stimulus and measured response is using a mathematical representation of the haemodynamic response \( h(t) \) in a convolution with \( u(t) \), as follows:\(^{47}\)

\[
x_{hrf}(t) = \int_{-\infty}^{\infty} h(\tau) u(t-\tau) d\tau,
\]

or, with discrete samples:

\[
x_{hrf}[k] = x_{hrf} = \sum_{\tau} h[\tau] u[k-\tau] \label{eq:2.9}
\]

Several models of the HRF \( h(t) \) in Eq. (2.8) have been proposed.\(^{48}\) A simple one is a gamma function or a difference of two gamma functions to model the slight intensity dip after the response has fallen back to zero. A standard example is the HRF available in the Matlab-toolboxes SPM’96 and SPM2:\(^{49}\)

\[
h(t) = \left( \frac{t}{s_1} \right)^{p_1} e^{-\frac{t-s_1}{q_1}} - \left( \frac{t}{s_2} \right)^{p_2} e^{-\frac{t-s_2}{q_2}},
\]

where \( t \) is time in seconds and \( s_j = p_j q_j, j = \{1, 2\} \) denotes the time at which peaks in the oxygen concentration occur, where the maximum occurs at \( s_1 \) and the minimum at \( s_2 \). This standard function \( h(t) \) with standard values \( (p_1 = 6, p_2 = 12, q_1 = q_2 = 0.9 \text{ seconds and } r = 0.35) \) as given by Glover [1999] and Worsley [2002] is shown in Figure 2.15.

Although this standard form of the HRF is being used widely, we have to remind that the HRF varies for different persons, tasks, brain regions and the presence of diseases.\(^{50}\) Also some remarks should be made with respect to the shape of the standard HRF. As mentioned in Section 2.2.2, the HRF starts with an initial dip due to instantaneous oxygen consumption (Figures 3.13 and 3.14),\(^{51}\) which does not exist in the standard models of the HRF.

Taking \( x_{hrf} \) in Eq. (2.6) as the HRF response modelled in Eqs. (2.8) –(2.10), \( \theta_{hrf} \) is related to the level or presence of activation during the measurement and does not depend on time. In general, \( \theta_{hrf} \) can be a vector, but in this thesis we will assume the \( \theta_{hrf} \) as a scalar, being the first element of the parameter vector \( \theta \). Hence, \( \theta_{hrf} \) is a voxel-wise constant to be estimated.

47. Friston [1998]; Lange [1999], p.285; Worsley [2002]
49. SPM Software [n.d.]; Worsley [2002]
50. Glover [1999]; Friston [1994], p.6; Parry [2002], p.12
51. Barinaga [1997]
Section 2.3  The General Linear Model for fMRI data

2.3.3 Trends
In fMRI data, some low frequency trend terms are observed.\[^{52}\] Two suggested explanations for these terms are physiological noise and subject motion. However, these phenomena do not seem to be the main cause of this drifting. The most likely cause of the drift terms are slight changes in the local magnetic field in the scanner due to scanner instabilities.\[^{53}\]

Low frequency drift terms and offset can be represented in Eq. (2.6) as a polynomial or as a combination of sines and cosines, of which the order or number of parameters \( m \) should be estimated.\[^{54}\] Typically a low order polynomial is chosen.\[^{55}\] Choosing order \( m = 2 \), \( x_{k,\text{tr}} \theta_{\text{tr}} \) in Eq. (2.6) can be expressed as

\[
x_{k,\text{tr}} \theta_{\text{tr}} = \theta_{\text{tr},0} + \theta_{\text{tr},1} k + \theta_{\text{tr},2} k^2.
\]

The trend is assumed to be describable by a constant model during measurements, so no changes in parameters or model structure occur. Combining the parameters of the haemodynamic response function \( \theta_{\text{hrf}} \) with the trend parameters \( \theta_{\text{tr}} \), the resulting parameter vector becomes

\[
\theta = [\theta_{\text{hrf}}, \theta_{\text{tr}}]
\]

2.3.4 Noise
A main contribution to fMRI data comes from the noise. The noise in the data has several sources, such as cardiac and breathing cycles, subject motion and thermal noise from electrons in the human body and receiver coils. Cardiac and breathing cycles cause a periodic fluctuation in blood flow and oxygen level, resulting in disturbances in the BOLD contrast. When we consider subject motion, we have to take into account that the resolution of fMRI is approximately 1 millimetre. As a result, even small head motion causes a serious

\[^{52}\] Purdon [2001]
\[^{53}\] Smith [1999], p.532
\[^{54}\] Worsley [2002]
\[^{55}\] Friston [2004], p.10; Gautama [2004], p.1211; Jezzard [1999], p.83; Nan [1999], p.321
disturbance in the fMRI signal. Thermal noise disturbs directly the measurements of the e.m.f. by the antennas.

In fMRI literature, the noise is often assumed to be Gaussian white, which means that the errors $e_k$ are temporally independent and identically distributed (IID), having a normal probability density function $\mathcal{N}(0, \sigma^2_e)$. The error variance $\sigma^2_e$ for a specific voxel is assumed constant during the tests, but not necessarily the same for all voxels. In this study, we approach the fMRI data voxel-wise, assuming that no spatial correlations exist.

In practice, the noise in fMRI is not white but contains temporal, and perhaps also spatial, correlations. The correlation coefficient between scans can be as high as 0.4. Correlation arises from e.g. cardiac and breathing cycles, called physiological effects. When we assume the noise to be coloured (i.e., correlated) instead of white, we will write $v_k$ instead of $e_k$ in Eq. (2.6). The noise is still assumed to be Gaussian distributed. Several approaches to deal with the noise of fMRI are in use. Smoothing the data is the most commonly used and most early worked-out method, but prewhitening of the data using an autoregressive model of the noise is an important subject of research. First we say something about the smoothing operation, but thereafter we will focus on prewhitening the data.

**Smoothing**

FMRI time series can be assumed to exist of low frequency signal contributions and low and high frequency noise contributions. Therefore, temporal smoothing can be used to increase the signal-to-noise ratio (SNR). The optimal smoothing function or kernel is in principal related to the specific haemodynamic response. After smoothing the data, the assumption is made that the manifesting correlations are mainly caused by the smoothing operation. Therefore, the smoothing parameter is chosen in such a way that the correlation created by smoothing is much more (e.g. 3 seconds) than the correlation due to physiological compounds (assumed to be about 1 second). The added known correlation overrules the present unknown correlation. While smoothing adds correlation to the data, this method is named ‘precolouring’ the data.

For smoothing, the general linear model given in Eq. (2.7) is extended with a convolution matrix (a Gaussian kernel with certain width parameter $s$) $K$, resulting in

$$K_y = (KX)0 + Ke,$$  \hspace{1cm} (2.13)

where $K_y$ represents the temporally smoothed data and $KX$ can be renamed as the new design matrix $X_{sm}$.

In common fMRI analysis, the images are also smoothed spatially, to make the data less noisy at the cost of spatial resolution. Typically the resolution is lowered from $256 \times 256 \times 128$ to $64 \times 64 \times 16$ voxels. Notice that this also influences the statistical properties of the data, because typically the smoothing kernel is substantially greater than the low resolution voxel size. In our further analysis we skip this spatial smoothing, concentrating at a temporal voxel-wise approach to fMRI data.
Prewhitening

Instead of precolouring the data by smoothing, the model can be designed to describe the colouring or correlations in the data. In this case, the present correlations in the noise are captured in the model. The remaining residuals are assumed to be white distributed. Therefore, this method is also called ‘prewhitening’ the data.\(^{62}\) Now, the model from Eqs. (2.6) and (2.7) can be expressed as:

\[
y_k = x_{k-1} a_1 + x_k a_2 + v_k
\]  
(2.14)

and

\[
y = X \theta + v
\]
(2.15)

with \(v\) representing the coloured noise and containing a model of the noise. The model that is most simple and widely used to account for the correlations in the noise is the first order autoregressive model (abbreviated as ‘AR(1)’).\(^{63}\) The first order autoregressive model describes the current (at time \(k\)) error as a weighted sum of the previous error (at time \(k-1\)) and a white noise term:

\[
v_k + a_1 v_{k-1} = \varepsilon_k
\]  
(2.16)

or, alternatively with a different sign of the parameter \(a_1\):

\[
v_k = a_1 v_{k-1} + \varepsilon_k
\]  
(2.17)

where \(v_i\) is the coloured noise, \(\varepsilon_i\) is again temporally independent and identically distributed (IID) white noise and \(|a_1| < 1\) is the AR(1) parameter, which indicates the strength of the correlation between previous and present error. Further on, we use the representation of Eq. (2.16), because this one corresponds to the definition used in the Matlab toolbox ARMASA. Model (2.16) can be extended to an AR(p) model with \(p\) parameters:

\[
v_k = -a_1 v_{k-1} - a_2 v_{k-2} - \ldots - a_p v_{k-p} + \varepsilon_k
\]  
(2.18)

where more combinations with previous noise terms are made. Note that Eqs. (2.14) – (2.18) still assume a voxel-wise approach to the fMRI data. With other persons, Worsley give some indication of autoregressive noise parameters. For the first four orders, the parameters will vary between -0.4 an 0.1 (for model structure (4.18)), causing a noise spectrum with power at low frequencies.\(^{64}\)

2.3.5 Parameter estimation

The different approaches to fMRI data result in different methods of estimation of the model parameters. In what follows, these different estimation methods will be considered.

When the noise is assumed to be white, an estimate of \(\theta\) is commonly obtained by least squares fitting of the model to the fMRI time series data, described by Eq. (2.7). An

---

62. Gautama [2004], p.1204; Worsley [2002], p.4
63. Gautama [2005], p.1211; Maxim [2005], p.142; Solo [2001], p.27; Worsley [2002], p.4
64. Friman [2005], p.861; Maxim [2005], p.141; Rowe [2004], p.1083; Woolrich [2004], p.223; Worsley [2002], p.6; Zöllei [2003], p.176
expression of this ordinary least squares (OLS) unbiased estimator is given by:\textsuperscript{65}
\[
\hat{\theta}_{\text{OLS}} = (X^T X)^{-1} X^T y,
\]  
(2.19)

where \( \hat{\theta}_{\text{OLS}} \) is the best linear unbiased estimator (BLUE) of \( \theta \), if the errors are assumed to be uncorrelated with mean \( \mu = 0 \) and equal variance \( \sigma_e^2 \). By ‘best’, we mean that the estimator has the smallest variance with respect to other unbiased estimators. By ‘unbiased’, we mean that the expected value of the parameter estimator equals the true value of the parameter.

When the fMRI data is being smoothed, the new design matrix \( X_{\text{sm}} \) replaces the combination of the smoothing kernel \( K \) and the original design matrix \( X \). The least squares estimators of \( \theta \) becomes, similar to Eq. (2.19):
\[
\hat{\theta}_{\text{sm}} = (X_{\text{sm}}^T X_{\text{sm}})^{-1} X_{\text{sm}}^T K y.
\]  
(2.20)

This estimator is not fully optimal (minimal variance), but unbiased and, when the noise is coloured, generally more efficient than the OLS estimator described by Eq. (2.19).\textsuperscript{66} The variance of the errors, \( \sigma_e^2 \), is estimated by:\textsuperscript{67}
\[
\hat{\sigma}_{\text{e,sm}}^2 = \frac{(Ky - \hat{X}_{\text{sm}} \hat{\theta}_{\text{sm}})^T (Ky - \hat{X}_{\text{sm}} \hat{\theta}_{\text{sm}})}{\text{tr}(R_{\text{sm}} KK^T)}
\]  
(2.21)

with ‘\( \text{tr}(\cdot) \)’ denotes the sum of the elements on the main diagonal (upper left to lower right) of the matrix \( \cdot \). The matrix \( R_{\text{sm}} \) is the residual-forming \( n \times n \) matrix, given by:
\[
R_{\text{sm}} = I - X_{\text{sm}} (X_{\text{sm}}^T X_{\text{sm}})^{-1} X_{\text{sm}}^T,
\]  
(2.22)

with \( I \) the identity matrix. It can be shown that the trace of \( R_{\text{sm}} \) equals \( n-m \), with \( n \) the number of observations and \( m \) the dimension of \( \theta \). Thus, for \( K = I \), \( \sigma_{\text{e,sm}}^2 \) reduces to
\[
\hat{\sigma}_{\text{e,sm}}^2 = \frac{(y - \hat{X}_{\text{sm}} \hat{\theta}_{\text{sm}})^T (y - \hat{X}_{\text{sm}} \hat{\theta}_{\text{sm}})}{n-m}.
\]  
(2.23)

When we prewhiten the data, the noise model can be used to construct an efficient estimator of the explanatory parameters \( \theta \). For this purpose the covariance matrix of the noise is required. We can define the \( n \times n \) covariance matrix \( \mathbb{E}(\mathbb{v}\mathbb{v}^T) = \sigma_v^2 \mathbb{V} \) as:
\[
\sigma_v^2 \mathbb{V} = \sigma_v^2 \begin{pmatrix}
\rho(0) & \rho(1) & \ldots & \rho(n-1) \\
\rho(1) & \rho(0) & \ldots & \rho(n-2) \\
\vdots & \vdots & \ddots & \vdots \\
\rho(n-1) & \rho(n-2) & \ldots & \rho(0)
\end{pmatrix},
\]  
(2.24)

with \( \rho(k) = \mathbb{E}[\mathbb{v}_i \mathbb{v}_{i+k}] / \sigma_v^2 \) the correlations between two noise contributions over the time difference \( k \), with \( \mathbb{E}[\cdot] \) the expected value of \( [\cdot] \). By this definition \( \rho(0) = 1 \). In Eq. (2.24), \( \sigma_v^2 \) and \( \sigma_e^2 \) are the variances of \( \mathbb{v} \) and \( \mathbb{e} \) respectively. The gain of the autoregressive noise is defined as:

\textsuperscript{65} Hu [2005], p.747; Sijbers [2005], p.540
\textsuperscript{66} Friston [1995a]; Woolrich [2001]; Worsley [1995], p.174
\textsuperscript{67} Worsley [1995], p.174
If the parameters $a_1, \ldots, a_p$ of the noise model are known, they can be used to compute the elements of the matrix $V$ in Eq. (2.24). This can be done by using the Yule Walker relations:

$$\rho(-k) = \rho(k).$$

Given the AR parameters, the covariance matrix $\sigma_e^2 V$ is obtained from Eqs. (2.24) and (2.26) and it is used to weigh the OLS estimator $\hat{\theta}_{OLS}$ of Eq. (2.19). The so-called generalized least squares (GLS) estimator of $\theta$ results in:

$$\hat{\theta}_{GLS} = (X^T V^{-1} X)^{-1} X^T V^{-1} y.$$  

This result is equivalent to applying the matrix $V^{-1/2}$ to the fMRI data $y$ in Eq. (2.15), before applying an OLS estimator, which is expressed in the name ‘prewhitening’. The whole process could be used iteratively to get a more precise estimate of $\theta$. The resulting $\hat{\theta}_{GLS}$ is the BLUE estimator of $\theta$. The variance of $\hat{\theta}_{GLS}$ is given by

$$\sigma^2_{\theta, GLS} = \frac{(y - X\hat{\theta}_{GLS})^T V^{-1} (y - X\hat{\theta}_{GLS})}{n-m},$$

which has a $\chi^2$ distribution with $n-m$ degrees of freedom.

In practice, the parameters $a_1, \ldots, a_p$ in Eqs. (2.18) and (2.26) and the order $p$ of the AR model of the noise are unknown and thus are often estimated from the residuals after OLS estimation of the deterministic part of the signal. For a chosen order $p$, the parameters $a_1, \ldots, a_p$ can be estimated using, for instance, the Least Squares (LS) method. From these parameters $\hat{a}_1, \ldots, \hat{a}_p$, the corresponding matrix $\hat{V}$ is computed, using the Yule-Walker equations, and is then substituted as an estimator of $V$ in Eq. (2.27). The generalized least squares estimator $\hat{\theta}_{GLS}$ of $\theta$ then becomes:

$$\hat{\theta}_{GLS} = (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} y,$$

while also in Eq. (2.28) $\hat{V}$ should be used. Although the statistics of the variance of this estimator are not completely known, the variance is assumed to have a $\chi^2$ distribution with $n-m$ degrees of freedom, as it would have when $V$ is used instead of $\hat{V}$. The choice of the order $p$ of the autoregressive model can be done a priori or by using order selection methods (OSMs). Several order selection methods will be briefly reviewed in Section 3.2. As mentioned in Chapter 1, our hypothesis about order selection is that order selection will give lower performance than selecting the true order, but will give better results than choosing a false order.

68. Kay [1998], p.541
69. Gautama [2004], p.1203
70. Van den Bos [1998], p.PS0115
2.3.6 Hypothesis tests for activation detection

The most common goal of an fMRI activation study is to determine whether activation is present at a particular region in an image. Because of the very noisy signal, statistical hypothesis tests have to be used to distinguish between inactive (null hypothesis $H_0$) and active (alternative hypothesis $H_1$) voxels. Because no a-priori knowledge about the presence of activation is present, the distribution of measured data without activation (the null-distribution) is used as basis for the hypothesis test.\footnote{Chatfield [1996], p.134; Friston [1995b]} A decision has to be made between the null hypothesis $H_0$ of no activation and the alternative hypothesis $H_1$ that activation is present:

$$
\begin{cases}
H_0: \theta_{\text{ref}} = 0 \\
H_1: \theta_{\text{ref}} \neq 0.
\end{cases}
$$

(2.30)

Because the estimator $\hat{\theta}_{\text{ref}}$ is a random variable, a test statistic with known distribution function under $H_0$ is used to decide between $H_0$ and $H_1$. The value of this test statistic expresses a measure of probability that no activation is present (i.e. $H_0$ has to be accepted) at the corresponding location in the brain. In practice, $H_1$ is accepted when the test statistic exceeds a user specific threshold.\footnote{Lukic [2002], p.73}

As explained in Section 2.2.2, the explanatory and the confounding contributions are estimated by the estimator $\hat{\theta}$, where $\hat{\theta}$ is a vector valued random variable. There exist several statistical tests for brain activity, based on the evaluation of $\hat{\theta}_{\text{ref}}$, the activation-related element of $\hat{\theta}$. For smoothed data, the test statistic most commonly used is given by: \footnote{Worsley [1995], p.174; Worsley [1995], p.174; Friston [2004], sec.IV; Lukic [2002], p.74; Rowe [2002a,b]}

$$
T_{sm} = \frac{\hat{\theta}_{\text{ref}}}{\sqrt{\left( X_{\text{sm}}^T X_{\text{sm}} \right)^{-1} X_{\text{sm}}^T K K^T X_{\text{sm}} (X_{\text{sm}}^T X_{\text{sm}})^{-1} \hat{\sigma}^2_{\text{ref}, sm}}}.
$$

(2.31)

with $\left[\cdot\right]_{11}$ denoting the upper left element of the matrix $\left[\cdot\right]$. $T_{sm}$ has a Student’s-$t$ distribution under $H_0$, where the denominator represents the square root of the variance of $\hat{\theta}_{\text{ref}}$. An expression for the number of the degrees of freedom of $T_{sm}$ under $H_0$ is given by:\footnote{Friston [2004], sec.IV; Lukic [2002], p.74; Rowe [2002a,b]}

$$
d.o.f_{sm} = \frac{2 \text{E} \left[ \hat{\sigma}^2_{\text{sm}} \right]^2}{\text{Var} \left[ \hat{\sigma}^2_{\text{sm}} \right]} = \frac{\left( \text{tr} \left( R_{\text{ref}, sm} K K^T \right) \right)^2}{\text{tr} \left( R_{\text{ref}, sm} K K^T R_{\text{ref}, sm} K K^T \right)}.
$$

(2.32)

For prewhitened data with parameter estimator $\hat{\theta}_{\text{ref}, p}$, the $t$-test and the $F$-test are most common.\footnote{Friston [2004], sec.IV; Lukic [2002], p.74; Rowe [2002a,b]} The $t$-test statistic to assess the significance of the parameter $\hat{\theta}_{\text{ref}}$ is given by:

$$
T_t = \frac{\hat{\theta}_{\text{ref}}}{\sqrt{\hat{\sigma}^2_{\text{ref, p}}}},
$$

(2.33)

where $\hat{\sigma}^2_{\text{ref, p}}$ is the first element of the matrix $(X^T \hat{\Phi}^{-1} X)^{-1}$ in Eq. (2.29). The product $\hat{\sigma}^2_{\text{ref, p}}$ is the estimated variance of $\hat{\theta}_{\text{ref}}$ being the first element of $\hat{\theta}_{\text{ref}, p}$. Approximately,
T, has a Student-t distribution under H_0 with n-m degrees of freedom.\(^{76}\)

The F-test is an alternative test statistic, quadratically related to the t-test:

\[
T_F = \frac{\hat{\theta}_{hrf}^2}{\hat{\Sigma}^{1/2}}.
\]

(2.34)

Approximately, T, has a F distribution under H_0 with n-m degrees of freedom.\(^{77}\)

A test statistic T can be used in a two-sided and a one-sided test. If it is not known whether \(\theta_{hrf}\) is positive or not during activity, a two-sided test has to be used. A two-sided test tests the null hypothesis H_0 that the parameter \(\theta_{hrf}\) equals zero against the alternative hypothesis H_1 that \(\theta_{hrf}\) is not zero (positive or negative). A one-sided test assumes \(\theta_{hrf}\) to be positive under H_1. The one-sided t-test decides H_1 when \(T_t > \alpha\), while the two-sided t-test decides H_1 when \(T_t < \alpha_a\) or \(T_t > \alpha_b\), with \(\alpha, \alpha_a\) and \(\alpha_b\) user specified thresholds. The F-test is a two-sided test and decides H_1 if \(T_F > |\alpha|\), with \(\alpha\) again a user specific threshold.\(^{78}\)

The performance of a test can be characterized by the false alarm rate (FAR or \(P_f\)) and the detection rate. The FAR is the probability that the test will decide H_1 while H_0 is true, denoted as Pr(H_1|H_0). This is also known as a Type-I error. When the statistical distribution of the test statistic under H_0 is known, it is possible to set (a) threshold(s) that correspond(s) to a specific constant false alarm rate (CFAR). The detection rate (\(P_d\)) is the probability that the test will decide H_1 when indeed H_1 is true, denoted as Pr(H_1|H_1). This is also called the power of the test. Furthermore, the probability that the test will decide H_0 while H_1 is true, denoted as Pr(H_0|H_1), is called a Type-II error.\(^{79}\) Hence, \(P_d = 1 - \text{Type-II error}\). The performance of a test can be used to compare several test with each other. For example, if two tests have the same FAR, the test with the highest detection rate is regarded as a better test.

\(^{76}\) Chatfield [1996], p.333
\(^{77}\) Chatfield [1996], p.334
\(^{78}\) Sijbers [2005], p.541
\(^{79}\) Chatfield [1996], p.159; Lukic [2002], p.73
Concluding remarks Chapter 2

In this chapter, we have modelled the measured data from fMRI as time series. These time series can modelled by a General Linear Model (GLM), as in Eq. (2.7):

\[ y = X\theta + e , \]

with \( y \) the measured data, \( X \) the ‘design matrix’, including the shape of the contributions of activation (\( x_{hrf} \)) and trend (\( x_{tr} \)), \( \theta \) the parameter vector related to these contributions and \( e \) the disturbing (white) noise. Noise comes from cardiac and breathing cycles, subject motion and from thermal fluctuations of electrons in the human body and receiver coils. We assume that the noise can be modelled by an autoregressive model, so our voxel-wise description of our measurement becomes Eq. (2.14), combined with Eq. (2.18):

\[ y_k = x_{k,hrf}\theta_{hrf} + x_{k,tr}\theta_{tr} - a_1y_{k-1} - \cdots - a_py_{k-p} + e_k , \]

with \( k \) the equidistantly distributed moments of observation and \( p \) the order of the autoregressive noise model. Using a model, we can estimate the model parameters. Using the GLM approach, this can be done by first estimating the autoregressive parameters, from which the covariance matrix of the noise is obtained, using the ordinary least squares method. With the estimated covariance matrix the parameter \( \theta \) can be obtained by the generalized least estimator of Eq. (2.29):

\[ \hat{\theta}_{hrf} = (X^T\hat{\Theta}^{-1}X)^{-1}X^T\hat{\Theta}^{-1}y . \]

When the parameters are estimated, we can test the activation-related parameter \( \theta_{hrf} \) using our formulated null hypothesis \( H_0 \) (no activation) and alternative hypothesis \( H_1 \) (activation), Eq. (2.30):

\[ \begin{cases} H_0: \theta_{hrf} = 0 \\ H_1: \theta_{hrf} \neq 0 . \end{cases} \]

Because due to the noise, \( \hat{\theta}_{hrf} \) is a stochastic variable, we use a test statistic to test \( \hat{\theta}_{hrf} \) for activity. For instance, we can use the Student-t test statistic as in Eq. (2.33):

\[ T_t = \frac{\hat{\theta}_{hrf}}{\hat{\Sigma}_{hrf}^{1/2}} . \]

If the value of the test statistic stays below a user specific threshold, we decide that no activation was present in the observations. However, if the value of the test statistic exceeds the threshold, we decide that activation was present in the signal. The performance of a test can be characterized by the false alarm rate (FAR or \( P_f \)) and the detection rate (\( P_d \)).

The false alarm rate (\( P_f \)) is the probability that the test will decide \( H_1 \) while \( H_0 \) is true, often denoted as \( \text{Pr}(H_1|H_0) \).

The detection rate (\( P_d \)) is the probability that the test will decide \( H_1 \) when indeed \( H_1 \) is true, often denoted as \( \text{Pr}(H_1|H_1) \).
Chapter 3

New mathematical approach:
Generalized Likelihood Ratio Test

After explaining the common mathematical approach to fMRI data, in this chapter we elaborate a new likelihood-based approach. We still use the model structure from Section 2.3.1, but our parameter estimation method and test statistic will be different.

3.1 Likelihood-based parameter estimation

In Section 2.3, we have explained the approach to fMRI data commonly used, resulting in the formulation of the t-test statistic. In this section, we will introduce a new approach: likelihood-based parameter estimation. Maximum likelihood (ML) estimation is based on the distribution of the observations. Having a model structure describing our observations, this section describes the approach to estimate with maximum likelihood the model parameters from the measured data.

3.1.1 Maximum likelihood estimation in general

The backgrounds and derivation of the maximum likelihood estimation method are as follows. Assume that we have done a number of measurements y(1), y(2), …, y(n). Because of the stochastic noise, the observations are random variables, having probability density functions p(y_i, β) with β the vector of model parameters. As in Section 2.2.1, we can write y = (y(1), y(2), …, y(n))^T, and denote the joint probability function of the observations as p(y, β). The joint probability density function of the set of observations defines the probability of occurrence of the set of observations. For independent observations, the joint PDF is given by multiplying the PDF’s of the individual observations:

\[ p(y, β) = \prod_{i=1}^{n} p(y_i, β) = p(y_1, β) \cdot p(y_2, β) \cdot \cdots \cdot p(y_n, β). \]  

(3.1)

In fact, in the joint PDF, the parameter β is taken fixed, while the observations are indeterminate variables. However, if we substitute our measurements in the joint PDF p(y, β) and regard the resulting function as a function of β, the resulting function is called the likelihood function, and will be denoted as l(y, β).\(^{80}\) So we consider l(y, β) as a function of β, instead of p(y, β) as a function of y. If l(y, β_1) > l(y, β_2) with β_1 and β_2 two possible

80. Nowak [2004], p.1
parameter vectors, we may say that $\beta_1$ is a more plausible or likely value of $\beta$, since $\beta_1$ gives a higher probability to the observations $y$ than $\beta_2$. The method of maximum likelihood estimation is based on the principle that we choose the value $\hat{\beta}$ as an estimator of $\beta$ which results in a maximum likelihood value, formally

$$\hat{\beta} = \arg \max_{\beta} \ell(y|\beta).$$

(3.2)

The maximum likelihood estimator has desirable properties. First, it becomes a minimum variance unbiased estimator as the number of observations increases. By unbiased, we mean that the expected value of the parameter estimator equals the true value of the parameter. By minimum variance, we mean that the estimator has the smallest variance with respect to other unbiased estimators. Asymptotically, the maximum likelihood estimator $\hat{\beta}$ converges to $\beta$ as

$$\hat{\beta} \sim N(\beta, F^{-1}(\beta)).$$

(3.3)

with $N(\mu, \sigma^2)$ denoting a normal distribution with mean $\mu$ and variance $\sigma^2$ and $F(\beta)$ the Fisher information matrix, evaluated at the true value $\beta$. The Fisher Information matrix is given by

$$F(\beta) = -\frac{\partial^2 \ln \ell(y, \beta)}{\partial \beta \partial \beta^\top}.$$  

(3.4)

The inverse of the Fisher information matrix is known as the Cramér and Rao (CR) lower bound. If $\hat{\beta}$ is an unbiased estimator of $\beta$, it follows from the CR inequality

$$\text{cov}(\hat{\beta}) \succeq F^{-1}(\beta)$$

(3.5)

that the diagonal elements of $F^{-1}(\beta)$ define the lower bound to the variances of the elements of $\hat{\beta}$.

In many cases it turns out that it is more convenient to use the (natural) logarithm of the likelihood function. This is also the case in our likelihood function as given in the next section. Since the logarithm is a monotonically increasing function, maximizing the logarithm of the likelihood function is equivalent to maximizing the likelihood function. We will use the natural logarithm in this thesis and write the so-called log-likelihood function as

$$L = L(y, \beta) = \ln(\ell(y, \beta)).$$

(3.6)

To compute the maximum likelihood estimates of the parameters, first the likelihood function as in Eq. (3.6) has to be constructed. We will consider the likelihood function in the next section.

81. Burnham [2004], p.14
82. Priestley [1994], p.308
83. Priestley [1994], p.306; Bos [1998], p.0210
84. Bos [1998], p. PS0210
3.1.2 Parameter estimation with likelihood function

Recall from Section 2.3.1 the General Linear Model (GLM) for fMRI data:

\[
y_k = x_{k,a} \theta_{a} + x_{k,e} \theta_{e} + e_k,
\]

with the assumption that the errors \( e_k \) are Gaussian distributed, and thus also the observations \( y_k \). Assume that our autoregressive noise model \( v \) to take account of coloured noise, as in Eq. (2.15), is of order \( p \). Then our total GLM becomes

\[
y_k = x_{k,a} \theta_{a} + x_{k,e} \theta_{e} - a_1 v_{(k-1)} - ... - a_p v_{(k-p)} + e_k,
\]

Using Bayes’ theorem, the joint probability density function (PDF) of the fMRI data can be derived from the PDF of the first \( p \) measurements and the conditional PDF of the remaining \( n-p \) measurements, given the first \( p \) measurements. The Bayes’ theorem says: \(^{85}\)

“If of two subsequent events the probability of the 1st be \( a/N \), and the probability of both together be \( P/N \), then the probability of the 2d on supposition the 1st happens is \( P/a \).”

or, mathematically

\[
Pr(b|a) = \frac{Pr(a\&b)}{Pr(a)} \text{ or } Pr(a\&b) = Pr(a) \cdot Pr(b|a),
\]

with \( Pr(\cdot) \) the probability that \( (\cdot) \) occurs. With respect to our joint probability density function with parameter vector \( \beta = (\theta^T, a^T, \sigma^2_e)^T \), this yields:

\[
p(y|\theta, a, \sigma^2_e) = p(y_1, ..., y_p|\theta, a, \sigma^2_e) \cdot p(y_{p+1}, ..., y_n|y_1, ..., y_p, \theta, a, \sigma^2_e),
\]

where both \( p(y_1, ..., y_p | \theta, a, \sigma^2_e) \) and \( p(y_{p+1}, ..., y_n | y_1, ..., y_p, \theta, a, \sigma^2_e) \) are multivariate Gaussian PDFs. For the first \( p \) measurements, the PDF becomes \(^{86}\)

\[
p(y_1, ..., y_p | \theta, a, \sigma^2_e) = \left( \frac{1}{2\pi \sigma^2_e} \right)^{\frac{p}{2}} |V_p|^{\frac{1}{2}} \exp \left( -\frac{1}{2\sigma^2_e} (y_p - X_{1p} \theta)^T V_p^{-1} (y_p - X_{1p} \theta) \right),
\]

while for the remaining \( n-p \) measurements, the conditional PDF, given the first \( p \) measurements, becomes

\[
p(y_{p+1}, ..., y_n | y_1, ..., y_p, \theta, a, \sigma^2_e) = \left( \frac{1}{2\pi \sigma^2_e} \right)^{\frac{n-p}{2}} \times \exp \left( -\frac{1}{2\sigma^2_e} \sum_{k=p+1}^{n} \left( y_k - x_k \theta + a_1 (y_{k-1} - x_{k-1} \theta) + ... + a_p (y_{k-p} - x_{k-p} \theta) \right)^2 \right),
\]

with \( y_k, x_k, \theta, a \) and \( \sigma^2_e \) as in Eq. (3.8) and \( V_p \) the \( p \times p \) covariance matrix of the parameters \( a \). The \( p \times p \) covariance matrix \( V_p \) is given by, analogous to Eq. (2.24):

85. Bayes [1763], p.5; Nowak [2004], p.2  
86. Priestley [1994], p.350, p.347
The PDF as in Eq. (3.11) corresponds to \(Pr(a)\) at the right hand side of Eq. (3.9), while Eq. (3.12) corresponds to \(Pr(b|a)\) at the right hand side of Eq. (3.9). The conditional PDF of Eq. (3.12) can be used to obtain an approximation to the likelihood function, by substituting all the observations \(y_1, ..., y_n\). To obtain an exact likelihood function, we use the total joint probability density function in Eq. (3.10), given as

\[
p(y; \theta, a, \sigma_e^2) = \left( \frac{1}{2\pi\sigma_e^2} \right)^\frac{n}{2} |V_p|^{-\frac{1}{2}} \exp \left( -\frac{Q_1(y; \theta, a)}{2\sigma_e^2} \right),
\]

with

\[
Q_1(y; \theta, a) = \sum_{i=1}^{p} \left[ V_p^{-1} \right]_{ii} \sum_{j=1}^{n} \left[ (y_j - x_j \theta)^2 + Q_2(y; \theta, a) \right],
\]

\[
Q_2(y; \theta, a) = \sum_{k=p+1}^{n} \left[ y_k - x_k \theta + a (y_k - x_k \theta) + ... + a (y_k - x_k \theta) \right]^2,
\]

with \([V_p^{-1}]_{ij}\) the \((i,j)^{th}\) element of the inverse covariance matrix \(V_p\). From this joint probability density function, the likelihood function can be obtained, as explained above, by regarding the observations as given and the parameters \(\theta\), \(a\) and \(\sigma_e^2\) as variables. To get the maximum likelihood values of \(\theta\), \(a\) and \(\sigma_e^2\), the likelihood function has to be maximized. Priestley gives an expression for the noise variance \(\sigma_e^2\) at the maximum of the exact likelihood function:

\[
\sigma_e^2 = \frac{1}{n} Q_1(y; \theta, a).
\]

By this expression, \(\sigma_e^2\) can be eliminated from Eq. (3.14), yielding the so-called concentrated likelihood function

\[
L(y; \theta, a) = \left( \frac{n}{2\pi Q_1} \right)^{\frac{n}{2}} |V_p|^{-\frac{1}{2}} \exp \left( -\frac{Q_1(y; \theta, a)}{2\sigma_e^2} \right).
\]

It follows from Eq. (3.17) that the natural logarithm of the concentrated likelihood function is given by:

\[
L(y; \theta, a) = \frac{n}{2} \ln \left( \frac{n}{2\pi Q_1} \right) - \frac{1}{2} \ln |V_p| - \frac{n}{2}.
\]

As we recall from Eqs. (2.24) and (2.26), \(V_p\) depends on the first \(p-1\) parameters of \(a\). Substituting the measured observations in the expression (3.18), the resulting function is only related to the unknown parameters \(\theta\) and \(a\). This function can be used to estimate the parameters by finding the maximum likelihood value with respect to these parameters by
(non-linear) maximizing procedures. Using this expression to estimate the parameters, all the parameters are estimated simultaneously. Hence, this method has two major advantages by comparison with the common parameter estimation method as described in Section 2.3.5. First, we use the exact likelihood function, which describes the statistics of the data, including the temporal correlation structure of the noise. Using the exact likelihood functions, no approximations are made. Second, it is no longer necessary to estimate the parameters \( \theta \) and \( a \) separately and no prewhitening step (in OLS sense) is needed. \(^8\)

### 3.2 Generalized Likelihood Ratio Test for activation detection

In our new likelihood approach, we still use the generalized linear model (GLM) to model our measured data. As explained in Section 2.3.6, we have to test our null hypothesis \( H_0 \) against our alternative hypothesis \( H_1 \) if we want to decide whether there is activation present in our voxel-wise measured data:

\[
\begin{cases}
H_0: \theta_{hrf} = 0 \\
H_1: \theta_{hrf} \neq 0 .
\end{cases}
\]

When the probability density functions (PDFs) of our test statistic under \( H_0 \) and \( H_1 \) are unknown, we use a composite hypothesis test, which accommodates unknown parameters. \(^9\)

The Generalized Likelihood Ratio Test (GLRT) has the advantage that the PDFs under \( H_0 \) or \( H_1 \) are allowed to be not completely known. Also no prior knowledge about the parameters is necessary, as it is for the also well known Bayesian approach.

The Generalized Likelihood Ratio Test uses the likelihood functions as derived in Section 3.1 under the assumptions \( H_0 \) (no activation) or \( H_1 \) (activation) to test the data on activation. This chapter discusses the how we can test our likelihood function for activation.

#### 3.2.1 Hypothesis test and test statistic

As its name implies, the Generalized Likelihood Ratio Test (GLRT) investigates the ratio of likelihood values under both hypotheses (\( \theta_{hrf} = 0 \) and \( \theta_{hrf} \neq 0 \)). For testing activation, the GLRT replaces the unknown parameters of interest by their maximum likelihood estimators (MLEs). The Generalized Likelihood Ratio is given by

\[
\lambda(y) = \frac{\sup_{\theta, a, \sigma_0^2} l(\theta, a, \sigma_0^2; y)}{\sup_{\theta_1, \ldots, \theta_{12}, \theta_{21}, \ldots, \theta_{a}; \sigma_0^2} l(\theta_1, \ldots, \theta_{12}, \theta_{21}, \ldots, \theta_{a}; \sigma_0^2; y)} .
\]

with in our case \( \theta_i \) represents \( \theta_{hrf} \) and where the nominator denotes the maximum value of the likelihood function under the assumption that \( H_1 \) is true and the denominator denotes the maximum value of the likelihood function under the assumption that \( H_0 \) is true. \(^9\)

As one can see, in the denominator, \( \theta_i \) is set to zero, which corresponds in our case to the assumption

---

89. Sijbers [2005], p. 546
90. Kay [1998], p.191
91. Kay [1998], p.200
\( \theta_{\text{act}} = 0 \). With Eq. (3.6), we can reformulate Eq. (3.20) as
\[
\lambda_{\text{th}}(y) = \frac{L_{\text{max}|H_0}(y; \theta, a, \omega^2)}{L_{\text{max}|H_1}(y; \theta, a, \omega^2)} = \frac{L_{\text{max}|H_0}}{L_{\text{max}|H_1}},
\]
(3.21)
with \( L \) the log-likelihood function as in Section 3.1.2. The GLRT rejects the null hypothesis if \( \lambda(y) > \lambda_0 \), where we have to specify the threshold \( \lambda_0 \). If we modify our test statistic (3.20) to
\[
T_{LR} = 2 \ln(\lambda(y)),
\]
(3.22)
or, alternatively with the use of (3.21) to
\[
T_{LR} = 2(L_{\text{max}|H_1} - L_{\text{max}|H_0}),
\]
(3.23)
we know that its distribution is asymptotically \( \chi^2 \) with 1 degree of freedom (d.o.f.) when \( H_0 \) is true.\(^9\) We will write our new threshold \( 2\ln(\lambda_0) \) as \( \alpha_0 \). Given the distribution of \( T_{LR} \) under \( H_0 \), a false alarm rate \( P_f \) can be set. For instance, when we accept a false alarm rate of 5%, our threshold \( \alpha_0 \) becomes 3.84 (see Figure 3.1).

As mentioned above, for the GLRT, the maximum values of the likelihood function under \( H_0 \) and \( H_1 \) are needed. To find these maximum values, we have to use non-linear optimization procedures, which exist in the Matlab optimization toolbox. Our likelihood-based hypothesis test becomes
\[
\begin{cases}
H_0 \text{ if } T_{LR} \leq \alpha_0 \\
H_1 \text{ if } T_{LR} > \alpha_0
\end{cases}
\]
(3.24)

\(^{92}\) Fan [2001], p.156
\(^{93}\) Fan [2001], p.160; Kay [1998], p.206
3.2.2 The false alarm rate and the detection rate

As explained in Section 2.3.6, the performance of a test is related to its false alarm rate and its detection rate. The false alarm rate $P_f$ is defined as the probability that the test will decide $H_1$ when $H_0$ is true, while the detection rate $P_d$ is the probability that the test will decide $H_1$ when indeed $H_1$ is true. In practice, these rates can only be obtained if it is known whether there is activation or not. In simulation experiments, however, these rates can be evaluated exactly. To compare different methods, the values of the $P_f$ and $P_d$ have to be assessed for all methods.

We decide to activation if the test statistic exceeds a certain threshold $\alpha_0$. As mentioned in the previous section, we can set the false alarm rate using the knowledge about the asymptotic distribution of the test statistic under $H_0$, which is a $\chi^2$ distribution with 1 d.o.f (Figure 3.1). If we want to account for a finite number of measurements, we can set our false alarm rate in the following way. Suppose we have generated $n$ simulated data sets with data without activation. For each of these data sets we can compute the GLRT statistic. If we sort all these $n$ test statistics ascending, for instance the value of the $0.05n$th test statistic can be taken as the threshold $\alpha_c$ that results in $P_f = 0.05$. Of course this threshold value can vary from the $(0.05n-1)$th to the $(0.05n+1)$th statistic, all resulting in $P_f = 0.05$, but it turns out that for large $n$ this varying choice of the threshold gives only low variation in the detection rate. For a given $P_f$ and corresponding threshold, the detection rate can be computed if we generate data with activation. The detection rate $P_d$ is the ratio between the number of test statistics exceeding the threshold and the total number of data sets considered. Generally, we can not say anything about the detection rate a priori, unless we know the distribution of the test statistic under $H_1$. Note that if we have data without activation, the detection rate is equal to the false alarm rate.

3.3 Noise model order selection criteria

In the analysis of the likelihood value described in the previous section, it is assumed that the order $p$ of the autoregressive noise model is known. In general, this order is not known, but has to be assumed or estimated. Choosing a certain model order is a trade-off between model bias and model variance. A model order which is too low will result in a low variance in the parameters, but a high bias, while a too high model order will result in a lower bias, but a higher variance. In literature, for fMRI data, often a model structure with fixed order $p = 1$ (AR(1)) is chosen, but also higher orders are found. Instead of selecting a model order a priori, it is also possible to select the model order from the data, using an order selection criterion (OSC). In this section, we briefly discuss three different order selection criteria: the Akaike Information Criterion (AIC)\(^{94}\), a criterion based on Kullback’s symmetric divergence (KICc)\(^{95}\) and the Minimum Description Length (MDL)\(^{96}\). All these criteria use the Maximum Likelihood Estimator (MLE) in the decision of the model order. All criteria take account of the fact that the likelihood value for the data will increase with an increasing number of model parameters (for nested models), by incorporating penalty factors. All analyses using the order selection criteria are done under the assumption that activation was present ($H_1$).

---

94. Akaike [1974]
95. Seghouane [2004], p.3316
96. Rissanen [1978]; Rissanen [2005]
3.3.1 Akaike’s Information Criterion

Suppose a collection of data \( y = (y_1, ..., y_n)^T \) has been generated according to an unknown probability density function (PDF) \( p(y, \beta_0) \). We try to find a parametric model that provides a suitable approximation for \( p(y, \beta_0) \). Let \( \mathcal{M}_k = \{ p(y, \beta_k) \mid \beta_k \in \mathbb{B}_k \} \) denote a parametric family, where \( \beta_k \) is a \( k \) dimensional vector with the model parameters.\(^97\) Let \( \hat{\beta}_k \) denote the vector of estimated parameters, obtained by maximizing the likelihood function \( l(y, \beta_k) \) over \( \mathbb{B}_k \), and let \( p(y, \hat{\beta}_k) \) denote the corresponding fitted model. The collection \( \mathcal{M}_k \) with \( k = 1, 2, ..., k_{\text{max}} \) consists of nested families, so \( \mathbb{B}_1 \subset \mathbb{B}_2 \subset ... \subset \mathbb{B}_{k_{\text{max}}} \). The discrepancy between the generating unknown model \( p(y, \beta_0) \) and the approximate model \( p(y, \beta_k) \) can be expressed in the Kullback-Leibler (KL) information function or discrepancy. The KL information function can be regarded as showing the loss of information. The KL directed divergence is given by\(^98\)

\[
I(\beta_0; \hat{\beta}_k) = E_{\beta_0} \{ \ln(p(y, \beta_0)) \} - E_{\hat{\beta}_k} \{ \ln(p(y, \hat{\beta}_k)) \} \bigg| \hat{\beta}_k, \beta_0,
\]  

(3.25)

with \( p(y, \beta) \) the probability density function of the data \( y \) with parameters \( \beta \), \( \beta_0 \) the true data generating parameter vector, \( \hat{\beta}_k \) the estimated model parameter vector and \( E_{\beta_0} \{ \cdot \} \) the expectation value of \( \{ \cdot \} \) with respect to \( p(y, \beta_0) \). The KL discrepancy equals zero if the models are the same, that is to say, if the estimated model is equal to the true model. For the purpose of order selection, only a comparison of the values of models with different \( \hat{\beta}_k \) has to be made, thus the common true model term in Eq. (3.25) can be discarded. The other term, related to the estimated model, is called ‘the relative KL information’.\(^99\)

\[
I(\beta_0; \hat{\beta}_k) = -E_{\hat{\beta}_k} \{ \ln(p(y, \hat{\beta}_k)) \} \bigg| \hat{\beta}_k, \beta_0.
\]  

(3.26)

Akaike argues that using the maximum likelihood estimator of \( \beta_k \), the relative Kullback-Leibler information of Eq. (3.26) can be estimated in an asymptotically unbiased manner as

\[
\hat{I}_r = -\ln(p(y; \hat{\beta}_{k,\text{ML}})) + k.
\]  

(3.27)

with \( \hat{\beta}_{k,\text{ML}} \) the maximum likelihood estimator (MLE) of \( \beta_k \).\(^100\) Therefore, rewriting the first term with the arguments as in Section 3.1.2, Akaike’s Information Criterion (AIC) becomes

\[
\text{AIC} = -2L(y; \hat{\beta}_k) + 2k,
\]  

(3.28)

where the factor 2 comes from historical reasons.\(^101\) The AIC has to be minimized for identification or order selection purposes. The \( \hat{\beta}_k \) which minimizes the criterion is regarded as the best trade off between model fit and model complexity. From Eq. (3.28) it can be seen that when the maximum likelihood value of two models is the same, the parameter vector with the smallest number of parameters is chosen. The factor 2 before the number of parameters \( k \) is known as the penalty factor. As we will see, other criteria use other penalty factors, giving more or less weight to the number of parameters.

We can make another small remark with respect to the ‘A’ in the criterion. Akaike himself
denotes: ‘IC stands for information criterion and A is added so that similar statistics, BIC, DIC etc. may follow.’ But after him, other criteria as BIC and KIC are named to their postulaters (Bayes, Kullback) and thus also AIC is named ‘Akaike Information Criterion’. Although Akaike adds the factor $k$ to his criterion to correct for bias in the estimated distance between real system and estimated model, the AIC is only asymptotically unbiased, while Akaike assumes that the number of observations $n$ tends to infinity. Therefore some variants have been constructed. One of them is the corrected AIC (AICc), posited by Hurvich:

$$AICc = -2L(y; \hat{\beta}_n) + \frac{2nk}{n-k-1},$$  \hspace{1cm} (3.29)

with $n$ the length of the time series and $k$ the total number of parameters to be estimated. AICc is found to provide a better model order selection than AIC for both linear regression and autoregressive models. As one can see, for $n \rightarrow \infty$, Eqs. (3.28) and (3.29) are equivalent.

Another adaption to AIC is made by Broersen. Considering the risks of under and over estimation of the model order, he proposed a Finite sample Information Criterion (FIC) with a higher penalty factor than AIC, replacing the factor 2 by 3:

$$FIC = -2L(y; \hat{\beta}_n) + 3k.$$  \hspace{1cm} (3.30)

Broersen defined the number of samples as finite if the order of the parameter vector $\beta_k$ to be estimated is greater than $0.1n$. The same expression as information criterion is also proposed by Cavanaugh for large sample time series.

### 3.3.2 Corrected Kullback Information Criterion

The Akaike Information Criterion is based on Kullback’s directed divergence between two models. The corrected Kullback Information Criterion proposed by Seghouane is based on Kullback’s symmetric divergence, which is defined as

$$I_{\Delta}(p_{\hat{\beta}_n}, p_{\hat{\beta}_0}) = I(p_{\hat{\beta}_n}, p_{\hat{\beta}_0}) + I(p_{\hat{\beta}_0}, p_{\hat{\beta}_n}).$$  \hspace{1cm} (3.31)

Using Eq. (3.25), the Kullback’s symmetric divergence (3.31) can be written as

$$E_{\hat{\beta}_n}\{\ln(p(y, \hat{\beta}_n))\} - E_{\hat{\beta}_n}\{\ln(p(y, \hat{\beta}_0))\} + E_{\hat{\beta}_0}\{\ln(p(y, \hat{\beta}_n))\} - E_{\hat{\beta}_0}\{\ln(p(y, \hat{\beta}_0))\}. \hspace{1cm} (3.32)$$

It can be shown that the symmetric divergence performs better than a directed divergence in a setting where both underfitted and overfitted models are used to estimate the order of...
the data describing model.\textsuperscript{111} Seghouane proposes the KICc as an exactly unbiased estimator of $L(\hat{\beta}_0, \hat{\beta}_k)$, assuming that the true system is present in the model set. He defines the KICc as

$$\text{KICc} = -2L(y; \hat{p}_k) + 2\frac{kn}{n-k-1} - n\psi\left(\frac{n-k-1}{2}\right) + n \ln \frac{n}{2},$$

(3.33)

where $\psi(\cdot)$ denotes the Psi-function or digamma-function, defined as

$$\psi(x) = \int_0^x \left(\frac{e^t - e^{-x}}{t - 1 - e^{-x}}\right) dt.$$  \hspace{1cm} (3.34)

The KICc is unbiased for overfitted or exactly specified models, but for underfitted models, the KICc is biased with a bias that depends on the amount of underfit. If an approximation for the Psi-function is made, the KICc becomes\textsuperscript{112}

$$\text{AKICc} = -2L(y; \hat{p}_k) + \frac{k(3n-k-1)}{n-k-1} + \frac{k-1}{n-k-1}. $$

(3.35)

As one can see, AKICc depends on the number of parameters and the length of the time series. Asymptotically, AKICc tends to AIC, just as AICc, but for a finite number of observations, it behaves different from AIC and AICc. A plot of the different behaviours is given in the Figures 3.2 to 3.5.

### 3.3.3 Minimum Description Length

Starting with AIC, Rissanen is in doubt with respect to the origin of Akaike’s penalty factor and the need for the linear term for the number of parameters.\textsuperscript{113} Rissanens method searches for the shortest description of the joint distribution of the noise. This is the case when the marginal distributions of $e_k$ in Eq. (3.8) are independent. Without going deeper into detail, we give the MDL criterion for autoregressive processes:\textsuperscript{114}

$$\text{MDL} = -2L(y; \hat{p}) + \sum_{i=1}^p \ln \left(\frac{\partial^2 L(y; \hat{p})}{\partial \hat{\theta}_i^2}\right) + (p+1)\ln(n+2) + 2\ln(n+1),$$

(3.36)

with $p$ the order of the autoregressive noise and $n$ the number of observations. For computational reasons, we ignore the second term in Eq. 3.36. Also the last term can be dropped, according to Rissanen.\textsuperscript{115} If we extend the MDL to our model, including the trend parameters, our approximate MDL criterion becomes\textsuperscript{116}

\textsuperscript{111} Seghouane [2004], p.3317
\textsuperscript{112} Seghouane [2004], p.3317
\textsuperscript{113} Rissanen [1978], pp.465, 468
\textsuperscript{114} Rissanen [1978], p.471
\textsuperscript{115} Rissanen [1978], p.471
\textsuperscript{116} Kay [1998], p.225
\[ \text{AMDL} = -2L(y; \hat{\theta}) + kn(n + 2), \] (3.37)

with \( k \) the total number of parameters (including the \( p \) autoregressive parameters) and \( n \) the number of our observations. In fact, in MDL the penalty factor contains the estimated parameters themselves, but these parameters are absent in the AMDL. However, the wanted strong penalization of over parameterization for larger time series is still present, as can be seen in Figures 3.2 to 3.5. As one can see from these figures, AICc and AKICc tend to be the same as AIC and FIC, respectively, for an increasing number of observations.
Figure 3.2 | Penalty factors of different order selection criteria for time series of 20 observations for an increasing number of parameters $k$.

Figure 3.3 | Penalty factors of different order selection criteria for time series of 50 observations for an increasing number of parameters $k$. 
Figure 3.4 | Penalty factors of different order selection criteria for time series of 100 observations for an increasing number of parameters $k$.

Figure 3.5 | Penalty factors of different order selection criteria for time series of 150 observations for an increasing number of parameters $k$. 
Concluding remarks Chapter 3

In this chapter, starting from our fMRI time series model as in Eq. (3.8),
\[ y_k = x_{k,\text{arf}}^\theta + x_{k,\text{hrf}}^\theta + \cdots + \theta_{k-p} v_{k-p} + \varepsilon_k, \]
we have introduced in Section 3.1 a new mathematical approach to deal with fMRI signals. This new approach is based on the likelihood function of the measured data and the model parameters. The log-likelihood function is given in Eq. (3.18):

\[ L(y; \theta, \sigma) = \frac{n}{2} \ln \left( \frac{n}{2\pi Q_1} \right) - \frac{1}{2} \ln |V_p| - \frac{n}{2}, \]

with \( n \) the number of observations, \( V_p \) the \( p \times p \) covariance matrix of the AR(p) noise model and \( Q_1 \) given in Eq. (3.15).

If we have defined our model and all the parameters are estimated by maximizing the log-likelihood function, we are able to test our null hypothesis (no activation) against our alternative hypothesis (activation). Because the estimator of the activation-related parameter \( \theta_{\text{hrf}} \) can be regarded as a stochastic variable with an unknown distribution under \( H_1 \), we use the Generalized Likelihood Ratio test statistic to test our hypothesis. The Generalized Likelihood Ratio Test (GLRT) has the advantage that the PDFs under \( H_0 \) or \( H_1 \) are allowed to be not completely known. Also no prior knowledge about the parameters is necessary. The GLR test statistic is given by Eq. (3.23):

\[ T_{\text{LR}} = 2(L_{\text{max}}|H_1) - L_{\text{max}}|H_0, \]

with \( L_{\text{max}}|H_1 \) the maximum log-likelihood value under \( H_1 \). Using this test statistic and an allowed false alarm rate, the detecting rate of the test can be obtained (see Section 3.2.2).

Because the order of the noise model (and therefore the order of the whole model) is generally unknown, we briefly discussed some order selection criteria in Section 3.3, namely three criteria based on Akaike's information criterion (Section 3.3.1), one criterion based on the symmetric Kullback's information criterion (Section 3.3.2) and one based on the Minimum Description Length (Section 3.3.3). These five criteria have different penalty factors for different numbers of model parameters and numbers of observations.
Chapter 4

Likelihood-based detection in simulation

In this chapter, we discuss the method to apply the theory from Chapter 3 in simulations, using the computer program Matlab.

4.1 Creation of simulated data

As explained in Section 3.2.2, simulated data is needed to assess the false alarm rate and detection rate of a test. To evaluate the performance of the Generalized Likelihood Ratio Test (GLRT) as derived in Section 3.2, we chose our contributions to the simulation data based on fMRI literature. All our simulations are based on the three contributions given in Section 2.3. These three contributions are: (1) the haemodynamic response, (2) trends and (3) noise. In our study, we focus on the influence of different noise contributions to the detection of activation, so we assume fixed models for both activation and trend contributions. The length of our fMRI time series used is 100 observations, equidistantly distributed in time. We use the symbol \( k \) as a discrete time unit, with \( 1 \leq k \leq 100 \) for time series with 100 measurements. During the simulation of one fMRI time series, the model structure is assumed to be fixed.

4.1.1 Haemodynamic response function

The haemodynamic response function (HRF) represents the blood oxygen level response to a stimulus (see Section 2.2.2). The HRF convolves with the stimulus function to form the activation contribution. Although several models for the HRF are proposed, we chose the widely used function given by Worsley (Figure 2.16).\(^\text{117}\) This function is given by Eq. (2.10). The sampled HRF used is given in Figure 4.1. During our 100 measurements, we apply three times a box-car stimulus at \( k = 10, k = 30 \) and \( k = 50 \) with signal amplitude 0.5 and length of 10 time-steps (see Figure 4.2). The result of the convolution with our HRF in time domain is given in Figure 4.3. The estimated spectrum of our activation signal is given in Figure 4.4. After the construction of the shape of the activation signal, the level of brain activation can be modelled by multiplying the activation signal by the parameter \( \theta_{hrf} \). This method corresponds to an external stimulus with constant level with a varying response-level in a voxel. In our analysis, the parameter \( \theta_{hrf} \) is chosen varying in 10 steps between 0 and 1, thus we have 10 levels of activation.

\(^{117}\) Worsley [2002], p.3
Section 4.1  Creation of simulated data

Figure 4.1 | The analytical HRF (curve) given by Worsley [2002] and the used sampled values (dots).

Figure 4.2 | The applied box-car on-off stimuli used in the simulation experiments. The measured activation signal is a convolution of these stimuli with the HRF of Figure 4.1.
Figure 4.3 | The fMRI activation signal of 100 observations as a convolution of the HRF with a certain stimulus.

Figure 4.4 | Estimated spectrum of the fMRI activation signal. The spectrum is obtained using the Fast Fourier Transform of one period of the activation signal extended with 1,000 zeros.
4.1.2 Trend model

Due to several reasons (as explained is Section 2.3.3) we have to account for trends in the fMRI data. As explained in Section 2.3.3, in literature these trends are modelled by, e.g., polynomials or sines and cosines. In our simulations, we have chosen a second order polynomial:

$$x_k \theta_{hrf} = 0.1 - 0.01k + 10^{-5}k^2.$$  \hspace{1cm} (4.1)

If the fMRI data were not disturbed by noise, the measured data would always be the same. With an activation level of 1 ($\theta_{hrf} = 1$), the fMRI observations would be as in Figure 4.5, with the associated spectrum given in Figure 4.6. The estimated spectrum is obtained using the Matlab toolbox ARMASA.\(^{118}\)

\(^{118}\) ARMASA, developed by P.M.T. Broersen. Available from MathWorks, MATLAB Central > File Exchange > Signal Processing > Spectral Analysis > ARMASA.

http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=1330&objectType=file
The sampled values of the activation signal (see Figure 4.3) form the first column of the design matrix $X$ in Eq. (2.7), while the polynomial trend model (see Eq. (4.1)) fills the next three columns. The design matrix $X$ times the trend vector $\theta$ forms the deterministic part of the fMRI data. Further, the fMRI data is disturbed by stochastic, autoregressively modelled, noise, which is described in the next subsection.

### 4.1.3 Autoregressive noise

To construct autoregressive noise with given parameters, the Matlab toolbox ARMASA is used. With the function `gendata.m`, a realisation of an autoregressive and moving average noise can be produced with an variance $\sigma^2 = 1$. We use this function, assuming that our noise can be modelled by the autoregressive noise model

$$v_k = -a_1 v_{k-1} - a_2 v_{k-2} - \ldots - a_p v_{k-p} + e_k,$$  \hspace{1cm} (4.2)

with $a_1, \ldots, a_p$ the AR-parameters and $e_k$ Gaussian white noise. During our simulations we varied the autoregressive model structure and the parameters. We examined fifteen noise models with varying order from zero (equal to white noise) to three. Worsley\textsuperscript{119} gives some empirically obtained boundaries for the autoregressive parameters, so we can use some realistic noise models. Note that Worsley uses the opposite noise model definition, described by Eq. (2.17) instead of (2.16). The parameter vectors used are given in Table 4.1.

<table>
<thead>
<tr>
<th>signal #</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>- 0.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>- 0.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>- 0.4</td>
<td>- 0.3</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>- 0.4</td>
<td>- 0.1</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>- 0.4</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>- 0.4</td>
<td>0.3</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>0.4</td>
<td>- 0.3</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>0.4</td>
<td>- 0.1</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>0.4</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>0.4</td>
<td>0.3</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>- 0.4</td>
<td>- 0.3</td>
<td>- 0.1</td>
</tr>
<tr>
<td>14</td>
<td>0.4</td>
<td>0.3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

\textsuperscript{119} Worsley [2002], p.6
In the Figures 4.7 to 4.10, two realisations are given of AR(1) noise with parameters $a_1 = 0.4$ and $a_1 = -0.4$. The upper two graphics (4.7 and 4.8) give the realisations in the time domain for 50 observations, while the lower two (4.9 and 4.10) gives their spectrum. All other spectra are given in Appendix A. The combination of the design matrix $X$, deterministic parameters $\theta$ and the noise $v$ gives the simulated fMRI data set $y$.

4.2 Model estimation and activation detection

In the previous section, we have described the creation of our data. Now, using the likelihood-based approach of Chapter 3, we will describe the estimation of the parameters of the models used. Because the computation of the likelihood value is very time-intensive, we gave attention to the time needed to compute the different elements in the likelihood function. Several functions of the ARMA toolbox are optimized for our specific computation. In this section, these functions are marked with a star (*). All computations are made using Matlab version 6, release 12.120

### 4.2.1 Maximum likelihood estimation

Given an fMRI data set of observations, the design matrix and the orders of the trend and the autoregressive noise model, the maximum likelihood estimator (MLE) of the corresponding parameters can be found by the Matlab function `findmllhpars`.

This function uses a non-linear Matlab optimization function to maximize the exact likelihood function of Eq. (3.18). First, the function `findmllhpars` computes the ordinary least squares (OLS) parameters of the haemodynamic response function and the trend and the autoregressive parameters from the residuals. This is done, to get a start position of the non-linear optimization near the expected maximum. The OLS parameters can be computed according to Eq. (2.19) by:

$$
\mathbf{t}_{\text{OLT}} = \hat{\theta}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.
$$

(4.3)

With this start vector, we can construct the following residual signal:

$$
\mathbf{v}_{\text{AR}} = \mathbf{y} - \mathbf{X}\mathbf{t}_{\text{OLT}}.
$$

(4.4)

From the toolbox ARMASA, we use the function `sig2ar` to obtain the set of $p$ AR-parameters from the residual signal (4.4), with $p$ the AR-order under investigation. Note that an autoregressive signal can be represented by its AR-parameters, but also by reflection coefficients (RCs). For stationary autoregressive signals, these reflection coefficients are smaller than 1. We use this property to keep the optimization process stable. Therefore, we transform the AR-parameters to RC-values. The start values $\mathbf{t}_{\text{start}}$ from Eq. (4.3) and the reflection coefficients from the residual signal are used to maximize the log-likelihood function, given by Eq (3.18):

$$
I(\mathbf{y}; \hat{\theta}, \mathbf{a}) = \frac{n}{2} \ln \left( \frac{n}{2\pi \mathbf{Q}} \right) - \frac{1}{2} \ln |\mathbf{V}_{\text{p}}| - \frac{n}{2}.
$$

(4.5)

with $n$ the number of observations, $\mathbf{V}_{\text{p}}$ the covariance matrix and $\mathbf{Q}$, a function of $\mathbf{y}$, $\hat{\theta}$ and $\mathbf{a}$. With these start values $\mathbf{t}_{\text{start}}$, we use the Matlab minimizing algorithm `fminsearch` from the optimization toolbox to minimize the self written function `mllh_ar`, which compute the log-likelihood value based on the reflection coefficients as input.

In the function `mllh_ar`, the reflection coefficients are converted back to autoregressive parameters by the function `fastrc2ar`. From the autoregressive parameters, the correlation matrix Eq. (3.13) is constructed:

$$
\mathbf{V}_{\text{p}} = G \begin{pmatrix}
\rho(0) & \rho(1) & \cdots & \rho(p-1) \\
\rho(1) & \rho(0) & \cdots & \rho(p-2) \\
\vdots & \vdots & \ddots & \vdots \\
\rho(p-1) & \rho(p-2) & \cdots & \rho(0)
\end{pmatrix}.
$$

(4.6)

with $G$ the power gain of the noise, obtained from the reflection coefficients by the function `rcgain`. As explained in Section 2.3, the correlation coefficients can be obtained by the Yule-Walker equations:

---

121. Broersen [2000], p.3551
Section 4.2 Model estimation and activation detection

Although the ARMASA toolbox contains a function to obtain the correlation coefficients from the autoregressive parameters (arma2cor), we programmed an alternative function ar2cor*. The function ar2cor computes the correlation coefficients from the Yule-Walker equations, using an algorithm from Linear Algebra, obtained by rewriting the Yule-Walker relations in Echelon form matrix. For AR(1) signals, our function ar2cor* is about 200 times faster than arma2cor.m and for AR(3) signals, our function computes the correlation coefficients 100 times faster than the ARMASA function. However, our function is only able to handle autoregressive signal of maximum order 5 due to specific computations for each order.

From the measured data, the design matrix and the autoregressive parameters under study (with the constructed correlation matrix), the (logarithm of the) likelihood function can be computed using Eqs. (3.15) to (3.18).

Because later on we want to test the null hypothesis $H_0$ (no activation) against the alternative hypothesis $H_1$ (activation), the likelihood value is computed twice, once under the assumption of activation (the activation parameter is estimated) and once under the assumption of no activation (the activation parameter is not estimated, but assumed to be zero). This results in two different MLEs per set of observations. Note that the maximum likelihood value under $H_0$ always should be smaller than under $H_1$, because in the latter an extra parameter can be used to estimate the signal.

4.2.2 Order selection

As mentioned at the beginning of the previous section, the orders of the trend and noise models have to be known to compute the likelihood value. That can be seen for example in Eqs. (4.5) and (4.6). For our analysis, we assume the order of the trend to be known (actually second order), but the order of the noise has to be estimated. Because in literature sometimes the order of the noise is assumed fixed at a certain value, we will do the same, comparing the results of the assumptions of order zero (AR(0)) to order five (AR(5)) to each other, resulting in six estimated models. But in addition, we also use order selection methods to get the order of the noise. We use five (slightly) different order selection methods, three based on the Akaike Information Criterion, one based on the symmetric Kullback-Leibler Information Criterion and one based on Minimum Description Length, as described in Section 3.3. All these order selection methods use the likelihood values as obtained from the fixed order estimation, and add a method specific penalty factor.

From the eleven methods we get eleven sets of likelihood values (both under $H_0$ and $H_1$). These can be used for the computation of the test statistics.

4.2.3 Test statistics, hypothesis test and detection

From the, per method, two selected likelihood values under the null hypothesis and the alternative hypothesis, the value of the test statistic can be computed. Therefore, we use the formula of Eq. (3.23):

$$ T_{LR} = 2(L_{\text{max}}|H_1 - L_{\text{max}}|H_0) $$

with $L_{\text{max}}|H_1$ and $L_{\text{max}}|H_0$ the respective log-likelihood values. Each method of noise modelling
results in a different test statistic value. These values are used to test our hypothesis. As explained in Section 3.2.1, the value should be larger than a certain threshold value $\alpha_0$ to decide activation. This threshold can be found by using the knowledge of the theoretical asymptotically valid $\chi^2$-distribution of $T_{LR}$ under $H_0$. In this study, we have set our false alarm rate to 5%, so from the $\chi^2$ distribution, a threshold for our test statistic of 3.841 is taken. For a false alarm rate of 10% or 1%, thresholds of 2.706 and 6.635 should be taken respectively. We can set our threshold $\alpha$ using the command `chi2inv`.

Now we ‘detect’ an activation in our measured signal when the test statistic exceeds this threshold. Since we have eleven test statistics for each measured time series, we can compare the performances of these methods, because in simulation, we know whether an activation was present in the signal.

### 4.2.4 Multiple runs

Because our signals are seriously disturbed by stationary stochastic noise, we have to compare the results of our methods in a statistical way. Therefore, we repeat the whole procedure described above a 1000 times. On an AMD Athlon 1600+ (1.4 Ghz) computer, a session of 1000 runs takes about an hour with the Matlab functions used. It turns out that compiling the function using a program as C++ will reduce the computation time up to forty times.

The 1000 runs result in the same number of test statistic values per order selection method, which are all compared to the threshold value. From these 1000 decisions, a detection rate can be calculated, dividing the number of threshold-exceeding values through the total number of runs (i.e. 1000).

The value of detection rate is obtained for all 11 order selection methods, and also for all 10 levels of activation and 14 types of originating noise (see Table 4.1). Using the command `bar`, we can visualize the value of the detection rate for the different methods, activation levels and generating noise processes. This is done using the written Matlab m-file `getpdpicts`.

Because the distribution of the test statistic is only asymptotically $\chi^2$, the false alarm rate during the simulation might not be exactly the specified false alarm rate. Remember from Section 3.2.2 that it is also possible to define the threshold $\alpha_c$ in such a way that the false alarm rate satisfies the given value. When the threshold $\alpha_c$ is defined, the computation of the detection rates is the same as above.

This section has described the method to estimate a model for the measured data and to detect activity. A statistical approach is described, taking into account the stochastic properties of the noise. From this point, we are able to analyse the results.
Concluding remarks Chapter 4

In this chapter, we have described the method to obtain simulated fMRI data and how to estimate the model parameters (Section 4.1). To create simulated fMRI data, we use sampled values of the haemodynamic response function, described by Glover and Worsley (see Figure 4.1), a box-car on-off stimulus signal, a second order trend model and fifteen different low order autoregressive noise models (for noise parameters, see Table 4.1). These fifteen different models are used to generate simulated data with 100 observations per time series.

In Section 4.2, we have discussed the likelihood-based method to estimate the parameters of the model from the simulated data. Starting with a model structure and least squares estimated parameters, we use non-linear optimizing procedures to maximize the value of the likelihood function Eq. (4.5), considering the observed data and the model structure:

\[
L(y; \theta, \alpha) = \frac{n}{2} \ln \left( \frac{n}{2\pi Q_l} \right) - \frac{1}{2} \ln \left| \psi \right| - \frac{n}{2}
\]

We use six model structures with an increasing number of noise parameters, while the order of haemodynamic response and trend are constant. In addition to these six likelihood values, five order selection methods are used to select the most probable model structure. Based on the value of the test statistic (Eq. (4.8))

\[
T_{\text{LR}} = 2(L_{\text{max}} - L_{\text{max-1}})
\]

we decide whether activation is present in the data or not. Because of the statistical behaviour of the results, every evaluation is repeated 1,000 times to average the results.
Chapter 5

Results, discussion and conclusions

In this chapter, we will discuss all the results obtained from the simulations as described in Chapter 4. We will discuss the performances of the GLRT itself, together with the results of order selection methods and the comparison between the GLRT and the t-test.

5.1 Performances of Generalized Likelihood Ratio Test

In this section, the results of the method described in Section 4.2 are given and discussed. In Section 2.3, we have defined the performance of a test by its false alarm rate and detection rate. The false alarm rate is the probability that a test decides that activation is present, when in fact no activation is present in the signal. The detection rate is the probability that the test decides correctly that activation is present. These two rates will be used as the basis for our evaluation of the test performance.

Our data from simulations consist of three contributions: activation, trend and noise. As pointed out in Section 4.1, in this study, we took a constant and known order of the trend (second order). The real order of the autoregressive noise source is supposed to be unknown and has to be assumed or estimated. In our analysis, we estimated the signal parameters using six autoregressive noise models, with increasing order from zero to five. From these six models, we are able to investigate the consequences of order mismatch.

5.1.1 False alarm rate

The false alarm rate (FAR) of a test is related to the distribution of the test statistic under the assumption that there is no activation ($H_0$). In the Generalized Likelihood Ratio approach, this distribution is asymptotically a $\chi^2_1$ distribution, having one degree of freedom. A plot of this distribution is given in Figure 3.1. In this study, we intended to set our false alarm rate to 5%, so from the $\chi^2_1$ distribution, a threshold for our test statistic of 3.841 is taken. When the value of the test statistic, computed by Eq. (3.23), exceeds this threshold, the signal is assumed to contain activation.

The model used for our investigation is given in Eq. (3.8). The haemodynamic response is described by the sampled values as in Figure 4.1, while in our study the level of activation (that is to say the parameter $\theta_{hrf}$) varies from zero to one. The trend related parameters are fixed, as given in Eq. (4.1). The vector $\theta$ is given in Eq. (2.12). We have used different noise generating systems, as given in Table 4.1. In Figure 5.1, two examples of the experimentally obtained false alarm rates (FARs), are given for two different underlying autoregressive noise processes. While we assume that for estimation the right model order of our data
generating process is unknown, the FARs are given for estimation with different noise model orders. More plots are given in Appendix B, Section a.

![Graphs showing FAR for different noise model orders](image)

**Figure 5.1** Two examples of experimental false alarm rates for the Generalized Likelihood Ratio Test. On the left (a), the generating noise process was white, while on the right (b), the generating noise process was autoregressive of first order with parameter $a_1 = -0.4$. The order of the trend in the signal is assumed to be known, while the order of the noise is unknown. On the horizontal axis, the different orders for noise model estimation are given. On the vertical axis, the false alarm rate is given. The confidence intervals are denoted by the red markers. More figures and values are given in Appendix B (Section a).

At first sight, we can see that the FAR is almost always larger than our intended 5%. We can also see that the FAR increases for estimation with larger order models. However, some values differ from this general behaviour (such as the value for white noise estimation in Figure 5.1-b). This occurrence will be explained below. The red bars in Figure 5.1 denote the confidence intervals of the values. From these confidence intervals, we can see within which region the value of the FAR will be found for other evaluations. Because all values of FARs are obtained from the same data, the differences in the found values are significant, although the confidence intervals are larger than these differences.

The reason that the FAR generally does not reach the 5%, is related to the finite number of samples during measurement. When we extend our measurements to time series of 1,000 data points instead of 100, the FARs get closer to or reaches our intended value. Two examples, one for signal 0 and one for signal 3 (see Table 4.1 for noise model parameters) are given in Figure 5.2.

The increasing values of the FAR for higher model order estimation can be explained by the increasing number of parameters. As we have already seen in Section 3.2, an increasing number of parameters results in a higher variance of the estimators of the parameters. A higher variance of the parameters results in a higher variance of the likelihood values (both under $H_0$ and $H_1$), which, on its turn, results in a higher variance of the test statistic. Finally, a higher variance of the test statistic results in more threshold-exceeding values, thus more decisions that activation is present are made and a higher FAR is obtained. These relations can be obtained from the stored data (on CD-ROM in Appendix C), and are shown in Figure 5.3 for the above given examples.
Figure 5.2] Two examples of the false alarm rate for the GLR approach for an increasing number of observations. At the left (a), the original data generating system was white, while on the right (b) noise was generated by a first order autoregressive process with parameter $\alpha_1 = -0.4$. The false alarm rate comes closer to our intended false alarm rate (i.e. 0.05) for more observations. The confidence intervals are denoted with the red markers.

Figure 5.3] The effects for higher model order estimation to the variances of the estimators of the activation-related parameter $\theta_{hrf}$ (left), the log-likelihood value $L(H_1)$ (middle) and the test statistic $T_{LR}$ (right) for two examples: white noise (a) and first order autoregressive noise with parameter $\alpha_1 = -0.4$ (b). The dashed lines in (b) represent the range in (a).
One important remark should be made with respect to our observations. The argumentation given above holds for estimation with right model order or too high model order, but not for estimations with too low model order. While estimating with a too low model order, irregularities to very high or low 'detection' occur, resulting in extraordinary values of the false alarm rate. This can be seen for instance in Figure 5.3-b, where the variance of the test statistic is very large for the estimation with a noise model order zero. We can also see the extraordinary values recurring in the detection rates (section 5.1.2 and Appendix B). This behaviour can be explained when we look at the spectra of the related noise signals. These spectra are given in Appendix A. When the noise model order is too low, the estimators of the trend parameters (including the activation-related parameter) have to compensate for the contributions that can not be described by the estimators of the noise parameters. For noise signals with a spectrum that looks like the spectrum of the activation signal, it is very attractive to assign the information which can not be described by the estimated noise model to the activation-related parameter. While the behaviour of the noise has to be estimated by the deterministic part of the model (especially the activation-related parameter \( \theta_{\text{hrf}} \)), the variance in these parameters will increase. This can be seen in Figure 5.3-b for a zero order estimation of a first order autoregressive process (signal 3 in Table 4.1). Analogous to the explanation above for the increasing of the variance for increasing number of parameters, this very large variance results in a higher false alarm rate and detection rate. For noise spectra which are not similar to the spectrum of the activation signal, the assignment of noise information to the activation-related parameter will be small. Therefore, no decision to detection will be made.

This explanation can be verified by investigation of the product of the activation signal with noise signals for different parameters. Analogous to Eq. (2.19), it can be derived that the variance of \( \hat{\theta} \) is related to the variance of the product \( X^T y \), or, rewriting \( y = X\theta + v \), to the variance of the product \( X^Tv \). As can be seen in Figure 5.4, this product varies much more for noise signals with a spectrum that looks like the spectrum of the activation signal (a) than for noise signals which spectra do not (b). It can also be verified from the saved data on the CD-ROM that the variance of the estimated \( \theta_{\text{hrf}} \) is much larger for those noise spectra which correspond to the spectrum of the activation signal than for those which do not.

![Figure 5.4](image1.png)  
*Figure 5.4* | The product of the activation signal \( x_{\text{hrf}} \) and two different AR(1)-noise signals \( v \) for 1,000 evaluations. The variance in this product is five times larger for the noise signal with parameter -0.4 (a) than that for the noise signal with parameter 0.4 (b). As can be seen from Appendix A, the power density of the signal in (a) corresponds to the power density of the spectrum of the response signal, while the spectrum of the signal in (b) do not.
5.1.2 Uncorrected detection rate

While false alarm rate denotes the probability of ‘detection’ when no activation is present, detection rate denotes the probability of detection when activation is present. Because the level of activation can vary, we have to investigate the performance of our test for different levels of brain activity.

Figure 5.5 | Two examples of detection rates for estimating with an increasing model order (from AR(0) to AR(5)) and for an increasing level of brain activity (horizontal axis). In (a), the noise was white, while in (b), the noise of...
the fMRI signal was a first order autoregressive process with parameter \( \alpha_1 = -0.4 \). The threshold, corresponding to a test false alarm rate of 5\%, is derived from the theoretical \( \chi^2 \) distribution of the test statistic under \( H_0 \). The \( \chi^2 \) distribution has 95\% of its values below 3.841, thus the theoretical threshold is 3.841. The real experimental false alarm rates – with activation level equal to zero – are given in Figure 5.1.

levels of activation (the value of \( \theta_{hrf} \)). With no activation taken as bottom-line, the level of activation increases from zero to one. In Figure 5.5, two examples of the detection rates are given for the same signals as in Figure 5.1, while more figures are given in Appendix B (Section b). As in Figure 5.1, for all levels of activation, we have estimated the observed signal using several autoregressive noise models of increasing order.

In Figure 5.5, we see the same behaviour as in Figure 5.1 for estimation with an increasing noise model order. Also the increase of the detection rate for an increasing level of activation does not surprise us. The results given in this section can be regarded as the detection probability during a real fMRI measurement, using the GLRT approach. It should be noticed that although the detection rate for estimating the signal in (b) with white noise is very high, also the false alarm rate is very high (see Figure 5.1), so it is still difficult to see whether this approach performs well. To compare the performances for several methods of estimation, we will correct for the differences in false alarm rate.

### 5.1.3 Detection rate with false alarm rate correction

To compare detection rates of multiple tests having the same false alarm rate, we have to correct the non-asymptotic behaviour of our false alarm rates, as is described in Section 3.3.2. We get detection rates as given in Figure 5.6 for again two examples.

From Figure 5.6 and Appendix B (Section c), we can see that for the corrected detection rates, no general behaviour is observed with respect to the model order during estimation. For some noise signals, a too high noise model order will lower the detection rate drastically (up to 15\% for, e.g., signals 0 and 9), while for some other signals the opposite appears (for example signal 5 and 13). Therefore, we can hardly say which order should be taken to estimate fMRI signals, if the real order is unknown. However, in literature, this is done frequently, mostly choosing beforehand white or first order autoregressive models. The effects of order selection criteria will be discussed in Section 5.2.

### 5.1.4 If everything is known

In practical fMRI, the real parameters of the data generating system are not known. Therefore, we have to estimate them. Unlike in practice, in simulation we know everything, especially whether activation is present or not. With our knowledge of the data generating model, we can compare the above described results with the results we would obtain if we were able to use this knowledge. Kay\(^{122}\) gives a distribution of the likelihood-based test statistic for large data records. The probability density function of the GLR test statistic under \( H_0 \) and \( H_1 \) becomes (asymptotically):

---

122. Kay [1998], p.205
Figure 5.6] Two examples of corrected detection rates for estimating with increasing noise model order (from AR(0) to AR(5)) and for an increasing level of brain activity (horizontal axis). In (a), the noise was white, while in (b), the noise of the fMRI signal was first order autoregressive with parameter $\alpha_1 = -0.4$. To correct the detection rates for not-adjusted false alarm rates, the false alarm rate is set to 0.05, using the real values of the test statistic under $H_0$, as elaborated in Section 3.3.2.
Section 5.1 Performances of Generalized Likelihood Ratio Test

Under $H_1$, the test statistic has a non-central $\chi^2$ distribution, with non-centrality parameter $\zeta$. The non-centrality parameter $\zeta$ depends on the true parameters and the Fisher information matrix. The Fisher information matrix is given in Eq. (3.4), while its elements can be computed. It can be shown that the Fisher information matrix can be simplified to

$$F = F(\theta, \sigma_0^2, \sigma_1^2) = \begin{pmatrix} F_{00} & 0 & 0 \\ 0 & F_{11} & 0 \\ 0 & 0 & F_{22} \end{pmatrix}$$

with

$$F_{00} = -E \left[ \frac{\partial^2 \log p(y, \theta, \sigma_0^2, \sigma_1^2)}{\partial \theta^2} \right]$$

From the Fisher information matrix of (5.2) with 4 elements in $\theta$, the non-centrality parameter $\zeta$ can be calculated by

$$\zeta = \theta_{00} \left[ F_{00}(1,1) - F_{00}(1,[2:4]) F_{00}(2:4,2:4) F_{00}(2:4,2:4)^{-1} F_{00}(2:4,1) \right] \theta_{00}^T,$$

with $F_{00}(1,[2:4])$ denoting the first column and the second to the fourth row of the matrix $F_{00}$. Note that the other elements of Eq. (5.2) cancel out. The change in the shape of the distribution function for $\zeta = 10$ and $\zeta = 40$ is illustrated in Figure 5.7. Note that the Fisher information matrix depends on the (number of) measurements and thus also the parameter $\zeta$ depends on the length of the time series.

If we set our threshold $\alpha$ (under $H_0$), we can use the known distribution under $H_1$ to obtain the theoretical detection rate. For instance, when we set our false alarm rate to 5%, the detection rate under $H_1$ will be 88% for $\zeta=10$. In practice, for finite samples the detection rate will not reach these theoretical values because the asymptotic requirements are not satisfied. However, the theoretical values can give a measure of the asymptotically highest attainable performance of the Generalized Likelihood Ratio Test. The values of $\zeta$ and the corresponding detection rates (having 100 measurements per time series) are given in Appendix B (Section d), while again two examples are given in Figure 5.8.

123. Personal communication A.J. den Dekker [2006]
Figure 5.7 | Three $\chi^2$ distributions with different non-central parameters. The blue curve denotes the normal $\chi^2$ distribution (non-central parameter $\zeta=0$), while the red and green curves denote the $\chi^2(\zeta)$ distributions with non-central parameter $\zeta=10$ and $\zeta=40$, respectively. For the $\chi^2$ distribution, the threshold for 5% false alarm is given by the dotted line. The corresponding detection rates (with the same threshold) for the other two distributions are 88% and (nearly) 100%.

Figure 5.8 | Two examples of the increasing detection rate for an increasing level of activation. In dark, the experimentally obtained detection rate is given for estimation with a right order of the noise model, while in the light bars the asymptotic value is given.

5.1.5 Conclusions

For an increasing number of model parameters, the false alarm rate will also increase, because the variance of the parameters is larger when more parameters have to be estimated. This increment of the false alarm rate can also be recognized in the detection rates. Another remark with respect to the false alarm rate should be made: for measurements with around 100 observations, the false alarm rate will be higher than the intended false alarm rate. Third, the false alarm rate will drastically increase when our model structure is too low in combination with a noise spectrum with a same behaviour as the spectrum of the activation signal.
The detection rate will not always attain 100%, but according to the asymptotic theory, this is what we expect. If we correct the detection rate for the behaviour of the false alarm rate to compare the different tests with each other, no clear distinction can be made between estimation with a model structure of right order or wrong order.

5.2 Performances of Generalized Likelihood Ratio Test with order selection

In this section, the influence of order selection methods on the detection rate will be investigated. When the true order of a data generating system is not known, we can simply assume an order or estimate the order by order selection criteria. As mentioned in Chapter 1, our hypothesis about order selection is that order selection will give lower performance than choosing the true order, but will give better results than choosing a false order. Which orders were selected by the different order selection criteria during this study can be found on the CD-ROM (Appendix C) in the variable ‘p_all’ in the saved data files.

5.2.1 Loss in comparison with right order

In Figure 5.5, two examples are given in which the detection rates under the right order are compared with the detection rates obtained when the model order is selected by several order selection criteria (OSCs). More figures can be found in Appendix B (section e).

![Graph showing detection rate comparison](image-url)
Figure 5.9 | Examples of detection rates with order selection criteria from two different (autoregressive) noise processes. The adjusted false alarm rate was 5%. In (a), the noise generating process was white (AR(0)), while in (b), the noise generating process was AR(1) with parameter -0.4. The detection rates of five order selection criteria (OSCs), with different penalty factors (see Section 3.3), are shown, along with the detection rate for a noise model with right model order. The vertical red bars denote the confidence intervals of the found values. Note that the values at the same level of activation are based on the same data set, so it cannot be said that differences in detection rate between the different OSC are not significant.

As we can see in the figures, mostly the decrement of the detection rate using order selection methods is very low with respect to the right order. In fact, as can be seen in Appendix B, the detection may even be higher when order selection methods are used. However, sometimes, e.g. in signal 9, the decrement in detection could be 9% absolutely, or 22% relatively for certain order selection criterion.

5.2.2 Profit of order selection with respect to assuming false order

We have formulated our hypothesis that the use of order selection will give better results than assuming a false order for the noise model. In Figure 5.10, two examples are given, more can be found in Appendix B (Section f).
Figure 5.10: Detection rates obtained using order selection criteria along with the detection rates obtained using noise models with a false model order. The adjusted false alarm rate was 5%. Two examples are given for two different (autoregressive) noise generating processes, more figures can be found in Appendix B (Section f). In (a), the generating noise process was white (AR(0)), while in (b), the generating noise process was AR(1) with parameter -0.4. The detection rates of five order selection criteria (OSCs), with different penalty factors (see Section
3.3), are shown, in comparison with the detection rates of two (a) or three (b) noise models with false order. The black horizontal lines denote the detection rate of the right order noise model. The red vertical bars denote the confidence intervals of the values found. Note that the values at the same level of activation are based on the same data set, so differences in detection rate between the different methods are significant.

From the results, we can see that most times order selection methods perform better in terms of detection rate than selecting a too high noise model order (see Figure 5.10-a). Assuming a low noise model order, e.g. zero or first order, can give good results, but the detection rate can also be much lower than the rates obtained by the test that includes order selection, as can be seen with noise signals 5 and 13.

5.2.3 Conclusions
Using order selection methods will perform almost as well as selecting a right order noise model. In the case that the right model order is unknown, sometimes it could be better to select a low order (e.g. zero or one) model structure, rather than using order selection methods. However, it could happen that selecting a low order model structure gives worse results than using order selection methods. No specific order selection method can be selected that gives the overall best performance.

5.3 A comparison of the Generalized Likelihood Ratio Test with the common $t$-test

In this section, the performances of the GLRT and the commonly used $t$-test will be compared. This comparison will be done in terms of the false alarm rate $P_f$ and the detection rate $P_d$.

5.3.1 False alarm rate

In Figure 5.11, the experimentally obtained false alarm rates (FARs) are given for the two-sided Generalized Likelihood Ratio Test (GLRT) and the two-sided $t$-test. The underlying model is given by Eq. (3.8), while fifteen different noise generating processes are considered (see Table 8.1). The false alarm rate is obtained while no activation is present in the observations. The chosen trend parameters of Eq. (5.8) are $\theta_t = [0.1, -0.01, 1 \cdot 10^{-5}]$. The test statistic of the GLRT is given by Eq. (3.23), while the $t$-test statistic is given by Eq. (2.33). All simulations are done with 100 data point per time series ($n = 100$) and the results are averaged over 1,000 realisations.
Section 5.3  A comparison of the GLRT with the common $t$-test

Figure 5.11  A comparison of experimentally obtained false alarm rates, based on the GLR-test (blue) and the $t$-test (red), for different noise generating processes (see Table 4.1 for noise model parameters). The intended false alarm rate was 0.05. The vertical green lines denote the confidence intervals of the found values, based on 1,000 simulations. Note that the values at the same noise signal are based on the same data set, so it cannot be said that the differences are not significant.

As we can see, the false alarm rate of the $t$-test is close to the false alarm rate of the GLRT. There are no big differences. Based on these results, we cannot say which test we prefer. Note that the intended false alarm rate of 5% is reached by none of the processes. From these results we expect that there will also be a gap between the intended and obtained false alarm rate in practical measurements. To compare different tests with respect to detection rate, we apply a correction to the threshold of the test statistic, as described in Section 3.2.2 to obtain equalized false alarm rates.

5.3.2 Detection rate

In Figure 5.12, two examples of comparisons of (corrected) detection rates between the GLRT and the $t$-test are given. The model used for the data generating process is given by Eq. (3.8). An increasing level of activation is present and the trend parameters of Eq. (3.8) are $\theta_t = [0.1, 0.01, 1 \cdot 10^{-5}]$. For the parameters of the fifteen noise generating processes, see Table 4.1. All simulations are made with 100 data point per time series ($n = 100$) and the results are averaged over 1,000 realisations.
Figure 5.12 | Two examples of comparisons of corrected detection rates for model estimation with the right order, based on the GLR test (blue) and the t-test (red), for models with different noise generating processes. On the left (a), the noise generating model was white, while on the right (b) the process was first order autoregressive with parameter $\alpha_1 = -0.4$. More drawings are given in Appendix B (Section g).

As can be seen in Figure 5.12 and in Appendix B (Section g), with respect to the (corrected) detection rate, there are only small differences between the GLR-test and the t-test. For white noise, no differences are found between the performance of the GLRT and the performance of t-test. In general, it is hard to say which test performs better in terms of detection rate.

5.3.3 Conclusions

Comparing the two-sided tests results of the GLRT and the t-test, we can see that the false alarm rates and the detection rates of the GLRT stay close to those of the t-test. Therefore, it is not possible to make a clear conclusion which test performs better. However, while the t-test does not need a optimizing procedure – contrary to the GLRT –, the t-test will provide results much faster than the GLRT.
Section 5.3  A comparison of the GLRT with the common t-test
Chapter 6

Conclusions and recommendations

In Chapter 1, we formulated the questions that we wished to address during this study:

What is the performance of a likelihood-based hypothesis test for brain activation detection from fMRI data in terms of detection rate and false alarm rate; and how does this performance hold in comparison with that of the conventional $t$-test-based method for activation detection?

After asking these questions together with some additional sub-questions, we followed the next four steps. In Chapter 2 we discussed the physical backgrounds of fMRI and the current mathematical approach, using the General Linear Model (GLM). Chapter 3 discusses our new approach: likelihood-based detection for fMRI data disturbed with coloured noise, of which the practical elaboration is given in Chapter 4. The results of our approach are given and discussed in Chapter 5. In this chapter, using the obtained results, we will answer the questions posed in Chapter 1.

Limited data length

The first question that arose alongside of our problem definition was the matter how both methods (GLRT and $t$-test) perform for practical finite fMRI sample lengths. The theoretical asymptotic properties are given in Sections 2.3 and 3.2 but, in practice, we cannot measure for an infinite length of time, nor repeat the evaluation infinitely. For both test methods, we can conclude that for finite time series length the false alarm rate will be higher than intended, while we can conclude for the GLRT that measurements made over a longer period of time will give better results (see Figure 5.2). The (corrected) detection rates are also lower than the asymptotically achievable values, as can be seen in Figure 5.8. In fMRI, a time series length of typically 80 – 120 data points is obtained (see Section 2.3.1), which is too small to let the results fully correspond with the asymptotic theory.

Unknown fMRI noise processes

Because the noise processes in fMRI data differ from voxel to voxel and are unknown in practice, we evaluated different noise models. It turns out that the performance of the GLRT differs for different types of noise. When the spectrum of the noise signal resembles the spectrum of the activation signal, the activation is difficult to detect. Unfortunately, this is the case with fMRI signals, since often the autoregressive noise parameters vary between -0.4 and 0.1, as mentioned in Section 2.3.4, so that the noise signals have spectra with most power at low frequencies, just as the spectrum of the activation signal (compare with Section 4.1 and Appendix A). For these signals, the detection rate will be low.
**Order selection methods**

In this study, we investigated the results for five different order selection methods, based on the Akaike’s information criterion, the symmetric Kullback’s information criterion and the Minimum Description Length. Using order selection methods will perform almost as well as selecting the right order noise model, while the loss in performance is small. Moreover, in general, the usage of order selection methods gives better results than selecting a model order a priori. Therefore, order selection should be used for fMRI measurements. From the obtained results, no specific order selection method can be indicated which provides the best overall performance.

To summarize, we can say that the performance of the Generalized Likelihood Ratio Test for practical fMRI data will behave like the asymptotic performance, but will not reach its level. Only small differences are found between the performances of the GLRT and the t-test. Knowledge of the statistical distribution of the test statistic in the absence of activation allows us to set a threshold that results in a specific false alarm rate. However, it has been found that the typical fMRI time series lengths are too short to allow the usage of the results from the asymptotic theory. Furthermore, while the model for the underlying noise process is unknown a priori, order selection methods should be used to avoid unintentionally low detection rates.

**Recommendations**

Because of the high false alarm rate for short periods of measurements and low detection rates more research should be done into fMRI modelling and parameter testing. We would like to give the following eight recommendations for further research and fMRI usage in practice. First we want to give four recommendations in connection with further research:

1. Investigate the shape of the haemodynamic response function. Commonly, a standard form is chosen, as given by Glover and Worsley (see Section 2.3.2), but it is pointed out that several models can be taken (see Section 2.2.2). Moreover, in this study we have assumed that the contribution made by the haemodynamic response function in the data can be estimated with one parameter. It would be better to estimate all the parameters of the haemodynamic response function (as for example in Eq. (2.10)) instead of one activation-related parameter $\theta_{hrf}$. Notice that in the case that the HRF is modelled with more parameters, attention should be given to the increasing variance.

2. Investigate the influence of unknown trend model structures. In this study, we have assumed that the model structure of the trend is known (second order polynomial), but in practice it is not known. For the use of the order selection method, a method should be developed to select the orders of trend and noise together, because in the General Linear Model approach, these two models are not nested.

3. Further research should be done into the statistical properties of the General Likelihood Ratio Test (GLRT) for practical fMRI data. It turned out that in simulations the method used does not attain the properties which are expected from asymptotic theory. To be certain that the practical usage of the GLRT will achieve the asymptotic properties (see Section 5.1.4 and Figure 5.2), larger data sets and more realisations should be investigated.

4. In this thesis, we focussed on a false alarm rate of 5%, which is large for practical FMRI
research. More simulations should be done to be able to investigate the effects of smaller false alarm rates on the detection rate.

From the results of this study, we will also provide some recommendations with respect to the data analysed from the fMRI scans.

5. First of all, the use of the theoretical properties of the tests is very precarious in practical circumstances. As can be seen in Section 5.3.1, in simulation, the value of the false alarm rate of both the $t$-test and the GLRT is much higher than the intended value. Due to these unintended values of the false alarm rate, the detection rates have to be corrected, which could only be done if the real present false alarm rate is known. Procedures should be developed to obtain the present false alarm rate.

6. Rather than the a priori choosing of a specific model order, order selection methods should be used.

7. The speed of the optimizing procedure should be improved. On an AMD Athlon 1600+ personal computer, evaluating 1000 voxels takes about one hour, which should be decreased for practical FMRI research.

8. Research should be done into tests with an obtained false alarm rate near the intended value and high detection rates for fMRI related signals.
Appendices
Appendix A

Spectra of activation signal and noise models used

In this section, the spectra will be given of the used activation signal and the noise models. The activation signal is the same for all experiments. For the noise signals, we know the autoregressive parameters for the different models, so we can give an exact plot of the noise spectra by the ARMASA function `arma2psd`.

Activation signal

![FMRI activation signal and its frequency spectrum](image1)

*Figure A.1* | The fMRI activation signal (with activation parameter $\theta_{hrf} = 1$) in the time domain (left) and its frequency spectrum (right).

Noise signals

![Spectrum of noise signal](image2)

*Figure A.2* | Spectrum of noise signal 0 (see Table 4.1 for autoregressive parameters).
Figure A.2 | Spectrum of noise signal 1.

Figure A.2 | Spectrum of noise signal 2.

Figure A.2 | Spectrum of noise signal 3.

Figure A.2 | Spectrum of noise signal 4.

Figure A.2 | Spectrum of noise signal 5.

Figure A.2 | Spectrum of noise signal 6.

Figure A.2 | Spectrum of noise signal 7.

Figure A.2 | Spectrum of noise signal 8.
Appendix A  Spectra of signal and noise

Figure A.2 | Spectrum of noise signal 9.

Figure A.2 | Spectrum of noise signal 10.

Figure A.2 | Spectrum of noise signal 11.

Figure A.2 | Spectrum of noise signal 12.

Figure A.2 | Spectrum of noise signal 13.

Figure A.2 | Spectrum of noise signal 14.
Appendix B

Graphical illustrations of the performance analysis of the Generalized Likelihood Ratio Test

In this appendix, the performance of the Generalized Likelihood Ratio Test (GLRT) will be given as obtained during simulation experiments. The performance of a test can be expressed in the false alarm rate and the detection rate. This appendix is divided into the following sections:

a. False alarm rate (page 88)
 b. Detection rate (91)
 c. Corrected Detection rates (99)
 d. Theoretical GLRT detecting rate (107)
 e. Loss of order selection with respect to right order estimation (110)
 f. Profit of order selection with respect to false order estimation (118)
 g. Comparison between the GLRT and the $t$-test (126)

a. GLRT false alarm rate

In this section, graphics are given of the obtained false alarm rates of the GLRT, while on theoretical knowledge of the $\chi^2$ distribution of the test statistic a 5% false alarm rate was intended. These results are used in Section 5.1.1. For an expression of the used test statistic, see Section 3.2.1.

The model used for the data generating process is given by Eq. (3.8). No activation is present ($\theta_{hrf} = 0$) and the trend parameters of Eq. (3.8) are $\theta_{tr} = [0.1, -0.01, 1 \times 10^{-5}]$. For the parameters of the fifteen noise generating processes, see Table 4.1. All simulations are done with 100 data point per time series ($n = 100$) and the results are averaged over 1,000 realisations. The red bars denote confidence intervals (see Section 5.1.1).

![False alarm rate for noise signal 0](image-url)
Appendix B  GLRT performance analysis

Fig. B.a.2 | False alarm rate for noise signal 1 for different estimated models (order 0–5).

Fig. B.a.3 | False alarm rate for noise signal 2 for different estimated models (order 0–5).

Fig. B.a.4 | False alarm rate for noise signal 3 for different estimated models (order 0–5).

Fig. B.a.5 | False alarm rate for noise signal 4 for different estimated models (order 0–5).

Fig. B.a.6 | False alarm rate for noise signal 5 for different estimated models (order 0–5).

Fig. B.a.7 | False alarm rate for noise signal 6 for different estimated models (order 0–5).

Fig. B.a.8 | False alarm rate for noise signal 7 for different estimated models (order 0–5).

Fig. B.a.9 | False alarm rate for noise signal 8 for different estimated models (order 0–5).
Likelihood-based brain activation detection

Fig. B.a.10 | False alarm rate for noise signal 9 for different estimated models (order 0–5).

Fig. B.a.11 | False alarm rate for noise signal 10 for different estimated models (order 0–5).

Fig. B.a.12 | False alarm rate for noise signal 11 for different estimated models (order 0–5).

Fig. B.a.13 | False alarm rate for noise signal 14 for different estimated models (order 0–5).

Fig. B.a.14 | False alarm rate for noise signal 13 for different estimated models (order 0–5).

Fig. B.a.15 | False alarm rate for noise signal 14 for different estimated models (order 0–5).
b. GLRT detection rate

In the following, graphics are given of the detection rates of the GLRT under an intended false alarm rate of 5%. These results are used in Section 5.1.2. For an expression of the used test statistic, see Section 3.2.1. The used model for the data generating process is given by Eq. (3.8). An increasing level of activation $\theta_h$ is present and the trend parameters of Eq. (3.8) are $\theta = [0.1, 0.01, 1 \cdot 10^5]$. For the parameters of the fifteen noise generating processes, see Table 4.1. All simulations are done with 100 data point per time series ($n = 100$) and the results are averaged over 1,000 realisations.

Figure B.b.1 | Detection rate for noise signal 0, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.
Figure B.b.2 | Detection rate for noise signal 1, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.

Fig. B.b.3 | Detection rate for noise signal 2, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.
Figure B.b.4 | Detection rate for noise signal 3, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.

Figure B.b.5 | Detection rate for noise signal 4, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.
Figure B.b.6 | Detection rate for noise signal 5, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.

Figure B.b.7 | Detection rate for noise signal 6, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.
Figure B.8 | Detection rate for noise signal 7, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.

Figure B.9 | Detection rate for noise signal 8, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.
Figure B.b.10 | Detection rate for noise signal 9, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.

Figure B.b.11 | Detection rate for noise signal 10, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.
Figure B.b.12 | Detection rate for noise signal 11, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.

Figure B.b.13 | Detection rate for noise signal 12, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.
**Figure B.b.14** Detection rate for noise signal 13, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.

**Figure B.b.15** Detection rate for noise signal 14, for increasing activation level, different estimated autoregressive noise model orders and intended false alarm rate of 5%.
c. Corrected GLRT Detection rates

In the following, graphics are given of the corrected detection rates of the GLRT under an adjusted false alarm rate of 5%. These results are used in Section 5.1.3. For an expression of the used test statistic, see Section 3.2.1. Correction to the detection rate is based on experimental test statistics, as described in Section 3.2.2. The used model for the data generating process is given by Eq. (3.8). An increasing level of activation is present and the trend parameters of Eq. (3.8) are \( \alpha = [0.1, 0.01, 1 \times 10^{-5}] \). For the parameters of the fifteen noise generating processes, see Table 4.1. All simulations are made with 100 data point per time series (\( n = 100 \)) and the results are averaged over 1,000 realisations. The red bars denote confidence intervals (see Section 5.1.1).

![Figure B.c.1](image)

**Figure B.c.1** | Detection rate for noise signal 0 (see Table 4.1), for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.
Figure B.c.2 | Detection rate for noise signal 1, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.

Figure B.c.3 | Detection rate for noise signal 2, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.
Figure B.c.4 | Detection rate for noise signal 3, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.

Figure B.c.5 | Detection rate for noise signal 4, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.
Figure B.c.6 | Detection rate for noise signal 5, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.

Figure B.c.7 | Detection rate for noise signal 6, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.
Figure B.c.8 | Detection rate for noise signal 7, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.

Figure B.c.9 | Detection rate for noise signal 8, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.
Figure B.c.10 | Detection rate for noise signal 9, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.

Figure B.c.11 | Detection rate for noise signal 10, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.
Figure B.c.12 | Detection rate for noise signal 11, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.

Figure B.c.13 | Detection rate for noise signal 12, for increasing activation level, different estimated autoregressive noise model orders and adjusted false alarm rate of 5%.
**Figure B.c.14** Detection rate for noise signal 13, for increasing activation level, different estimated autoregressive noise model orders and *adjusted* false alarm rate of 5%.

**Figure B.c.15** Detection rate for noise signal 14, for increasing activation level, different estimated autoregressive noise model orders and *adjusted* false alarm rate of 5%.
d. Theoretical GLRT detecting rate

In this section, the influence of knowledge about the asymptotic distribution of the GLR test statistic under $H_1$ is used to calculate the asymptotic attainable detection rates. First, the values of the non-centrality parameter $\zeta$ as in Eq. (5.1) are given, for the fifteen different noise generating processes (see Table 4.1) and for increasing level of activation (from zero to one. Thereafter, comparisons are made between experimental obtained detection rates and the asymptotically obtainable detection rates for time series with 100 observations. These results are used in Section 5.1.4.

Figure B.d.1 | Values of non-centrality parameter $\zeta$ of the $\chi^2(\zeta)$ distribution, for increasing level of activation (from 0 to 1) for each of the fifteen different noise signals (see Table 4.1) for time series length of 100 observations.

Figure B.d.2 | Practical and theoretical attainable detection rates for fMRI data with noise process 0 (see Table 4.1).

Figure B.d.3 | Practical and theoretical attainable detection rates for fMRI data with noise process 1 (see Table 4.1).
Figure B.d.4 | Practical and theoretical attainable detection rates for fMRI data with noise process 2 (see Table 4.1).
Figure B.d.5 | Practical and theoretical attainable detection rates for fMRI data with noise process 3 (see Table 4.1).
Figure B.d.6 | Practical and theoretical attainable detection rates for fMRI data with noise process 4 (see Table 4.1).
Figure B.d.7 | Practical and theoretical attainable detection rates for fMRI data with noise process 5 (see Table 4.1).
Figure B.d.8 | Practical and theoretical attainable detection rates for fMRI data with noise process 6 (see Table 4.1).
Figure B.d.9 | Practical and theoretical attainable detection rates for fMRI data with noise process 7 (see Table 4.1).
Figure B.d.10 | Practical and theoretical attainable detection rates for fMRI data with noise process 8 (see Table 4.1).
Figure B.d.11 | Practical and theoretical attainable detection rates for fMRI data with noise process 9 (see Table 4.1).
Figure B.d.12 | Practical and theoretical attainable detection rates for fMRI data with noise process 10 (see Table 4.1).

Figure B.d.13 | Practical and theoretical attainable detection rates for fMRI data with noise process 11 (see Table 4.1).

Figure B.d.14 | Practical and theoretical attainable detection rates for fMRI data with noise process 12 (see Table 4.1).

Figure B.d.15 | Practical and theoretical attainable detection rates for fMRI data with noise process 13 (see Table 4.1).

Figure B.d.16 | Practical and theoretical attainable detection rates for fMRI data with noise process 14 (see Table 4.1).
Loss of detection from order selection methods with respect to right order estimation

In the following, graphics are given with the corrected detecting rates of the right order for the GLRT, together with the detection rates of methods with order selection. The order selection methods are discussed in Section 3.3. The adjusted false alarm rate is 5%. The red vertical lines give the confidence intervals (see Section 5.1.1). The results given in this section are used in Section 5.2.1.

Figure B.e.1 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 0 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.e.2 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 1 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.e.3 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 2 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.e.4 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 3 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.e.5 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 4 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.e.6 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 5 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.e.7 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 6 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.e.8 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 7 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.e.9 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 8 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.e.10 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 9 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.e.11 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 10 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.e.12 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 11 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.e.13 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 12 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.e.14 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 13 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.e.15 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 14 (see Table 4.1) for right order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
f. Profit of detection from order selection methods with respect to false order estimation

In the following, graphics are given for the corrected GLRT detection rate for models with right order, together with the detection rates of methods with order selection. The adjusted false alarm rate is 5%. The red vertical lines with the dots give the confidence intervals, while the black horizontal lines give the detecting rate for the right order (see Section B-e). The results of this section are used in Section 5.2.2.

Figure B.f.1: The comparison of corrected detection rates from the GLRT for fMRI data with noise process 0 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.f.2 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 1 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.f.3 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 2 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
The comparison of corrected detection rates from the GLRT for fMRI data with noise process 3 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

The comparison of corrected detection rates from the GLRT for fMRI data with noise process 4 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.f.6 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 5 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.f.7 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 6 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.f.8 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 7 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.f.9 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 8 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.f.10 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 9 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.f.11 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 10 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.f.12 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 11 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.f.13 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 12 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
Figure B.f.14 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 13 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.

Figure B.f.15 | The comparison of corrected detection rates from the GLRT for fMRI data with noise process 14 (see Table 4.1) for different false order estimation and order selection methods. The order selection methods are discussed in Section 3.3. The vertical red lines denote the confidence intervals.
**g. Comparison between the GLRT and the \( t \)-test**

**False alarm rate**

In the Figure B.g.1, the false alarm rates are given of the GLRT in comparison with the \( t \)-test. The intended false alarm rates, based on the theoretical properties of the test statistic, are 5%. The model used for the data generating process is given by Eq. (3.8). No activation is present (\( \theta_{h\text{rf}} = 0 \)) and the trend parameters of Eq. (3.8) are \( \theta_t = [0.1, -0.01, 1\cdot 10^{-5}] \). For the parameters of the fifteen noise generating processes, see Table 4.1. All simulations are made with 100 data point per time series (\( n = 100 \)) and the results are averaged over 1,000 realisations. These results are used in Section 5.3.1.

![Comparison GLRT and t-test](image)

**Figure B.g.1** | False alarm rates of GLRT and \( t \)-test for different noise signals (see Table 4.1). The vertical green lines denote the confidence intervals, based on 1,000 observations (see also the description in Section 5.1.1).

**Detection rate**

In the following, the corrected detection rates are given of the GLRT in comparison with the \( t \)-test. The adjusted false alarm rate is 5%. Correction to the detection rate is based on experimental test statistics, as described in Section 3.2.2. The model used for the data generating process is given by Eq. (3.8). No activation is present (\( \theta_{h\text{rf}} = 0 \)) and the trend parameters of Eq. (3.8) are \( \theta_t = [0.1, -0.01, 1\cdot 10^{-5}] \). For the parameters of the fifteen noise generating processes, see Table 4.1. All simulations are made with 100 data point per time series (\( n = 100 \)) and the results are averaged over 1,000 realisations. These results are used in Section 5.3.2.
Figure B.g.2 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 0 (see Table 4.1).

Figure B.g.3 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 1.

Figure B.g.4 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 2.

Figure B.g.5 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 3.

Figure B.g.6 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 4.
Figure B.g.7 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 5.

Figure B.g.8 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 6.

Figure B.g.9 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 7.

Figure B.g.10 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 8.

Figure B.g.11 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 9.

Figure B.g.12 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 10.
Appendix B  GLRT performance analysis 129

Figure B.g.13 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 11

Figure B.g.14 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 12

Figure B.g.15 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 13

Figure B.g.16 | Comparison between detection rates of GLRT and t-test for fMRI data with noise process 14
Appendix C

**CD-ROM with programs, data files and digital version of thesis**

All used Matlab programs, stored data files and figures in this thesis can be found on the CD-ROM, together with pdf-versions of this thesis and the bibliography. The files are managed in the following structure:

<table>
<thead>
<tr>
<th>Directory</th>
<th>File-description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bibliography</td>
<td>All available pdf-files of the used papers, for scientific use only.</td>
</tr>
<tr>
<td>thesis</td>
<td>This thesis with all the graphics</td>
</tr>
<tr>
<td>simulation</td>
<td>All self-written Matlab functions together with the toolbox ARMASEL. The stored data files can also be found here</td>
</tr>
</tbody>
</table>
No CD-ROM found?

Contact with W. Brouwer
willem@wbrouwer.nl
+31703220708
Bibliography


MRI on WebCT. [online, n.d.] School of Physics and Astronomy, University of Nottingham; Retrieved from the University of Nottingham website: http://www.nottingham.ac.uk/physics/ugrad/courses/mod_home/f33ab5/mri/ (23 May 2006).


