Simulation of fMRI data: A statistical approach

Marijke Welvaert

Promotor: Prof. dr. Yves Rosseel

Proefschrift ingediend tot het behalen van de academische graad
van Doctor in de Psychologie

2013

Do not put your faith in what statistics say until you have carefully considered what they do not say

William W. Watt

Table of Contents

Een welgemeende dank u ...
Sincere thanks ...

... aan mijn promotor

Yves, je bent reeds mijn promotor sinds oktober 2006. Eerst voor mijn masterthesis in de Master of Statistical Data Analysis en daarna voor mijn doctoraatsthesis. Het spreekt voor zich dat ik uitermate dankbaar ben voor alle begeleiding die ik in die bijna 7 jaar mocht ontvangen. Toch wil ik een bijzondere dankjewel uitspreken voor de – mag ik zeggen niet-altijd-conventionele – manier waarop je me op het rechte pad hielp en hield. Ik denk maar aan het advies om *eens goed op café te gaan* of de stapel dvd's met series wanneer het onderzoek op een ijsberg was gelopen of ik het wandelpad ergens tussen de dichte bebossing verloren was of me, zoals nu weer, teveel liet verleiden tot literaire beeldspraak. Ook apprecieerde ik ten zeerste de vrijheid die je me gaf bij het afbakenen van mijn onderzoeksthema, mijn tijdsindeling en bij momenten ook om mijn onderwijstaak te organiseren. Jouw liberale *bewind* als promotor heeft er mede voor gezorgd dat ik mijn extra-curriculaire activiteiten naar hartelust kon behouden. Dus, merci!

... aan mijn begeleidingscommissie

Prof. dr. Beatrijs Moerkerke, dr. Ruth Seurinck en prof. dr. Tom Verguts, in de eerste plaats dankjewel om al die jaren als lid van mijn doctoraatscommissie te fungeren. Tijdens onze jaarlijke bijeenkomsten zijn we meer dan eens een discussie gestart die uiteindelijk het doctoraat niet haalde, maar jullie feedback heb ik altijd ter harte genomen.

Bedankt om me jaar na jaar gunstig advies te geven om door te gaan.

... to my collaborators

Bieke, Geert, Joke, Karsten, Ruth and Yves, it was my pleasure and honour to write at least one paper with you. I know my deadlines were sometimes challenging and I possess the gift of delaying everything until the last minute. Thank you for your invaluable input to make the papers theoretically sound and for the numerous corrections in my writing.

... to my colleagues

Some of you I only knew shortly, many of you were with me all these years and for the *younger* ones it is all still yet to begin. Thank you for teaching, researching, lunch- & teabreaking and celebrating the *fin de saisons* with me. Thanks to Amelie, Alessandro, Haeike, Maarten, GuoRong and Roma for sharing an office with me. Having three or four of us in the room was not always that easy but we managed anyhow to keep the working atmosphere. Special thanks to Franky for staying calm after the *soup accident* and saving my laptop, and also thank you Isabelle, for letting me crash your office when I was in need for a break or a late lunch.

... aan mijn vrienden

De meisjes, de leden van de *driehoeksverhouding*, het Tropical Heat-team, ..., bedankt voor al jullie steun de afgelopen jaren. Om begrip te tonen wanneer mijn agenda zo vol zat dat ik soms weken niks van me liet horen en om me daarna weer met open armen te ontvangen. Wanneer het allemaal een beetje veel begon te worden, keek ik reikhalzend uit naar een spelletjes/filmpjes-samenkomst met veel te veel zoetigheid, of naar een gezellig glas whiskey dikwijls voorafgegaan door heerlijk gekokkerel, of naar die week in augustus dat ik met mijn tent naar Esse afzak. Arne, bedankt voor het recept van de *adaptieve smoothies*. Bedankt om me te omringen met jullie onvoorwaardelijke vriendschap!

... aan de muzikanten van Fadièse, Harmonie Eendragt Maekt Magt, Har-

monie Evergem Guy Duijck, KF Kunst en Broedermin, het Gents Universitaire Harmonie Orkest (of kortweg GUHO), het saxensemble van de Poel en de andere *muzieken* waar ik sporadisch meespeelde
Jullie hebben misschien niet echt een rechtstreekse bijdrage geleverd aan mijn doctoraat, maar toch vond ik dat jullie niet in het lijstje mochten ontbreken. Dankzij ons samenspel had ik de ontspanning telkens voor het grijpen en kon ik mij uitleven in mijn grote passie. We veroorzaakten kippenvel bij het publiek, presteerden buitengewoon op tornooien en gingen ook wel eens samen iets drinken, maar bovenal kon ik met jullie cornet/contrabas/saxofoon spelen. Tijdens de repetities en concerten met jullie kon ik mijn hoofd volledig leegmaken en dat heeft zeker zijn effect niet gemist op mijn werk.

... aan mijn ouders

Mama en papa, jullie leerden me van kindsbeen af om zelfstandig in het leven te staan, om door te zetten en af te maken wat je begonnen bent. Jullie moedigden me steeds aan om mijn dromen na te streven, maar met telkens een gezonde portie bescheidenheid en realiteitszin in het achterhoofd. Bedankt om me op te voeden tot de persoon die ik geworden ben!

Marijke
Gent, Maart 2013

1 | General introduction

This doctoral dissertation is about simulating fMRI data and bridges the scientific fields of psychology, statistics, neuroimaging and informatics. Before setting out the specific research questions and motivation, crucial concepts that are basic in one field, but maybe less known in another, are explained. First, we will explain the notion of simulation, followed by a discussion on the origin and properties of fMRI data. Then, the position of this dissertation within a statistical framework is explained. This general introduction ends with an overview of the different chapters in this dissertation.

1.1 Simulation

In general, simulation is defined as the imitation of the working of a real-world process or system. Simulation can be performed in many contexts. Take, for example, flight simulators that are used for training pilots using lifelike experiences in a controlled environment or emergency simulations to test safety procedures. In this dissertation, the focus lies on computational simulation within a statistical validation framework. The computational aspect refers to the use of computers and their computational power to model the system under investigation, while the statistical validation refers to the ability of simulation to gain insight in the performance of statistical models. In this context, a specific form of simulation is Monte Carlo simulation.

Monte Carlo methods are often used in computer simulation of physical and mathematical systems. These methods rely on repeated random sampling to compute their results and are specifically used in cases that are technically intractable (i.e. their solution involves prohibitively expensive labour costs). The first Monte Carlo studies of a statistical procedure that have been documented were performed by Erastus Lyman de Forest in the 1870s. de Forest studied ways to smooth time series using a simulation based on cards that were drawn from a box (Stigler, 1978). With the introduction of high-speed computers, simulation studies have gained wide interest for their flexibility and accessibility.

Monte Carlo simulation studies are equivalent to experimental studies. Therefore, designing and analysing a Monte Carlo experiment is very similar

to the design and analysis of any other scientific experiment. Gentle (2005) and Robert & Casella (2005) discuss in detail how a Monte Carlo study has to be designed and analysed. As for any experiment, the design of a Monte Carlo study is very important, because this design determines the validity of the conclusions drawn from the Monte Carlo experiment (Skrondal, 2000). Since only a limited number of conditions can be investigated, the *external validity*, generalising results beyond a given experiment, is important. This external validity depends also on the quality of the generated data. Conclusions drawn from a simulation study can only be generalised if the data generation is representative for the real data. Also the *precision* or statistical efficiency becomes an issue because the results of stochastic simulations are by nature always more or less unreliable. Finally, the available *computer resources and time set* puts its limits on Monte Carlo experimentation. To illustrate the Monte Carlo technique, an example of a standard Monte Carlo experiment will now be examined. Suppose we have a linear regression model under standard Gauss-Markov assumptions

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{with} \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2).$$

Both unbiasedness and efficiency of the parameter estimates can be proven analytically. However, when we want to test robustness against violation of the normality assumption, direct solutions can be challenging or even impossible. Therefore, we will conduct a hypothetical Monte Carlo experiment. The anatomy of this experiment is described based on the steps formulated by Skrondal (2000).

Statement of the research problem

We start with a statement of the research problem. For example, we want to investigate the bias on the estimation of β_1 when the residuals are not normally distributed. The bias is defined as the difference between the real parameter value and the estimated parameter value, i.e. $\beta_1 - \hat{\beta}_1$

Experimental plan

Based on this research problem, an experimental plan has to be developed. In this experimental plan, it is decided which factors will probably influence the outcome and need to be manipulated (i.e. the conditions of the simulation experiment). Possible factors in this case are, for example, the distribution of the error term (F, Gamma, Exponential, Uniform, ...), the number of observations, and small or medium effects for X.

Simulation

During the simulation, data will be generated for all the conditions in the experimental plan. To generate the data Y_i, values have to be chosen for β_0, β_1, X_i and σ_ε^2 as well as the distribution of ε_i. While β_0, β_1 and X_i will account for the fixed part of the data generation, drawing from a distribution with variance σ_ε^2 will define the random component of the simulated data. The following description of a hypothetical simulation algorithm illustrates the data generating process. Choose parameter values $\beta_0 = 10$, $\beta_1 = 2$, $\mathbf{X} = (1\ 2.6\ 7\ 3.5\ 11.4)'$ and $\sigma_\varepsilon^2 = 4$.

> *Draw:* e_i from a χ^2 distribution with $df = 2$
> *Compute:* $\mathbf{Y} = 10 + 2 \times (1\ 2.6\ 7\ 3.5\ 11.4)' + \mathbf{e}$

The random number generation is implemented in many statistical software packages like, for example, R.

Estimation

For the simulated data, the parameters in the linear regression model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ have to be estimated, namely β_0, β_1 en σ_ε^2. Then, the estimate $\hat{\beta}_1$ can be compared with the known value of β_1. It is crucial to understand that in order to determine the bias of this estimate the real value of β_1 has to be known. We will call this the *ground truth*.

Replication

A Monte Carlo simulation is based on numerous replications of the same process. Therefore, both the simulation step and the estimation step will be replicated N times. At the end, we will have an estimate for the bias of β_1 for each simulation run and the expected bias will be calculated as the average of these values for each condition in the Monte Carlo experiment. Consequently, it is possible to compare the expected bias values for each condition and draw conclusions concerning the bias on the estimation of β_1 in a simple regression model when the normality assumption is violated.

In conclusion, a Monte Carlo simulation is an excellent tool to infer statistical properties when analytical derivations are challenging or in complex situations (e.g. fMRI data).

1.2 fMRI data

Magnetic resonance imaging (MRI) is a medical imaging method that is used widely in clinical settings because of its ability to provide non-invasive high-resolution images of body structures. The principle of magnetic resonance has been discovered independently by Felix Bloch (Bloch et al., 1946) and Edward Purcell (Purcell et al., 1945), who were rewarded for their contribution to science with the Nobel Prize in Physics in 1952. It took until the late 1970s for MRI images to become more or less standard practice in medical applications. Until then computerised tomography (CT) was very popular to create high-resolution images of the human body. One drawback for using CT imaging is that it requires concentrated X-ray exposure. Because MRI images can provide the same kind of information without this exposure, they became increasingly popular.[1]

In 1992, Ogawa et al. (1992) and Kwong et al. (1992) discovered the possibility of using the MRI principle to capture the working brain. Functional MRI (fMRI) was born and caused an explosion of research dedicated to the

[1] Around the same period *nuclear magnetic resonance* (NMR), the former name of MRI, fell into disfavour, because the word *nuclear* was associated with health risks, which was completely unjustified since NMR does not use ionising radiation.

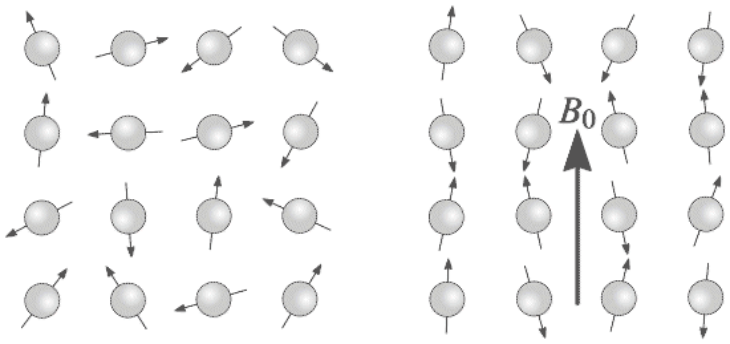

Figure 1.1 – (A) Protons in free space with random orientations. (B) Protons in a magnetic field with orientations aligned to the magnetic field. (source: `wikidoc.org`)

function of the brain, resulting in about 4000 fMRI related publications in 2012 alone. In order to understand the structure of fMRI data, we will briefly discuss MRI physics, the neurovascular coupling, and the sources of noise in the fMRI signal.

1.2.1 A brief introduction to MRI physics

The MRI-scanner is mainly a powerful magnet in which the subject is positioned alongside the z-axis that is parallel to the direction of the main magnetic field $\mathbf{B}_0$ (Figure 1.2, top). This magnetic field has the purpose to align the protons in the human body. These protons present random orientations in their natural state (Figure 1.1A), but when placed in a magnetic field the protons align with or against the magnetic field (Figure 1.1B). A first crucial concept is that, dependent on the direction of the alignment with the magnetic field, along or against it, the protons will be in low-energy or high-energy state, respectively, and more will be at the low-energy state. The sum of all the proton vectors is the net magnetisation, $\mathbf{M}_0$. Secondly, the protons will show a spinning motion around the axis of the magnetic field (Figure 1.1B). This phenomenon is known as *precession* and the resonant frequency of the spinning, the *Larmor frequency*, is specific to the atomic nucleus and the magnitude of the magnetic field in the MR scanner. Thirdly,

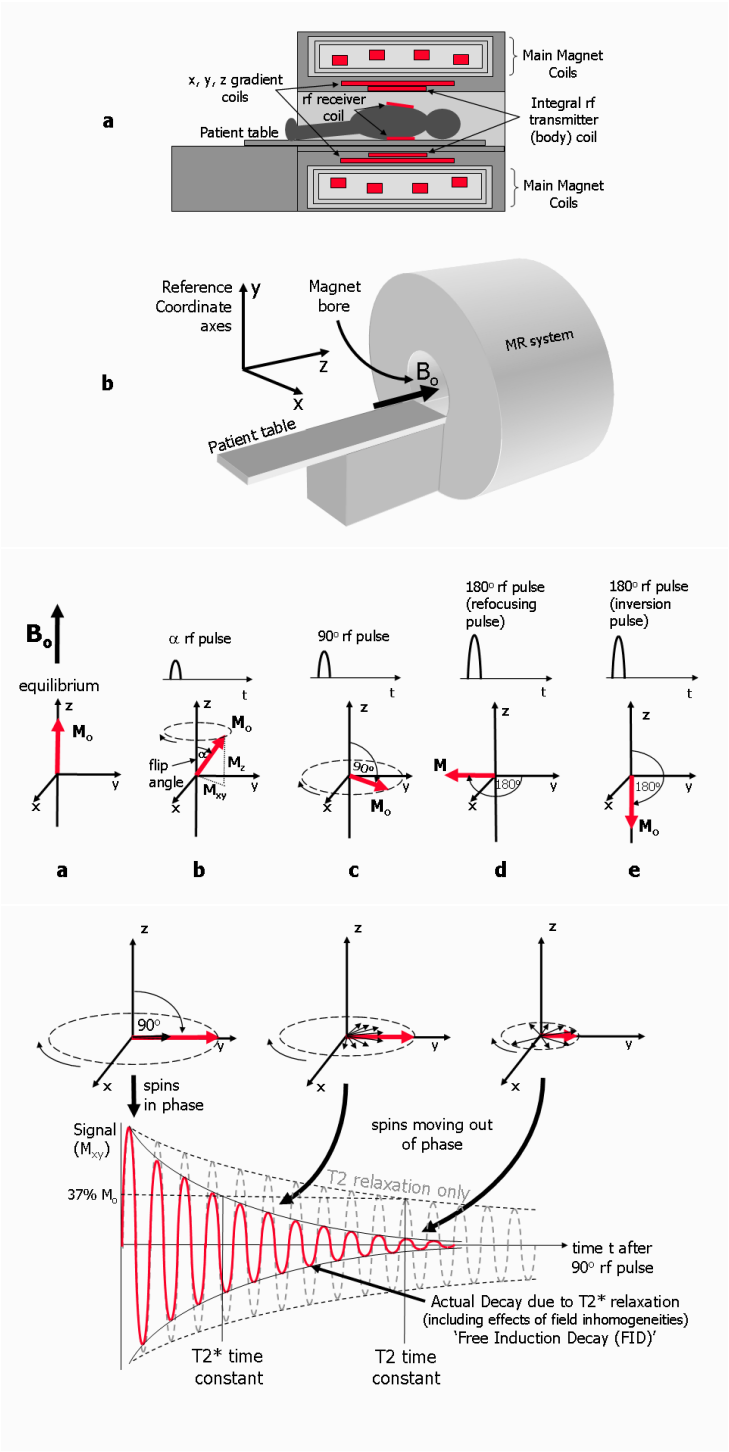

Figure 1.2 – Illustration of the MRI scanner, RF pulses and T2⋆ relaxation (Ridgway, 2010).

exciting the system with electromagnetic pulses can change the precession angle (Figure 1.2, middle).

Putting the three concepts together, by exciting the protons with an electromagnetic pulse, the spins in low-energy state will jump to the high-energy state because they absorb the additional energy from the pulse. As soon as the pulse is turned off, some of the high-energy protons will again go to the low-energy state, thereby releasing the absorbed energy until a natural equilibrium is reached. This energy release can be captured by a *radio frequency (RF) coil*.

Thus, what is measured in MRI is the energy that is discharged during the relaxation of $\mathbf{M}_0$. There are different angles to look at the relaxation rate and for fMRI in particular the transversal relaxation is of greatest importance. This means that the relaxation process is evaluated in the (x, y) plane. Take the case where an 90° pulse excites the system (Figure 1.2, bottom). Before excitation, the equilibrium state, $\mathbf{M}_0$ is parallel to the longitudinal direction (i.e. z-direction) and will present zero transverse magnetisation. After excitation, the direction of $\mathbf{M}_0$ is flipped perpendicular to the z-axis and the net magnetisation in the transverse plane will be maximal. Over time, $\mathbf{M}_0$ returns to its equilibrium and the transverse magnetisation relaxes to its original state. This transverse relaxation is referred to as T_2 relaxation. However, of more significance to fMRI is the $T_2^\star$ relaxation, which is very similar to T_2. $T_2^\star$ relaxation stems from the dephasing effect of the individual spins. Immediately after the excitation pulse, the nuclei will all spin in phase, but due to variations in $\mathbf{B}_0$ the spins start to move out of phase, meanwhile decreasing the sum of all spins (i.e. the strength of the net magnetisation $\mathbf{M}_0$). So in the end, what is measured during fMRI is the $T_2^\star$ relaxation effect, which is in general smaller than the T_2 relaxation rate.

By adding RF or gradient pulses, the spatial location of the process can be manipulated and for each location, the $T_2^\star$ relaxation is written in the frequency domain, also known as *k-space*. After applying Fourier transformation, a complex MRI image is obtained that has a real part and an imaginary part. These complex data can then be transformed to magnitude data and phase data. In practice, the magnitude data will mostly be used as the final

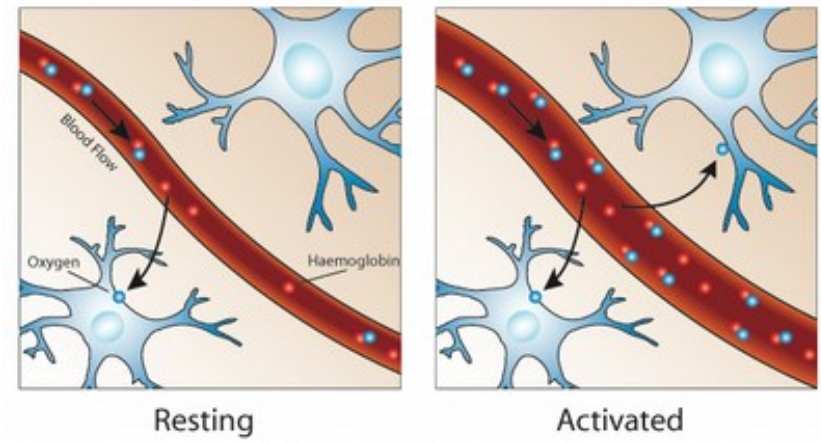

Figure 1.3 – Illustration of the neurovascular coupling and the difference in oxygenation of the blood when in rest or during activation. (source: `psychcentral.com`)

fMRI data.

1.2.2 From neural activity to the BOLD response

Neuroimaging researchers want to investigate the function of the brain and in particular the properties of the neural activity of brain cells. Unfortunately, firing neurons are not directly measurable using the MRI principle. Instead, the vascular response to the neural activity is measured. This neurovascular coupling is mainly driven by the excessive oxygen supply in response to neural activity. The main energy source of a firing neuron is oxygen. This oxygen is supplied by the haemoglobin present in the blood. Whenever neural activity occurs in the brain, the blood flow increases in order to deliver more oxygen to the neurons (Figure 1.3). So, there is an increase of oxygenated haemoglobin and it is the variation in the oxygenation of the blood that is picked up by the MR scanner. Because blood deoxygenation affects magnetic susceptibility, MR pulse sequences sensitive to $T_2^\star$ will show more MR signal where blood is highly oxygenated and less MR signal where blood is highly deoxygenated. The result is the measurement of the Blood Oxygenation Dependent Level (BOLD). This response is a correlate of the neural activity and has some

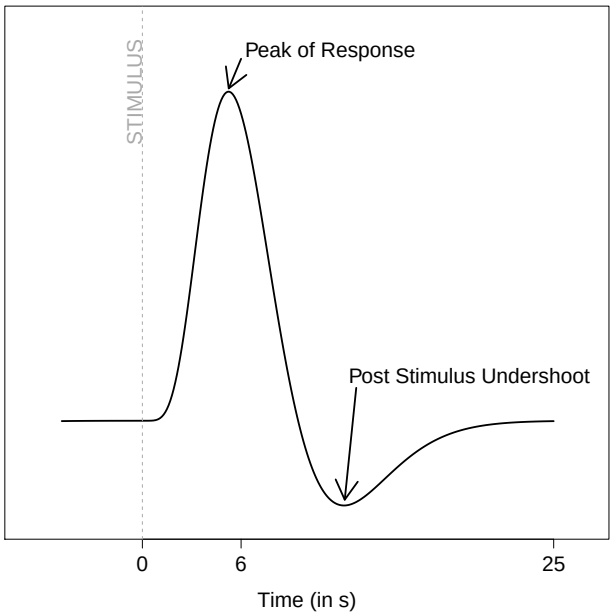

Figure 1.4 – Illustration of the typical characteristics of the BOLD response.

typical characteristics. The peak of the response is delayed and occurs around 6s after the stimulus has been presented (Figure 1.4). After the peak, the BOLD response returns to baseline, but before reaching this baseline a post-stimulus undershoot can be observed. However, the specific form of the response varies a lot across regions and within and between subjects. The exact mechanisms of the neurovascular coupling are still subject to debate, but it is accepted that the BOLD response is a plausible, yet indirect measure of neural activity in the brain (Handwerker et al., 2012).

1.2.3 Signal and noise in fMRI data

It is said that looking for BOLD signal changes in fMRI data is like looking for *a needle in a haystack*, because the fluctuations of the BOLD response are rather small and the data consist mainly of noise. Noise is here a collective noun for all unwanted fluctuations in the fMRI signal. This noise is characterised by both temporal and spatial features and stems from different

sources (see Greve et al., 2012, for a recent review).

Thermal noise

Thermal noise is caused by thermal motion of electrons within the subject and within the scanner. This thermal motion generates changes in the signal intensity over time and the higher the temperature in the scanner room, the greater the fluctuations of the signal. There is also a linear relationship between thermal noise and the magnetic field strength (Edelstein et al., 1986). For example, thermal noise of 3T data will be twice as large as compared to 1.5T data. These fluctuations are completely random and unrelated to space or time. The contribution of thermal noise to the fMRI data can easily be decreased by minimising the room temperature. Remaining thermal noise can also be eliminated by data averaging.

System noise

Variations and instabilities in the scanner hardware (e.g. nonlinearities in the gradient fields) cause system noise, which is most commonly recognised as scanner drift. Over time, the measured fMRI signal is subjected to low-frequency drift that can be present as low-frequency fluctuations in the signal but also as a linear decrease or increase of the overall signal level. Because system noise is related to the hardware, it is hard to control it during data acquisition. However, low-frequency drift removal is standard practice in fMRI data analysis and several algorithms are available.

Physiological noise

Physiological noise is inherently linked to the scanning of living subjects. Heart beat causes blood pulses through the arteries and the veins resulting in regular fluctuation in blood flow. Similarly, inhaling and exhaling air will influence the basic oxygen level in the blood. fMRI signal fluctuations related to this heart and respiratory rate are called physiological noise. This noise source can be dependent on place or time and is not necessarily random. Figure 1.5 demonstrates that physiological noise is especially troublesome

for fMRI since it is more present in grey matter (Krüger & Glover, 2001). On the other hand, in the phantom images, the influence of physiological noise is negligible. Measurement of heart beat and respiratory rate can be used to filter their contributions out of the fMRI data, but the measurement equipment has to be adapted to the scanner environment, which can be expensive.

Motion noise

When the subject is not lying completely still in the scanner, motion noise can be of great influence. The problem is that motion alters the location in space of the voxels over time, whereby it can be hard to derive time series of particular voxels. Figure 1.5d illustrates that motion noise will manifest around the edges of the brain. Motion noise is rarely random and often correlated with the task that the subject is performing (e.g. small movements each time a response button has to be pressed). Head motion is often limited by fixing the head using foam blocks. However, even then, chest movements related to the respiratory system can cause motion noise. Since the impact of motion noise for fMRI data is severe, removal of this noise source is a standard aspect of the data analysis process.

Non-task-related noise

Spontaneous neural activity and non-task-related neural activity can cause unwanted fluctuations in the BOLD signal. For example, while the subject is performing a task, his/her brain reacts to the acoustic noise of the scanner, or wonders about appointments later that day. This non-task-related noise can be greatly diminished by an optimised experimental design.

Task-related noise

fMRI data contain also noise that is correlated to the stimulus. This can be for example spontaneous activation due to the cognitive processes or movements caused by behavioural responses. Hyde et al. (2001) demonstrated that the power frequencies of resting voxels showed intense low-frequency

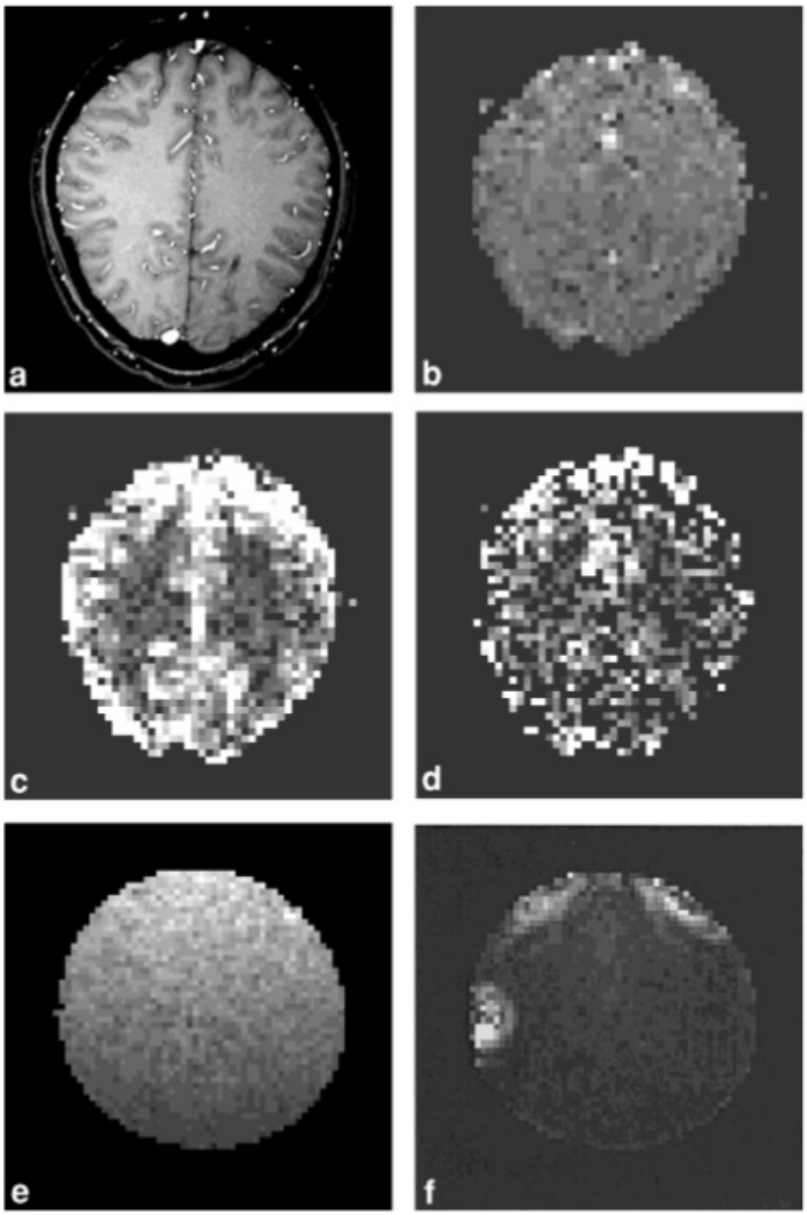

Figure 1.5 – Illustration of the distribution of noise in the brain of one subject: (a) anatomical image, (b) noise from all sources, (c) physiological noise caused by blood flow variations and metabolic processes, (d) motion noise, (e) noise from all sources in a phantom and (f) physiological noise in a phantom (From Krüger & Glover, 2001)

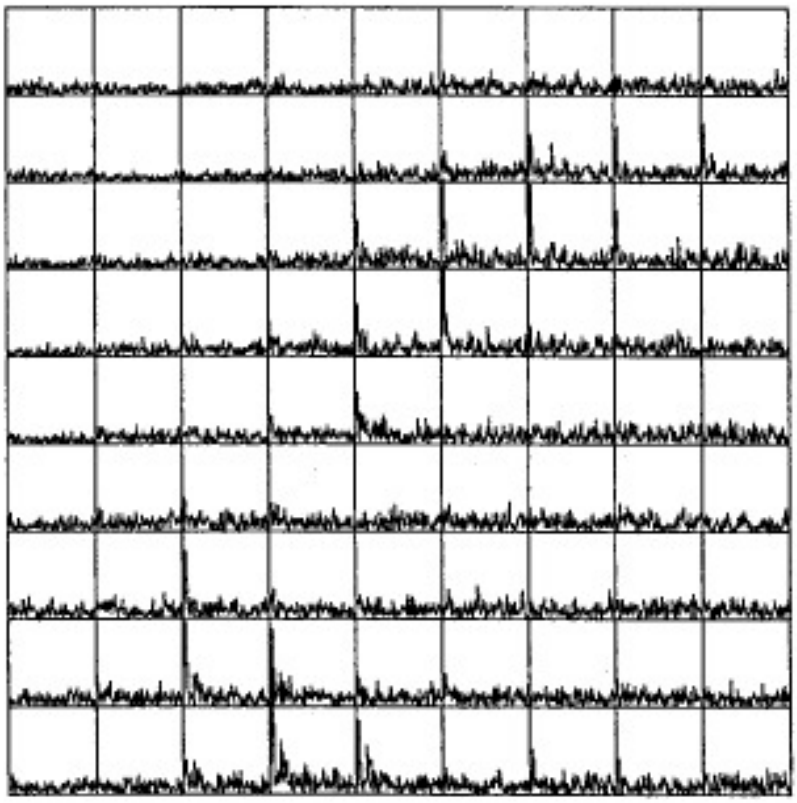

Figure 1.6 – 9×9 voxel array from a bilateral finger-tapping task. Frequency domain time series of resting-state voxels are presented. Intense low-frequency peaks (e.g. left side of the time series in voxel 4 and 5 of the last row) are observed. These frequency peaks in the resting-state voxels correspond to the frequency of the task activation (From Hyde et al., 2001)

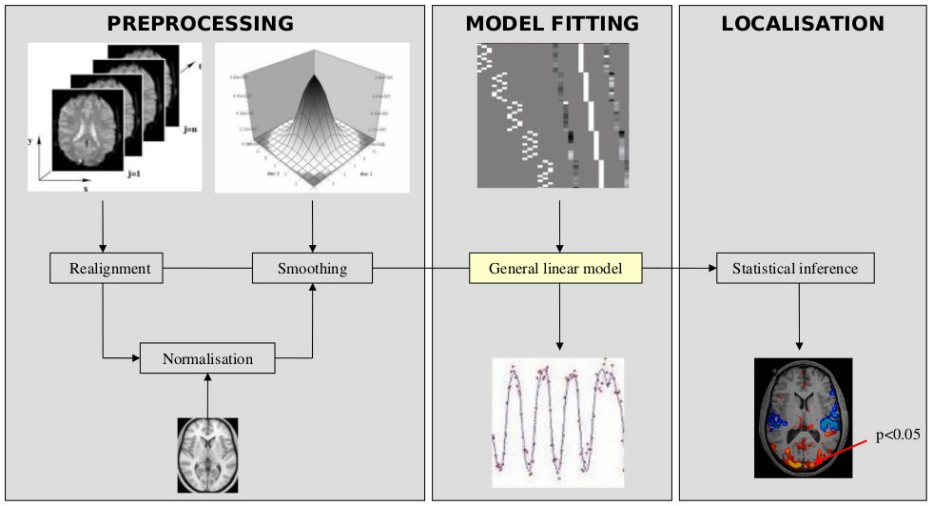

Figure 1.7 – Example of an fMRI analysis pipeline used to localise active voxels related to an experimental task

peaks that were coincidal with the task frequency (Figure 1.6), meaning that performance of the task resulted in additional noise in non-active voxels. Controlling this noise source is non-trivial and there are no standard practices to account for task-related noise.

1.2.4 fMRI data analysis: a short overview

Regardless of the specific analysis method that is used, analysing fMRI data is a pipeline process. In this pipeline, different steps involving different methods are performed. Roughly, the steps can be subdivided in preprocessing, estimation and inference (Figure 1.7). Over the years, some standard procedures have been proposed for each section in the pipeline, but the development of analysis methods for fMRI data is still ongoing research. Here, we will give a short overview of the most common procedures to localise activation (see Friston et al., 2007, for a detailed discussion).

Data preprocessing

The preprocessing of the fMRI data has several goals. Largely it involves noise reduction, but also preparing the data for inter-subject comparison. First, the influence of movements is reduced by realigning the images. Each fMRI volume is realigned with a reference image (e.g. the first image) using affine transformations. Additionally, it is possible to use the resulting estimated motion parameters later in the estimation step. Secondly, the fMRI images are normalised to a common space. Using a set of basis functions, the individual brain images are transformed until they fit a standard template. One example of such a template is the MNI152 template, which is the result of averaging 152 adult brains. Normalisation of the images makes it possible to compare the results across subjects. Finally, the realigned and/or normalised images can be smoothed. Spatial smoothing is an averaging process in which the voxels values are combined according to a Gaussian weighted kernel. The result is a blurring of the data that comes with a cost of losing some spatial resolution. However, the data averaging has the beneficial result that the noise in the data is also reduced.

Model estimation

The classical approach to localise active voxels is a mass-univariate General Linear Model (GLM) estimation. For each voxel in the brain, a GLM is fitted to the time series of that voxel. The estimation is independent for each voxel, but the GLM is based on a common design matrix. In this design matrix, the experiment is modelled using regressors that indicate when a stimulus in a particular condition has been presented. These stimulus regressors are convolved with a haemodynamic response function (HRF) to model the expected BOLD response based on the experimental design. Optionally, temporal and dispersion derivatives can be added to account for mismodelling in the delay and dispersion of the HRF. In addition to the stimulus regressors, the motion parameters estimated in the realignment step are often added to the design matrix, as well as other possible confounders (e.g. reaction times or error trials). A major difference with the classical GLM is that, during the model

estimation for fMRI data, temporal auto-correlation is taken into account.

Statistical inference

Once the model is estimated, conclusions with regard to where active voxels are located can be inferred based on the parameter estimates. For a given contrast, a corresponding statistic is calculated for each voxel, creating a parametric statistical map (SPM). Due to the mass-univariate approach, a standard thresholding of this statistical map will yield inflated Type I error rates caused by a huge multiple testing problem. The classical Bonferroni correction will be too conservative because the voxel-wise tests are not completely independent. A possible correction for this problem is controlling the Family Wise Error (FWE) rate (i.e. the probability of making even one Type I error in the family). By using properties from Random Field Theory (RFT), an estimation of the total number of independent tests is made and the inference threshold is adjusted accordingly.

fMRI analysis software

The most popular software package for fMRI data analysis is, by far, SPM[2]. This Matlab toolbox is freely available and provides an intuitive GUI interface. The software is supported by a well documented manual and numerous courses organised by the developers. Other competitive open-source software packages are FSL[3] and AFNI[4]. Both FSL and AFNI are comprehensive C libraries of functions for fMRI data analysis that are controllable through a GUI interface. The software packages provide extensive online documentation and courses are frequently organised by the developers. All three open-source software packages have the additional advantage that they are easily scriptable and can be used to develop customised analysis protocols that can be carried out automatically.

[2] `http://www.fil.ion.ucl.ac.uk/spm/`
[3] `http://www.fmrib.ox.ac.uk/fsl/`
[4] `http://afni.nimh.nih.gov/afni/`

A critical note

Despite the available software and the well-documented pipeline process, the researcher performing an fMRI data analysis is confronted with numerous options and it not clear yet how these choices influence the final result. Are the defaults necessarily optimal or is there another more suitable value/option? For example, during the preprocessing phase, one can choose between linear and non-linear realignment. The number of degrees of freedom for the realignment is also an option. Which one is better? In the spatial smoothing step, the smoothing effect is largely determined by the choice of the width of the smoothing kernel. How should one choose this value? When constructing the design matrix, the researcher is again confronted with several questions: Is it useful to add the estimated movement parameters? Could there be a benefit of including a *junk* predictor? Which HRF is best suitable for the data? Is voxel-wise correction more appropriate than cluster-based multiple testing correction? Is it better to control the number of false negatives or rather the number of false positives?

There are no standard answers to these questions and a lot of *expert knowledge* is expected from the researcher who is applying these techniques. It might not be realistic to expect the thorough statistical knowledge that is needed to make insightful decisions on these matters to be generally present. Therefore, the *methods researchers* have a major role to play in educating the *applied researchers* on *best-practices*. However, this is impossible without a comprehensive understanding of the effects that the choices have on the end result.

1.3 A statistical approach

In this dissertation, a statistical approach to the simulation of fMRI data is adopted. This is in contrast to a physical approach. In both fields, simulations of fMRI data are commonplace but the underlying research question is quite different. For example, Drobnjak et al. (2006) investigated the interaction between rigid-body motion artifacts and $\mathbf{B}_0$ inhomogeneities. MRI-based

simulations were carried out in which the Bloch equations (Bloch, 1946) were solved numerically and the size of motion distortion and $\mathbf{B}_0$ inhomogeneity was varied. In this simulation scenario, the correct modelling of the MRI physics underlying the fMRI data acquisition was especially important.

fMRI simulation studies within a statistical context are mainly driven by the question for model validation or model comparison. In this approach, the crucial aspect of fMRI data generation will be to capture the components in the data that will have an influence on the outcome of a statistical model. The physical underlying mechanisms are of lesser importance, but the challenge is to develop models that can accurately represent the fMRI data components. For example, Smith et al. (2011) conducted a large-scale simulation study to evaluate and compare connectivity analysis methods with respect to their capability to detect and induce the directionality of connections between brain regions. In this study, the influence of the MRI physics on the measurement of the brain network is not of interest, since, first of all, this influence will be equal for all statistical techniques that were evaluated, and second, data across subjects will most likely be acquired using the same pulse sequence or scanner. On the other hand, noise components in the fMRI data that possibly hinder accurate detection of the brain network should be modelled, but the characteristics of these sources can be defined without explicit modelling of the MRI physics.

1.4 Motivation and outline

Due to the complexity of fMRI data and the variety of available statistical methods, thorough validation of these methods is a must. In order to be able to validate a statistical method, the *ground truth* of the data has to be known. In the case of fMRI data this ground truth is hard to come by and simulations offer an easy and priceless validation tool. The original goal of this doctoral thesis proposal was to validate statistical methods for fMRI data using Monte Carlo simulations. The validation studies would have a twofold purpose. First, feedback could be given on the validity, efficiency and robustness of the statistical methods, and second, guidelines could be

provided on the optimal choices for the analysis parameters. However, the search for an appropriate way to simulate fMRI data failed and it became clear that there is still confusion in the literature on how fMRI data should be simulated. Consequently, the focus of the doctoral project shifted to the simulation of fMRI data *an sich*.

This dissertation has multiple goals, which correspond to the respective chapters. First, an investigation of the fMRI simulation literature was carried out (Chapter 2). In this review, current simulation studies were assessed with a specific focus on how the fMRI data were generated. The most common data generating processes were evaluated based on their correspondence to real fMRI data. This literature survey raised a number of issues: (1) There is no consensus on how to define signal-to-noise ratio for fMRI data; (2) fMRI simulation studies are carried out using *ad hoc* and *in-house* simulation scripts; (3) a great discrepancy was observed between the structure of the noise that is present in fMRI data and the model for fMRI noise that is used in simulations. These observations led to the following four studies in this dissertation.

Chapter 3 is a comment on the definition of signal-to-noise (SNR) and contrast-to-noise (CNR) ratio for fMRI data. In this comment, the comparability between fMRI studies is challenged based on the variety of SNR and CNR values that are reported. This variety stems from the use of several definitions that differ in how the signal of interest is operationalised. The advantages and drawbacks of the definitions are pointed out and tools are provided to increase the transparency of fMRI studies.

Chapter 4 meets the lack of simulation software for fMRI data. An R package, **neuRosim**, is presented that bundles functions to generate fMRI data. Special attention is paid to how the simulated data are representative for real data and how the different components are modelled. The software package provides fast, flexible and intuitive simulation of fMRI data. All the simulation studies presented in this dissertation used **neuRosim** for the data generation.

In chapter 5 the noise model of simulated fMRI data is investigated. In a series of simulations, the impact of different noise models on the conclusions of

fMRI simulation studies is assessed. Special attention is given to the specific effect of adding physiological noise to the data generating process and proof is provided that ignoring this noise source can lead to biased conclusions.

In the final study (chapter 6), a demonstration is provided how simulation studies can be used for the validation of analysis methods for fMRI data. In this application of fMRI simulation, spatial smoothing techniques and inference methods are systematically evaluated based on their effect on the sensitivity and the specificity of an activation detection analysis.

Chapter 7 provides an overview of the main findings and conclusions in each study. The implications of these results are discussed in a more general framework and some topics for future research are suggested.

Chapters 4 and 5 are published papers in the Journal of Statistical Software and the Journal of Neuroscience Methods, respectively. Chapters 2, 3, and 6 are submitted. Full bibliographic details of these chapters are listed in the bibliography at the end of this dissertation.

References

Bloch, F. (1946). Nuclear induction. *Physical Review*, *70*(7–8), 460–474.

Bloch, F., Hansen, W., & Packard, M. (1946). Nuclear induction. *Physical Review*, *69*, 127.

Drobnjak, I., Gavaghan, D., Süli, E., Pitt-Francis, J., & Jenkinson, M. (2006). Development of a fMRI simulator for modelling realistic rigid-body motion artifacts. *Magnetic Resonance in Medicine*, *56*(2), 364–380.

Edelstein, W., Glover, G., Hardy, C., & Redington, R. (1986). The intrinsic signal-to-noise ratio in NMR imaging. *Magnetic Resonance in Medicine*, *3*, 604–618.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., & Penny, W. (Eds.). (2007). *Statistical parametric mapping: The analysis of functional brain images*. Massachussets, USA: Academic Press.

Gentle, J. (2005). *Elements of computational statistics*. New York, USA: Springer.

Greve, D., Brown, G., Mueller, B., Glover, G., & Liu, T. (2012). A survey of the sources of noise fMRI. *Psychometrika*.

Handwerker, D., Gonzalez-Castillo, J., D'Esposito, M., & Bandettini, P. (2012). The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage*, *62*(2, SI), 1017-1023.

Hyde, J., Biswal, B., & Jesmanowicz, A. (2001). High-resolution fMRI using multislice partial k-space GR-EPI with cubic voxels. *Magnetic Resonance in Medicine*, *46*, 114–125.

Krüger, G., & Glover, G. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *46*, 631–637.

Kwong, K., Belliveau, J., Chesler, D., Goldberg, I., Weisskoff, R., Poncelet, B., et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 5675–5679.

Ogawa, S., Tank, D., Menon, R., Ellermann, J., Kim, S., Merkle, H., et al. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 5951–5955.

Purcell, E., Torrey, H., & Pound, R. (1945). Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, *69*, 37–38.

Ridgway, J. (2010). Cardiovascular magnetic resonance physics for clinicians: part i. *Journal of Cardiovascular Magnetic Resonance*, *12*.

Robert, C., & Casella, G. (2005). *Monte carlo statistal methods. second edition.* New York, USA: Springer-Verlag.

Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, *35*, 137–167.

Smith, S., Miller, K., Salimi-Khorshidi, G., Webster, M., Beckmann, C., Nichols, T., et al. (2011). Network modelling methods for FMRI. *NeuroImage*, *54*, 875–891.

Stigler, S. (1978). Mathematical statistics in the early states. *Annals of Statistics*, *6*, 239–265.

2 | A review of fMRI simulation studies

Marijke Welvaert & Yves Rosseel
Submitted manuscript

Abstract

Simulation studies that validate statistical techniques for fMRI data are challenging due to the complexity of the data. Therefore, it is not surprising that no common data generating process is available (i.e. several models can be found to model BOLD activation and noise). Based on a literature search, a database of simulation studies was compiled. The information in this database was analysed and critically evaluated focusing on the parameters in the simulation design, the adopted model to generate fMRI data and on how the simulation studies are typically reported. Some striking findings are discussed and some guidelines are provided that could improve the quality of fMRI simulation studies.

2.1 Introduction

Twenty years ago, functional magnetic resonance imaging (fMRI) was founded as a method to measure brain activity (Kwong et al., 1992; Ogawa et al., 1992). In these past twenty years, this technique has been used increasingly and has pioneered the search to map and connect the brain that caused a world-wide collaboration of scientists from different disciplines. Engineers and physicists are intrigued by the acquisition of the fMRI data, while physicians and psychologists are challenged to adapt their behavioural experimental protocols to the scanner environment. Last but not least, the analysis of fMRI data has been, and still is, a topic of numerous discussions among statisticians. The latter are confronted with the fact that the data acquired through fMRI have no ground truth. This ground truth is needed to ensure validation of the statistical methods that are used to analyse the data and to assess statistical properties such as sensitivity, specificity, bias and robustness. Despite great efforts to develop mechanical models (Brosch et al., 2002) or measuring neural activity with intracranial EEG (David et al., 2008), simulations are probably still the most feasible way to establish the ground truth of fMRI data.

NeuroImage, one of the flagship journals in the neuroimaging community, celebrated the 20th anniversary of the first fMRI publications with a special volume that consisted of 103 reviews about the early beginnings, developments in acquisition, software, processing and methodology, and prospectives for the future (Bandettini (Editor), 2012). Although the advances in statistical methods for fMRI data are discussed in several of these reviews, simulations *an sich* are not mentioned. In general, it appears that simulation studies are still not standard practice for fMRI methods validation. A possible explanation is that it can be quite challenging to simulate fMRI data. Not only is the coupling between the neural activity and the Blood Oxygenation Dependent Level (BOLD) not completely understood (Handwerker et al., 2012), fMRI data are also characterised by a great deal of noise coming from multiple sources (Greve et al., 2012). Consequently, no common data generating process for fMRI data is available and the data generation

in fMRI simulation studies is mostly defined *ad hoc*.

The goal of this review is to provide an overview of the most common data generation methods used in fMRI simulation studies. An established and accepted data generating process does not yet exist and therefore an investigation of the existing published models is called for. Especially the validity of these data generating methods is analysed and the overall reporting and conducting of fMRI simulation studies is critically reviewed. The rest of the paper is organised as follows: In the Methods section the article selection criteria are reported that were applied to establish a database of fMRI simulation studies literature, and the focus points of the article evaluation are discussed. The Results section focuses on different aspects of the simulation studies, namely, the goals of the studies, the experimental design under investigation, the simulation parameters and the data generation models. Finally, in the Discussion, guidelines and best practices are provided to increase the reliability and generalisability of fMRI simulation studies.

2.2 Methods

2.2.1 Article selection

Articles were selected from the Web of Science database using the following query: "*fmri* AND *simulation* AND (*statistics* OR *data analysis*)". By excluding articles labelled as reviews or proceedings, this search resulted in 318[1] hits. All these articles were manually inspected on content and relevance. This screening resulted in excluding articles based on the following criteria: the conducted simulations were for another modality (e.g. PET, EEG, MEG, ...); no time series were simulated (e.g. inference methods are often validated on simulated statistical maps); non-human fMRI was simulated; and no simulation study was conducted (e.g. papers presenting simulation software). After exclusion, the remaining 119 articles were taken into account in this analysis. Full bibliographic details of our sample can be found in the appendix. These articles were published in 39 peer-reviewed academic

[1] Result as of January, 1st 2013

Table 2.1 – Overview of journals in the survey. Full details of the included studies can be found in the appendix.

Journal title	Number of articles
NeuroImage	37
Human Brain Mapping	11
IEEE Transactions on Medical Imaging	10
Magnetic Resonance Imaging	7
IEEE Transactions on Biomedical Engineering	6
Journal of Magnetic Resonance Imaging	6
Magnetic Resonance in Medicine	4
Other	38

journals (Table 2.1) over a period of 16 years (Figure 2.1). In this sample, most simulation studies were published in NeuroImage (37), Human Brain Mapping (11), IEEE Transactions on Medical Imaging (10), Magnetic Resonance Imaging (7), IEEE Transaction on Biomedical Engineering (6) and the Journal of Magnetic Resonance Imaging (6).

During our article selection, we focused on simulation studies conducted to validate or compare analysis procedures for BOLD-fMRI data. In order to perform this validation, a data generating process is resulting in ariticial data reflect to some degree the characteristics of real measured fMRI data. From a statistical perspective, scanning parameters that influence magnetic properties of the data (e.g. flip angle) are of less importance since they mainly have an effect on the signal-to-noise ratio. For instance, when these scanning parameters are optimised, the baseline signal might increase while the noise level decreases. The crucial aspect is to determine the components in the data that are expected to have an effect on the data analysis and model these components while generating the simulated fMRI data.

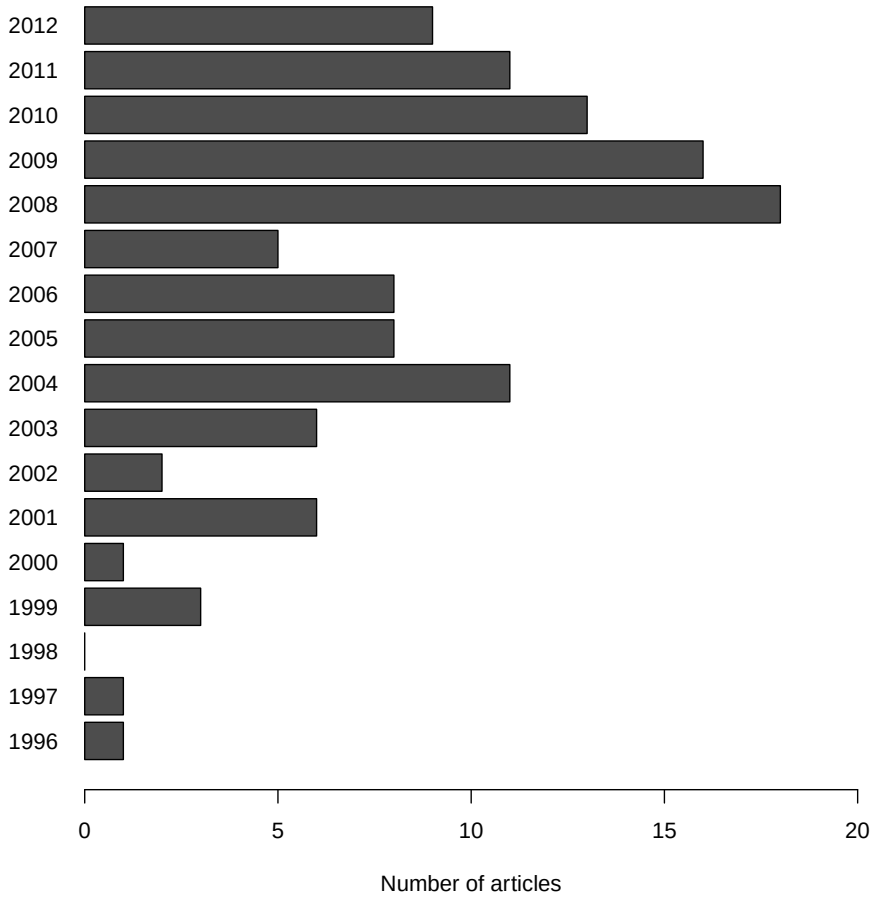

Figure 2.1 – Overview of number of articles for each publication year included in the survey

2.2.2 Article evaluation

In the present study, we analysed the sections describing the simulation study for the selected papers. Where necessary the appendices or supplementary materials were also included and whenever there was still missing information after screening these sections, the whole paper was searched for this information. Only the reported methodology was evaluated (i.e. no authors were contacted for more information).[2] For each study we evaluated the goal of the simulation study, the simulation parameters and the data generating process. In the case that multiple simulation studies were present in the article, this information was retrieved from the most complex case that was described. In the following section, summarised results are presented. For a detailed results list on the individual study level, the reader is referred to Table 2.3.

2.3 Results

2.3.1 Study goals

Simulation studies are conducted to evaluate statistical models based on a given experimental design. For each article we assessed which statistical technique was validated. Six categories of statistical models were distinguished (see Figure 2.2, left panel). Most simulation studies (25.2%) are conducted for signal decomposition models like Principal Components Analysis (PCA), Independent Component Analysis (ICA) and Wavelet analysis. This group of methods is closely followed by General Linear Model (GLM) analysis, Likelihood Ratio Tests (LRT) and *t*-tests (23.5% of the selected studies). 11.8% of the simulation studies investigate properties of classification techniques using for example Support Vector Machines or cluster analysis. Methods that are

[2]There might be a discrepancy between the conducted and reported simulation studies (e.g. not all details are mentioned), however, to ensure reproducible science all critical elements should be reported. It may not always be feasible to report everything in the main text, but academic journals allow for crucial content to be described in appendices or through online supplementary materials.

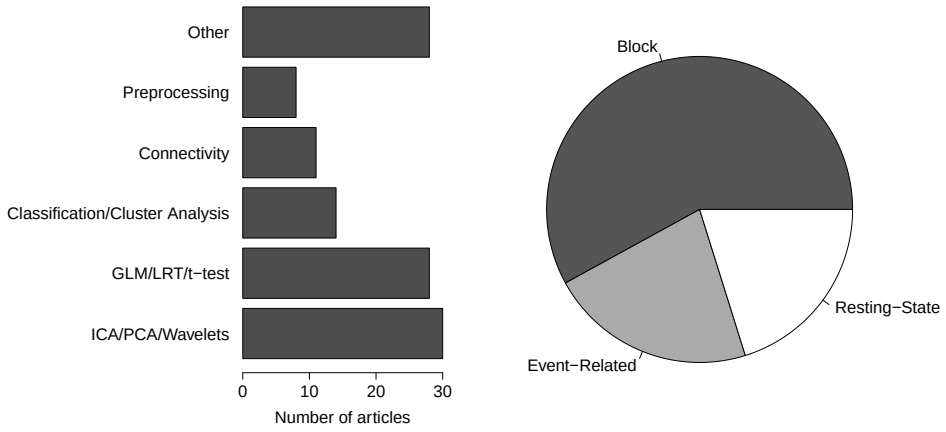

Figure 2.2 – Statistical models (left) and experimental designs (right) investigated in the selected articles.

less represented in our sample are connectivity analyses (9.2%) or preprocessing methods like motion correction and spatial smoothing (6.7%). All studies that did not use any of the previous methods (23.5%) were gathered in a rest category. In this category are included, for example, HRF estimation methods, spatio-temporal models, bootstrapping and nonparametric techniques.

2.3.2 Experimental design

The methods described above are validated using a given experimental design (Figure 2.2, right panel). The majority of simulation studies (58%) report using a block design for the generation of the BOLD activity. When this design is not used, modelled activation is based on an event-related design (21.8%) or it concerns a resting-state study (20.2%).

2.3.3 Simulation parameters

The general goal of a simulation study is to research a certain outcome (e.g. power, bias, ...) under several conditions (e.g. noise level, HRF variability, ...). The most common method to achieve this goal is by conducting a Monte

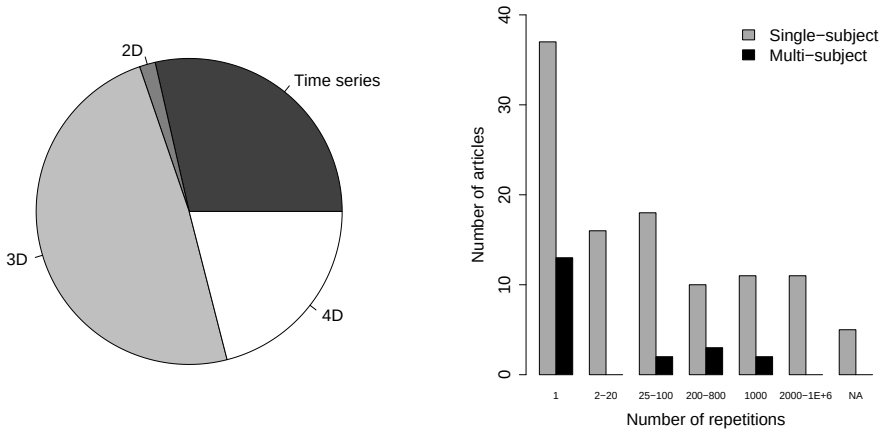

Figure 2.3 – Overview of the dimensions of the simulated data (left) and the number of replications for single-subject and multi-subject simulations (right).

Carlo experiment. The simulation reports in our database were evaluated on the dimensions of the simulated data, the number of replications and parameter variation.

Data dimensions

fMRI data have in essence four dimensions (i.e. coordinates in an xyz-space and time). However, the majority of articles in our sample (48.7%) published results for 3D data where time series are simulated for all voxels in a single slice (Figure 2.3, left panel), while 21% considered full 4D fMRI data. On the other hand, 28.6% of the articles reported simulating fMRI time series only with no spatial context. In this case, mass-univariate techniques were mostly evaluated that also regard fMRI data as being multiple measurements of single time series. A very small proportion (1.7%) considered two-dimensional data. This was reported exclusively in an ICA validation context, where the fMRI data are organised as *voxels* $\times$ *timepoints*.

Table 2.2 – Proportions of studies reporting parameter variation and justification of the chosen parameter values

	Justification of value	
Parameter variation	No	Yes
No	20.2%	10.9%
Yes	32.8%	36.1%

Replications

83.2% of the selected articles considered single-subject data, while the remaining 16.8% simulated data for multiple subjects. In these last studies, the number of subjects that was simulated corresponded typically with sample sizes reported in real fMRI studies (e.g. 4 to 20 subjects) and the data for these subjects were mostly simulated once (with a few notable exceptions, see Figure 2.3, right panel). For the single-subject simulation studies, the number of repetitions was higher in the majority of the studies, but still 37.4% of the articles reported only 1 replication of the simulated data for each setting of the manipulated parameters. It should be noted that simulating 3D or 4D datasets without any spatial correlations is equal to the simulation of fMRI time series with n replications where n is the number of voxels. This was true for 22 of the 37 studies that reported using 1 replication. However, for the remaining studies conclusions are based on 1 realisation of the data. Two studies reported simulating time series just once for each setting of the simulation parameters.

Parameter variation

Other possible parameters taken into account in the simulations were, for example, strength of the modelled activation, number of time points, noise level, repetition time (TR), etc. The relevance of a simulation study depends highly on the representativeness of these chosen parameter values. To ensure that the parameters are characteristic for fMRI data, it is recommended that a range of values is evaluated. Additionally, a justification is expected on why

specific values of certain parameters are chosen. In our sample both requirements were assessed (see Table 2.2 for an overview). A study was classified as using varying parameters as soon as more than one value of a specific parameter was considered. Whenever a reason for choosing a specific parameter value was reported, the simulation study was evaluated positive on the justification of the chosen parameters. About one third of the studies reported a variation in the values and gave a justification for their choices. Frequently reported variations were several noise levels and activation strengths that were taken into account. As for the choices of the values, authors mostly justified these as being realistic values in real fMRI data or being estimated from real data. However, 32.8% of the studies reporting variation of the parameters did not give any justification, 10.9% did justify the choice of the parameter values but only used one specific value for each parameter, while one fifth of the studies in our sample (20.2%) did neither.

2.3.4 Data generation models

Of all simulation studies investigated, 84% were pure synthetic simulations while the other 16% adopted a hybrid simulation strategy. In hybrid simulations, a resting-state dataset is acquired and synthetically generated activation is added to these data. As such, knowledge of the ground truth is assured while the noise is representative for real data. However, manipulation of the noise in the simulated fMRI data is not possible and replicating the data will be a costly process. Therefore, in most simulations the fMRI data are generated completely artificially.

All synthetic simulation studies adopted an additive data generation model (e.g. Bellec et al., 2009) in which three main components can be distinguished: (1) a baseline signal, (2) BOLD activation and (3) noise. However, half of the studies did not report using a baseline for the data, so we could assume that this is zero for these studies. For the other half, 47% used a static baseline, for example a constant when simulating time series and a template slice or volume that was repeated for each time point in the case of simulating 3D or 4D fMRI data. A minority of the studies (3%) used a

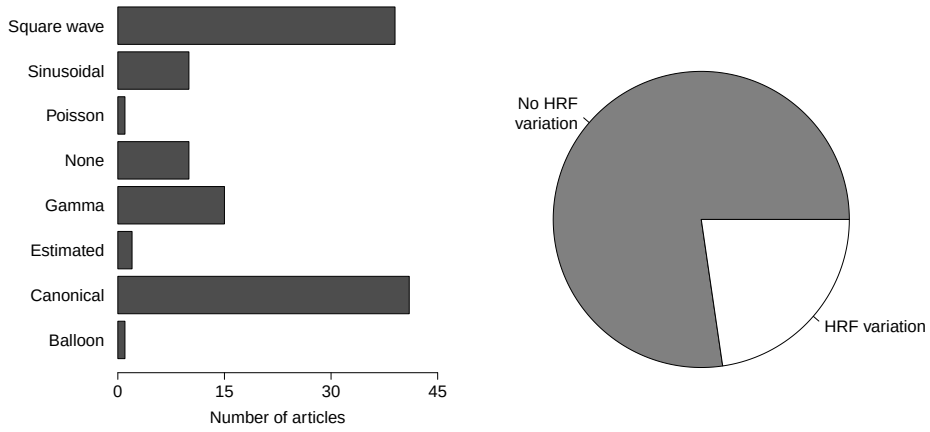

Figure 2.4 – Overview of the different HRF functions used in the simulation studies (left) and whether HRF variability was taken into account (right).

varying baseline, meaning that the baseline values were varied over time, e.g. to model thermal shifts (Backfrieder et al., 1996).

BOLD response

An important component in the simulated fMRI data is the BOLD response because this signal defines the ground truth in the simulation studies. Despite the fact that the coupling between the neural activation and the BOLD response is still not completely understood (Handwerker et al., 2012), several models are available to generate a haemodynamic response function (HRF). See Figure 2.4 for an overview of the used models in the selected articles. Those methods are, for example, a gamma function (Boynton et al., 1996; Cohen, 1997), a difference of two gamma functions, also known as the canonical HRF (Friston et al., 1998; Glover, 1999) or the Balloon model (Buxton et al., 1998, 2004). Usage of these models was respectively reported in 12.6%, 34.5% and 0.8% of the articles. Nevertheless, 32.8% of the reported simulation studies disregarded any BOLD characteristics and chose a square wave (i.e. a boxcar function) to represent the BOLD activation in the simulated fMRI data. When no experimental task was simulated, resting-state activation was predominantly modelled as a set of sinusoidal functions (8.4% of

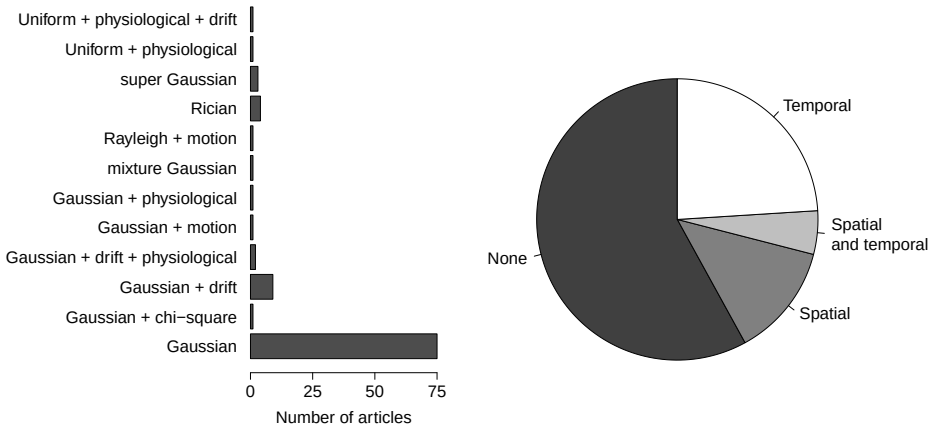

Figure 2.5 – Overview of the noise models in the synthetic simulation studies (left) and the reported use of correlated noise (right).

the total sample), although 8.4% of the selected studies did not simulate any BOLD activation. The shape of the HRF varies immensely from brain region to brain region and also from subject to subject. 22.7% of the simulation studies reported modelling this variation in the HRF parameters, while the majority (77.3%) considered a fixed HRF in all simulations (Figure 2.4, right panel).

Noise model

Noise is not only characteristic for fMRI data but also ensures generalisability of the conclusions based on simulations. All simulation studies incorporated some noise generating process (see Figure 2.5, left panel, for an overview). The vast majority of the synthetic simulation studies (i.e. 75%) selected the noise randomly from a Gaussian distribution. An additional 9% added also some drift function to this noise, while about 7% of the studies considered a skewed noise distribution (e.g. Rician or super Gaussian distribution). The remainder of the studies used a very specific noise model (for example by adding physiological noise, using a uniform distribution or adding motion correlated noise), because they focused on the effects of these noise sources. fMRI noise is also known to be spatially and temporally correlated. However,

58% of the selected articles did not report modelling any correlations in the noise (Figure 2.5, right panel). Temporal correlation (24%) was almost exclusively modelled as an autoregressive autocorrelation process. Typically this process was of order 1, but there are exceptions that used a model order of 3 or 4. Spatial correlations (13%) were typically created by spatial smoothing of the generated noise. A small fraction of the studies (i.e. 5%) models both spatial and temporal correlations.

2.4 Discussion

Whenever statistical models are validated based on simulations, the model that is used for the data generation is of utmost importance. In this paper, a survey was conducted to list currently used data generation models. Based on 119 research articles we described the simulation type, use and justification of simulation parameters and the different components in the fMRI data generating process. The survey results showed that current fMRI simulation studies sometimes lack a thorough experimental manipulation. The parameters in the simulation study (e.g. noise level, TR, HRF delay, ...) are not always varied, while representative values of some of these parameters are not known. Further, the number of replications is a major topic of concern. It was surprising to observe that the conclusions of some of the simulation studies were based on only one replication of the random data generating process. The external validity of these simulation can be questioned. In general, the goal of a simulation study is to provide stable results. Therefore, it is hard to belief that only one replication will suffice.

Model-based versus data-based simulation

In the majority of the reported simulation studies, the fMRI data were generated based on the same model as the model that was being validated (e.g. generating time series from a VAR model to evaluate Granger causality). As such, the simulation is entirely model-based and the assumptions of the model under investigation are completely met. Consequently, the conclusions

of these simulation studies give only partial information on the applicability of these models as an analysis tool for fMRI data, since fMRI data generally do not meet the assumptions of most statistical models. A better practice would be to start from the data themselves and to define a data generating process that models the different sources that are present in fMRI data. By using data-based simulations, the properties of the analysis techniques can be assessed in more realistic circumstances.

In this context, it should be noted that the data generating proces used in most current simulation studies is not compatible with the knowledge on how fMRI data are constructed. For instance, it is well-known that the BOLD response is the result of a haemodynamic coupling to neural activity. Although the precise dynamics are perhaps still debatable, there is consensus about the BOLD signal being a delayed response with varying dynamics over the brain regions and between subjects. Additionally, there could be nonlinearities in the signal. Therefore, it came to us as a big surprise that about one third of the reported simulation studies in our database did not model any of these characteristics and used a simple boxcar function to distinguish stimulus induced activation from rest. About the same number did model the slow emergence of the BOLD signal by using a canonical HRF, but only a small fraction (i.e. two studies) did also model BOLD nonlinearities by means of the Balloon model. In the case of spontaneous neural activation (for example in resting-state studies), BOLD fluctuations were mostly modelled through sinusoidal functions with frequencies that are commonly observed in resting-state studies. However, describing these spontaneous fluctuations by sinusoids stems from the tradition to use ICA to analyse these data and is again more compatible with the model under investigation than being representative for the data. Further, variability of the BOLD response was taken into account only in about one fifth of the simulation studies. With regard to modelling BOLD activation, in a data-based simulation context at least some form of HRF should be used that takes into account the basic characteristics of the BOLD signal, while any variation of the parameters of this model will enhance the generalisability of the simulation results.

The generation of fMRI noise causes also a discrepancy between simulated

and real fMRI data. The noise in fMRI consists of several sources (Lazar, 2008; Greve et al., 2012), for example thermal noise, motion related noise, physiological noise and task-related noise. Nevertheless, the vast majority of simulation studies investigated here have only used a white Gaussian noise model to generate fMRI noise ignoring its multiple-source character. In some cases, spatial or temporal correlations are added. Again, this noise model is consistent with many of the statistical models for fMRI data (e.g. GLM). Unfortunately, the Gaussian noise model only accounts for a fraction of the noise in real data. One solution is to use hybrid simulations in which using real noise acquired in a resting-state study increases the realistic character of the simulated data. However, it is impossible to manipulate noise related parameters and unwanted activation in resting-state data can influence the simulation results. Moreover, multiple replications (i.e. acquiring resting-state data from multiple subjects) are costly. Perhaps the better solution is to model more than only Gaussian noise (i.e. thermal noise) and also include, as has been demonstrated in several simulation studies, motion noise, physiological noise, signal drift, etc. In some simulation studies, the results will not be altered under a full noise model. It may not always be necessary to include all noise sources (e.g. if a certain noise source is removed or the influence of a source is assumed to be equal in all conditions), but this should be motivated at least. To assure generalisability of the simulation results, a more complex noise model, compared to the one that is generally adopted now, might be imperative.

Guidelines for simulating fMRI data

Based on these results we present some guidelines to improve the reliability and generalisability of fMRI simulation studies.

1. All parameters for which a value is chosen in the simulation experiments should be thoroughly justified. If a single value is not agreed upon, a range of values should be evaluated (see Bellec et al., 2009; Park et al., 2012; Penny, 2011; Sturzbecher et al., 2009, for some examples).

2. The conditions in the simulation study, (e.g. statistical model, parameter values,...), have to be combined in an experimental design. The construction of this experimental design in essential (Skrondal, 2000). Factors that can be considered in the experiment are, for example, variations of parameter levels, analysis methods and number of replications. The most complete design is the full-factorial design, although there might be reasons to adopt fractional designs. Based on the experimental design, the simulation experiment will have external validity (i.e. its results can be generalised beyond a given experiment).

3. A Monte Carlo experiment has to be repeated to exclude random influences on the simulation results. Therefore, a sufficient number of replications of the experiment has to be performed. In the case of time series simulations, at least 10000 replications might be necessary, while for the simulation of 3D or 4D fMRI data a total of 100 might be enough. In general, the more replications, the better. For example, Sturzbecher et al. (2009) generated 10000 replications of 3D datasets, and Park et al. (2012) simulated 4D multi-subject datasets to represent twin data using 500 replications of each paired dataset. In practice, this number can be limited due to time or computational constraints. Whenever in doubt, the convergence of the results should be tested.

4. The simulated task-related activation signal should reflect known properties of the BOLD response. This includes, but is not limited to, response delay, nonlinearities and inter-region and -subject variability. Either the canonical HRF or the Balloon model can be used (see Johnston et al., 2008, for an example using the Balloon model).

5. fMRI noise is partially white (i.e. system noise) and this part can be modelled by random Gaussian noise. However, additionally one should account for (residual) motions, heart rate and respiratory rate fluctuations, task-related noise and spatial and temporal correlations (see, for example, Bellec et al., 2009; Fadili et al., 2001; Schippers et al., 2011).

6. If either the BOLD model or the noise model is simplified, this should

be duly motivated.

Conclusion

The use of simulation studies to validate statistical techniques for fMRI data should be highly encouraged, because simulation experiments are a fast and costless tool to assess the quality and applicability of the analysis techniques. However, our survey of the fMRI simulation literature raised several concerns with respect to simulation studies as they are conducted now. The observation that the number of fMRI simulation studies seems to decrease the last couple of years is troubling. Furthermore, it was demonstrated that the data generating process used to simulate fMRI data is often model-based and parameter variation in the data generating process is not standardly implemented.

A possible reason for the absence of a common fMRI data generation model might be the lack of established software packages. Current simulation studies are mainly conducted using in-house software routines that have no common programming language and are not widely available. Recently, developments to fill this gap have resulted in the release of software packages that provide a flexible and fast framework for fMRI simulations (Welvaert et al., 2011; Erhardt et al., 2012). Using these software packages can be an important step in the right direction. Additionally, by taking into account the different sources present in fMRI data and adopting a complete simulation design with sufficient replications, conclusions from fMRI simulation studies can be expected to be more reliable.

Researchers that conduct fMRI simulation studies are encouraged to implement the guidelines presented in this paper in order to increase the reliability and generalisability of the conclusions from simulation studies.

Table 2.3 – Detailed results for the analysis of the fMRI simulation database. The ID numbers refer to the references in the appendix.

ID	Model	Design	dim.	nS	rep	parV	parJ	HRFm	HRFv	type	Noise model	Noise correlations
1	Mutual Information	block	4D	no	30	yes	no	gamma	yes	synthetic	Gaussian	temporal
2	ICA	rest	3D	yes	1	yes	yes	canonical	no	synthetic	Rician	none
3	Cortical Surface Mapping	rest	2D	no	500	yes	no	square wave	no	synthetic	Gaussian	spatial
4	PCA	block	3D	no	1	yes	no	square wave	no	synthetic	Gaussian	none
5	ICA	block	4D	no	1	no	yes	sinusoidal	no	synthetic	Uniform + physiological	none
6	Parametric Bootstrap	block	4D	yes	1	yes	yes	canonical	yes	synthetic	Gaussian + drift + physiological	none
7	cluster analysis	rest	3D	yes	1	yes	no	none	–	synthetic	Gaussian	temporal
8	GLM	ER	1D	no	1	yes	yes	gamma	yes	synthetic	Gaussian + motion	none
9	ICA	rest	3D	no	1	no	yes	sinusoidal	no	synthetic	Gaussian	none
10	spatial smoothing	block	3D	no	1	no	no	square wave	no	synthetic	Gaussian	spatial & temporal
11	wavelets	ER	1D	no	10^5	no	no	canonical	no	synthetic	Gaussian	none
12	wavelets	block	1D	no	10^5	yes	yes	canonical	no	synthetic	Gaussian	none
13	ICA	rest	1D	no	30	yes	no	sinusoidal	no	synthetic	Gaussian	none
14	GLM	block	1D	no	1	yes	no	canonical	yes	synthetic	Gaussian	none
15	Spatio-temporal	rest	3D	no	100	yes	no	sinusoidal	no	synthetic	Gaussian	spatial
16	BOLD estimation	ER	1D	no	200	yes	no	canonical	no	synthetic	Gaussian + drift	temporal
17	BOLD estimation	ER	3D	no	200	yes	no	canonical	no	synthetic	Gaussian + drift	none
18	ICA	block	3D	no	1	no	yes	square wave	no	synthetic	Gaussian	none
19	ICA	block	3D	no	1	yes	no	none	–	synthetic	Gaussian	none
20	cluster analysis	block	3D	no	1	yes	no	square wave	no	synthetic	Gaussian	none
21	t-test	block	4D	no	1	no	yes	square wave	no	hybrid	–	–
22	preprocessing	block	3D	no	500	yes	no	canonical	no	synthetic	Gaussian	spatial
23	classification	block	4D	no	1	yes	yes	canonical	yes	synthetic	Gaussian	temporal
24	non-parametric	block	1D	no	10^4	yes	yes	canonical	no	synthetic	Gaussian	temporal
25	LRT	block	1D	no	1000	yes	yes	canonical	no	synthetic	Gaussian	temporal
26	wavelets	block	3D	no	4	yes	yes	square wave	no	synthetic	Gaussian	spatial
27	Granger causality	rest	1D	no	10^5	yes	no	canonical	no	synthetic	Gaussian	temporal

ID – paper identification number; dim. – data dimension; nS – Multiple subjects?; rep – Number of replications
parV – Parameter variation?; parJ – Parameter justification; HRFm – HRF model; HRFv – HRF variation?

Table 2.3 – *Continued*

ID	Model	Design	dim.	nS	rep	parV	parJ	HRFm	HRFv	type	Noise model	Noise correlations
28	GLM	block	1D	yes	1000	yes	yes	square wave	no	synthetic	Gaussian	none
29	cluster analysis	block	3D	no	1	yes	yes	square wave	no	synthetic	Gaussian	none
30	ICA	rest	4D	no	2	yes	no	none	–	synthetic	Gaussian	none
31	cluster analysis	block	3D	no	1	yes	yes	Poisson	no	synthetic	Gaussian + drift + physiological	none
32	GLM	block	4D	yes	1	yes	no	square wave	no	hybrid	–	–
33	Granger causality	rest	1D	no	200	no	no	gamma	yes	synthetic	Gaussian	none
34	adaptive thresholding	block	3D	no	500	no	yes	square wave	no	synthetic	Gaussian	spatial
35	GLM	ER	1D	no	10^4	yes	yes	canonical	no	synthetic	Gaussian	temporal
36	Bayesian inference	ER	3D	no	NA	yes	no	canonical	yes	synthetic	Gaussian	none
37	ICA	ER	4D	no	1	yes	yes	gamma	yes	hybrid	–	–
38	cluster analysis	none	3D	yes	1	no	yes	estimated	no	hybrid	–	–
39	ICA	rest	3D	yes	1	yes	yes	sinusoidal	no	synthetic	Gaussian	none
40	cluster analysis	block	3D	no	100	yes	no	square wave	no	synthetic	Gaussian	spatial
41	ICA	block	4D	no	10	yes	no	gamma	yes	hybrid	–	–
42	STAP algorithm	block	4D	no	7	yes	no	square wave	no	hybrid	–	–
43	cluster analysis	block	3D	no	1	yes	no	gamma	yes	hybrid	–	–
44	connectivity	block	4D	no	15	yes	no	canonical	no	synthetic	Gaussian	none
45	BOLD estimation	ER	1D	no	100	yes	yes	Balloon	no	synthetic	Gaussian	none
46	denoising	ER	3D	no	1	yes	no	square wave	yes	hybrid	–	–
47	mixed effects	block	1D	no	500	no	yes	square wave	no	synthetic	Gaussian	spatial & temporal
48	connectivity	block	1D	yes	1	yes	no	gamma	yes	synthetic	none	none
49	GLM	ER	3D	no	1	yes	yes	canonical	no	synthetic	Gaussian	temporal
50	fractal scaling	ER	3D	no	1	yes	yes	gamma	no	hybrid	–	–
51	GLM	block	3D	no	1	no	no	square wave	no	synthetic	Gaussian	none
52	ICA	block	3D	no	100	yes	yes	square wave	no	synthetic	Uniform + physiological + drift	temporal
53	ICA	block	3D	yes	1	no	no	gamma	yes	synthetic	Gaussian	none
54	ICA	ER	4D	yes	1	yes	no	canonical	yes	hybrid	–	–
55	GLM	ER	3D	yes	400	yes	no	square wave	no	synthetic	Gaussian	spatial
56	GLM	ER	3D	yes	1000	yes	yes	square wave	no	synthetic	Gaussian + Chi-square	spatial

ID – paper identification number; dim. – data dimension; nS – Multiple subjects?; rep – Number of replications
parV – Parameter variation?; parJ – Parameter justification; HRFm – HRF model; HRFv – HRF variation?

Table 2.3 – *Continued*

ID	Model	Design	dim.	nS	rep	parV	parJ	HRFm	HRFv	type	Noise model	Noise correlations
57	ICA	rest	3D	no	1	no	no	sinusoidal	yes	synthetic	super Gaussian	spatial
58	ICA	rest	4D	no	1	yes	yes	sinusoidal	yes	synthetic	super Gaussian	spatial
59	cluster analysis	ER	3D	no	1	yes	no	square wave	yes	synthetic	Gaussian	none
60	PCA/ICA	block	2D	no	30	no	no	canonical	no	synthetic	mixture Gaussian	temporal
61	t-test	block	4D	no	50	yes	no	square wave	no	hybrid	–	–
62	ICA	rest	3D	no	1	yes	yes	canonical	no	synthetic	Gaussian	none
63	spatial smoothing	block	3D	no	1000	no	no	square wave	no	synthetic	Gaussian	none
64	change-point theory	block	3D	no	1	no	no	square wave	no	synthetic	Gaussian	temporal
65	GLM	block	3D	no	1000	yes	yes	square wave	no	synthetic	Gaussian	spatial
66	residual analysis	block	1D	no	1000	yes	no	canonical	no	synthetic	Gaussian	none
67	ICA	block	3D	no	50	yes	no	canonical	no	synthetic	Gaussian	none
68	GLM	block	1D	no	10^4	yes	yes	square wave	no	synthetic	Gaussian + drift	none
69	cluster analysis	ER	3D	no	1	yes	no	gamma	no	synthetic	Gaussian	none
70	spatio-temporal	block	3D	no	1	yes	no	canonical	no	synthetic	Gaussian	none
71	correlation	block	1D	no	1000	yes	yes	gamma	no	synthetic	Gaussian + drift	none
72	BOLD estimation	ER	1D	no	1000	yes	no	canonical	no	synthetic	Gaussian + drift	none
73	ICA	rest	3D	yes	1	yes	no	sinusoidal	no	synthetic	super Gaussian	none
74	spectral analysis	rest	3D	no	1000	yes	yes	sinusoidal	yes	synthetic	Gaussian	temporal
75	LRT	block	3D	no	1	no	no	square wave	no	synthetic	Gaussian	none
76	spatio-temporal	block	3D	no	1	no	no	sinusoidal	no	hybrid	–	–
77	spatial decomposition	block	4D	yes	500	yes	yes	square wave	no	synthetic	Gaussian	none
78	t-test	block	1D	no	10^4	yes	yes	square wave	no	synthetic	Gaussian	none
79	GLM	block	3D	no	100	yes	no	square wave	no	hybrid	–	–
80	connectivity	ER	1D	no	1000	yes	yes	canonical	no	synthetic	Gaussian	none
81	BOLD estimation	ER	3D	no	80	yes	yes	gamma	yes	synthetic	Gaussian + drift	none
82	spatio-temporal	block	3D	no	1	no	yes	gamma	no	hybrid	–	–
83	spatio-temporal	block	3D	no	1	no	no	gamma	no	hybrid	–	–
84	conditional maximisation	ER	1D	no	100	yes	yes	canonical	no	synthetic	Gaussian	none
85	connectivity	ER	1D	no	25	yes	yes	canonical	yes	synthetic	Gaussian	temporal

ID – paper identification number; dim. – data dimension; nS – Multiple subjects?; rep – Number of replications
parV – Parameter variation?; parJ – Parameter justification; HRFm – HRF model; HRFv – HRF variation?

Table 2.3 – *Continued*

ID	Model	Design	dim.	nS	rep	parV	parJ	HRFm	HRFv	type	Noise model	Noise correlations
86	cluster analysis	rest	3D	no	500	no	no	none	–	synthetic	Gaussian	spatial
87	connectivity	block	1D	no	1000	no	no	none	–	synthetic	Gaussian	temporal
88	Support Vector Machine	block	3D	yes	100	no	yes	square wave	no	synthetic	Gaussian	none
89	Granger causality	rest	1D	no	200	yes	yes	canonical	yes	synthetic	Gaussian + physiological	temporal
90	ICA	rest	3D	no	50	no	yes	none	–	synthetic	Gaussian	none
91	ICA	rest	4D	yes	1	no	yes	none	–	synthetic	Gaussian	none
92	LRT	block	1D	no	NA	no	no	square wave	no	synthetic	Rician	none
93	GLM	block	1D	no	NA	no	no	square wave	no	synthetic	Rician	none
94	LRT	block	1D	no	10^4	yes	no	square wave	no	synthetic	Gaussian	temporal
95	connectivity	block	1D	no	20	yes	yes	canonical	no	synthetic	Gaussian	none
96	GLM	ER	3D	no	10^5	yes	yes	canonical	yes	synthetic	Gaussian	none
97	permutation tests	block	3D	no	NA	no	no	square wave	no	synthetic	Gaussian	spatial & temporal
98	cluster analysis	block	4D	no	1	no	no	canonical	no	synthetic	Gaussian	none
99	cluster analysis	block	4D	no	1	no	no	canonical	no	synthetic	Gaussian	none
100	spatial smoothing	block	3D	no	1	yes	no	canonical	no	synthetic	Gaussian	spatial & temporal
101	STAP algorithm	block	3D	no	15	no	no	square wave	no	hybrid	–	–
102	STAP algorithm	block	3D	no	100	no	no	square wave	no	hybrid	–	–
103	ICA	rest	3D	yes	800	yes	yes	canonical	no	synthetic	Rician	spatial
104	connectivity	rest	1D	no	NA	no	no	none	–	synthetic	Gaussian	temporal
105	ICA	block	4D	no	1	yes	yes	canonical	yes	hybrid	–	–
106	adaptive mixture modelling	ER	3D	no	100	yes	yes	canonical	no	synthetic	Gaussian + drift	none
107	GLM	block	4D	yes	1	no	yes	gamma	no	synthetic	Gaussian	temporal
108	robust regression	rest	1D	no	2000	yes	yes	none	–	synthetic	Gaussian	none
109	GLM	block	4D	yes	1	yes	yes	canonical	yes	hybrid	–	–
110	GLM	block	3D	no	1000	yes	yes	square wave	no	synthetic	Gaussian	none
111	connectivity	rest	4D	no	100	yes	no	none	–	synthetic	Gaussian	none
112	Canonical Variates Analysis	block	4D	no	100	no	no	canonical	yes	synthetic	Gaussian	spatial & temporal
113	dimension estimation	block	1D	no	20000	yes	no	square wave	no	synthetic	Gaussian	temporal
114	spatial smoothing	block	3D	no	1	no	no	canonical	no	synthetic	Gaussian	none

ID – paper identification number; dim. – data dimension; nS – Multiple subjects?; rep – Number of replications
parV – Parameter variation?; parJ – Parameter justification; HRFm – HRF model; HRFv – HRF variation?

Table 2.3 – *Continued*

ID	Model	Design	dim.	nS	rep	parV	parJ	HRFm	HRFv	type	Noise model	Noise correlations
115	classification	ER	1D	no	1000	yes	no	canonical	no	synthetic	Gaussian	temporal
116	BOLD estimation	ER	4D	no	1	yes	yes	estimated	no	synthetic	Gaussian + drift	temporal
117	BOLD estimation	ER	4D	no	500	yes	yes	canonical	no	synthetic	Gaussian + drift	temporal
118	cluster analysis	block	3D	no	1	no	yes	square wave	yes	synthetic	Gaussian	none
119	BOLD estimation	block	1D	yes	100	yes	yes	canonical	yes	synthetic	Gaussian	temporal

ID – paper identification number; dim. – data dimension; nS – Multiple subjects?; rep – Number of replications
parV – Parameter variation?; parJ – Parameter justification; HRFm – HRF model; HRFv – HRF variation?

References

Backfrieder, W., Baumgartner, R., Sámal, M., Moser, E., & Bergmann, H. (1996). Quantification of intensity variations in functional MR images using rotated principal components. *Physics in Medicine and Biology*, *41*(8), 1425–1438.

Bandettini (Editor), P. (2012). 20 years of fMRI. *NeuroImage*, *62*, 575–1324.

Bellec, P., Perlbarg, V., & Evans, A. (2009). Bootstrap generation and evaluation of an fMRI simulation database. *Magnetic Resonance Imaging*, *27*, 1382–1396.

Boynton, G., Engel, S., Glover, G., & Heeger, D. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *The Journal of Neuroscience*, *16*(13), 4207–4221.

Brosch, J., Talabave, T., Ulmer, J., & Nyenhuis, J. (2002). Simulation of human respiration in fMRI with a mechanical model. *IEEE Transactions on Biomedical Engineering*, *49*, 700–707.

Buxton, R., Uludăg, K., Dubowitz, D., & Liu, T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, *23*, S220–S233.

Buxton, R., Wong, E., & Frank, L. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine*, *39*, 855–864.

Cohen, M. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage*, *6*, 93–103.

David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., et al. (2008). Identifying neural drivers with functional MRI: An electrophysiological validation. *PLoS Biology*, *6(12)*, e315.

Erhardt, E., Allen, E., Wei, Y., Eichele, T., & Calhoun, V. (2012). SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability. *NeuroImage*, *59*, 4160–4167.

Fadili, M., Ruan, S., Bloyet, D., & Mazoyer, B. (2001). On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. *Medical Image Analysis*, *5*(1), 55–67.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*, 30–40.

Glover, G. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*, 416–429.

Greve, D., Brown, G., Mueller, B., Glover, G., & Liu, T. (2012). A survey of the sources of noise fMRI. *Psychometrika*.

Handwerker, D., Gonzalez-Castillo, J., D'Esposito, M., & Bandettini, P. (2012). The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage*, *62*(2, SI), 1017-1023.

Johnston, L., Duff, E., Mareels, I., & Egan, G. (2008). Nonlinear estimation of the BOLD signal. *NeuroImage*, *40*(2), 504–514.

Kwong, K., Belliveau, J., Chesler, D., Goldberg, I., Weisskoff, R., Poncelet, B., et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 5675–5679.

Lazar, N. (2008). *The statistical analysis of functional mri data.* Berlin, Germany: Springer Verlag.

Ogawa, S., Tank, D., Menon, R., Ellermann, J., Kim, S., Merkle, H., et al. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 5951–5955.

Park, J., Shedden, K., & Polk, T. (2012). Correlation and heritability in neuroimaging datasets: a spatial decomposition approach with application to an fMRI study of twins. *NeuroImage*, *59*(2), 1132–42.

Penny, W. (2011). Comparing Dynamic Causal Models using AIC, BIC and Free Energy. *NeuroImage*, *59*(1), 319–330.

Schippers, M., Renken, R., & Keysers, C. (2011). The effect of intra- and inter-subject variability of hemodynamic responses on group level Granger causality analyses. *NeuroImage*, *57*(1), 22–36.

Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, *35*, 137–167.

Sturzbecher, M., Tedeschi, W., Cabella, B., Baffa, O., Neves, U., & De Araujo, D. (2009). Non-extensive entropy and the extraction of BOLD spatial information in event-related functional MRI. *Physics in Medicine and Biology*, *54*(1), 161–174.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., & Rosseel, Y. (2011). neuRosim: An R package for generating fMRI data. *Journal of Statistical Software*, *44*, 1–18.

Appendix: Bibliographic details of the selected articles

[1] Afshin-Pour, B., Soltanian-Zadeh, H., Hossein-Zadeh, G.A., Grady, C., Strother, S., 2011. A mutual information-based metric for evaluation of fMRI data-processing approaches. Human Brain Mapping 32, 699–715.

[2] Allen, E., Erhardt, E., Wei, Y., Eichele, T., Calhoun, V., 2012. Capturing inter-subject variability with group independent component analysis of fMRI data: A simulation study. NeuroImage 59, 4141–59.

[3] Andrade, A., Kherif, F., Mangin, J., Worsley, K., Paradis, A., Simon, O., Dehaene, S., Le Bihan, D., Poline, J., 2001. Detection of fMRI activation using cortical surface mapping. Human Brain Mapping 12, 79–93.

[4] Backfrieder, W., Baumgartner, R., Sámal, M., Moser, E., Bergmann, H., 1996. Quantification of intensity variations in functional MR images using rotated principal components. Physics in Medicine and Biology 41, 1425–1438.

[5] Bai, P., Shen, H., Huang, X., Truong, Y., 2008. A supervised singular value decomposition for independent component analysis of fMRI. Statistica Sinica 18, 1233–1252.

[6] Bellec, P., Perlbarg, V., Evans, A., 2009. Bootstrap generation and evaluation of an fMRI simulation database. Magnetic Resonance Imaging 27, 1382–1396.

[7] Bellec, P., Rosa-Neto, P., Lyttelton, O., Benali, H., Evans, A., 2010. Multi-level bootstrap analysis of stable clusters in resting-state fMRI. NeuroImage 51, 1126–1139.

[8] Birn, R., Cox, R., Bandettini, P., 2004. Experimental designs and processing strategies for fMRI studies involving overt verbal responses. NeuroImage 23, 1046–1058.

[9] Biswal, B., Ulmer, J., 1999. Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. Journal Of Computer Assisted Tomography 23, 265–271.

[10] Brezger, A., Fahrmeir, L., Hennerfeind, A., 2007. Adaptive Gaussian Markov random fields with applications in human brain mapping. Journal of the Royal Statistical Society: Series C (Applied Statistics) 56, 327—-345.

[11] Cabella, B., Sturzbecher, M., De Araujo, D., Neves, U., 2009. Generalized relative entropy in functional magnetic resonance imaging. Physica A: Statistical Mechanics and its Applications 388, 41–50.

[12] Cabella, B., Sturzbecher, M., Tedeschi, W., Baffa Filho, O., de Araa, D., Neves, U., 2008. A numerical study of the Kullback-Leibler distance in functional magnetic resonance imaging. Brazilian Journal of Physics 38, 20 – 25.

[13] Calhoun, V., Adali, T., 2006. Complex infomax: Convergence and approximation of Infomax with complex nonlinearities. Journal of VLSI Signal Processing Systems for Signal Image and Video Technology 44, 173–190.

[14] Calhoun, V., Adali, T., Stevens, M., Kiehl, K., Pekar, J., 2005. Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis. NeuroImage 25, 527–538.

[15] Calhoun, V., Stevens, M., Pearlson, G., Kiehl, K., 2004. fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. NeuroImage 22, 252–257.

[16] Casanova, R., Ryali, S., Serences, J., Yang, L., Kraft, R., Laurienti, P., Maldjian, J., 2008. The impact of temporal regularization on estimates of the BOLD hemodynamic response function: a comparative analysis. NeuroImage 40, 1606–1618.

[17] Casanova, R., Yang, L., Hairston, W., Laurienti, P., Maldjian, J., 2009. Evaluating the impact of spatio-temporal smoothness constraints on the BOLD hemodynamic response function estimation: an analysis based on Tikhonov regularization. Physiological Measurement 30, N37–N51.

[18] Chen, H., Yao, D., 2004. Discussion on the choice of separated components in fMRI data analysis by spatial independent component analysis. Magnetic Resonance Imaging 22, 827–833.

[19] Chen, H., Yao, D., Zhuo, Y., Chen, L., 2003a. Analysis of fMRI data by blind separation of data in a tiny spatial domain into independent temporal component. Brain Topography 15, 223–232.

[20] Chen, H., Yuan, H., Yao, D., Chen, L., Chen, W., 2006. An integrated neighborhood correlation and hierarchical clustering approach of functional MRI. IEEE Transactions on Biomedical Engineering 53, 452–458.

[21] Chen, N.K., Dickey, C., Yoo, S.S., Guttmann, C., Panych, L., 2003b. Selection of voxel size and slice orientation for fMRI in the presence of susceptibility field gradients: application to imaging of the amygdala. NeuroImage 19, 817–825.

[22] Churchill, N., Yourganov, G., Oder, A., Tam, F., Graham, S., Strother, S., 2012. Optimizing Preprocessing and Analysis Pipelines for Single-Subject fMRI: 2. Interactions with ICA, PCA, Task Contrast and Inter-Subject Heterogeneity. PLoS ONE 7, e31147.

[23] De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., Formisano, E., 2008. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. NeuroImage 43, 44–58.

[24] De Mazière, P., Van Hulle, M., 2007. fMRI bold signal analysis using a novel nonparametric statistical method. Journal of Magnetic Resonance 185, 138–151.

[25] Den Dekker, A., Poot, D., Bos, R., Sijbers, J., 2009. Likelihood-based hypothesis tests for brain activation detection from MRI data disturbed by colored noise: a simulation study. IEEE Transactions on Medical Imaging 28, 287–296.

[26] Desco, M., Hernandez, J., Santos, A., Brammer, M., 2001. Multiresolution analysis in fMRI: sensitivity and specificity in the detection of brain activation. Human Brain Mapping 14, 16–27.

[27] Deshpande, G., Sathian, K., Hu, X., 2010. Assessing and Compensating for Zero-Lag Correlation Effects in Time-Lagged Granger Causality Analysis of fMRI. IEEE Transactions on Biomedical Engineering 57, 1446–1456.

[28] Desmond, J., Glover, G., 2002. Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. Journal of Neuroscience Methods 118, 115–128.

[29] Dimitriadou, E., Barth, M., Windischberger, C., Hornik, K., Moser, E., 2004. A quantitative comparison of functional MRI cluster analysis. Artificial Intelligence in Medicine 31, 57–71.

[30] Esposito, F., Goebel, R., 2011. Extracting functional networks with spatial independent component analysis: the role of dimensionality, reliability and aggregation scheme. Current Opinion in Neurology 24, 378–385.

[31] Fadili, M., Ruan, S., Bloyet, D., Mazoyer, B., 2001. On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. Medical Image Analysis 5, 55–67.

[32] Gavrilescu, M., Shaw, M., Stuart, G., Eckersley, P., Svalbe, I., Egan, G., 2002. Simulation of the effects of global normalization procedures in functional MRI. NeuroImage 17, 532–542.

[33] Goebel, R., Roebroeck, A., Kim, D.S., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magnetic Resonance Imaging 21, 1251–1261.

[34] Gorgolewski, K., Storkey, A., Bastin, M., Pernet, C., 2012. Adaptive thresholding for reliable topological inference in single subject fMRI analysis. Frontiers in Human Neuroscience 6.

[35] Grinband, J., Wager, T., Lindquist, M., Ferrera, V., Hirsch, J., 2008. Detection of time-varying signals in event-related fMRI designs. NeuroImage 43, 509–520.

[36] Groves, A., Chappell, M., Woolrich, M., 2009. Combined spatial and non-spatial prior for inference on MRI time-series. NeuroImage 45, 795–809.

[37] Gu, H., Engelien, W., Feng, H., Silbersweig, D., Stern, E., Yang, Y., 2001. Mapping transient, randomly occurring neuropsychological events using independent component analysis. NeuroImage 14, 1432–1443.

[38] Guo, Y., 2010. A weighted cluster kernel PCA prediction model for multi-subject brain imaging data. Statistics and Its Interface 3, 103–111.

[39] Guo, Y., DuBois Bowman, F., 2008. Modeling dose-dependent neural processing responses using mixed effects spline models: with application to a PET study of ethanol. NeuroImage 40, 698–711.

[40] Heller, R., Stanley, D., Yekutieli, D., Rubin, N., Benjamini, Y., 2006. Cluster-based analysis of FMRI data. NeuroImage 33, 599–608.

[41] Hu, D., Yan, L., Liu, Y., Zhou, Z., Friston, K., Tan, C., Wu, D., 2005. Unified SPM-ICA for fMRI analysis. NeuroImage 25, 746–755.

[42] Huang, L., Thompson, E., Schmithorst, V., Holland, S., Talavage, T., 2009. Partially adaptive STAP algorithm approaches to functional MRI. IEEE Transactions on Biomedical Engineering 56, 518–521.

[43] Jahanian, H., Hossein-Zadeh, G.A., Soltanian-Zadeh, H., Ardekani, B., 2004. Controlling the false positive rate in fuzzy clustering using randomization: application to fMRI activation detection. Magnetic Resonance Imaging 22, 631–638.

[44] Joel, S., Caffo, B., Van Zijl, P., Pekar, J., 2011. On the relationship between seed-based and ICA-based measures of functional connectivity. Magnetic Resonance in Medicine 66, 644–657.

[45] Johnston, L., Duff, E., Mareels, I., Egan, G., 2008. Nonlinear estimation of the BOLD signal. NeuroImage 40, 504–514.

[46] Kadah, Y., 2004. Adaptive denoising of event-related functional magnetic resonance imaging data using spectral subtraction. IEEE Transactions on Biomedical Engineering 51, 1944–1953.

[47] Kang, H., Ombao, H., Linkletter, C., Long, N., Badre, D., 2012. Spatio-Spectral Mixed-Effects Model for Functional Magnetic Resonance Imaging Data. Journal of the American Statistical Association 107, 568–577.

[48] Kim, B., Yeo, D., Bhagalia, R., 2008. Comprehensive mathematical simulation of functional magnetic resonance imaging time series including motion-related image distortion and spin saturation effect. Magnetic Resonance Imaging 26, 147–159.

[49] Kim, E., Han, Y., Park, H., 2011. New fMRI analysis method for multiple stimuli using reference estimation. International Journal of Imaging Systems and Technology 21, 315—-322.

[50] Lee, J.M., Hu, J., Gao, J., Crosson, B., Peck, K., Wierenga, C., McGregor, K., Zhao, Q., White, K., 2008. Discriminating brain activity from task-related artifacts in functional MRI: fractal scaling analysis simulation and application. NeuroImage 40, 197–212.

[51] Lee, K., Tak, S., Ye, J., 2011a. A Data-Driven Sparse GLM for fMRI Analysis Using Sparse Dictionary Learning With MDL Criterion. IEEE Transactions on Medical Imaging 30, 1076–1089.

[52] Lee, S., Shen, H., Truong, Y., Lewis, M., Huang, X., 2011b. Independent Component Analysis Involving Autocorrelated Sources With an Application to Functional Magnetic Resonance Imaging. Journal of the American Statistical Association 106, 1009–1024.

[53] Lei, X., Qiu, C., Xu, P., Yao, D., 2010. A parallel framework for simultaneous EEG/fMRI analysis: methodology and simulation. NeuroImage 52, 1123–1134.

[54] LeVan, P., Gotman, J., 2009. Independent component analysis as a model-free approach for the detection of BOLD changes related to epileptic spikes: a simulation study. Human Brain Mapping 30, 2021–31.

[55] Li, Y., Gilmore, J., Wang, J., Styner, M., Lin, W., Zhu, H., 2012. Twin-MARM: two-stage multiscale adaptive regression methods for twin neuroimaging data. IEEE Transactions on Medical Imaging 31, 1100–12.

[56] Li, Y., Zhu, H., Shen, D., Lin, W., Gilmore, J., Ibrahim, J., 2011. Multiscale Adaptive Regression Models for Neuroimaging Data. Journal of the Royal Statistical Society Series B Statistical methodology 73, 559–578.

[57] Liao, R., Krolik, J., McKeown, M., 2005. An information-theoretic criterion for intrasubject alignment of FMRI time series: motion corrected independent component analysis. IEEE Transactions on Medical Imaging 24, 29–44.

[58] Liao, R., McKeown, M., Krolik, J., 2006. Isolation and minimization of head motion-induced signal variations in fMRI data using independent component analysis. Magnetic Resonance in Medicine 55, 1396–1413.

[59] Liao, W., Chen, H., Yang, Q., Lei, X., 2008. Analysis of fMRI Data Using Improved Self-Organizing Mapping and Spatio-Temporal Metric Hierarchical Clustering. IEEE Transactions on Medical Imaging 27, 1472–1483.

[60] Lin, F.H., Huang, T.Y., Chen, N.K., Wang, F.N., Stufflebeam, S., Belliveau, J., Wald, L., Kwong, K., 2005. Functional MRI using regularized parallel imaging acquisition. Magnetic Resonance in Medicine 54, 343–353.

[61] Lin, F.H., McIntosh, A., Agnew, J., Eden, G., Zeffiro, T., Belliveau, J., 2003. Multivariate analysis of neuronal interactions in the generalized

partial least squares framework: simulations and empirical studies. NeuroImage 20, 625–642.

[62] Lin, Q.H., Liu, J., Zheng, Y.R., Liang, H., Calhoun, V., 2010. Semiblind spatial ICA of fMRI using spatial constraints. Human Brain Mapping 31, 1076–1088.

[63] Lindquist, M., Wager, T., 2008. Spatial smoothing in fMRI using prolate spheroidal wave functions. Human Brain Mapping 29, 1276–1287.

[64] Lindquist, M., Waugh, C., Wager, T., 2007. Modeling state-related fMRI activity using change-point theory. NeuroImage 35, 1125–1141.

[65] Logan, B., Rowe, D., 2004. An evaluation of thresholding techniques in fMRI analysis. NeuroImage 22, 95–108.

[66] Loh, J., Lindquist, M., Wager, T., 2008. Residual analysis for detecting mis-modeling in fMRI. Statistica Sinica 18, 1421–1448.

[67] Long, Z., Chen, K., Wu, X., Reiman, E., Peng, D., Yao, L., 2009. Improved application of independent component analysis to functional magnetic resonance imaging study via linear projection techniques. Human Brain Mapping 30, 417–431.

[68] Lowe, M., Russell, D., 1999. Treatment of baseline drifts in fMRI time series analysis. Journal Of Computer Assisted Tomography 23, 463–473.

[69] Lu, N., Shan, B.C., Li, K., Yan, B., Wang, W., Li, K.C., 2006. Improved temporal clustering analysis method for detecting multiple response peaks in fMRI. Journal of Magnetic Resonance Imaging 23, 285–290.

[70] Luo, H., Puthusserypady, S., 2006. Spatio-temporal modeling and analysis of fMRI data using NARX neural network. International Journal of Neural Systems 16, 139–149.

[71] MacIntosh, B., Baker, S., Mraz, R., Ives, J., Martel, A., McIlroy, W., Graham, S., 2007. Improving functional magnetic resonance imaging motor

studies through simultaneous electromyography recordings. Human Brain Mapping 28, 835–845.

[72] Marrelec, G., Benali, H., Ciuciu, P., Pélégrini-Issac, M., Poline, J.B., 2003. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. Human Brain Mapping 19, 1–17.

[73] Moosmann, M., Eichele, T., Nordby, H., Hugdahl, K., Calhoun, V., 2008. Joint independent component analysis for simultaneous EEG-fMRI: principle and simulation. International Journal of Psychophysiology 67, 212–221.

[74] Müller, K., Neumann, J., Grigutsch, M., Von Cramon, D., Lohmann, G., 2007. Detecting groups of coherent voxels in functional MRI data using spectral analysis and replicator dynamics. Journal of Magnetic Resonance Imaging 26, 1642–1650.

[75] Nan, F., Nowak, R., 1999. Generalized likelihood ratio detection for fMRI using complex data. IEEE Transactions on Medical Imaging 18, 320–329.

[76] Ngan, S., Auffermann, W., Sarkar, S., Hu, X., 2001. Activation detection in event-related fMRI data based on spatio-temporal properties. Magnetic Resonance Imaging 19, 1149–1158.

[77] Park, J., Shedden, K., Polk, T., 2012. Correlation and heritability in neuroimaging datasets: a spatial decomposition approach with application to an fMRI study of twins. NeuroImage 59, 1132–42.

[78] Parrish, T., Gitelman, D., LaBar, K., Mesulam, M., 2000. Impact of signal-to-noise on functional MRI. Magnetic Resonance in Medicine 44, 925–932.

[79] Pendse, G., Borsook, D., Becerra, L., 2009. Enhanced false discovery rate using Gaussian mixture models for thresholding fMRI statistical maps. NeuroImage 47, 231–261.

[80] Penny, W., 2011. Comparing Dynamic Causal Models using AIC, BIC and Free Energy. NeuroImage 59, 319–330.

[81] Puthusserypady, S., Ratnarajah, T., Jue, R., 2010. Robust Estimation of HDR in fMRI using H-Filters. IEEE Transactions on Biomedical Engineering 57, 1133–1142.

[82] Quirós, A., Diez, R., Gamerman, D., 2010a. Bayesian spatiotemporal model of fMRI data. NeuroImage 49, 442–456.

[83] Quirós, A., Diez, R., Wilson, S., 2010b. Bayesian spatiotemporal model of fMRI data using transfer functions. NeuroImage 52, 995–1004.

[84] Rodriguez, P., 2010. Using conditional maximization to determine hyperparameters in model-based fMRI. NeuroImage 50, 472–478.

[85] Ryali, S., Supekar, K., Chen, T., Menon, V., 2011. Multivariate dynamical systems models for estimating causal interactions in fMRI. NeuroImage 54, 807–823.

[86] Salli, E., Aronen, H., Savolainen, S., Korvenoja, A., Visa, A., 2001. Contextual clustering for analysis of functional MRI data. IEEE Transactions on Medical Imaging 20, 403–414.

[87] Sato, J., Junior, E., Takahashi, D., De Maria Felix, M., Brammer, M., Morettin, P., 2006. A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality. NeuroImage 31, 187–196.

[88] Sato, J., Mourão Miranda, J., Morais Martin, M., Amaro, E., Morettin, P., Brammer, M., 2008. The impact of functional connectivity changes on support vector machines mapping of fMRI data. Journal of Neuroscience Methods 172, 94–104.

[89] Schippers, M., Renken, R., Keysers, C., 2011. The effect of intra- and inter-subject variability of hemodynamic responses on group level Granger causality analyses. NeuroImage 57, 22–36.

[90] Schmithorst, V., 2009. Higher-order contrast functions improve performance of independent component analysis of fMRI data. Journal of Magnetic Resonance Imaging 29, 242–249.

[91] Schmithorst, V., Holland, S., 2004. Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. Journal of Magnetic Resonance Imaging 19, 365–368.

[92] Sijbers, J., den Dekker, A., 2004. Generalized likelihood ratio tests for complex fMRI data. Medical Imaging 2004: Physiology, Function, and Structure from Medical Images 5, 652–663. Medical Imaging 2004 Conference, San Diego, CA, FEB 17-19, 2004.

[93] Sijbers, J., den Dekker, A., Bos, R., 2005. A likelihood ratio test for functional MRI data analysis to account for colored noise. Advanced Concepts for Intelligent Vision Systems 3708, 538–546. 7th International Conference on Advanced Concepts for Intelligent Vision Systems, Antwerp, BELGIUM, SEP 20-23, 2005.

[94] Sijbers, J., Den Dekker, A.J., 2005. Generalized likelihood ratio tests for complex fMRI data: a Simulation study. IEEE Transactions on Medical Imaging 24, 604–611.

[95] Stephan, K., Kasper, L., Harrison, L., Daunizeau, J., Den Ouden, H., Breakspear, M., Friston, K., 2008. Nonlinear dynamic causal models for fMRI. NeuroImage 42, 649–662.

[96] Sturzbecher, M., Tedeschi, W., Cabella, B., Baffa, O., Neves, U., De Araujo, D., 2009. Non-extensive entropy and the extraction of BOLD spatial information in event-related functional MRI. Physics in Medicine and Biology 54, 161–174.

[97] Suckling, J., Bullmore, E., 2004. Permutation tests for factorially designed neuroimaging experiments. Human Brain Mapping 22, 193–205.

[98] Sun, F., Morris, D., Babyn, P., 2009. The optimal linear transformation-based fMRI feature space analysis. Medical & Biological Engineering & Computing 47, 1119–1129.

[99] Sun, F., Morris, D., Lee, W., Taylor, M., Mills, T., Babyn, P., 2010. Feature-Space-Based fMRI Analysis Using the Optimal Linear Transformation. Information Technology in Biomedicine, IEEE Transactions on 14, 1279 –1290.

[100] Tabelow, K., Polzehl, J., Ulug, A., Dyke, J., Watts, R., Heier, L., Voss, H., 2008. Accurate Localization of Brain Activity in Presurgical fMRI by Structure Adaptive Smoothing. IEEE Transactions on Medical Imaging 27, 531–537.

[101] Thompson, E., 2006. A parallel approach to STAP implementation for fMRI data. Journal of Magnetic Resonance Imaging 23, 216–221.

[102] Thompson, E., Holland, S., Schmithorst, V., 2004. A STAP algorithm approach to fMRI: a simulation study. Journal of Magnetic Resonance Imaging 20, 715–722.

[103] Vahdat, S., Maneshi, M., Grova, C., Gotman, J., Milner, T., 2012. Shared and Specific Independent Components Analysis for Between-Group Comparison. Neural Computation 24, 3052–3090.

[104] Valdés-Sosa, P., Sánchez-Bornot, J., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., Canales-Rodríguez, E., 2005. Estimating brain functional connectivity with sparse multivariate autoregression. Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences 360, 969–981.

[105] Valente, G., De Martino, F., Filosa, G., Balsi, M., Formisano, E., 2009. Optimizing ICA in fMRI using information on spatial regularities of the sources. Magnetic Resonance Imaging 27, 1110–1119.

[106] Vincent, T., Risser, L., Ciuciu, P., 2010. Spatially adaptive mixture modeling for analysis of FMRI time series. IEEE Transactions on Medical Imaging 29, 1059–1074.

[107] Visscher, K., Miezin, F., Kelly, J., Buckner, R., Donaldson, D., McAvoy, M., Bhalodia, V., Petersen, S., 2003. Mixed blocked/event-related designs separate transient and sustained activity in fMRI. NeuroImage 19, 1694–1708.

[108] Wager, T., Keller, M., Lacey, S., Jonides, J., 2005. Increased sensitivity in neuroimaging analyses using robust regression. NeuroImage 26, 99–113.

[109] Wang, Z., 2009. A hybrid SVM-GLM approach for fMRI data analysis. NeuroImage 46, 608–615.

[110] Weeda, W., Waldorp, L., Christoffels, I., Huizenga, H., 2009. Activated region fitting: a robust high-power method for fMRI analysis using parameterized regions of activation. Human Brain Mapping 30, 2595–2605.

[111] Weeda, W., Waldorp, L., Grasman, R., Van Gaal, S., Huizenga, H., 2011. Functional connectivity analysis of fMRI data using parameterized regions-of-interest. NeuroImage 54, 410–416.

[112] Worsley, K., Poline, J., Friston, K., Evans, A., 1997. Characterizing the response of PET and fMRI data using multivariate linear models. NeuroImage 6, 305–319.

[113] Xie, X., Cao, Z., Weng, X., Jin, D., 2009. Estimating intrinsic dimensionality of fMRI dataset incorporating an AR(1) noise model with cubic spline interpolation. Neurocomputing 72, 1042–1055.

[114] Yue, Y., Loh, J., Lindquist, M., 2010. Adaptive spatial smoothing of fMRI images. Statistics and Its Interface 3, 3–13.

[115] Zhang, C., 2010. Statistical inference of minimum BD estimators and classifiers for varying-dimensional models. Journal of Multivariate Analysis 101, 1574 – 1593.

[116] Zhang, C., Lu, Y., Johnstone, T., Oakes, T., Davidson, R., 2008. Efficient modeling and inference for event-related fMRI data. Computational Statistics & Data Analysis 52, 4859–4871.

[117] Zhang, C., Yu, T., 2008. Semiparametric detection of significant activation for brain fMRI. Annals of Statistics 36, 1693–1725.

[118] Zhang, J., Tuo, X., Yuan, Z., Liao, W., Chen, H., 2011. Analysis of fMRI Data Using an Integrated Principal Component Analysis and Supervised Affinity Propagation Clustering Approach. IEEE Transactions on Biomedical Engineering 58, 3184–3196.

[119] Zhang, T., Li, F., Beckes, L., Brown, C., Coan, J., 2012. Nonparametric inference of the hemodynamic response using multi-subject fMRI data. NeuroImage 63, 1754–1765.

3 | On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data

Marijke Welvaert & Yves Rosseel

Submitted manuscript

Abstract

Signal-to-noise ratio, the ratio between signal and noise, is a quantity that has been well established for MRI data but is still subject of ongoing debate and confusion when it comes to fMRI data. fMRI data are characterised by small activation fluctuations in a background of noise. Depending on how the signal of interest is identified, signal-to-noise for fMRI data is reported by using many different definitions. Since each definition comes with a different scale, interpreting and comparing SNR values for fMRI data can be a very challenging job. In this paper, we provide an overview of existing definitions. Further, the relationship with activation detection power is investigated. Reference tables and conversion formulae are provided to facilitate comparability between fMRI studies.

3.1 Introduction

In science and engineering, the signal-to-noise ratio (SNR) is a measure that compares the level of a desired signal to the level of background noise. For data acquired through magnetic resonance imaging (MRI), this quantification is typically used to allow comparison between imaging hardware, imaging protocols and acquisition sequences. In this context, SNR is conceptualised by comparing the signal of the MRI image to the background noise of the image (Parrish et al., 2000; Edelstein et al., 1986). Mathematically, the SNR is the quotient of the (mean) signal intensity measured in a region of interest (ROI) and the standard deviation of the signal intensity in a region outside the anatomy of the object being imaged (i.e. a region from which no tissue signal is obtained). By optimising, for example, field of view, scan parameters, magnetic field strength and slice thickness, the SNR of MRI images can be increased because this optimisation reduces the background noise.

Translating SNR of MRI images to fMRI images is not as straightforward as it may seem. First of all, the noise in fMRI images does not correspond to the background noise of MRI images. In fMRI images, system noise effects the image as well as noise stemming from the subject (i.e. cardiac and respiratory pulsations, motion) and the task that is performed. Using time series outside the brain as noise measurement only, will not be sufficient to capture the noise data (Parrish et al., 2000; Krüger & Glover, 2001; Tabelow et al., 2009). Secondly, since the main goal of fMRI studies is to detect small fluctuations over a period of time, image SNR might not be suitable. Therefore, temporal SNR (tSNR), in which the (mean) signal over time is taken into account, can be used to determine the SNR of fMRI time series (Triantafyllou et al., 2005).

How to define SNR for MRI and fMRI data is documented quite well from a physical perspective. Several studies have demonstrated the dependence on scanning parameters and illustrated the necessary conditions to obtain higher SNR (e.g. Krüger & Glover, 2001; Krüger et al., 2001; Parrish et al., 2000). However, in the end, one is interested in how well the experimentally induced

activation can be detected. From a statistical perspective, it is not entirely clear how the SNR measurements relate to this detection power, because the small activation fluctuations (typically around 1–5%) cannot be derived from the mean signal based on a static image or time series. So for fMRI data, using the contrast-to-noise ratio (CNR) of the time series instead of (t)SNR is more preferred because CNR compares a measure of the activation fluctuations to the noise (Hyde et al., 2001).

To retrieve the range of possible values of SNR and CNR, we looked at the reported values of SNR and CNR in fMRI studies. NeuroImage published in 2012 about 458 fMRI studies. Of these studies, 50 mentioned the role of SNR/CNR for their experiment or method, while only 18 papers also reported SNR or CNR values. Reported SNR values ranged from 0.35 to 203.6. Many authors explicitly reported tSNR values ranging from 4.42 to 280, while in a few other cases CNR values were reported that varied from 0.01 to 1.8.

Since the determination of the SNR and CNR of real data can be a demanding job and is not standardly reported, we also looked at the SNR/CNR values that were reported in simulation studies. In simulation studies, the range of the reported SNR/CNR values was determined based on the fMRI simulation database from Welvaert & Rosseel (2013). The reported values varied widely across studies and were almost exclusively labelled as SNR. For example, the SNR for the simulations varied from 1 to 10 in one study, while the range was 0.01 to 1 in another, and in yet other studies, we found SNR values that could be negative, for instance, from -13 to 30.

Both in the experimental and simulation studies, the reported values demonstrated a range that was much wider than can be explained by natural variation only. There is only one reason that could account for the found variation, namely, the use of different definitions to calculate SNR or CNR. Indeed, several definitions can be found in the literature, especially for CNR. All these CNR measurements model some form of relative signal change, related to the contrast of interest, relative to the noise level. However, there is no consensus on how this contrast of interest should be conceptualised. Therefore, the scale of the CNR definitions varies widely and this makes comparing studies very hard. Furthermore, it is not clear how the CNR def-

initions are related to the sensitivity (or power) to detect activation. In this paper, an overview of the SNR and CNR definitions for fMRI data that are most commonly reported is presented. The advantages and drawbacks of these definitions are discussed, as well as the relationship between the definitions and the power to detect activation. In addition, conversion strategies between the definitions are derived that will enable comparison between different fMRI studies.

3.2 SNR and CNR definitions for fMRI data

Both SNR and CNR definitions have in common that a signal measure is compared to the noise level. The distinction between SNR versus CNR and the differences between the CNR definitions will be the result of how the signal measure is defined. While discussing the definitions, we will consider fMRI time series as the result of an addition of an activation signal time course and a noise signal time course. The activation signal time course, denoted as S, contains both the baseline signal and the possible fluctuations in the signal due to the experimental task. In general, S can be calculated as the average haemodynamic response function (HRF) of the fMRI time series in a certain ROI (see for example Huettel et al., 2001). The noise signal, N, will typically be the composition of several noise sources such as system noise, physiological noise and task-related noise. When referring to the noise signal, we implicitly take into account all these sources, ignoring the specific influence or distribution of these sources (see Krüger & Glover, 2001, for an extensive discussion). To calculate N from the fMRI series in an ROI, the contribution of the activation signal can be reduced by subtracting the average HRF from the time series (Huettel et al., 2001).

In the overview of the SNR and CNR definitions below, we will focus on those definitions that were found in the literature database from Welvaert & Rosseel (2013). In fMRI simulation studies, values for SNR/CNR are often chosen to give an indication of the strength of the modelled signal relative to the modelled noise. Six different definitions were found in total. We will discuss their definition and whether they should be referred to as SNR or

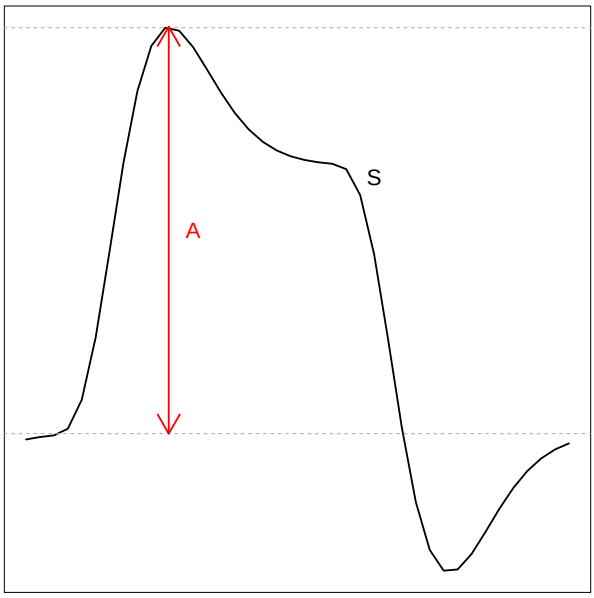

Figure 3.1 – Illustration how the amplitude A is determined from an activation signal S.

CNR. Note that, although in most papers these formulae were labelled as SNR, the majority of them are in fact CNR measurements.

Definition 1 (SNR)

The first definition models SNR based on the mean signal of the fMRI time series and the standard deviation of the noise in the time series (Nan & Nowak, 1999; Chen & Yao, 2004),

$$\frac{\bar{S}}{\sigma_N}.$$

As such, the global signal level, comprised of the baseline and activation, is related to the noise.

Definition 2 (CNR)

Joel et al. (2011); Den Dekker et al. (2009); Valente et al. (2009); Lindquist & Wager (2008) and De Martino et al. (2008), for example, used a CNR definition in which an amplitude measurement is related to the standard deviation of the noise,

$$\frac{A}{\sigma_N}.$$

The amplitude of the signal is generally defined as the absolute difference between the baseline of the signal and the signal peak (Figure 3.1).

Definition 3 (CNR)

The previous definition of the CNR can also be transformed in decibel (dB) scale, which is a common scale in signal processing (Marrelec et al., 2003; Suckling & Bullmore, 2004; Vincent et al., 2010),

$$10\log_{10}\left(\frac{A^2}{\sigma_N^2}\right).$$

Definition 4 (CNR)

Another possibility is to model the strength of the signal based on the standard deviation of the activation signal (Churchill et al., 2012; Esposito & Goebel, 2011; Penny, 2011; Schippers et al., 2011; Stephan et al., 2008; De Mazière & Van Hulle, 2007; Calhoun et al., 2005),

$$\frac{\sigma_S}{\sigma_N}.$$

This definition is also implemented in the DCM simulator (Friston et al., 2003) and is a very intuitive measurement of CNR because the ratio of the fluctuations of both activation signal and noise is calculated.

Definition 5 (CNR)

Lee et al. (2011) and Zhang (2010) used the ratio of the variances,

$$\frac{\sigma_S^2}{\sigma_N^2},$$

which is of course equal to the square of Definition 4.

Definition 6 (CNR)

Again, the ratio of the standard deviations is also found in dB scale (Casanova et al., 2008; Sturzbecher et al., 2009; Lin et al., 2010; Ryali et al., 2011; Bellec et al., 2010; Cabella et al., 2009),

$$10 \log_{10} \left(\frac{\sigma_S^2}{\sigma_N^2} \right).$$

Comments on the definitions

Definition 1 is a pure SNR measurement and is directly related to tSNR (Triantafyllou et al., 2005). Since the baseline levels in fMRI are typically quite high (e.g. around 800) and the signal fluctuations are very small, no real information about the activation signal strength is included in this definition, which makes it possibly not very suitable for fMRI data. In fact, the higher the baseline value of the data, the less impact the activation signal will have on the value of the SNR. Therefore, based on the SNR value of a certain voxel it will not be possible to distinguish active from non-active voxels.

In contrast, the remainder of the definitions all include some measurement of the activation signal strength. Therefore, these definitions are referred to as CNR formulae. It should be clear that, in theory, the value for these CNR definitions will always be 0 for non-active voxels and > 0 for active voxels. Consequently, theoretically it would be possible to detect active voxels based on their CNR value. In practice however, the activation signal is stricto sensu unknown and it may be complicated to calculate CNR values for single

voxels.

For the CNR definitions, two different sets can be distinguished; the first set (Definition 2–3) focuses on the amplitude of the activation signal, A, while the second set (Definition 4–6) incorporates the standard deviation of the activation as the signal of interest. With regard to the first set, these formulae can be interpreted as definitions of effect size based on means or differences between means, like for example Cohen's d (Cohen, 1988). As such it is a direct indication of the strength of the signal.

In the case of a block activation signal (Figure 3.1), the determination of the amplitude A is quite straightforward. However, this is not the case in, for example, an event-related design. In this experimental design, it is typical that multiple events will cause several peaks in the signal and the timing of the stimuli will have an effect on the height of the peak. In this case, the amplitude of the signal could be either the difference between the baseline and the maximal height of the signal, or the mean amplitude over all peaks. In contrast, calculating the standard deviation of the activation signal, σ_S, is independent of the experimental design (i.e. block or event-related designs).

So far, the definitions described above were only discussed based on a single condition experiment. As soon as multiple conditions are considered in a experiment, it is not quite clear anymore how to calculate the SNR or CNR of the fMRI data. One option could be to determine the SNR/CNR for each condition separately, which would be valid when distinct regions are activated by the conditions. Another option could be to first create an expected activation signal based on a contrast between the conditions, and then to calculate the SNR/CNR of the contrast signal in the same manner as for single condition time series. In this way, the signal of interest is directly based on the contrast that will be tested.

In essence all of these definitions have the same denominator (i.e. σ_N) so that differences are just scaling differences based on the definition of the activation signal. One desirable property for an SNR or CNR definition of fMRI time series would be that it is closely related to the activation detection power. If the SNR/CNR is high, then the power should be high too (keeping everything else constant). Secondly, the scaling differences make it hard to

compare the values of the discussed definitions. In the remainder of this paper, we will present some tools that will enable comparison among the different definitions and further, we will shed some light on the relationship with activation detection power.

3.3 Comparing the SNR and CNR values

Due to the fact that there is no consensus on how to define the SNR or CNR for fMRI data, interpreting a value can be an almost impossible job. Dependent on how the SNR/CNR is calculated, the values will be on a different scale. This impedes comparability between fMRI studies and consequently delays convergence of conclusions. In order to facilitate the comprehension of SNR and CNR values, three reference tables were assembled (Table 3.1–3.3), based on three experimental designs. The designs are (1) a block design, (2) an event-related (ER) design, and (3) a contrast between two conditions. These experimental designs serve as basic templates. More complex designs can be partially reduced to one of these three design types based on the specific research hypotheses at hand (i.e. a specific contrast or the effect of a specific predictor). An activation signal of 200s was modelled for each design. The block design consisted of alternating task and rest blocks that lasted 20s each. For the ER design, 25 events were randomly distributed over the whole time series. For the contrast, two alternating block conditions of 20s each were modelled with a rest period of 20s after each sequence AB and the effect of condition A was twice as high as the effect of condition B. The baseline value of the time series was considered fixed at 100 and we chose three levels of percent signal change, 1%, 2% and 5% respectively. The standard deviation of the noise was allowed to vary between 0.1 and 10. For all levels of these parameters, the SNR or CNR according to the six definitions was calculated and the results are presented in Table 3.1, Table 3.2 and Table 3.3. Note that for the ER design the amplitude was defined as the maximal amplitude (i.e. amplitude of the highest peak). In the case of the contrast design, the SNR and CNR values were calculated based on the contrast signal that was the difference of the activation signals of the two conditions. The

Table 3.1 – Reference table for the different SNR/CNR definitions based on a block design. Coloured cells indicate values that are within the range reported in fMRI simulation studies.

% Sig. ch.	σ_N	$\frac{\bar{S}}{\sigma_N}$	$\frac{A}{\sigma_N}$	$10\log\left(\frac{A^2}{\sigma_N^2}\right)$	$\frac{\sigma_S}{\sigma_N}$	$\frac{\sigma_S^2}{\sigma_N^2}$	$10\log\left(\frac{\sigma_S^2}{\sigma_N^2}\right)$	Power
1	0.1	1003	10	20	4.46	19.85	12.98	1.00
	0.2	502	5	14	2.23	4.96	6.96	1.00
	0.5	201	2	6	0.89	0.79	-1.00	1.00
	1	100	1	0	0.45	0.20	-7.02	0.99
	2	50	0.5	-6	0.22	0.050	-13.04	0.58
	5	20	0.2	-14	0.089	0.0079	-21.00	0.14
	10	10	0.1	-20	0.045	0.0020	-27.02	0.07
2	0.1	1007	20	26	8.91	79.42	19.00	1.00
	0.2	503	10	20	4.46	19.85	12.97	1.00
	0.5	201	4	12	1.78	3.18	5.02	1.00
	1	101	2	6	0.89	0.79	-1.00	1.00
	2	50	1	0	0.45	0.20	-7.02	0.99
	5	20	0.4	-8	0.18	0.032	-14.98	0.42
	10	10	0.2	-14	0.089	0.0079	-21.00	0.15
5	0.1	1017	50	34	22.28	496.35	26.96	1.00
	0.2	508	25	28	11.14	124.09	20.94	1.00
	0.5	203	10	20	4.46	19.85	12.98	1.00
	1	102	5	14	2.23	4.96	6.96	1.00
	2	51	2.5	8	1.11	1.24	0.94	1.00
	5	20	1	0	0.45	0.1985	-7.02	0.99
	10	10	0.5	-6	0.22	0.0496	-13.04	0.59

Table 3.2 – Reference table for the different SNR/CNR definitions based on an ER design. Coloured cells indicate values that are within the range reported in fMRI simulation studies.

% Sig. ch.	σ_N	$\frac{\bar{S}}{\sigma_N}$	$\frac{A}{\sigma_N}$	$10\log\left(\frac{A^2}{\sigma_N^2}\right)$	$\frac{\sigma_S}{\sigma_N}$	$\frac{\sigma_S^2}{\sigma_N^2}$	$10\log\left(\frac{\sigma_S^2}{\sigma_N^2}\right)$	Power
1	0.1	1002	10	20	3.07	9.41	9.74	1.00
	0.2	501	5	14	1.53	2.35	3.72	1.00
	0.5	200	2	6	0.61	0.38	-4.24	0.99
	1	100	1	0	0.31	0.094	-10.26	0.67
	2	50	0.5	-6	0.15	0.024	-16.28	0.22
	5	20	0.2	-14	0.06	0.0038	-24.24	0.08
	10	10	0.1	-20	0.03	0.00094	-30.26	0.06
2	0.1	1004	20	26	6.14	37.64	15.76	1.00
	0.2	502	10	20	3.07	9.41	9.74	1.00
	0.5	201	4	12	1.23	1.51	1.78	1.00
	1	100	2	6	0.61	0.38	-4.24	0.99
	2	50	1	0	0.31	0.094	-10.26	0.75
	5	20	0.4	-8	0.13	0.015	-18.23	0.17
	10	10	0.2	-14	0.06	0.0038	-24.24	0.08
5	0.1	1010	50	34	15.34	235.26	23.72	1.00
	0.2	505	25	28	7.67	58.81	17.69	1.00
	0.5	202	10	20	3.07	9.41	9.74	1.00
	1	101	5	14	1.54	2.35	3.72	1.00
	2	51	2.5	8	0.77	0.59	-2.31	0.99
	5	20	1	0	0.31	0.15	-10.26	0.64
	10	10	0.5	-6	0.15	0.024	-16.28	0.21

Table 3.3 – Reference table for the different SNR/CNR definitions based on a contrast. Coloured cells indicate values that are within the range reported in fMRI simulation studies.

% Sig. ch.	σ_N	$\frac{\bar{S}}{\sigma_N}$	$\frac{A}{\sigma_N}$	$10\log\left(\frac{A^2}{\sigma_N^2}\right)$	$\frac{\sigma_S}{\sigma_N}$	$\frac{\sigma_S^2}{\sigma_N^2}$	$10\log\left(\frac{\sigma_S^2}{\sigma_N^2}\right)$	Power
1	0.1	1001	10.56	20.47	3.02	9.14	9.61	1.00
	0.2	501	5.28	14.45	1.51	2.28	3.59	1.00
	0.5	200	2.11	6.49	0.60	0.37	-4.37	0.96
	1	100	1.06	0.47	0.30	0.09	-10.39	0.46
	2	50	0.53	-5.55	0.15	0.02	-16.41	0.15
	5	20	0.21	-13.51	0.06	0.0037	-24.37	0.07
	10	10	0.11	-19.53	0.03	0.0009	-30.39	0.05
2	0.1	1003	21.12	26.49	6.05	36.56	15.63	1.00
	0.2	501	10.56	20.47	3.02	9.14	9.61	1.00
	0.5	201	4.22	12.51	1.21	1.46	1.65	1.00
	1	100	2.11	6.49	0.60	0.37	-4.37	1.00
	2	50	1.06	0.47	0.30	0.091	-10.39	0.92
	5	20	0.42	-7.49	0.12	0.015	-18.35	0.27
	10	10	0.21	-13.51	0.06	0.004	-24.37	0.10
5	0.1	1007	52.79	34.45	15.12	228.50	23.59	1.00
	0.2	504	26.40	28.43	7.56	57.12	17.57	1.00
	0.5	201	10.56	20.47	3.02	9.14	9.61	1.00
	1	101	5.28	14.45	1.51	2.28	3.59	1.00
	2	50	2.64	8.43	0.76	0.57	-2.43	0.99
	5	20	1.06	0.47	0.30	0.09	-10.39	0.47
	10	10	0.53	-5.55	0.15	0.02	-16.41	0.16

amplitude of this contrast signal was calculated as the difference between the maximum and the minimum.

The results in Tables 3.1–3.3 demonstrate that the SNR definition (Definition 1) is highly dependent on the value of the baseline, since the formula is based on the mean signal strength. Additionally, the obtained values are almost invariant to changes in the activation signal strength and the experimental design.

The CNR definitions based on the amplitude of the signal (Definition 2 and Definition 3) are also partially determined by the baseline since the (maximal) amplitude of the signal will always correspond to the % signal change relative to the baseline. However, given the relative % signal change of the activation or contrast signal, the amplitude is constant over experimental designs. This is not true for the CNR definitions based on the standard deviation of the activation signal (Definition 4, Definition 5 and Definition 6). Although these CNR definitions are completely independent from the baseline, the activation standard deviation will be influenced by the number of events in an ER design or by the length of the epochs in a block design. The reference tables (Table 3.1, 3.2 and 3.3) illustrate this variation, but the close range of these CNR values over the designs indicates that this variation is rather small. Therefore, the reference tables presented here provide a tool to roughly compare and interpret the values for the different SNR/CNR definitions.

Of course, the conversion of one definition to another can also be solved analytically in some cases. Given the percent signal change p of the activation signal, the amplitude of the signal will be defined as

$$A = \frac{p \times 100}{b},$$

with b the baseline of the activation signal. A CNR value c calculated based on Definition 2 or Definition 4 can be converted to a CNR value in dB, c' using

$$c' = 10 \log_{10}(c^2).$$

Vice versa, a dB CNR value c' can be back transformed to the CNR in the

original scale, c, by

$$c = 10^{c'/20}.$$

Since the standard deviation of the activation signal (as in Definition 4–6) will be partially determined by the experimental design, there is no direct way to go from the percent signal change to the standard deviation. To compare these CNR values, either the reference tables, listed here, can be used to provide a rough estimate, or the values have to be calculated specifically for each design.

3.4 The relationship with detection power

There is no discussion on the fact that SNR or CNR is somehow related to activation detection power. Indeed, the higher the signal or the lower the noise (i.e. higher values for the SNR/CNR), the higher the power will be. Naively, one could expect that, when, for example, SNR$= 5$ and the power$= 0.30$, the power will increase to 0.60 for data with an SNR of 10. In other words, one may expect an approximate linear relationship between SNR/CNR values and the power to detect activation. In order to establish the approximate relationship between activation detection power and the SNR/CNR definitions, 10^4 time series were simulated for each design and for each level of activation strength and noise in the reference tables. Time series were generated by adding random Gaussian noise to the convolved activation signal. The empirical power was determined by fitting a standard GLM model to each of the simulated time series. In both the block and the ER design, the power was assessed by testing $H_0 : \beta_1 = 0$. For the contrast design, $H_0 : \beta_A - \beta_B = 0$ was tested. Power results are presented in the last column of the reference tables (Tables 3.1–3.3). Note that these results represent maximal power values. In real fMRI data, the power will be smaller due to the influence of non-white noise.

Looking at the results, we can immediately conclude that the simple rule *"twice as much signal will double the power"* is not valid. In general, the power will be lower for the time series that contain more noise, but

their is no clear linear relationship with the SNR or CNR values. However, comparing the power values for the different designs, overall lower values can be observed for the ER design notwithstanding equal activation strengths and noise levels. This lower power is in itself not that surprising, but this can only be predicted based on the CNR definitions that use the standard deviation of the activation signal, since the SNR/CNR values for the other definitions are constant over the designs. Additionally, in the lower power cases, the CNR values of Definition 4 are within the same range, indicating that these CNR values can be used as a rough estimate of activation detection power.

3.5 Discussion

fMRI data are often characterised by their SNR or CNR. SNR measurements are, for example, used to compare scanner hardware or the quality of scanning sequences, while CNR can be indicative of the quality (i.e. detectability) of the contrast of interest. In this paper, an overview was provided of common SNR and CNR definitions in an fMRI time series context. It was established that the literature lacks consensus on how to define SNR/CNR for fMRI data. Consequently, reported SNR and CNR values are hard to compare, possibly hindering the convergence of conclusions based on fMRI studies.

Based on how the signal of interest is defined, an explicit distinction was made between SNR and CNR. SNR compares the global signal level to the amount of noise and can be applied to either MRI images or task-related and resting-state fMRI (e.g. tSNR). The main purpose of determining the SNR of the data will be to assess the quality of the data (e.g. influence of noise). However, when applied to task-related fMRI data, the SNR of the data will most likely miss out on the small fluctuations present in the activation signal that are caused by the task. Therefore, in the case of these particular data, in which the signal of interest is a specific contrast that models the influence of certain conditions, it would be better to consistently use the concept of CNR. The CNR value will also give an indication of the quality of the data in terms of noise, but additionally it contains information on the strength of

the activation signal for a specific task. This information can be related to activation detection sensitivity.

A sceptical reader would argue that it might be meaningless to capture the information present in 4D fMRI data, which are characterised by very high inter- and intra-subject and -scanner variability, in one single number (either SNR or CNR). Indeed, for real data, SNR or CNR values are seldom reported. Moreover, screening of the simulation database discussed in Welvaert & Rosseel (2013) teaches us that no less than 62.2% of the simulation studies avoid reporting an SNR/CNR value. Instead, they reported separate parameters for the activation strength and the noise level. A second problem might be that the same value of SNR/CNR can indicate different levels of activation strength and noise, which can have a different impact on the detection accuracy. Despite the justly scepticism, determining the SNR or CNR of fMRI data can still hold useful information, because it provides an assesment of the quality of the data at a glance. However, we recommend to calculate the values only for small regions that are likely to have the same value of SNR/CNR based on anatomy or function. For simulation studies in particular, it would be interesting to report the SNR/CNR of the simulated value along with the specific values of activation strength and noise level. As such, generalising the conclusions from these studies to real data will be facilitated.

To conclude, although a CNR measure based on the standard deviation of the activation signal (Definition 4) could be a candidate, pushing for a common SNR/CNR definition may be preliminary right now because the measurement depends very much on how the signal of interest is defined. The tables presented in this chapter are a reference allowing easy comparison from one definition to another. The ability to compare the values that are reported in fMRI studies, either based on real or simulated data, will facilitate the convergence of fMRI based knowledge.

References

Bellec, P., Rosa-Neto, P., Lyttelton, O., Benali, H., & Evans, A. (2010). Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *NeuroImage*, *51*(3), 1126–1139.

Cabella, B., Sturzbecher, M., De Araujo, D., & Neves, U. (2009). Generalized relative entropy in functional magnetic resonance imaging. *Physica A: Statistical Mechanics and its Applications*, *388*(1), 41–50.

Calhoun, V., Adali, T., Stevens, M., Kiehl, K., & Pekar, J. (2005). Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis. *NeuroImage*, *25*(2), 527–538.

Casanova, R., Ryali, S., Serences, J., Yang, L., Kraft, R., Laurienti, P., et al. (2008). The impact of temporal regularization on estimates of the BOLD hemodynamic response function: a comparative analysis. *NeuroImage*, *40*(4), 1606–1618.

Chen, H., & Yao, D. (2004). Discussion on the choice of separated components in fMRI data analysis by spatial independent component analysis. *Magnetic Resonance Imaging*, *22*(6), 827–833.

Churchill, N., Yourganov, G., Oder, A., Tam, F., Graham, S., & Strother, S. (2012). Optimizing Preprocessing and Analysis Pipelines for Single-Subject fMRI: 2. Interactions with ICA, PCA, Task Contrast and Inter-Subject Heterogeneity. *PLoS ONE*, *7*(2), e31147.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage*, *43*(1), 44–58.

De Mazière, P., & Van Hulle, M. (2007). fMRI bold signal analysis using a novel nonparametric statistical method. *Journal of Magnetic Resonance*, *185*(1), 138–151.

Den Dekker, A., Poot, D., Bos, R., & Sijbers, J. (2009). Likelihood-based hypothesis tests for brain activation detection from MRI data disturbed by colored noise: a simulation study. *IEEE Transactions on Medical Imaging*, *28*(2), 287–296.

Edelstein, W., Glover, G., Hardy, C., & Redington, R. (1986). The intrinsic signal-to-noise ratio in NMR imaging. *Magnetic Resonance in Medicine*, *3*, 604–618.

Esposito, F., & Goebel, R. (2011). Extracting functional networks with spatial independent component analysis: the role of dimensionality, reliability and aggregation scheme. *Current Opinion in Neurology*, *24*(4), 378–385.

Friston, K., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.

Huettel, S., Singerman, J., & McCarthy, G. (2001). The effects of aging upon the Hemodynamic Response measured by functional MRI. *NeuroImage*, *13*, 161–175.

Hyde, J., Biswal, B., & Jesmanowicz, A. (2001). High-resolution fMRI using multislice partial k-space GR-EPI with cubic voxels. *Magnetic Resonance in Medicine*, *46*, 114–125.

Joel, S., Caffo, B., Van Zijl, P., & Pekar, J. (2011). On the relationship between seed-based and ICA-based measures of functional connectivity. *Magnetic Resonance in Medicine*, *66*(3), 644–657.

Krüger, G., & Glover, G. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *46*, 631–637.

Krüger, G., Kastrup, A., & Glover, G. (2001). Neuroimaging at 1.5t and 3.0t: Comparison of oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *45*, 595–604.

Lee, S., Shen, H., Truong, Y., Lewis, M., & Huang, X. (2011). Independent Component Analysis Involving Autocorrelated Sources With an Application to Functional Magnetic Resonance Imaging. *Journal of the American Statistical Association*, *106*(495), 1009–1024.

Lin, Q., Liu, J., Zheng, Y., Liang, H., & Calhoun, V. (2010). Semiblind spatial ICA of fMRI using spatial constraints. *Human Brain Mapping*, *31*(7), 1076–1088.

Lindquist, M., & Wager, T. (2008). Spatial smoothing in fMRI using prolate spheroidal wave functions. *Human Brain Mapping*, *29*(11), 1276–1287.

Marrelec, G., Benali, H., Ciuciu, P., Pélégrini-Issac, M., & Poline, J. (2003). Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. *Human Brain Mapping*, *19*(1), 1–17.

Nan, F., & Nowak, R. (1999). Generalized likelihood ratio detection for fMRI using complex data. *IEEE Transactions on Medical Imaging*, *18*(4), 320–329.

Parrish, T., Gitelman, D., LaBar, K., & Mesulam, M. (2000). Impact of signal-to-noise on functional MRI. *Magnetic Resonance in Medicine*, *44*(6), 925–932.

Penny, W. (2011). Comparing Dynamic Causal Models using AIC, BIC and Free Energy. *NeuroImage*, *59*(1), 319–330.

Ryali, S., Supekar, K., Chen, T., & Menon, V. (2011). Multivariate dynamical systems models for estimating causal interactions in fMRI. *NeuroImage*, *54*(2), 807–823.

Schippers, M., Renken, R., & Keysers, C. (2011). The effect of intra- and inter-subject variability of hemodynamic responses on group level Granger causality analyses. *NeuroImage*, *57*(1), 22–36.

Stephan, K., Kasper, L., Harrison, L., Daunizeau, J., Den Ouden, H., Breakspear, M., et al. (2008). Nonlinear dynamic causal models for fMRI. *NeuroImage*, *42*(2), 649–662.

Sturzbecher, M., Tedeschi, W., Cabella, B., Baffa, O., Neves, U., & De Araujo, D. (2009). Non-extensive entropy and the extraction of BOLD spatial information in event-related functional MRI. *Physics in Medicine and Biology*, *54*(1), 161–174.

Suckling, J., & Bullmore, E. (2004). Permutation tests for factorially designed neuroimaging experiments. *Human Brain Mapping*, *22*(3), 193–205.

Tabelow, K., Piëch, V., Polzehl, J., & Voss, H. (2009). High-resolution fMRI: Overcoming the signal-to-noise problem. *Journal of Neuroscience Methods*, *178*, 357–365.

Triantafyllou, C., Hoge, R., Krueger, G., Wiggins, C., Potthast, A., Wiggins, G., et al. (2005). Comparison of physiological noise at 1.5t, 3t and 7t and optimization of fMRI acquisition paramters. *NeuroImage*, *26*, 243–250.

Valente, G., De Martino, F., Filosa, G., Balsi, M., & Formisano, E. (2009). Optimizing ICA in fMRI using information on spatial regularities of the sources. *Magnetic Resonance Imaging*, *27*(8), 1110–1119.

Vincent, T., Risser, L., & Ciuciu, P. (2010). Spatially adaptive mixture modeling for analysis of FMRI time series. *IEEE Transactions on Medical Imaging*, *29*(4), 1059–1074.

Welvaert, M., & Rosseel, Y. (2013). A review of fMRI simulation studies. *Manuscript submitted to NeuroImage*.

Zhang, C. (2010). Statistical inference of minimum BD estimators and classifiers for varying-dimensional models. *Journal of Multivariate Analysis*, *101*(7), 1574 – 1593.

4 | neuRosim: An R package for generating fMRI data

Marijke Welvaert, Joke Durnez, Beatrijs Moerkerke,
Geert Verdoolaege & Yves Rosseel
Journal of Statistical Software (2011), 44(11), 1–18

Abstract

Studies that validate statistical methods for fMRI data often use simulated data to ensure that the ground truth is known. However, simulated fMRI data are almost always generated using in-house procedures because a well-accepted simulation method is lacking. In this article we describe the R package ***neuRosim****, which is a collection of data generation functions for neuroimaging data. We will demonstrate the possibilities to generate data from simple time series to complete 4D images and the possibilities for the user to create his own data generation method.*

4.1 Introduction

Despite optimization of experimental designs and significant improvements in scanner technology, fMRI data still contain a considerable amount of noise. Statistics are needed to infer information from the data. However, a major problem is that the ground truth of fMRI data (i.e. *where* and *when* the activation is located) is unknown and can only be measured with very invasive techniques (i.e. intracranial EEG) that are almost always unethical to perform with humans (David et al., 2008). Therefore, when researchers try to establish the validity of a new statistical method, or when they want to assess the sensitivity and the specificity of an existing method, they need to know the ground truth. As a solution, simulation studies have gained great interest as a validation tool because in these studies, the data themselves are generated under a known model.

Although the neccesity of knowing the ground truth is acknowlegded, a standard simulation procedure for fMRI data is lacking. In the literature, two major categories of computational simulations can be distinguished, namely (1) generating time series based on an experimental design and (2) simulating the magnetic signal by solving the Bloch equations (Bloch, 1946). Unfortunately, the first category in itself has no common method. Most researchers model the activation in the time series as the convolution of a haemodynamic response function and a stimulus vector. Additionally, some noise is added ranging from pure random Gaussian noise (Lei et al., 2010; Liao et al., 2008; Lin et al., 2010), over temporally correlated noise (Grinband et al., 2008; Locascio et al., 1997; Bullmore et al., 1996; Purdon & Weisskoff, 1998) to *real* noise derived from empirically acquired resting state scans (Bianciardi et al., 2004; Lange, 1999; Weibull et al., 2008; Lee et al., 2008; Lange et al., 1999; Hansen et al., 2001; Skudlarski et al., 1999). Furthermore, all simulations are done using in-house software routines. As a consequence, convergence of the simulation methods is impossible as long as fMRI simulators are not available. In contrast, the second method (Drobnjak et al., 2006), using the Bloch equations, is embedded in a simulator as part of the software package **FSL** (S. Smith et al., 2004). However, the simulator is rarely used for

validation studies. Probably, this is due to the fact that solving the Bloch equations is computationally very intensive and it takes, for example, about a month to generate a 4D dataset of 100 scans including all artefacts using a PC with a 3.4 GHz processor. By developing our package **neuRosim**, we want to respond to the current lack of fMRI simulators. Our package is by no means intended to provide *the* fMRI data generation method. The aim of the package is to provide a tool for simulating fMRI data that can initiate the search for more established and validated simulation methods for fMRI data such that the results of simulation studies can be generalized.

The R package **neuRosim** is created with two types of users in mind. The first type is the practical researcher who uses the fMRI scanner as a tool to acquire data that hopefully support his theory. This researcher normally would not think of generating fMRI data. However, by generating some data before the actual scanning process is started, this researcher can check the effectiveness of his design without almost any cost, both in time and money. In this way, the most effective design for a particular research question can be tested and adjusted.[1] Secondly, the more theoretical researcher (e.g. a statistician) can validate both existing and new methods based on the generated data. Because the data generation in **neuRosim** is fairly fast, the generation process can easily be embedded in large simulation studies.

fMRI data are in fact the result of a Fourier transformation of the k-space and are, as a result, complex-valued data (Rowe & Logan, 2004). However, in most fMRI studies the data analysis is done for the magnitude data and not for the phase data. In the current version of **neuRosim**, only the generation of fMRI magnitude data is considered. Therefore, all assumptions that are made to model the data apply only to the characteristics of magnitude data. The generation of magnitude fMRI data is seen as an additive source problem (Bellec et al., 2009) in which two main sources are distinguished, namely (1) the activation caused by an experimental design or resting state activation, and (2) the noise. **neuRosim** contains several functions to model both sources. These functions are regarded as low-level functions, meaning that

[1] It should be noted that **AFNI** also contains algorithms for design optimization in the function `3dDeconvolve` without the need for data (Cox, 1996).

they generate only a specific part of the data and are mostly used as building blocks to construct higher-level functions. For beginning users, it will be more convenient to start with the high-level functions that are described in section 4.3. However, advanced users can use the high-level functions as a basis for their completely customized simulations. In section 4.2, we will give an overview of the different models in the low-level functions.

Further, it should be noted that the data generated by **neuRosim** are considered to be pre-processed data. This implies that several artefacts (e.g., head motion, magnetic field inhomogeneity) that are normally removed during the pre-processing stage of the data are not explicitly modelled. However, it is possible to incorporate some residual effects of these artefacts under the assumption that the artefacts are not completely removed by the pre-processing analyses. For example, **neuRosim** data can contain task-related noise that can account for residual head movements.

4.2 Features and examples of low-level functions

4.2.1 Experimental activation and design

To generate BOLD activation, **neuRosim** uses a stimulus function that is part of the experimental design. A BOLD response is only generated if the function indicates the presence of a stimulus. Block designs, as well as event-related designs (or a combination of both) can be defined based on the onsets and the durations of the task as defined by the user. The function `stimfunction` uses these arguments to generate a 0-1 valued time vector where 1 indicates that the stimulus is present. Note that for a single event, the duration of the stimulus should be defined as 0. For example, to generate a stimulus function for a 20s ON/OFF block design of 200s with a microtime resolution of 0.1s:

```
R> totaltime <- 200
R> onsets <- seq(1, 200, 40)
R> dur <- 20
```

```
R> s <- stimfunction(totaltime = totaltime, onsets = onsets,
+     durations = dur, accuracy = 0.1)
```

The resulting stimulus function is shown as a dashed line in Figure 4.1. To simulate the BOLD signal caused by the task, the stimulus function is convoluted with a haemodynamic response function (HRF). The role of the microtime resolution is to ensure a high-precision convolution with the specified HRF. In the current version of **neuRosim**, three different response functions are implemented.

1. The stimulus function is convoluted with a gamma-variate HRF as implemented in the function `gammaHRF` with a user-defined full width at half maximum (FWHM) value (Buxton et al., 2004). The function is defined as

$$h(t) = \frac{1}{k\tau_h(k-1)!}\left(\frac{t}{\tau_h}\right)^k e^{-t/\tau_h} \tag{4.1}$$

with $k = 3$. To provide the desired FWHM, the time constant τ_h is given by $\tau_h = 0.242 \times \text{FWHM}$ (Buxton, Uludağ, Dubowitz, & Liu, 2004, p. S227).

```
R> gamma <- specifydesign(totaltime = 200,
+     onsets = list(onsets), durations = list(dur),
+     effectsize = 1, TR = 2, conv = "gamma")
```

To modulate the strength of the activation in each condition, the argument `effectsize` in the function `specifydesign` should be specified. The values, provided in this argument, are used to increase (values larger than 1) or decrease (values smaller than 1) the amplitude of the generated BOLD response.

2. The stimulus function is convoluted with a double-gamma HRF via `canonicalHRF`, which models an initial dip and an undershoot of the BOLD signal (Friston et al., 1998),

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} e^{\left(-\frac{t-d_1}{b_1}\right)} - c\left(\frac{t}{d_2}\right)^{a_2} e^{\left(-\frac{t-d_2}{b_2}\right)}. \tag{4.2}$$

where a_1 and a_2 model the delay of the response and the undershoot relative to the onset, b_1 and b_2 model the dispersion of the response and the undershoot, c models the scale of the undershoot, and d_1 and d_2 model the time to peak of the response and the undershoot. The default values of the parameters are $d_i = a_i b_i$, $a_1 = 6$, $a_2 = 12$, $b_i = 0.9$ and $c = 0.35$ (Glover, 1999).

```
R> canonical <- specifydesign(totaltime = 200,
+     onsets = list(onsets), durations = list(dur),
+     effectsize = 1, TR = 2, conv = "double-gamma")
```

3. The stimulus function is used as the input for the balloon model implemented in the `balloon` function (Buxton et al., 2004). The solving of the differential equations in the model is based on the Runge-Kutta solver in the R package **deSolve**. The parameters of the model can be modulated via the `param` argument, which should be a list containing values for all the parameters in the model. If not specified, the default values as described by Buxton et al. (2004) are used.

```
R> balloon <- specifydesign(totaltime = 200,
+     onsets = list(onsets), durations = list(dur),
+     effectsize = 1, TR = 2, conv = "Balloon")
```

The spatial location of the activation is specified as regions using the function `specifyregion`. A region can be modelled in three ways, namely (1) as a cube, (2) as a sphere or (3) manually. The first two forms can be modelled by defining two arguments, namely the coordinates of the center of the region and the distance from the center to the edge of the region in voxels. For example, to define an activated sphere (the result is displayed in Figure 4.2):

```
R> a <- specifyregion(dim = c(64, 64), coord = c(20, 20),
+     radius = 10, form = "sphere", fading = 0.5)
```

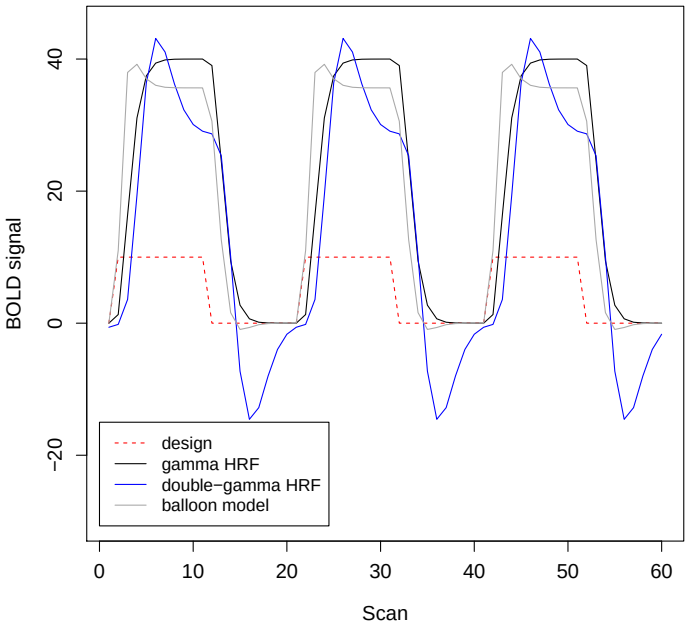

Figure 4.1 – The BOLD signals based on the three convolution functions for a 20s ON/OFF block design.

To define the form manually, the coordinates of all voxels that are part of the region should by specified as a matrix with colums corresponding to their (x,y)-coordinates.

```
R> coord <- matrix(c(rep(20, 20), rep(26:30, each = 2),
+     20:27, 20:27, rep(28, 6), 21:40, 30:21, rep(31, 8),
+     rep(40, 8), 33:38), ncol = 2, byrow = FALSE)
R> head(coord)
```

```
     [,1] [,2]
[1,]   20   21
[2,]   20   22
[3,]   20   23
[4,]   20   24
[5,]   20   25
[6,]   20   26
```

```
R> b <- specifyregion(dim = c(64, 64), coord = coord,
+     form = "manual")
```

The resulting activated slice is shown in Figure 4.2.

Additionally, it is possible to differentiate the strength of the measured activation between voxels in the activated region. This can be the case if, for example, the BOLD response to a certain stimulus is of different size in some parts of the activated region. A first method to include this variability is to divide the activated region into seperate subregions and specify separate parameters of the HRF for each subregion in `specifydesign`. The subregions can than be merged together using the high-level function `simprepSpatial` (see section 3). Secondly, if the region is defined as a cube or a sphere, the `fading` option can be used to require that the region has the largest effect in the center and smaller activation towards the edges (see Logan & Rowe, 2004). This fading of the BOLD response is modelled as an exponential decay depending on the distance of the activated voxel to the center of the region. The decay rate λ can vary between 0 and 1 with 0 meaning no decay and 1 indicating the strongest decay. In 3D this corresponds to

$$A(i,j,k) = \left(3e^{\frac{(i-i')^2+(j-j')^2+(k-k')^2}{\lambda}} + 3\right)/6 \tag{4.3}$$

where (i',j',k') are the (x,y,z)-coordinates of the voxel in the center of the region, λ is the decay rate and the activation is scaled to be 1 in the center of the region. An example of an activated sphere with fading ($\lambda = 0.5$) is presented in Figure 4.2.

4.2.2 Noise

The noise present in fMRI data is caused by different sources, such as for example the scanner and the subject. **neuRosim** offers a bundle of functions to model noise from one of these sources. The noise functions can be divided into four categories, namely (1) white noise, (2) coloured noise, (3) temporal noise and (4) spatial noise. The white noise (modelled by the function `systemnoise`) represents the system noise that is part of the fMRI data.

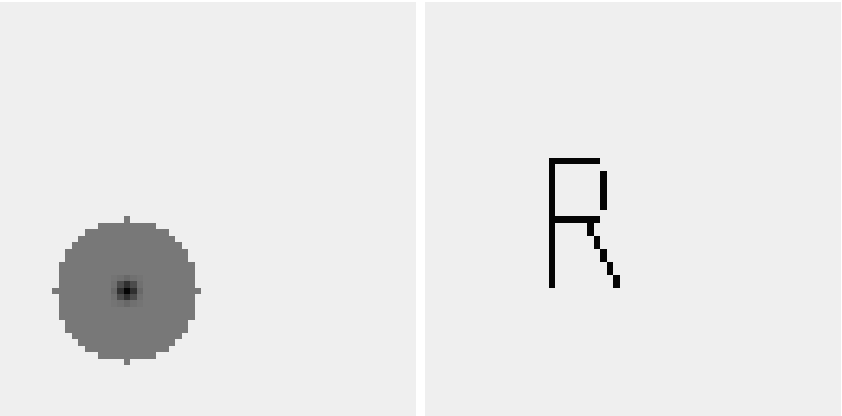

Figure 4.2 – Example of an activated slice: on the left, the activation is modelled as a sphere, on the right, the activated voxels are defined manually.

Two types of system noise are considered: (1) system noise that is Rician distributed and (2) system noise that is Gaussian distributed. The former is applicable for fMRI magnitude data with low signal-to-noise ratio (SNR), while the latter can be used for higher SNR (about more than 10) (Haacke et al., 1999; Gudbjartsson & Patz, 1995) . The standard deviation of the noise is user-defined or can be based on the desired SNR defined by the user. In all noise functions, average SNR is defined as

$$\text{SNR} = \frac{\bar{S}}{\sigma_N} \tag{4.4}$$

where $\bar{S}$ represents the average magnitude of the signal, and σ_N stands for the standard deviation of the noise (Krüger & Glover, 2001). For example (the resulting time series is plotted in Figure 4.3),

```
R> n.white <- systemnoise(dim = 1, nscan = 100, sigma = 15,
+     type = "rician")
```

Coloured noise depends on either the signal, the timing or the location. **neuRosim** contains three types of signal-dependent noise, (1) low-frequency drift, (2) physiological noise and (3) task-related noise.

- Low-frequency drift, generated by `lowfreqdrift`, is a consequence of system noise (A. Smith et al., 1999) that can be attributed to slow

fluctuations in the scanner hardware (Lazar, 2008). The drift is modelled as a basis of discrete cosine functions. The number of functions is determined by the frequency of the drift with a default value of 128s. For example (the resulting time series is plotted in Figure 4.3),

```
R> n.low <- lowfreqdrift(dim = 1, nscan = 100, TR = 2,
+     freq = 120)
```

- Physiological noise (`physnoise`) is defined as possible cardiac and respiratory artefacts and as such accounts for the variability in the signal that is caused by the heart beat and respiratory rate. These artefacts are often categorized as low-frequency drift. However, we choose to model the physiological noise separately because it is shown that the frequency of these artefacts is often higher than the scanner fluctuations (A. Smith et al., 1999). The physiological noise is modelled as sine and cosine functions with user-defined frequencies. Default values are 1.17 Hz and 0.2 Hz for heart beat and respiratory rate respectively (Biswal et al., 1996). For example (the resulting time series is plotted in Figure 4.3),

```
R> n.phys <- physnoise(dim = 1, nscan = 100, sigma = 15,
+     TR = 2)
```

- Task-related noise accounts for spontaneous neural activity due to the experimental task (Hyde et al., 2001) and is operationalized by adding random noise only where and when activation is present. The distribution of this noise can be either Gaussian or Rician. Additionally, the task-related noise can be interpreted as residual noise from head motion that is not removed in the pre-processing stage. For example (the resulting time series is plotted in Figure 4.3),

```
R> n.task <- tasknoise(act.image = s, sigma = 15)
```

Temporal noise accounts for the fact that fMRI data are repeated measurements (Purdon & Weisskoff, 1998). The function `temporalnoise` generates

noise based on an auto-regressive model of order p (AR(p)) defined as

$$\varepsilon_t = \Sigma_{i=1}^{p} \rho_i \varepsilon_{t-i} + \chi_t \tag{4.5}$$

with $\chi_t \sim N(0, \sigma^2)$. For example, the generate temporally correlated noise of order 2 (the resulting time series is plotted in Figure 4.3),

```
R> n.temp <- temporalnoise(dim = 1, sigma = 15,
+     nscan = 100, rho = c(0.4, -0.2))
```

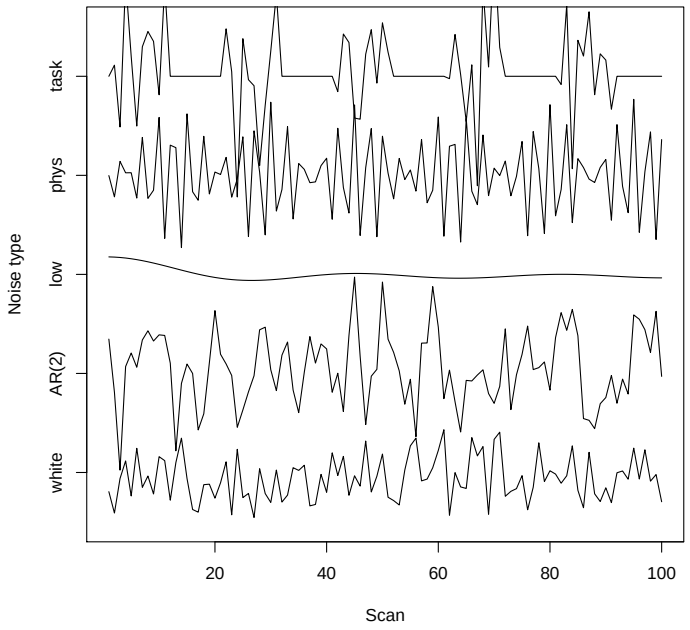

Figure 4.3 – Time series of the noise structures in **neuRosim**

Finally, spatial noise models the spatial dependencies in fMRI data (Logan & Rowe, 2004). Of course, voxels are arbitrary units and neighbouring voxels are more likely to be correlated than voxels that are further apart. The function `spatialnoise` incorporates three types of spatial noise models, namely (1) an autoregressive correlation structure, (2) a Gaussian random field and (3) a Gamma random field. The first structure correlates the voxels

with each other based on random Gaussian or Rician noise. The strength of the correlation depends on the value of the auto-correlation coefficient (default value is `rho=0.75`) and the distance between the voxels. If spatial correlation based on random fields is chosen, the full-width-half-maximum (FWHM) of the kernel, which is used to generate the random field, should be provided (default is `FHWM=4`). Additionally, if the method is `gammaRF`, the shape (default is `gamma.shape=6`) and rate (default is `gamma.rate=1`) parameter of the Gamma distribution should be defined as additional arguments. For example, to generate spatially correlated noise for a 20×20 slice:

```
R> d <- c(20, 20)
R> n.corr <- spatialnoise(dim = d, sigma = 15,
+     nscan = 100, method = "corr", rho = 0.7)
R> n.gaus <- spatialnoise(dim = d, sigma = 15,
+     nscan = 100, method = "gaussRF", FWHM = 4)
R> n.gamma <- spatialnoise(dim = d, sigma = 15,
+     nscan = 100, method = "gammaRF", FWHM = 4,
+     gamma.shape = 3, gamma.rate = 2)
```

Figure 4.4 displays the correlation matrices for the generated slices. To generate these images, all voxels were ordered and the correlation matrix of the generated time series was calculated. Therefore, the diagonal represents the perfect correlation of each voxel with itself. We see that voxels that are close to this diagonal, representing neighbouring voxels, are also highly correlated. The block diagonal structure, which can be observed clearly with the Gaussian random field structure (Figure 4.4b), is the result of reducing the two-dimensional structure of the slice.

Additionally, all noise functions include the functionality that a template or mask can be provided. As such, the noise is only generated for those voxels that are included in the mask. This would allow the user to make for example a distinction between the noise source in the grey matter, the white matter or in the cerebrospinal fluid.

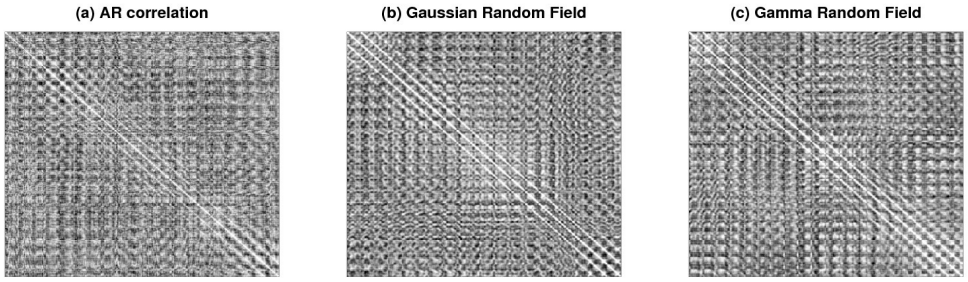

Figure 4.4 – Correlation images for (a) an autoregressive correlation structure, (b) a Gaussian random field and (c) a Gamma random field.

4.3 Examples of high-level functions

The aim of the high-level functions is to allow the user to generate fMRI data efficiently and transparantly. The functions are developed such that they can easily be implemented in a simulation environment. Of course, these functions have limits in their use. Therefore, we refer users who desire more functionalities to the low-level functions.

4.3.1 Generating fMRI time series

The `simTSfmri` function generates fMRI time series for a specified design matrix and with an additive noise structure. The field of the design matrix should be prepared with the `simprepTemporal` function, to ensure that all arguments are in the correct format. As an example, we will generate a time series for a block design with two conditions. The experiment lasts 100 scans with TR=2 and the first condition has activation blocks of 20s, while the second condition had activation blocks of 7s:

```
R> TR <- 2
R> nscan <- 100
R> total <- TR * nscan
R> os1 <- seq(1, total, 40)
R> os2 <- seq(15, total, 40)
R> dur <- list(20, 7)
```

```
R> os <- list(os1, os2)
R> effect <- list(3, 10)
R> design <- simprepTemporal(totaltime = total,
+     onsets = os, durations = dur, effectsize = effect,
+     TR = TR, hrf = "double-gamma")
```

Figure 4.5 presents the resulting activation from this design in dashed lines. The following arguments should be specified to ensure a complete definition of the design matrix: the total duration of the experiment in seconds (`total`), the onsets of each condition represented as a list (`onsets`), the duration of the stimulus in each condition represented as a list (`durations`), the repetition time in seconds (`TR`) and the form of the HRF (either `"gamma"`, `"double-gamma"` or `"balloon"`). The noise can be either of the structures described in section 2, but it is also possible to add a mixture of noise. The different noise components are then weighted with a vector of weights specified by the user. The weights can vary between 0 and 1, however, the weights should sum to one. For example, we will add a mixture of noise to our above specified design. The mixture contains Rician system noise, temporal noise of order 1, low-frequency drift, physiological noise and task-related noise and has a baseline value of 10:

```
R> w <- c(0.3, 0.3, 0.01, 0.09, 0.3)
R> ts <- simTSfmri(design = design, base = 10, SNR = 2,
+     noise = "mixture", type = "rician", weights = w,
+     verbose = FALSE)
```

The resulting time series are plotted in Figure 4.5.

4.3.2 Generating fMRI volumes

The function `simVOLfmri` is built to generate complete fMRI datasets (i.e. 3D for a slice and 4D for a volume). In this function, some spatial properties of the data are introduced. For this function, not only a design matrix –defining *when* the activation occurs– has to be specified, but also a region –defining *where* the activaton takes place– should be provided. Similarly as

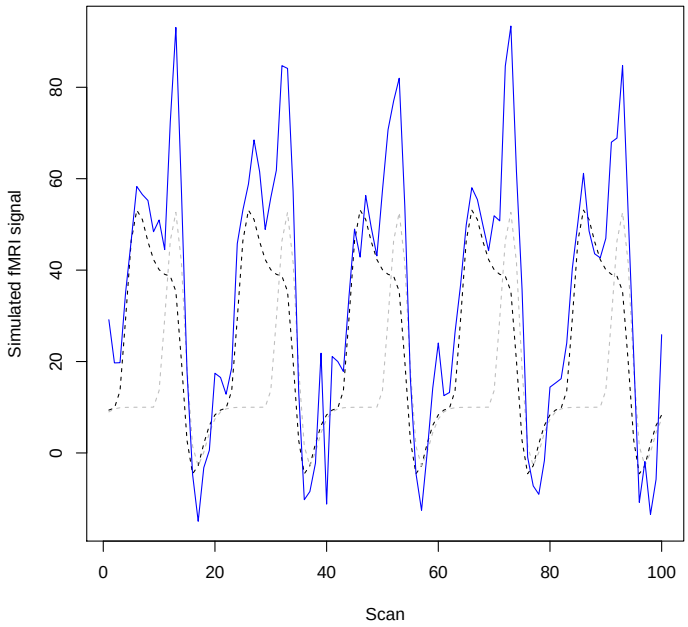

Figure 4.5 – Generated time series (in blue) based on an experiment with two conditions (dashed lines).

for the design matrix, a preparation function (`simprepSpatial`) is needed to ensure that all arguments that define the region of activation are in the correct format. Suppose that we wish to simulate 2 activated regions that are part of a small network. We need to call the `simprepSpatial` function as follows:

```
R> regions <- simprepSpatial(regions = 2,
+     coord = list(c(10, 5, 24), c(53, 29, 24)),
+     radius = c(10, 5), form = "sphere")
```

The arguments that should be provided in the function are: the number of activated regions (`regions`), a list of coordinates specifying the regions (`coord`), the radius of the region (`radius`, not needed if the region is defined manually) and the shape of the region (`form`) The implemented shapes are `cube` and `sphere`. For any other shape, the coordinates of all voxels in the region should be entered manually (see section 2 for an example). Further,

we will generate the activation in both regions following the same design matrix as for the generation of the time series.

```
R> onset <- list(os, os)
R> duration <- list(dur, dur)
R> effect1 <- list(2, 9)
R> effect2 <- list(14, 8)
R> design2 <- simprepTemporal(regions = 2,
+     onsets = onset, durations = duration,
+     TR = TR, hrf = "double-gamma",
+     effectsize = list(effect1, effect2),
+     totaltime = total)
```

We can now generate an fMRI dataset corresponding to this very simple two-region network. Again, we will add a mixture of noise with the additional possibility that we can add spatially correlated noise.

```
R> w <- c(0.3, 0.3, 0.01, 0.09, 0.1, 0.2)
R> data <- simVOLfmri(dim = c(64, 64, 64),
+     base = 100, design = design2,
+     image = regions, SNR = 10, noise = "mixture",
+     type = "rician", weights = w, verbose = FALSE)
```

The result is a 4D fMRI dataset. To analyze the data with standard fMRI data analysis software like **SPM**, **FSL**, **AFNI**,. . . , the dataset can be exported as a NIfTI file using for example the function `nifti.image.write` of the R package **Rniftilib** or the function `writeNIfTI` from the R package **oro.nifti**. Note that with `simTSfmri` and `simVOLfmri` it is also possible to simulate data that contain only activation or only noise.

4.3.3 Simulating and analyzing a 4D fMRI dataset

To further demonstrate the functionalities of the package, we present a more real-life example. Consider the data from a repetition priming experiment

performed using event-related fMRI (Henson et al., 2002)[2]. The data are the result of a 2×2 factorial study with factors *fame* and *repetition* where famous and non-famous faces were presented twice against a checkerboard (Henson et al., 2002, for more details, see). An orthographic overview of the measured data is given on the left side of Figure 4.6. To generate data using **neuRosim** that are representative for this study, we start by defining the design. First we define some parameters like the dimension of the image space, the number of scans and TR. Then, since we simulate an event-related design, we also assign the onsets for each condition.

```
R> dim <- c(53, 63, 46)
R> nscan <- 351
R> TR <- 2
R> total.time <- nscan * TR
R> onsets.N1 <- c(6.75, 15.75, 18, 27, 29.25, 31.5, 36,
+     42.75, 65.25, 74.25, 92.25, 112.5, 119.25, 123.75,
+     126, 137.25, 141.75, 144, 146.25, 155.25, 159.75,
+     162, 164.25, 204.75, 238.5) * TR
R> onsets.N2 <- c(13.5, 40.5, 47.25, 56.25, 90, 94.5,
+     96.75, 135, 148.5, 184.5, 191.25, 202.5, 216, 234,
+     236.25, 256.5, 261, 281.25, 290.25, 303.75, 310.5,
+     319.5, 339.75, 342) * TR
R> onsets.F1 <- c(0, 2.25, 9, 11.25, 22.5, 45, 51.75,
+     60.75, 63, 76.5, 78.75, 85.5, 99, 101.25, 103.5,
+     117, 130.5, 150.75, 171, 189, 227.25, 265.5, 283.5,
+     285.75, 288, 344.25) * TR
R> onsets.F2 <- c(33.75, 49.5, 105.75, 153, 157.5, 168.75,
+     177.75, 180, 182.25, 198, 222.75, 240.75, 254.25,
+     267.75, 270, 274.4, 294.75, 299.25, 301.5, 315,
+     317.25, 326.25, 333, 335.25, 337.5, 346.5)
```

[2]The use of the dataset is with permission from the corresponding author and can be downloaded from his personal webpage (`http://www.mrc-cbu.cam.ac.uk/people/rik.henson/personal/`)

Next, we have to specify which voxels are activated. We will consider 5 regions. The first three are general regions that activate when faces are presented, the fourth region is only activated if famous faces are shown, while in the last region adaptation to the repetition of faces is modelled.

```
R> region.1A.center <- c(13, 13, 11)
R> region.1A.radius <- 4
R> region.1B.center <- c(40, 18, 9)
R> region.1B.radius <- 6
R> region.1C.center <- c(10, 45, 24)
R> region.1C.radius <- 3
R> region.2.center <- c(15, 16, 31)
R> region.2.radius <- 5
R> region.3.center <- c(12, 16, 13)
R> region.3.radius <- 5
```

In each region, the same design matrix will be considered. However, the effect size in each condition will vary over conditions.

```
R> onsets <- list(onsets.N1, onsets.N2, onsets.F1, onsets.F2)
R> onsets.regions <- list(onsets, onsets, onsets, onsets,
+     onsets)
R> dur <- list(0, 0, 0, 0)
R> dur.regions <- list(dur, dur, dur, dur, dur)
R> region.1a.d <- list(160.46, 140.19, 200.16, 160.69)
R> region.1b.d <- list(140.51, 120.71, 160.55, 120.44)
R> region.1c.d <- list(120.53, 120.74, 140.02, 100.48)
R> region.2.d <- list(-0.24, 10.29, 80.18, 160.24)
R> region.3.d <- list(200.81, 50.04, 240.6, 50.83)
R> effect <- list(region.1a.d, region.1b.d, region.1c.d,
+     region.2.d, region.3.d)
```

Additionally, we will consider a baseline image. The baseline value for each voxel is determined as the mean value of the measured time series of that voxel. Nonbrain voxels are defined as voxels with an average measured value less than 250 and are fixed to 0 in the baseline image.

```
R> library(oro.nifti)
R> Hensondata <- readNIfTI("preprocessed_face.nii.gz")
R> baseline <- apply(Hensondata@.Data, 1:3, mean)
R> baseline.bin <- ifelse(baseline > 250, 1, 0)
R> ix <- which(baseline == 1)
R> baseline[-ix] <- 0
```

Consequently, the anatomical structure of the brain will be incorporated in the simulated data. Now, we can use the functions `simprepTemporal` and `simprepSpatial` to prepare the temporal and spatial structure of our simulated 4D fmri data.

```
R> design <- simprepTemporal(regions = 5,
+     onsets = onsets.regions, durations = dur.regions,
+     hrf = "double-gamma", TR = TR, totaltime = total.time,
+     effectsize = effect)
R> spatial <- simprepSpatial(regions = 5,
+     coord = list(region.1A.center, region.1B.center,
+     region.1C.center, region.2.center, region.3.center),
+     radius = c(region.1A.radius, region.1B.radius,
+     region.1C.radius, region.2.radius, region.3.radius),
+     form = "sphere", fading = 0.01)
```

Finally, we can generate the dataset. Note that the values for the SNR and the temporal autocorrelation coefficients were estimated based on the real data.

```
R> sim.data <- simVOLfmri(design = design, image = spatial,
+     base = baseline, SNR = 3.87, noise = "mixture",
+     type = "rician", rho.temp = c(0.142, 0.108, 0.084),
+     rho.spat = 0.4, w = c(0.05, 0.1, 0.01, 0.09, 0.05, 0.7),
+     dim = c(53, 63, 46), nscan = 351, vee = 0,
+     template = baseline.bin, spat = "gaussRF")
```

An orthographic overview of the simulated data is given on the righthand side of Figure 4.6.

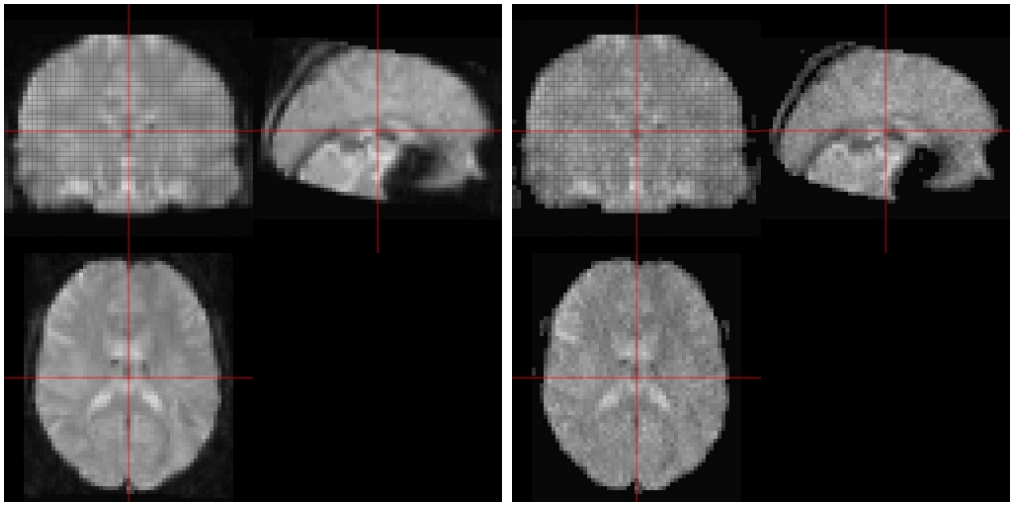

Figure 4.6 – Orthographic view of fMRI data for an event-related repetition priming study. On the left, the data measured by Henson et al. (2002) and on the right, the data simulated by **neuRosim**

Next, we analyzed the simulated data in SPM following the exact description given in the manual of SPM8 (Chapter 29). We considered three contrasts, namely: (1) the overall effect of faces versus baseline checkerboard, (2) the effect of famous faces and (3) the effect of repetition. The results were thresholded with $p < 0.05$ (uncorrected), just to demonstrate the detection of the activation. Figure 4.7 shows a comparison between some of the activated regions that are found in the real data (lefthand) and in the simulated data (righthand).

4.4 Conclusions and future work

neuRosim provides a flexible framework for generating fMRI data including a large variety of activation models and noise structures. High-level functions are available to simulate time series or full 4D data in an efficient and transparant way. For more advanced users, the low-level functions create the opportunity to build customized simulation functions. Currently, we are working on an extention of a resting state module such that in future updates it will be possible to have the same functionalities for the generation of resting

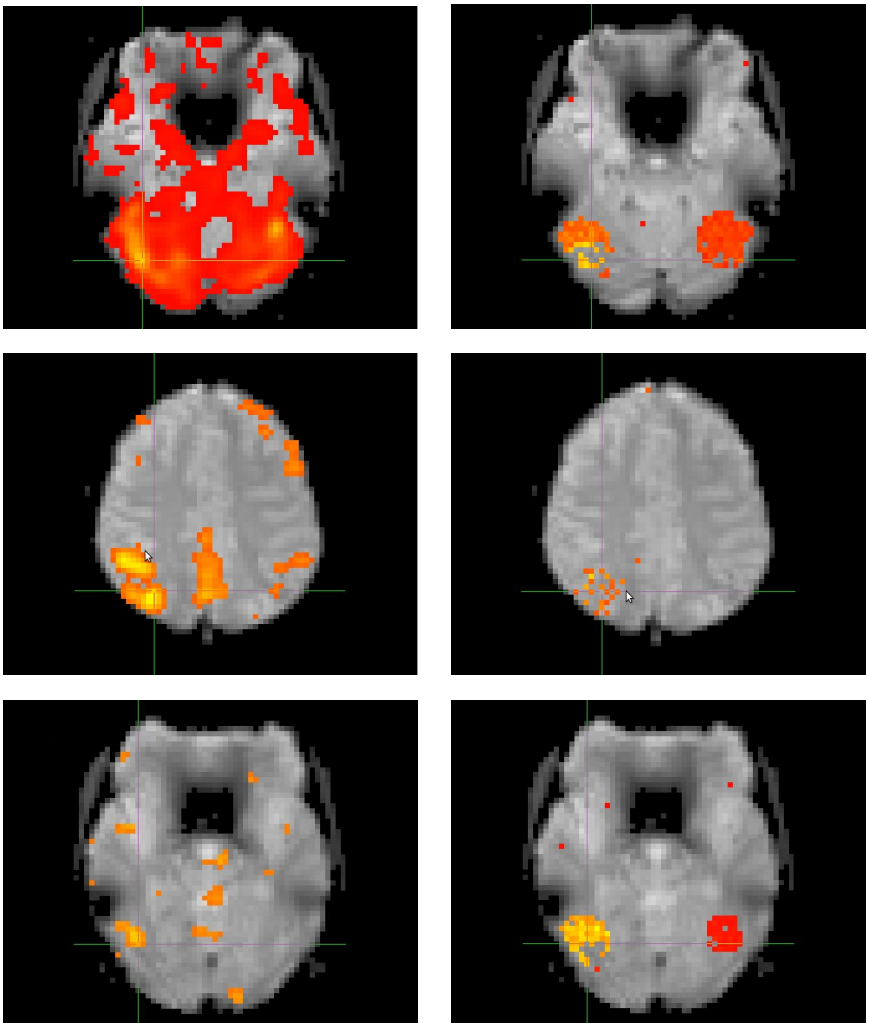

Figure 4.7 – Axial slice view of the activated voxels for the real (left) and simulated data (right): faces versus baseline contrast (top), famous versus non-famous contrast (middle), first versus second presentation contrast (bottom)

state data as for fMRI data. Other future plans are to include more neurobiological models, for example, the metabolic-hemodynamic model (Sotero & Trujillo-Barreto, 2008; Sotero et al., 2009) and spatiotemporal BOLD dynamics (Drysdale et al., 2010). To extent the generalizability of the data simulated by **neuRosim**, we plan to include the generation of complex-valued fMRI data consisting of both magnitude and phase data. To conclude, it is our hope that **neuRosim** will evolve to a general platform for simulating

fMRI data. Simulation studies should be a requisite to publish a statistical validation paper in the field of neuroscience. This will only be possible when standardized and trustworthy simulation methods using validated data generation techniques are available.

References

Bellec, P., Perlbarg, V., & Evans, A. (2009). Bootstrap generation and evaluation of an fMRI simulation database. *Magnetic Resonance Imaging*, *27*, 1382–1396.

Bianciardi, M., Cerasa, A., Patria, F., & Hagberg, G. (2004). Evaluation of mixed effects in event-related fMRI studies: Impact of first-level design and filtering. *NeuroImage*, *22*, 1351-1370.

Biswal, B., DeYoe, E., & Hyde, J. (1996). Reduction of physiological fluctuations in fMRI using digital filters. *Magnetic Resonance in Medicine*, *35*, 107–113.

Bloch, F. (1946). Nuclear induction. *Physical Review*, *70*(7–8), 460–474.

Bullmore, E., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., et al. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, *35*, 261–277.

Buxton, R., Uludağ, K., Dubowitz, D., & Liu, T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, *23*, S220–S233.

Cox, R. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.

David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., et al. (2008). Identifying neural drivers with functional MRI: An electrophysiological validation. *PLoS Biology*, *6(12)*, e315.

Drobnjak, I., Gavaghan, D., Süli, E., Pitt-Francis, J., & Jenkinson, M. (2006). Development of a fMRI simulator for modelling realistic rigid-body motion artifacts. *Magnetic Resonance in Medicine*, *56*(2), 364–380.

Drysdale, P., Huber, J., Robinson, P., & Aquino, K. (2010). Spatiotemporal bold dynamics from a poroelastic hemodynamic model. *Journal of Theoretical Biology*, *265*(4), 524-34.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*, 30–40.

Glover, G. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*, 416–429.

Grinband, J., Wager, T., Lindquist, M., Ferrera, V., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *NeuroImage*, *43*, 509–520.

Gudbjartsson, H., & Patz, S. (1995). The rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, *34*(6), 910–914.

Haacke, E., Brown, R., Thompson, M., & Venkatesan, R. (1999). *Magnetic resonance imaging: Principles and sequence design.* New York: John Wiley and Sons.

Hansen, L., Nielsen, F., Strother, S., & Lange, N. (2001). Consensus inference in neuroimaging. *NeuroImage*, *13*, 1212–1218.

Henson, R., Shallice, T., Gorno-Tempini, M.-L., & Dolan, R. (2002). Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex*, *12*, 178–186.

Hyde, J., Biswal, B., & Jesmanowicz, A. (2001). High-resolution fMRI using multislice partial k-space GR-EPI with cubic voxels. *Magnetic Resonance in Medicine*, *46*, 114–125.

Krüger, G., & Glover, G. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *46*, 631–637.

Lange, N. (1999). Statistical thinking in functional and structural magnetic resonance neuroimaging. *Statistics in Medicine*, *18*, 2401–2407.

Lange, N., Strother, S., Anderson, J., Nielsen, F., Holmes, A., Kolenda, T., et al. (1999). Plurality and resemblance in fMRI data analysis. *NeuroImage*, *10*, 282–303.

Lazar, N. (2008). *The statistical analysis of functional mri data.* Berlin, Germany: Springer Verlag.

Lee, J., Hu, J., Gao, J., Crosson, B., Peck, K., Wierenga, C., et al. (2008). Discriminating brain activity from task-related artifacts in functional MRI: Fractal scaling analysis simulation and application. *NeuroImage*, *40*, 197–212.

Lei, X., Qiu, C., Xu, P., & Yao, D. (2010). A parallel framework for simultaneous EEG/fMRI analysis: Methodology and simulation. *NeuroImage*, *52*, 1123–1134.

Liao, W., Chen, H., Yang, Q., & Lei, X. (2008). Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering. *IEEE Transactions of Medical Imaging*, *27*, 1472–1483.

Lin, Q.-H., Liu, J., Zheng, Y.-R., Liang, H., & Calhoun, V. (2010). Semiblind spatial ICA of fMRI using spatial constraints. *Human Brain Mapping*, *31*, 1076–1088.

Locascio, J., Jennings, P., Moore, C., & Corkin, S. (1997). Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. *Human Brain Mapping*, *5*, 168–193.

Logan, B., & Rowe, D. (2004). An evaluation of thresholding techniques in fMRI analysis. *NeuroImage*, *22*, 95–108.

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

Rowe, D., & Logan, B. (2004). A complex way to compute fMRI activation. *NeuroImage*, *23*, 1078–1092.

Skudlarski, P., Constable, R., & Gore, J. (1999). ROC analysis of statistical methods used in functional MRI: Individual subjects. *NeuroImage*, *9*, 311–329.

Smith, A., Lewis, B., Ruttimann, U., Ye, F., Sinnwell, T., Yang, Y., et al. (1999). Investigation of low frequency drift in fMRI signal. *NeuroImage*, *9*, 526–533.

Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219.

Sotero, R., & Trujillo-Barreto, N. (2008). Biophysical model for integrating neuronal activity, EEG, fMRI and metabolism. *NeuroImage*, *39*, 290–309.

Sotero, R., Trujillo-Barreto, N., Jiménez, J., Carbonell, F., & Rodriíguez-Rojas, R. (2009). Identification and comparison of stochastic metabolic/hemodynamic models (smhm) for the generation of the bold signal. *Journal of Computational Neuroscience*, *26*, 251–269.

Weibull, A., Gustavsson, H., Mattsson, S., & Svensson, J. (2008). Investigation of spatial resolution, partial volume effects and smoothing in functional MRI using artificial 3D time series. *NeuroImage*, *41*, 346–353.

5 | How ignoring physiological noise can bias the conclusions from fMRI simulation results

Marijke Welvaert & Yves Rosseel
Journal of Neuroscience Methods (2012), 211(1), 125–132

Abstract

Neuroimaging researchers use simulation studies to validate their statistical methods because it is acknowledged that this is the most feasible way to know the ground truth of the data. The noise model used in these studies typically varies from a simple Gaussian distribution to an estimate of the noise distribution from real data. However, although several studies point out the presence of physiological noise in fMRI data, this noise source is currently lacking in simulation studies. Therefore, we explored the impact of adding physiological noise to the simulated data. For several experimental designs, fMRI data were generated under different noise models while the signal-to-noise ratio was kept constant. The sensitivity and specificity of a standard statistical parametric mapping (SPM) analysis were determined by comparing the known activation with the detected activation. We show that by including physiological noise in the data generation process, the simulation results in terms of sensitivity and specificity drop dramatically. Additionally, we used the new proposed simulation model to compare a standard SPM analysis against the method proposed by Cabella et al. (2009). The results indicate

that the analysis of data containing no physiological noise yields a better performance of the SPM analysis. However, if physiological noise is included in the data, the sensitivity and specificity of the Cabella method are higher compared to the SPM analysis. Based on these results, we argue that the results of current simulation studies are likely to be biased, especially when analysis methods are compared using ROC curves.

5.1 Introduction

Neuroimaging researchers using functional magnetic resonance imaging (fMRI) depend highly on the statistical analysis of their data. This can be mainly attributed to two reasons. First, the response of interest, the blood oxygenated level dependent (BOLD) contrast, is embedded in a very noisy signal and second, the BOLD response itself shows high variability among brain regions, scanning session and subjects. Consequently, statistical methods are under constant development and new techniques are published on a regular basis. Validating these techniques, both old and new, should be considered a main issue in this fast developing research field. Validating a statistical method can have different purposes. One goal of validation studies can be to check if the method works properly in ideal circumstances, i.e. when the assumptions of the model hold. Another goal can be to investigate the robustness and statistical properties of the method in more realistic circumstances. Some studies look at test-reliability and reproducibility of results (see for example Schuyler et al., 2010), while others use resampling techniques (e.g. bootstrapping) and use the distributional properties of the resampled data as validation measures (see Bellec et al., 2009, for an example of using parametric bootstrapping). However, validation studies often need to know the ground truth of their data, and this is almost impossible for human data without very invasive procedures (David et al., 2008) or technically challenging constructions (Cheng et al., 2006; Brosch et al., 2002). As a solution, simulation studies have gained great interest as a validation tool because in these studies the data themselves are generated under a certain model representing the ground truth.

Based on a search of the literature, we noticed a discrepancy between the methods of generating fMRI noise that are currently used in simulation studies and the several noise sources that are known to be present in fMRI data. Based on this discrepancy we wanted to investigate if including all noise sources in the data generation process has an effect on the simulation results. In this paper we give an overview of data generating methods (section 2.1) and a detailed description of fMRI noise (section 2.2). We continue

with describing two simulation studies (section 3 and 4) in which we compared three types of fMRI noise generation by assessing the sensitivity and specificity of an activation detection analysis. In a third simulation study (section 5) we show how the different noise generation models can effect the results of a comparison between statistical analysis methods. Finally we give a discussion of the main results (section 6) and highlight some implications for fMRI simulation research. Additionally, full data generation details are described in Appendix.

5.2 Simulating fMRI data

5.2.1 Data generating methods

The greatest challenge of simulating fMRI data is how to generate the data realistically such that they resemble empirically acquired datasets as closely as possible. Taking into account the complexity of an fMRI signal, it is not surprising that a variety of data generating methods is present in the literature. Without being exhaustive, we will give a short overview and discussion of the simulation methods.

The currently most used method is the so-called *hybrid simulation* (Bianciardi et al., 2004; Lange, 1999; Weibull et al., 2008; Lee et al., 2008; Lange et al., 1999; Hansen et al., 2001; Skudlarski et al., 1999). This technique combines known activation with "real" noise. Resting-state noise is acquired in a standard scanning process and next, activated time series that are the result from the convolution of an experimental design with a known haemodynamic response function (i.e. canonical HRF) are added to the data. Consequently, the simulated data are very close to real datasets. However, these data contain a lot of unintended and often unknown factors, such as activity from the default mode network (Raichle & Snyder, 2007). Some authors tried to avoid this unwanted activity by constructing the noise from summary statistics based on the real data (Bellec et al., 2009) or by modelling the noise as the resampled residuals from a GLM-analysis of the real data (Havlicek et al., 2010).

Secondly, another series of methods define the activation as the convolution of an experimental design function with a known HRF. To construct fMRI time series, noise is added to the convoluted stimulus function. In contrast to the first described method, the noise is not based on real data, but is described as a stochastic process from an underlying known distribution. Two approaches can be distinguised. The first approach is to model the noise as purely white Gaussian noise (Lei et al., 2010; Liao et al., 2008; Lin et al., 2010). Hence, the assumption is made that the noise is independently and identically distributed, which is a major oversimplification of reality. The second method models the noise as a first-order autoregressive (AR) model, often combined with additive white noise (Grinband et al., 2008; Locascio et al., 1997; Bullmore et al., 1996; Purdon & Weisskoff, 1998). Consequently, this approach takes into account temporal correlations in fMRI data. An implementation of this approach can be found in the simulation module of DCM, as part of the SPM software (Friston et al., 2007). In this module it is possible to simulate time series for a brain network (S. Smith et al., 2011). The additive noise of the time series is implemented as a mixture of Gaussian white noise and AR(1) noise.

The final method (Drobnjak et al., 2006) is mainly specialised in MRI physics and uses the Bloch equations (Bloch, 1946) to simulate fMRI data. The POSSUM simulator was originally intended to evaluate motion correction algorithms, but can easily be used for other types of validation. The magnetic signal is calculated for each voxel on each timepoint while including several artifacts like B_0-field inhomogeneities and rigid-body motion. Noise is modelled as additive, independent, white Gaussian noise, although respiratory and cardiovascular noise can also be included. The simulator, as implemented in FSL (S. Smith et al., 2004) delivers very realistic data, however it is quite time-consuming and not often used for Monte Carlo simulation studies.

Almost all above described data generating methods (except the last method) have in common that the known activation is based on the generation of time series, while the main difference is in defining the noise. Because of the discrepancies in noise simulation, in this study, we will elaborate on the

use of noise models while generating data. That noise should be considered as an important factor in fMRI simulation studies is quite trivial. It is generally known that roughly 1% of the fMRI signal is desired activation. The rest of the signal can be classified as noise. Therefore, in the next section, we will focus on what causes fMRI noise.

5.2.2 Disentangling the noise

The noise present in fMRI data can be seperated in three major components, namely (1) thermal noise, (2) system noise and (3) subject and task-related noise (see Lazar, 2008, Chapter 3, for an overview). Thermal noise is the result of collisions between tissue electrons and the electronic components of the scanner. The number of collisions is related to both the room temperature in which the scanning process takes place, and the strength of the magnetic field (Edelstein et al., 1986). This type of noise, which is often referred to as white noise, has a random nature, meaning that it can be averaged out. This noise source is mostly operationalized by drawing randomly from a Gaussian distribution.

System noise can be attributed to fluctuations in the scanner hardware, such as inhomogeneities in the static magnetic field and instabilities in the gradient fields. These scanner instabilities are the main causes for low-frequency drift in the signal (A. Smith et al., 1999). A. Smith et al. (1999) showed that 13.7% to 68% of the voxels showed systematic changes in the measured signal. In addition, the low frequency drift can create unrelated patterns in adjacent voxels because the signal intensity in one voxel can have different fluctuations over time compared to a neighbouring voxel.

Since fMRI studies are often conducted with humans, subject- and task-related noise cannot be ignored. This type of noise is often referred to as coloured noise. A first major source of this coloured noise is head motion. Noise related to head motion cannot be regarded as random. Not only does the head move entirely, which creates extra spatial dependence between voxels, the movement is often induced by the experimental task.

The second source of task- and subject related noise, physiological noise,

often refers to respiration and heart-beat related noise, but is not restricted to it. Other possible sources are spontaneous neural activity due to the experimental task (Hyde et al., 2001) and fluctuations in the transverse relaxation rate, which are closely linked to brain physiology (Krüger & Glover, 2001). Several studies demonstrate that acknowledging physiological noise is unavoidable in fMRI studies. Not only is the noise dependent on the magnetic field strength (i.e. physiological noise increases with the signal strength), the contribution of physiological noise to the fMRI signal is higher than that of thermal and system noise (Krüger & Glover, 2001). In addition, physiological noise is not uniformally distributed over the brain but shows a clear spatial structure, namely, the noise is more pronounced in grey matter (Lund et al., 2006). Therefore, since functional studies are more and more conducted at higher field strengths and functional activity is almost exclusively contained in grey matter, physiological noise deserves a lot of attention.

As an alternative to this additive noise model, the observed MR signal can be considered approximate Rician. In general, it should be modeled as a square root of a mixture of noncentral (χ^2) distributed random variables, where the noncentrality parameter describes the mean effect including physiological noise and system noise. However, with higher SNR values this noncentral χ^2 distribution can be approximated by a Gaussian distribution. Since the SNR of fMRI is usually high enough, it is safe to assume an additive noise model (Haacke et al., 1999; Gudbjartsson & Patz, 1995). For simulation studies, the additive noise model has the additional advantage that different noise sources can be investigated seperately.

5.2.3 Current study

Based on the literature we can conclude that physiological noise is systematically present in fMRI data. However, in simulation studies it is mostly neglected. In fact, this is also the case for most of the statistical models used to analyze fMRI data. The result is that the statistical model is most likely misspecified. This leads to the estimation of projection parameters instead of the parameters of interest. This estimate will contain a systematic bias due

to the ignored physiological noise and consequently, the test statistics based on this estimate will be biased as well. In the end, this results in a lack of interpretability of the results (see Monti, 2011, for a review). Of course, when the simulation studies that validate these statistical models also neglect the physiological noise component, this bias will never be observed. In this paper, we show how this bias can be observed by introducing physiological noise in the data generation process. We will compare analysis results for datasets that were generated under different noise conditions and demonstrate that using a simple noise model leads to biased sensitivity and specificity levels.

5.3 Simulation study I

5.3.1 Simulation design and data analysis

To assess the impact of physiological noise on the sensitivity and specificity of detecting activation, we conducted a simulation study with a $3 \times 2 \times 2$ full-factorial design. We generated data based on a 20s ON/20s OFF block design where the BOLD response was simulated by convoluting a stimulus boxcar function with the canonical HRF (Friston et al., 1998). Further details on the data generation method can be found in Appendix. The first factor in the simulation design represents the noise condition. We simulated data with different types of noise and this noise layer was constructed as one of the following: (1) white noise only, (2) AR(1) noise only, and (3) a mixture of white noise, AR(1) noise and physiological noise. Example time series for the different noise conditions are presented in Figure 5.1. To ensure that the SNR was kept equal between the different noise conditions, some weighting parameters were introduced. Table 5.1 presents an overview of these weights per noise condition. The weights were based on the estimates of the standard deviation of the different noise components by Krüger & Glover (2001). Based on 6 subjects and an additive noise source model, they estimated mean standard deviations of white noise and two physiological noise components. We used these estimates to set the ratio between system/thermal noise (i.e. white and AR(a) noise) on the one hand and physiological noise

on the other hand. The ratio of the estimated mean standard deviations is equal to 0.333. So, physiological noise accounts for almost two thirds of the total noise variance. The ratio between heart and respiratory noise and task related noise was based on the ratio of the estimated standard deviation of the physiological noise components, namely 0.396.

The second factor of the simulation design represents the activation generation method (i.e. fMRI activation without noise). We simulated activation data based on the Bloch equations on the one hand, and we generated time series on the other hand. The third factor controls the signal-to-noise ratio. We included two levels of SNR in our simulation study, both high SNR datasets (SNR=10) and low SNR datasets (SNR=5) were generated. The activation based on the Bloch equations was generated using POSSUM (S. Smith et al., 2004). All other data were generated using neuRosim (Welvaert et al., 2011). The simulation study was conducted in R (R Development Core Team, 2010) and each cell of the simulation design was replicated 100 times.

All generated datasets were analyzed using a standard GLM analysis with SPM8 (Friston et al., 2007) using the canonical HRF with time and dispersion derivatives and an AR(1) autocorrelation model. The default setting for the high-pass filter was used (i.e. 128). The resulting F-map was used as the starting point for a ROC analysis in which both the sensitivity and specificity were evaluated. In the ROC analysis, the F-map was thresholded with alpha's ranging from 0.01 to 0.99. For each alpha, the average true positive rate (TPR) was determined as the number of correctly detected voxels compared to the number of active voxels in each simulation run. Similarly, the average false positive rate (FPR) was determined as the number of falsely selected voxels on the number of non-active voxels in each simulation run.

5.3.2 Results and discussion

The results of the ROC analysis are presented in Figure 5.2. The ROC curves are based on the average TPR and average FPR for each specified significance level calculated per condition over all datasets in the Monte Carlo simulation. As can be seen from the figure, the analysis of the white noise

Table 5.1 – Weights of the different noise sources to ensure constant SNR for each noise condition

	Noise condition		
	Condition 1	Condition 2	Condition 3
1. White noise	1	0	0.165
2. AR(1) noise	0	1	0.165
3a. Low-frequency drift	0	0	0.12
3b. Heart/Respiratory noise	0	0	0.156
3c. Task related noise	0	0	0.394

data with high SNR results in almost perfect ROC curves for both the Bloch equations method (Figure 5.2b) and the time series method (Figure 5.2d). The curves are situated completely in the upper left corner of the figure, meaning that high sensitivity and specificity is obtained. However, when temporal correlation is added to the data (noise condition 2), both the sensitivity and specificity decrease (blue line). Furthermore, in noise condition 3, the analysis of data including physiological noise results in a further drop of the sensitivity and specificity (Figure 5.2b and 5.2d, red line). This decrease in sensitivity is most pronounced for the time series method. For the analyses of the low SNR data (Figure 5.2a and 5.2c), we observe the same pattern with a general decreased level of the average TPR and FPR.

As a main finding from this simulation study we can state that including physiological noise in the data generation process has a significant impact on the results of the analysis. We observe a clear difference with the white noise model. Of course, this is not a suprising result. The white noise perfectly meets the assumptions of the conducted SPM analysis. Therefore, a decrease in power when physiological noise is added can be explained by violating these assumptions (i.e., not all physiological noise sources are accounted for in the statistical model). More specifically, the variance model of the GLM is incorrect so as a result, the assessment of the variances of the estimates will be inaccurate, inevitably leading to incorrect scaling of the test statistics. A similar logic is also applicable in the comparison between noise condition 2

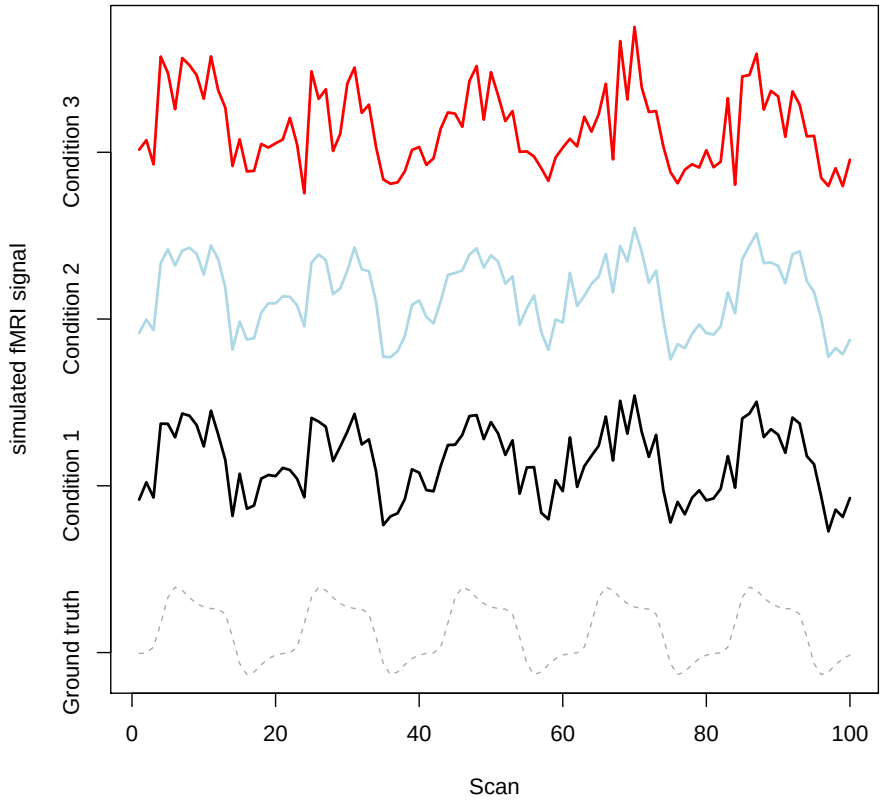

Figure 5.1 – Example time series for the different noise conditions and the design specified in simulation study I.

and 3, because the SPM model takes into account temporal correlations of the time series. Indeed, for both methods we see a further drop of the power and FPR in condition 3 compared to condition 2.

This simulation study used the powerful block design to create an activity pattern. However, in reality, fMRI studies often use event-related designs because they are more closely related to the designs of the behavioural experiments that often precede fMRI studies. We expect that the results of this simulation study can be extended to event-related designs. However, it is possible that this design suffers more from physiological noise because

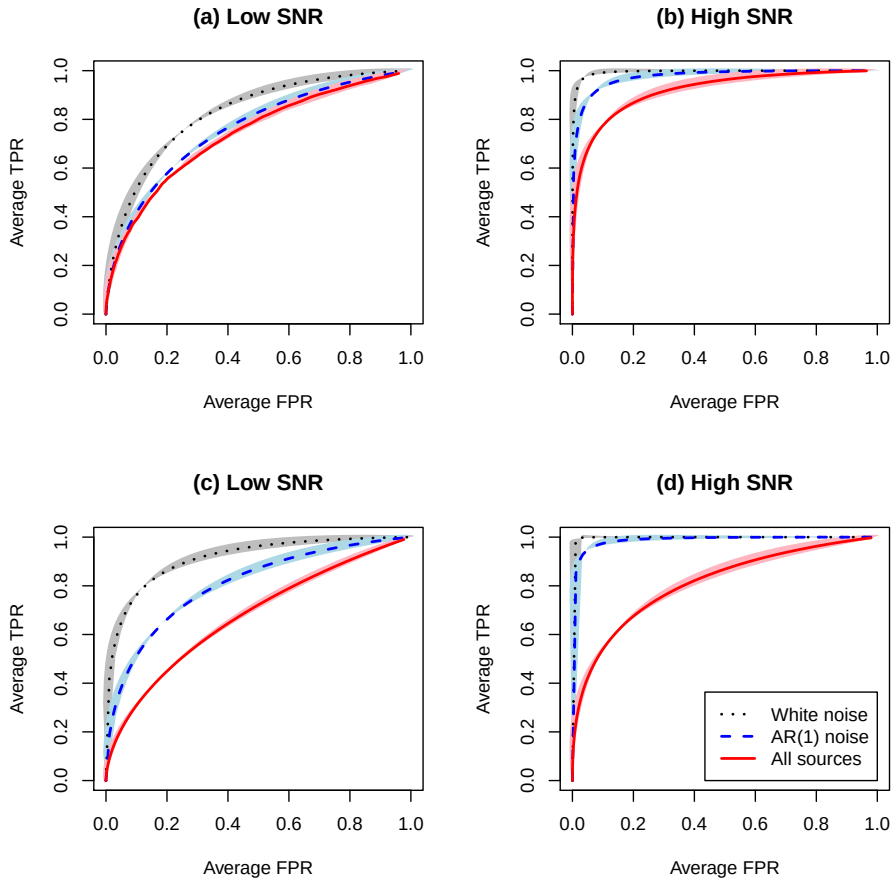

Figure 5.2 – ROC curves representing the average TPR and FPR in the three noise conditions in simulation study I. Shaded areas represent the 95% confidence intervals. The upper panel shows the results for the Bloch equations method and the lower panel contains the results for the time series method.

the frequency of events is more likely to interfere with the frequency of the physiological noise. Therefore, we conducted a second simulation study for an event-related fMRI experiment.

5.4 Simulation study II

5.4.1 Simulation design and data analysis

In this simulation study we used the same full-factorial design as in simulation study I, however, now applied to an event-related fMRI experiment. The experimental design was based on an event-related repetition priming experiment (Henson et al., 2002). We generated activation for the first 100 scans of this experiment (i.e. 34 events). Again we considered 3 noise conditions, 2 data generating methods and 2 SNR levels. Further details can be found in Appendix. A standard SPM analysis was carried out (same setting as in simulation study I) and ROC curves based on the average true positive rate and average false positive rate were determined.

5.4.2 Results and discussion

The results of the ROC analysis are presented in Figure 5.3. As in simulation study I, the ROC curves are based on the average TPR and FPR calculated per noise condition and over the different levels of SNR. We observe again a clear difference in the results for the physiological noise datasets compared to the white noise and temporal noise datasets. This means that we obtain lower sensitivity and specificity in condition 3. However, while in the first simulation study we also saw different results for the white and temporal noise conditions, here the obtained sensitivity and specificity seems to be equal in the high SNR conditions (Figure 5.3b and 5.3d). Only for data generated under low SNR including temporal autocorrelation results in a power decrease compared to the white noise only condition as can be seen from the non-overlapping confidence intervals in Figure 5.3a and 5.3c.

The hypothesis we made above, that an event-related design would suffer more from physiological noise, seems not valid. However, we demonstrated clearly that also for event-related designs ignoring physiological noise would result in biased sensitivity and specificity results. These results can be explained analogously as for simulation study I. Due to the incorrect variance

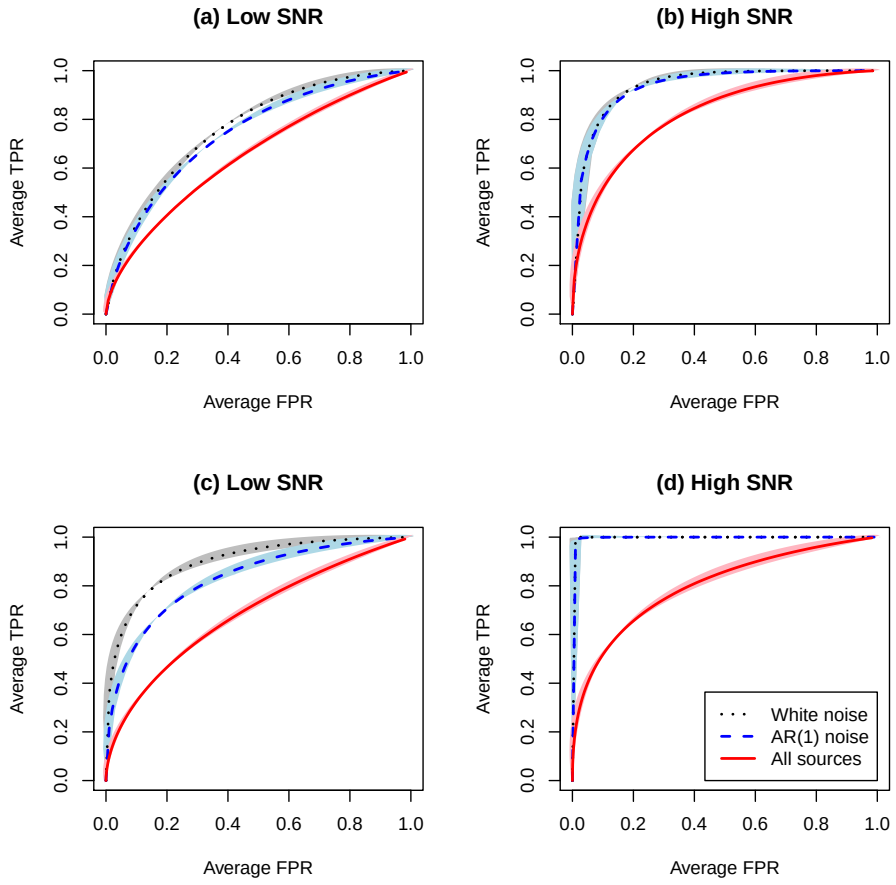

Figure 5.3 – ROC curves representing the average TPR and FPR in the three noise conditions in simulation study II. Shaded areas represent the 95% confidence intervals. The upper panel shows the results for the Bloch equations method and the lower panel contains the results for the time series method.

modeling, the scaling of the test statistics is biased.

So far, we only observed a power decrease due to the bias induced by the physiological noise. In general, this would mean that the results of current simulation studies provide us only with a polished version of the reality. However, a problem arises when the drop in sensitivity and specificity would not be equal between different analysis methods and so, when comparing these methods, using white noise only would result in wrong conclusions

about the validity of the methods. In particular ROC curves are a valuable tool to validate new analysis techniques against more established ones (see for example Luo & Puthusserypady, 2007; Lange et al., 1999; Hansen et al., 2001; Skudlarski et al., 1999). The method that would result in the highest ROC curve is most likely the best method to use. The question is now if the same method is chosen as the best one if physiological noise is incorporated in the data generation process. Therefore, in the final simulation study we compared the standard SPM analysis against an analysis method for event-related fMRI developed by Cabella et al. (2009). The latter technique is based on the generalized relative entropy of the time series and used the Kullback-Leibler divergence to distuinguish between activated epochs and resting epochs (see Cabella et al., 2009, for a full description). Again we used ROC curves to assess the sensitivity and specificity of both methods.

5.5 Simulation study III

5.5.1 Simulation design and data analysis

Following the simulations of Cabella et al. (2009), we generated time series for an event-related design with 24 events and an interstimulus interval (ISI) of 7 scans with $TR = 1.5$. The stimulus time course was convolved with the canonical HRF and noise was added according to the same conditions as in simulation studies I and II. The SNR was set to 10 and all other parameters were identical to the Cabella study. We performed 10 000 simulations for the activated time series to determine the average power and 10 000 simulations of noise only time series to assess the average FPR.

Each generated time series was analyzed using a standard SPM analysis. According to the Cabella method, we also computed the mean generalized relative entropy based on the sample average of the 24 epochs along the whole time course. This value was evaluated as a statistic for testing the null hypothesis that the time course is pure noise against the alternative hypothesis that the time course is composed of both BOLD signal and noise. For both analyses ROC curves were constructed based on the average TPR

and average FPR.

5.5.2 Results and discussion

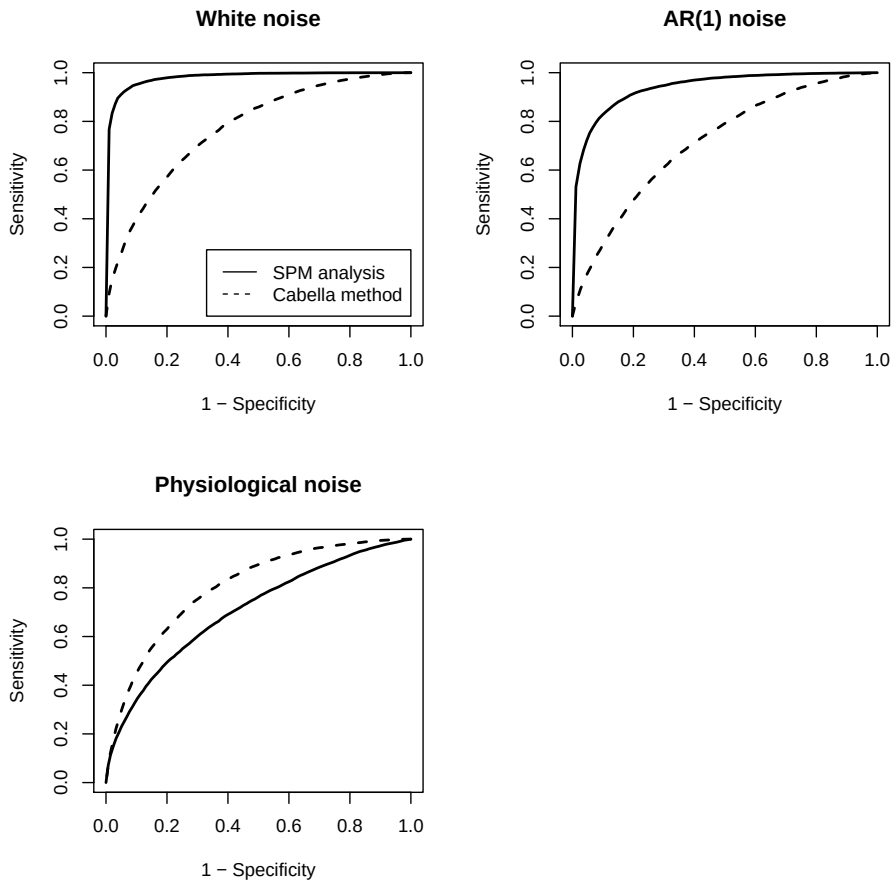

Figure 5.4 – ROC curves comparing the SPM analysis against the Cabella method for the three noise conditions in simulation study III. Shaded areas represent the 95% confidence intervals.

The results of the ROC analyses are presented in Figure 5.4. Based on either the white noise or the AR(1) noise condition, we would conclude that the SPM analysis outperforms the Cabella method in terms of achieving higher power combined with low FPR levels. However, when physiological noise is included in the data, the conclusion changes drastically. As was

observed in the previous simulation studies, the ROC curve for the SPM analysis drops dramatically, but this is not the case for the Cabella method. On the contrary, in the physiological noise condition the Cabella method results in higher sensitivity and specificity compared to the white noise and AR(1) noise conditions. When comparing the methods this results in exactly the opposite conclusion as made in the white noise condition. The Cabella method shows higher power and lower FPR compared to the SPM analysis.

The distinction between the two methods boils down to a difference in mean structure. In the GLM analysis the observed time series are compared to the convolved stimulus function, while in the Cabella method the distance between activation and resting periods in the observed time series is calculated. In general, the Cabella method is a more robust method than the standard GLM analysis and therefore suffering less from the physiological noise component. However, if only using a simple noise model in the simulation study, this method will not be chosen as the best method. Therefore, it is of great interest to developers of robust analysis techniques to use physiological noise in their simulation studies. Only this way, the benefits of their methods can be highlighted and more insight can be provided in how the methods behave in more realistic circumstances.

5.6 General discussion

Simulation studies conducted to validate fMRI data analysis methods often use a very simple noise model to generate fMRI data. The noise is typically Gaussian distributed and white (e.g. S. Smith et al., 2011). In case coloured noise is considered, this is limited to introducing temporal autocorrelation (e.g. Grinband et al., 2008). However, a number of studies have shown that fMRI data contain physiological noise (e.g. Hyde et al., 2001). Another approach of creating hybrid simulations (e.g. Weibull et al., 2008), mixing known activation with real measured noise, does include this type of noise, but suffers from the major disadvantage that the data can contain unwanted activity, so the ground truth is not entirely known. The question is if a simplification of the noise model is justified when assessing statistical

properties like sensitivity and specificity of activation detection. To address this issue, we conducted two simulation studies that compared the described noise models used for simulating fMRI data. The results show clearly that there is indeed a difference between the simple noise models and a model containing physiological noise. We demonstrated that including physiological noise in the data generation process results in a major decrease of the obtained power and FPR levels based on an SPM analysis both in block and event-related designs and for different levels of SNR. However, in simulation study III we demonstrated that this effect is dependent on the method. For example, using the Cabella method (Cabella et al., 2009) we observed the reverse effect. Including physiological noise resulted in higher sensitivity and specificity compared to the analysis of white noise/AR(1) only data. In statistics this phenomenon is very well known as model misspecification. In terms of the statistical model, model misspecification is caused by omitting a variable that is related to the dependent variable. As a consequence, the observed relationships in the misspecified models can be biased. In terms of the data generating model, this translates to: omitting an important aspect of the data while simulating will result in a biased assessment of the parameters of interest. Stated the other way around, if the statistical model is known to neglect a systematic component of the data, i.e. physiological noise, including this component in the data generation process of simulation studies would be necessary to have clear insight in the consequences of the misspecification of the statistical model. Although preprocessing techniques might reduce the influence of physiological noise (see for example Biswal et al., 1996), the use of these techniques is not widespread probably due to a lack of implementation in widely used software packages. Therefore, the search for more robust analysis methods driven by extensive validation research should be encouraged.

In our simulation studies we used ROC curves as a measurement to compare the results for the different data generating methods. These ROC curves are also an instrument to contrast analysis techniques (e.g. Lange et al., 1999; Luo & Puthusserypady, 2007; Hansen et al., 2001; Skudlarski et al., 1999). Our concern is that the size of the systematic drop in the ROC curve we saw in

this study, can be different for other analysis techniques as was demonstrated in simulation study III. Consequently, when comparing analysis methods, the difference between the ROC curves for these methods is unpredictable when physiological noise is considered in the simulated data. Therefore, including physiological noise in the simulation process would be highly beneficial, if not necessary, in order to have a thorough understanding of simulation studies that validate fMRI data analysis methods.

In summary, simulation studies that are used to assess statistical properties or to validate statistical methods for fMRI data analysis may suffer from model misspecification if physiological noise is ignored in the data generation process. In general, this will probably result in an overestimation of the sensitivity and the specificity of the analysis but in more particular cases this may result in a biased view on the performance of analysis techniques. Therefore, we recommend including physiological noise when simulating fMRI data in order to avoid the bias in the simulation results.

References

Bellec, P., Perlbarg, V., & Evans, A. (2009). Bootstrap generation and evaluation of an fMRI simulation database. *Magnetic Resonance Imaging*, *27*, 1382–1396.

Bianciardi, M., Cerasa, A., Patria, F., & Hagberg, G. (2004). Evaluation of mixed effects in event-related fMRI studies: Impact of first-level design and filtering. *NeuroImage*, *22*, 1351-1370.

Biswal, B., DeYoe, E., & Hyde, J. (1996). Reduction of physiological fluctuations in fMRI using digital filters. *Magnetic Resonance in Medicine*, *35*, 107–113.

Bloch, F. (1946). Nuclear induction. *Physical Review*, *70*(7–8), 460–474.

Brosch, J., Talabave, T., Ulmer, J., & Nyenhuis, J. (2002). Simulation of human respiration in fMRI with a mechanical model. *IEEE Transactions on Biomedical Engineering*, *49*, 700–707.

Bullmore, E., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., et al. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, *35*, 261–277.

Cabella, B., Sturzbecher, M., Araujo, D. de, & Neves, U. (2009). Generelized relative entropy in functional magnetic resonance imaging. *Physica A*, *388*, 41–50.

Cheng, H., Zhao, Q., Duensing, G., Edelstein, W., Spencer, D., Browne, N., et al. (2006). Smartphantom – an fMRI simulator. *Magnetic Resonance Imaging*, *24*, 301–313.

David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., et al. (2008). Identifying neural drivers with functional MRI: An electrophysiological validation. *PLoS Biology*, *6(12)*, e315.

Drobnjak, I., Gavaghan, D., Süli, E., Pitt-Francis, J., & Jenkinson, M. (2006). Development of a fMRI simulator for modelling realistic rigid-body motion artifacts. *Magnetic Resonance in Medicine*, *56*(2), 364–380.

Edelstein, W., Glover, G., Hardy, C., & Redington, R. (1986). The intrinsic signal-to-noise ratio in NMR imaging. *Magnetic Resonance in Medicine*, *3*, 604–618.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., & Penny, W. (Eds.). (2007). *Statistical parametric mapping: The analysis of functional brain images*. Massachussets, USA: Academic Press.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*, 30–40.

Grinband, J., Wager, T., Lindquist, M., Ferrera, V., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *NeuroImage*, *43*, 509–520.

Gudbjartsson, H., & Patz, S. (1995). The rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, *34*(6), 910–914.

Haacke, E., Brown, R., Thompson, M., & Venkatesan, R. (1999). *Magnetic resonance imaging: Principles and sequence design.* New York: John Wiley and Sons.

Hansen, L., Nielsen, F., Strother, S., & Lange, N. (2001). Consensus inference in neuroimaging. *NeuroImage*, *13*, 1212–1218.

Havlicek, M., Jan, J., Brazdil, M., & Calhoun, V. (2010). Dynamic granger causality based on kalman filter for evaluation of functional network connectivity in fMRI data. *NeuroImage*, *53*, 65–77.

Henson, R., Shallice, T., Gorno-Tempini, M.-L., & Dolan, R. (2002). Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex*, *12*, 178–186.

Hyde, J., Biswal, B., & Jesmanowicz, A. (2001). High-resolution fMRI using multislice partial k-space GR-EPI with cubic voxels. *Magnetic Resonance in Medicine*, *46*, 114–125.

Krüger, G., & Glover, G. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *46*, 631–637.

Lange, N. (1999). Statistical thinking in functional and structural magnetic resonance neuroimaging. *Statistics in Medicine*, *18*, 2401–2407.

Lange, N., Strother, S., Anderson, J., Nielsen, F., Holmes, A., Kolenda, T., et al. (1999). Plurality and resemblance in fMRI data analysis. *NeuroImage*, *10*, 282–303.

Lazar, N. (2008). *The statistical analysis of functional mri data.* Berlin, Germany: Springer Verlag.

Lee, J., Hu, J., Gao, J., Crosson, B., Peck, K., Wierenga, C., et al. (2008). Discriminating brain activity from task-related artifacts in functional MRI: Fractal scaling analysis simulation and application. *NeuroImage*, *40*, 197–212.

Lei, X., Qiu, C., Xu, P., & Yao, D. (2010). A parallel framework for simultaneous EEG/fMRI analysis: Methodology and simulation. *NeuroImage*, *52*, 1123–1134.

Liao, W., Chen, H., Yang, Q., & Lei, X. (2008). Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering. *IEEE Transactions of Medical Imaging*, *27*, 1472–1483.

Lin, Q.-H., Liu, J., Zheng, Y.-R., Liang, H., & Calhoun, V. (2010). Semiblind spatial ICA of fMRI using spatial constraints. *Human Brain Mapping*, *31*, 1076–1088.

Locascio, J., Jennings, P., Moore, C., & Corkin, S. (1997). Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. *Human Brain Mapping*, *5*, 168–193.

Lund, T., Madsen, K., Sidaros, K., Luo, W.-L., & Nichols, T. (2006). Non-white noise in fMRI: Does modelling have an impact? *NeuroImage*, *29*, 54–60.

Luo, H., & Puthusserypady, S. (2007). fMRI data analysis with nonstationary noise models: A bayesian approach. IEEE *Transactions on Biomedical Engineering*, *54*, 1621–1630.

Monti, M. (2011). Statistical analysis of fMRI time-series: A critical review of the GLM approach. *Frontiers in Human Neuroscience*, *5*, doi:10.3389/fnhum.2011.00028.

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

R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org` (ISBN 3-900051-07-0)

Raichle, M., & Snyder, A. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage*, *37*, 1083–1090.

Schuyler, B., Ollinger, J., Oakes, T., Johnstone, T., & Davidson, R. (2010). Dynamic causal modeling applied to fMRI data shows high reliability. *NeuroImage*, *49*, 603–611.

Skudlarski, P., Constable, R., & Gore, J. (1999). ROC analysis of statistical methods used in functional MRI: Individual subjects. *NeuroImage*, *9*, 311–329.

Smith, A., Lewis, B., Ruttimann, U., Ye, F., Sinnwell, T., Yang, Y., et al. (1999). Investigation of low frequency drift in fMRI signal. *NeuroImage*, *9*, 526–533.

Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., et al. (2004). Advances in functional and structural

MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219.

Smith, S., Miller, K., Salimi-Khorshidi, G., Webster, M., Beckmann, C., Nichols, T., et al. (2011). Network modelling methods for FMRI. *NeuroImage*, *54*, 875–891.

Weibull, A., Gustavsson, H., Mattsson, S., & Svensson, J. (2008). Investigation of spatial resolution, partial volume effects and smoothing in functional MRI using artificial 3D time series. *NeuroImage*, *41*, 346–353.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., & Rosseel, Y. (2011). neuRosim: An R package for generating fMRI data. *Journal of Statistical Software*, *44*, 1–18.

Appendix: Data generation details simulation study I and II

All data in both simulation studies were generated according to an additive simulation model (Bellec et al., 2009). More specifically, we considered data consisting of two separate layers, namely (1) an activation layer and (2) a noise layer. Depending on the cell determining the activation generating method, the activation layer was generated as follows. For the Bloch equations method, the activated regions were based on the standard activation template that is part of the FSL distribution (S. Smith et al., 2004). We generated 100 scans of a $64 \times 64 \times 12$ dataset using an EPI pulse sequence with $TE = 0.03$, $TR = 2.04$ and an isotropic voxel size of 3mm. For the time series method, a baseline image was generated based on the MNI152 anatomical template. We selected only the voxels that consisted of more than 34% grey matter. This image containing the percentages of grey matter for each voxel was multiplied with a baseline value of 40. We selected 3 activated regions in the grey matter. The convoluted time series representing the activation were added to the voxels in these regions. The activated time series were originally generated with $TR = 0.01s$ to ensure a high-precision convolution with the HRF. The effective time points used in the simulated fMRI time series were the result of downsampling with $TR = 2s$ after convolution.

While generating the noise layer, several noise sources were considered. A white noise component was constructed by drawing randomly from a normal distribution with zero mean and standard deviation σ_N for each timepoint and each voxel. The standard deviation of the noise is determined by the signal-to-noise ratio (SNR):

$$\text{SNR} = \frac{S}{\sigma_N} \tag{5.1}$$

where S stands for the magnitude of the signal (Krüger & Glover, 2001). Temporal correlation was modelled based on AR(1) noise that was generated by

$$\varepsilon_i = \rho\varepsilon_{i-1} + \chi_i \tag{5.2}$$

with $\chi_i \sim N(0, \sigma_\chi)$ (Purdon & Weisskoff, 1998). The autocorrelation coefficient ρ was set to 0.3 and σ_χ was calculated based on equation 5.1 and the additional requirement that

$$\sigma_N^2 = \frac{\sigma_\chi^2}{1 - \rho^2}. \tag{5.3}$$

Finally, physiological noise was taken into account by modelling three components. First, low frequency drift was modelled as the sum of a basis of discrete cosine functions (Friston et al., 2007) with frequencies ranging between 0.005 Hz and 0.015 Hz (A. Smith et al., 1999). Second, heartbeat and respiratory rate noise were modelled as sine and cosine functions respectively with frequency 1.17 Hz and 0.2 Hz (Biswal et al., 1996). Additionally, random Gaussian noise was added to create stochastic variability between voxels. Third, task related noise was operationalized by adding random Gaussian noise only in activation blocks to account for the instability of cognitive processes (Hyde et al., 2001). A combination of these noise sources as described in Table 5.1 resulted in the three noise conditions used in both simulation studies.

6 | Adaptive smoothing as inference strategy: more specificity for unequally sized or neighbouring regions

Marijke Welvaert, Karsten Tabelow[1], Ruth Seurinck & Yves Rosseel

Submitted manuscript

Abstract

Although spatial smoothing of fMRI data can serve multiple purposes, increasing the sensitivity of activation detection is probably its greatest benefit. However, this increased detection power comes with a loss of specificity when non-adaptive smoothing (i.e. the standard in most software packages) is used. By conducting simulation studies and analysing experimental data, we systematically investigated the effect of spatial smoothing on the power and number of false positives in two particular cases that are often encountered in fMRI research: (1) Single condition activation detection for regions that differ in size, and (2) multiple condition activation detection for neighbouring regions. Our results demonstrate that adaptive smoothing is superior in both cases because less false positives are introduced by the spatial smoothing process compared to standard Gaussian smoothing or FDR inference of unsmoothed data.

[1]Weierstrass Institute of Applied Stochastics and Analytics, Berlin, Germany

6.1 Introduction

For many neuroscientists, spatial smoothing of fMRI data has become an automatic preprocessing step. The purpose of this smoothing procedure can be threefold. First, spatial smoothing moderates intersubject variation in brain anatomy, especially when individual brains are transformed to a standard brain space in order to allow intersubject comparison. Second, the smoothing of fMRI data will increase the signal-to-noise ratio (SNR). Third, a voxel-based mass-univariate analysis of fMRI data calls for the need of multiple testing corrections. In this context, spatial smoothing enhances random field theory (RFT) based inference (see for example Worsley, 2003).

Although spatial smoothing, also referred to as spatial filtering, is mostly performed during the preprocessing stage of the analysis, it is actually a crucial step in the whole data analysis process because of its influence on the sensitivity and specificity of the activation detection analysis. The critical point here is that we need a useful estimate of the required width of smoothing. Depending on the goal of smoothing, guidelines vary substantially. To allow intersubject averaging, more smoothing might be necessary (e.g. a Gaussian kernel full-width half-maximum (FWHM) of 8 mm, which is often the default value in software packages) (Friston et al., 2007). In the context of increasing SNR and based on the matched filter theorem (Rosenfeld & Kak, 1982), one wants to take into account the size of the region of interest (ROI), which can vary roughly from 2 to 10 mm, or even more. Finally, the spatial smoothness needed for valid statistical inference can be rather small (around 4 mm will mostly be sufficient) (Friston et al., 1994). In terms of maintaining spatial structure, it would be better to smooth as little as possible. A smoothing kernel width that is at least twice the voxel size should be appropriate in almost all cases (Poldrack et al., 2011).

6.1.1 Non-adaptive smoothing

Spatial smoothing is applied by the convolution of each volume of the fMRI dataset and a Gaussian kernel. In practice this translates to the signal in

the volume being blurred by averaging the data over all voxels in a spherical region. Based on the size of the FWHM of the Gaussian kernel, local weights determine if a certain voxel is part of this region. In the case of non-adaptive smoothing, these weights are acting as indicators on how far the smoothing kernel reaches in space and are completely defined by the value of the FWHM.

Although spatial smoothing increases the sensitivity of activation detection, it also has a major disadvantage, namely the decrease of spatial resolution. This results in loss of information on the spatial extent and the shape of activated regions. Moreover, when the smoothing kernel is large compared to the activated area, the sensitivity of activation detection will decrease and false positives are introduced. Some algorithms were developed to overcome this increase of false positive rate and decrease of spatial resolution, such as, for example, scale space methods (Poline & Mazoyer, 1994) and non-linear filtering (Descombes et al., 1998). In this paper, we will focus on the most standard spatial smoothing method, referred to as Gaussian smoothing, as an example of non-adaptive smoothing.

6.1.2 No smoothing

In a reaction to the specificity loss caused by non-adaptive smoothing, some researchers omit the smoothing step and perform inference based on unsmoothed data. The result is that activation detection is probably highly specific, but on the other hand might also be overly conservative. First, there has to be enough signal/contrast in the data to be sensitive to the activation, and second, inference that corrects for multiple comparisons based on RFT will be conservative due to unsufficient smoothness in the data (Logan & Rowe, 2004). One solution to get more sensitive results out of unsmoothed data would be to use False Discovery Rate control (FDR) (Benjamini & Hochberg, 1995), since this method is known to be more sensitive (Logan & Rowe, 2004). We will consider both strategies as alternatives to spatial smoothing of the data.

6.1.3 Adaptive smoothing

A second option to overcome the drawbacks of non-adaptive spatial smoothing, and probably also of no smoothing, is to smooth adaptively. With this smoothing technique, the smoothing width is not chosen a priori but based on the data by using for example adaptive region growing, nonstationary spatial Gaussian Markov random fields or adaptive weights smoothing (e.g. Lu et al., 2003; Harrison et al., 2008; Yue et al., 2010; Tabelow et al., 2006; Polzehl et al., 2010). We will discuss two of these methods in more detail which will serve as examples of adaptive smoothing.

In 2006, Tabelow et al. introduced a structural adaptive smoothing procedure based on the propagation-separation approach (Polzehl & Spokoiny, 2006). Without using any prior anatomical knowledge, the methodology takes into account the size and shape of activated regions (Tabelow et al., 2006). During an iterative process adaptive weighting schemes at each location are determined based on the parameter estimates of the statistical parametric map (SPM). These weights separate areas of distinct parameter values in particular activated voxels from non-activated voxels. This avoids the blurring bias at the borders of these areas typically observed in non-adaptive Gaussian filtering (see Tabelow et al., 2006, for the technical details). Signal detection can then be performed, for instance, using thresholds based on Random Field Theory (Adler, 2000; Worsley, 1994, 2003). Additionally, Polzehl et al. (2010) presented a new version of the structural adaptive smoothing algorithm, namely structural adaptive segmentation. Their extended algorithm combines the estimation and smoothing step with the inference step based on multiscale tests. Both simulations and analysis of real data showed higher sensitivity and specificity compared to the original adaptive smoothing procedure (see Polzehl et al., 2010, for the technical details).

A major advantage of these adaptive smoothing techniques is that users only have to provide the maximum bandwidth of smoothing while the algorithm determines the optimal local weighting scheme. Moreover, when no spatial structure can be detected based on the functional activation, the smoothing procedure reduces to non-adaptive Gaussian smoothing such that

the SPM under the null hypothesis is again approximately a Gaussian Random Field.

It should be noted that spatial smoothing is usually part of the preprocessing pipeline. However, parameter estimation, with the exception of prewhitening effects, is not effected by the spatial smoothing and therefore the order of smoothing and parameter estimation is interchangeable (Tabelow et al., 2006). Since the adaptive smoothing procedures rely heavily on the SNR in the data, parameter estimation before smoothing is necessary as a variance and dimension reduction step. Therefore, in this paper, all smoothing procedures will be applied after parameter estimation.

6.1.4 Current study

Despite the fact that the disadvantages of non-adaptive spatial smoothing are well-known and that new improved methods (e.g. adaptive smoothing) have been introduced, Gaussian smoothing is still used almost exclusively, probably because it is widely available. In this paper, we address two particular situations that are applicable to fMRI research and for these situations we provide a systematic analysis of the dependence of the sensitivity and specificity on smoothing parameters in a combination with inference methods. By applying spatial smoothing, we will increase the SNR resulting in higher sensitivity, but, by comparing the results for different smoothing and inference methods, we will look for an optimal trade-off between gaining sensitivity and losing specificity.

In the first study (section 2), experimental tasks are considered that cause activation in multiple unequally sized regions. When we want to take into account the spatial extent of these regions, the choice of the value of the smoothing kernel width is non-trivial. Assuming that the actual size of the activated regions is known (e.g. based on anatomical structures, localizer tasks or previous analyses of similar tasks), the researcher is confronted with a choice of the FWHM value ranging from the size of the smallest region to the size of the largest region. The former will result in undersmoothing the larger regions, while choosing the latter will oversmooth the smaller regions.

The second study (section 3) investigates the effect of spatial smoothing on the specificity and sensitivity of activation detection when two neighbouring regions are activated due to different tasks, conditions or contrasts. Here, we definitely want to avoid oversmoothing because this could result in an unintended overlap between the activation regions.

For both studies, results for unsmoothed, non-adaptively smoothed and adaptively smoothed data are contrasted. For the unsmoothed data, RFT based Family-Wise Error (FWE) control and FDR control[2] are compared for two reasons. First, since the latter inference method is known to be more sensitive (Logan & Rowe, 2004), we will investigate its properties as an alternative to using spatial smoothing to augment the power. Secondly, due to unsufficient smoothness, RFT inference is known to be too conservative on unsmoothed data. For the smoothed data, inference is based on RFT (except for the adaptive segmentation algorithm which used multiscale tests). Other inference methods, like for example cluster-based inference are not used here as they have the same blurring problem as RFT when relying on a Gaussian filter. Using more realistic simulations (as opposed to Tabelow et al., 2006; Polzehl et al., 2010) the empirical power and false positive rate (FPR) are assessed at several contrast-to-noise ratio (CNR) values (section 2.1 and 3.1). As an illustration, the conditions from both studies are also applied on real data (section 2.2 and 3.2).

6.2 Study 1: regions with unequal size

6.2.1 Simulation study

Data generation and design

We simulated fMRI datasets of size $20 \times 20 \times 20 \times 107$ consisting of 2 mm^3 isotropic voxels in R (R Development Core Team, 2010) using the package **neuRosim** (Welvaert et al., 2011). The stimulus function was based on a

[2]We explicitly discard cluster-based FDR methods (e.g. Chumbley & Friston, 2009) because they are also dependent on RFT inference.

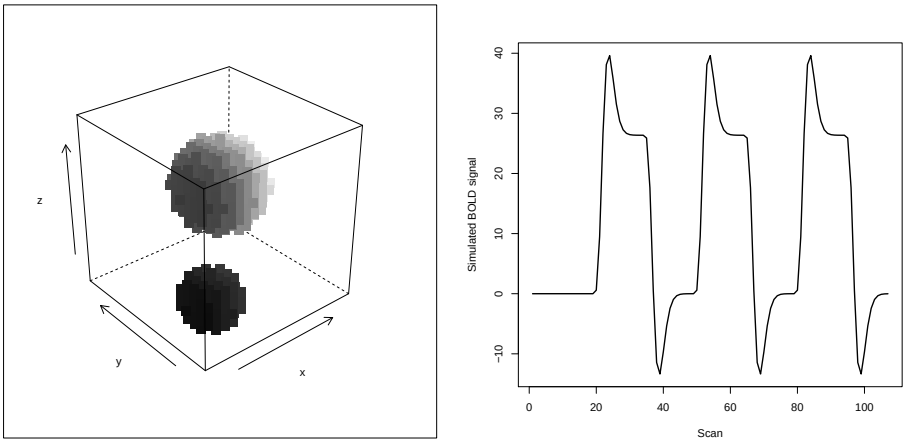

Figure 6.1 – Ground truth activation in simulation study 1. Left: location of the activated regions (grayscale values reflect position in space). Right: timecourse of the block design.

block design with 3 activation blocks that each lasted 15 scans (see Figure 6.1 for a display of the timecourse). We modelled 2 activated regions (spheres with diameter 6mm and 10mm respectively, see Figure 6.1). *Rich* noise was added including temporal correlations ($\rho = 0.3$), spatial ($\rho = 0.7$) correlations and physiological noise (i.e. noise due to heart rate, respiratory rate and task-related noise). Specifically, the noise consisted for 10% of white noise, 30% temporally correlated noise, 20% low frequency drift, 10% physiological noise, 10% task-related noise and 20% spatially correlated noise (see Welvaert & Rosseel, 2012, for more details on the noise generation). These raw data were then analysed in R using the package **fmri** (Tabelow & Polzehl, 2011). Six smoothing conditions were considered: (**Condition 1**) no spatial smoothing, (**Condition 2a**) Gaussian smoothing with the size of the smoothing kernel equal to the size of the smallest region (FWHM=6 mm), (**Condition 2b**) Gaussian smoothing with the size of the smoothing kernel equal to the average size of the regions, corresponding to the default value (FWHM=8 mm), (**Condition 2c**) Gaussian smoothing with the size of the smoothing kernel equal to the size of the largest regions (FWHM=10 mm), (**Condition 3**) structural adaptive smoothing with maximum bandwidth

Table 6.1 – Overview of the spatial smoothing and multiple corrections methods applied in simulation study 1.

Condition	Smoothing	Inference
1a	none	Random Field Theory
1b	none	False Discovery Rate
2a	non-adaptive (FWHM = 6 mm)	Random Field Theory
2b	non-adaptive (FWHM = 8 mm)	Random Field Theory
2c	non-adaptive (FWHM = 10 mm)	Random Field Theory
3	adaptive smoothing	Random Field Theory
4	adaptive segmentation	Multiscale Tests

FWHM=10 mm, and (**Condition 4**) structural adaptive segmentation with maximum bandwidth FWHM=10 mm. We also varied the CNR level (i.e., ratio of changes in the signal due to the experiment and fluctuations due to noise) of the data between 0.02 and 0.5, and each dataset was replicated 100 times. For each replication, we assessed the power and FPR of the detected activation based on a general linear model analysis including an AR(1) temporal correlation model conducted using **fmri** (Tabelow & Polzehl, 2011).

The resulting statistical parametric maps were smoothed according to the six smoothing conditions. So, in this context spatial smoothing is not considered to be part of the pre-processing steps (Tabelow et al., 2006). Sensitivity was measured by means of the average power, which was calculated as the ratio of correctly detected voxels and the total number of active voxels. Similarly, specificity was measured as the average false positive rate (FPR) obtained by taking the ratio of falsely detected voxels and the total number of non-active voxels. All results are corrected for multiple comparisons at $p < 0.05$ based on a Family-wise Error (FWE) correction using Gaussian Random Field theory in the no smoothing, non-adaptive smoothing, and adaptive smoothing conditions (Worsley, 2003), and using multiscale tests for the adaptive segmentation condition (Dümbgen & Spokoiny, 2001; Polzehl et al., 2010). As a comparison and because RFT might not work properly on unsmoothed data due to unsufficient smoothness, we also applied FDR based inference (Benjamini & Hochberg, 1995) on the non-smoothed data. Table

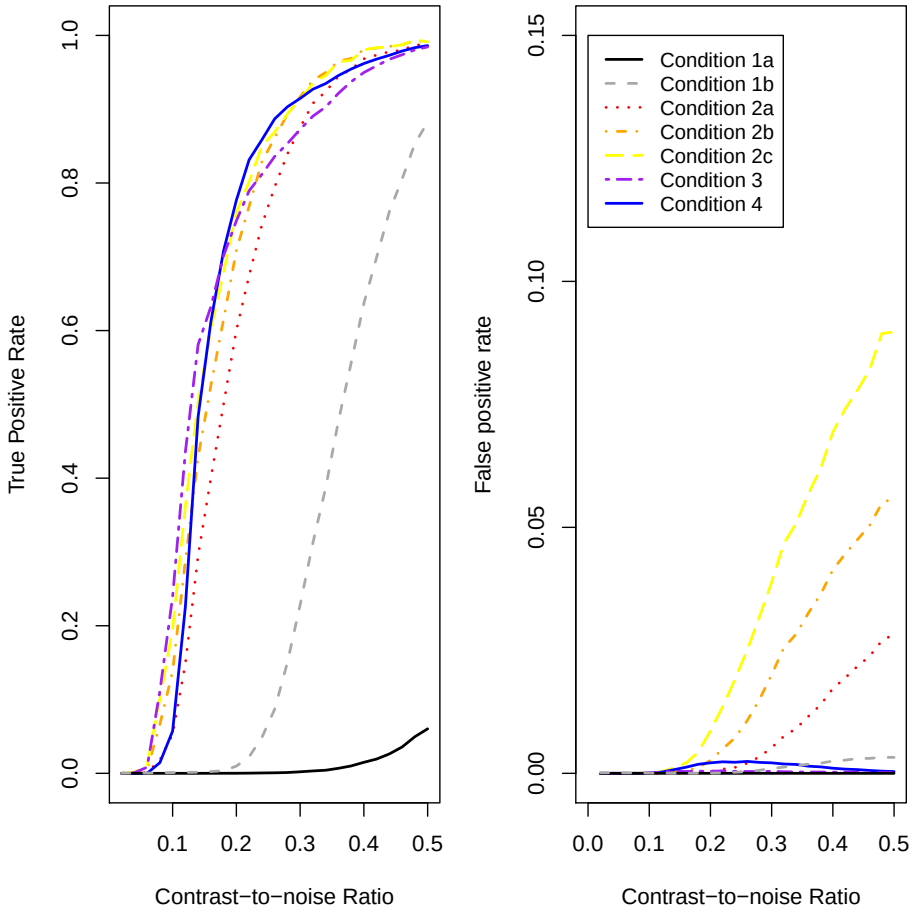

Figure 6.2 – Power (left) and FPR (right) results for the conditions in simulation study 1. Table 6.1 presents an overview of the conditions.

6.1 provides an overview of the combination of the smoothing conditions and the multiple corrections methods.

Results

The results are presented in Figure 6.2. The power results on the left hand side of the figure and the FPR results on the right hand show that there is indeed a difference between the smoothing conditions. In the no smoothing condition combined with RFT inference, the overall power and FPR levels are very low. This might be due to unsufficient smoothness in the data for

the algorithm to work properly. By comparison, using FDR control on the unsmoothed data already gives more power, so this will be our baseline condition to compare the smoothing methods against. By spatially smoothing the data we would like to obtain higher power, however, ideally, the FPR should stay as low as when no smoothing is applied. Looking at the power results first (Figure 6.2, left panel), we observe higher power values for all smoothing methods as expected. The differences between the conditions are rather small, although it seems that adaptive segmentation results in the highest sensitivity levels, while Gaussian smoothing with the smallest kernel (FWHM = 6 mm) provide a lower bound on the power results.

So, while the obtained power results are quite similar, more striking differences are observed for the FPR results (figure 6.2, right panel). For the Gaussian smoothing conditions, more false positives are observed with increasing kernel width and increasing CNR values. Since there are almost no false positives when no smoothing is applied, the increased FPR rate observed in the Gaussian smoothing conditions is completely attributable to the smoothing procedure. For example, with a medium-to-high CNR of 0.4, about 5% false positives are introduced by the Gaussian smoothing procedure in the case the default value of the FWHM is used (i.e. 8 mm). In contrast, the FPR results for the adaptive smoothing and adaptive segmentation techniques are very similar to the no smoothing condition; the number of false positives is very low. Also in the case of no smoothing with FDR based inference, the number of false positives appears to be well controlled.

Based on these simulation results, we can conclude that when there are multiple activated regions, higher sensitivity can be obtained by either adaptive smoothing, adaptive segmentation or Gaussian smoothing, compared to no smoothing (combined with either RFT or FDR). However, only the adaptive procedures (adaptive smoothing and adaptive segmentation) succeed in maintaining the specificity while this greatly decreases for the Gaussian smoothing conditions.

6.2.2 Real data

We applied the smoothing procedures as described in the simulation study above to experimental fMRI data from a passive viewing task used to localize hV5/MT+ involved in motion perception (Seurinck et al., 2011). The data were acquired on a 3T Siemens TRIO MR scanner (Siemens Medical Systems, Erlangen, Germany) with scanning parameters TR = 1940 ms, TE = 35 ms, flip angle = 80°, 28 slices, slice thickness = 3 mm with a distance factor of 17%, FOV = 244 mm, matrix = 64×64. We used 425 EPI images (corresponding to the first run) from a random subject in the dataset.

All images were motion corrected[3] and normalized to the MNI152 template using SPM8 (www.fil.ion.ucl.ac.uk/spm). Spatial smoothing was performed using the R package **fmri** (Tabelow & Polzehl, 2011) corresponding to the same conditions as in simulation study 1. In a GLM analysis, we tested the contrast *moving stimuli* versus *static stimuli*. We expect larger activation in hV5/MT+ for the moving stimuli compared to the static stimuli. Figure 6.3 shows the results for the different smoothing conditions. Since locator tasks have typically high CNR, the results for the no smoothing condition (Figure 6.3a) already show clear bilateral activation in hV5/MT+. Based on the activation detected in this condition, we distinguish two activated regions, one in the right hemisphere, which we will refer to as the small region, and one in the left hemisphere, which we will refer to as the large region. The size of the FWHM values for the spatial smoothing are in this case based on the sizes of the detected regions in these unsmoothed data. When using FDR controlled inference on the unsmoothed data (Figure 6.3b), we observe more sensitive results, but surpisingly also more activation outside the hV5/MT+ region which is improbable given the contrast. Non-adaptive (Gaussian) smoothing with a kernel width equal to the small region (FWHM = 6) gives slightly extended activation regions (Figure 6.3c). However, increasing the kernel width further results in loss of detected activation. For Gaussian smoothing with FWHM = 8 mm, the activation in the right hemi-

[3]The influence of spatial smoothing in this stage was minimized by setting the FWHM of the Gaussian smoothing kernel to 1 mm, before estimating the realignment parameters.

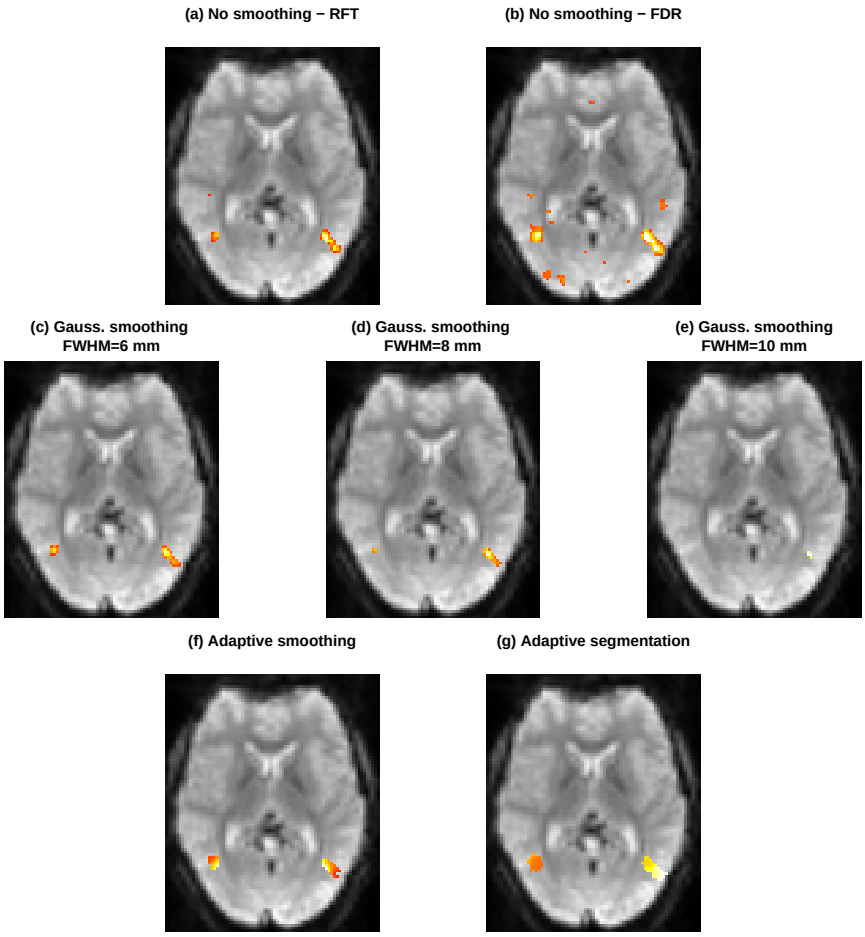

Figure 6.3 – Example slices (Left=Right orientation) showing hV5/MT+ activation in a visual localizer experiment (Seurinck et al., 2011). Brighter colours indicate stronger activation (based on the estimated signal of active voxels in case of adaptive segmentation, and on the p-values otherwise).

sphere is greatly diminished (Figure 6.3d), and with FWHM = 10 mm, it disappears completely (Figure 6.3e). For this kernel width, also in the left hemisphere only half of the activation survives the threshold. In contrast, the adaptive smoothing techniques (Figure 6.3f-g) can produce slightly larger hV5/MT+ activations with increased sensitivity compared to the no smoothing condition. This demonstrates that loss of activation due to oversmoothing is overcome with adaptive smoothing. Additionally, while sensitivity to activation is increased for both adaptive procedures, the highest sensitivity levels are observed for adaptive segmentation. Concerning the extra activation outside the hV5/MT+ regions that was detected in the unsmoothed - FDR condition, spatial smoothing (all methods) succeeds in decreasing this activation, resulting in clearly localized activation areas.

6.3 Study 2: neighbouring regions

6.3.1 Simulation study

Data generation and design

We simulated fMRI datasets of size $18 \times 18 \times 18 \times 120$ consisting of 2 mm^3 isotropic voxels in R (R Development Core Team, 2010) using the package **neuRosim** (Welvaert et al., 2011).

The experiment was a block design with 2 conditions. Both conditions contain 5 activation blocks of 10 scans which are alternately active. We modelled 2 activated regions (6×6 cubes) next to each other accounting for activation based on condition 1 and 2, respectively (see Figure 6.4), and we used the same noise model as in simulation study 1. We considered 4 conditions: (1) no smoothing with RFT inference, (2) no smoothing with FDR inference, (3) Gaussian smoothing with FWHM=6 mm (i.e. congruent with the size of the activated region), and (4) adaptive segmentation with the maximal bandwidth equal to 6 mm. The CNR varied between 0.02 and 1, which are values common in fMRI research, and we simulated 100 replications of each dataset. The data were analyzed using the R package **fmri** (Tabelow

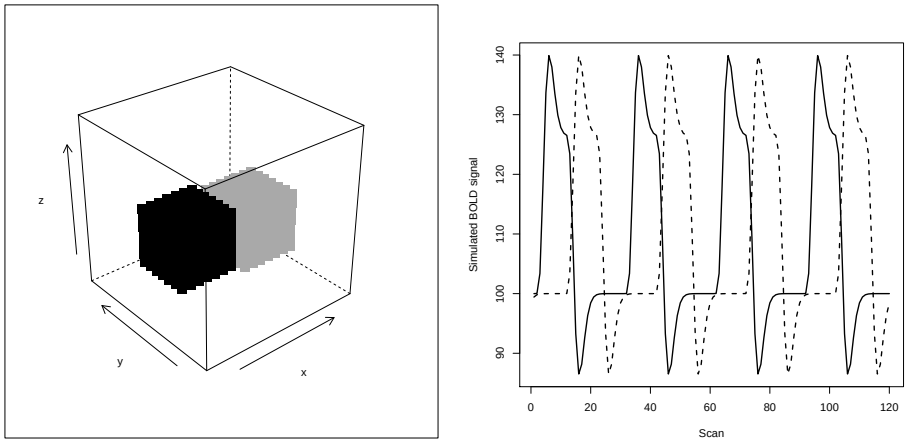

Figure 6.4 – Ground truth activation in simulation study 2. Left: location of the activated regions. Right: timecourses of the block design.

& Polzehl, 2011) and the power and FPR were determined for each activation condition separately.

Results

The results are shown in Figure 6.5. As in simulation study 1, in the no smoothing condition using RFT inference, both power and FPR are very low for all CNR levels. Again inference based on RFT might not be entirely correct because of unsufficient smoothness. The FDR based results show indeed more power, while keeping the FPR low. This condition will serve again as our baseline. Looking at the power results (figure 6.5, left panel), we see increasing power with increasing CNR levels for both smoothing methods and in both conditions. For the tests of both conditions, Gaussian smoothing results in slightly more power compared to adaptive segmentation for lower CNR values, but the reverse is the case for higher CNR values. Overall, the differences in sensitivity between both smoothing techniques are rather low. The FPR results (figure 6.5, right panel) demonstrate a better performance of adaptive segmentation. This smoothing procedure can control the number of false positives very well in both conditions. However, for the Gaussian smoothing condition, the FPR increases with increasing CNR values. Based

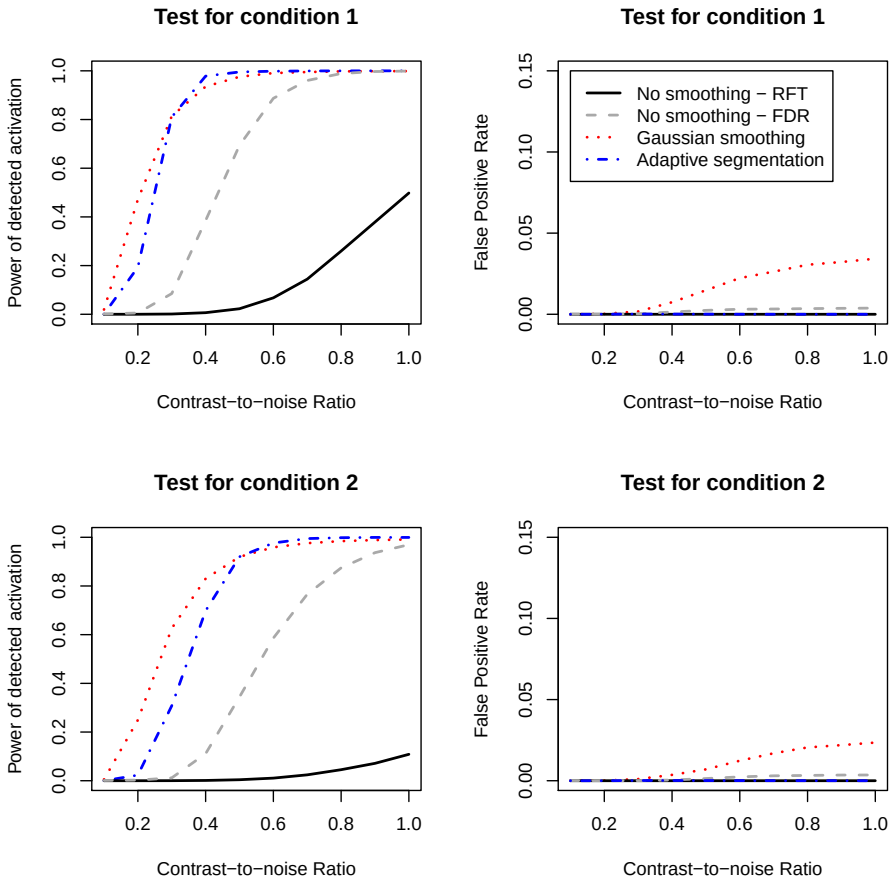

Figure 6.5 – Power (left panel) and FPR (right panel) results for each condition in simulation study 2.

on these simulation results we would again recommend adaptive smoothing as the better method to increase the sensitivity while maintaining specificity in the case of neighbouring regions.

6.3.2 Real data

Similar to the first simulation study, we demonstrate the impact of the smoothing procedures on the results from experimental fMRI data for the case of neighbouring regions. We used the same data as in the previous example (Seurinck et al., 2011), however now we focus on different contrasts. In a GLM analysis we tested the following two contrasts: (1) motion versus rest, and (2) static versus rest. These are again localizer contrasts indicated to find seperately active regions based on moving stimuli on the one hand and static stimuli on the other hand. We do not expect any overlap between the regions based on these contrasts. Similar to the second simulation study, we consider four conditions: (1) no smoothing with RFT inference, (2) no smoothing with FDR inference, (3) Gaussian smoothing with the FWHM matched to the size of the activated region (i.e. 8 mm), and (4) adaptive segmentation. Figure 6.6 shows results for the different smoothing procedures on the most representative slice. When the data are not smoothed (upper panel), RFT inference results in clearly distinct activation clusters for both conditions without any overlap. However, using FDR control shows a very different activity pattern. The static stimuli contrast (yellow) results in widespread activation in several areas and also a lot of overlap (green) is detected with the moving stimuli contrast (blue). Smoothing the data (lower panel) results in two clusters that show small overlap in the Gaussian smoothing condition, despite the fact that the value of the FWHM is slightly smaller than the size of the region (based on the unsmoothed data). In contrast, using adaptive segmentation enables us to localize two activation regions for the moving and static stimuli separately without any overlap.

Next to more specific activation detection in the adaptive segmentation condition, the sensitivity for activation strength is also higher in this condition. Figure 6.7 shows the probability density functions of the p values of

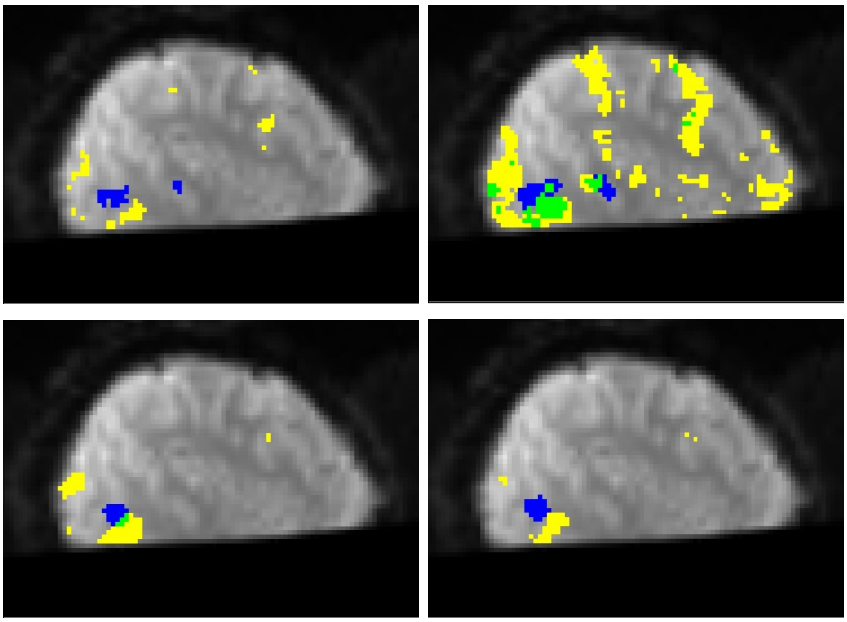

Figure 6.6 – Application of the smoothing conditions from simulation study 2 to the visual localizer data from Seurinck et al. (2011). Upper left: no smoothing (RFT); upper right: no smoothing (FDR); lower left: Gaussian smoothing; lower right: adaptive segmentation. Blue indicates activation for moving stimuli, yellow codes for activation for static stimuli and green shows overlap between the two contrasts.

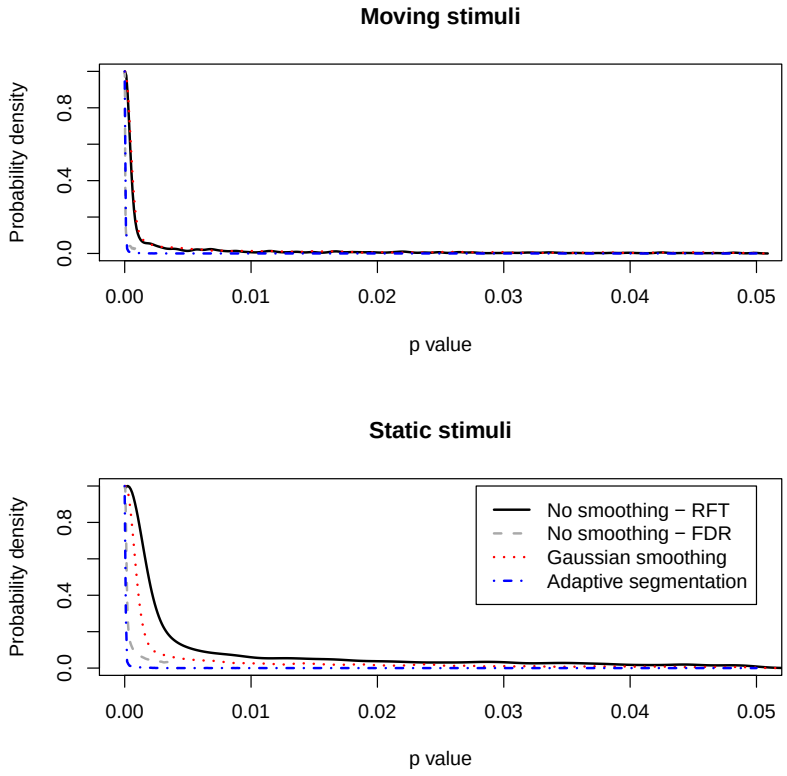

Figure 6.7 – Distribution of the p values of active voxels for moving stimuli and static stimuli respectively. Note that for the adaptive segmentation condition the scaled estimated effect is plotted instead of a p value.

active voxels for the non-smoothing and Gaussian smoothing conditions and the scaled estimated effect for the adaptive segmentation condition. Based on this figure, we can conclude that for both estimated contrasts the activation evidence is highest for the adaptive segmentation, followed by the unsmoothed data with FDR control.

As already demonstrated in the simulations, this real data example again indicates that adaptive segmentation is successful in maintaining the original shape of the activated region (i.e., increasing specificity compared to Gaussian smoothing), while increasing the overall sensitivity of activation detection at the same time.

6.4 Discussion

Despite the consensus on the benefits of spatial smoothing, the procedure is often used in a standard preprocessing pipeline using default values, especially with regard to the kernel width. For example, only 8% of studies that report using spatial smoothing, describe the reason why a particular value of smoothing kernel width is chosen (Carp, 2012). In case a justification is provided, the relation with the specific selected smoothing kernel is at least vague. Poldrack et al. (2011) give as a guideline that a smoothing kernel width of at least twice the voxel size would be appropriate in many cases (pp. 51–52). A second problem is that applying non-adaptive spatial smoothing comes at a cost of losing specificity. To accommodate this issue, the kernel width should be chosen wisely to achieve the desired increase in sensitivity while maintaining an acceptable specificity level. Given these confusing guidelines, the drawbacks of standard smoothing techniques and constant development of more advanced analyses, a systematic evaluation of the impact of spatial smoothing is called for.

In this paper we focused on two specific scenarios often encountered in fMRI research, namely (1) activation in unequally sized regions and (2) activation in neighbouring regions. For both cases several methods for smoothing and signal detection were systematically evaluated using realistic simulations. Based on the power results, we can conclude that all spatial smoothing pro-

cedures are successful in increasing the power of the activation detection analysis. Logan & Rowe (2004) have already demonstrated on unsmoothed that FDR based inference is more sensitive compared to RFT inference, however the sensitivity levels of the FDR method are not as high when compared to the smoothed data results. Although increased sensitivity is desirable, our FPR analyses have shown that differences between smoothing methods are quite dramatic for the specificity results. Applying non-adaptive smoothing will inevitably lead to more false positives. In the case of unequally sized regions, the chosen value of the FWHM can even increase this effect further. On the other hand, even when the FWHM can be matched to the size of the activated region, an overlap will be created between neighbouring regions. Both effects were not present when applying adaptive smoothing. Almost no false positives are introduced by these smoothing methods.

A demonstration on real experimental data further accords with these results. However, it should be noted that, while the FDR based inference of unsmoothed data in the simulation studies could control the FPR, the same analysis of the real data showed more activation clusters that seemed improbable given the localizer contrasts. Applying FDR inference on smoothed data might be a possible solution to control this unwanted activation, but the method requires independent tests to be valid which is not the case for smoothed data (although Benjamini & Yekutieli (2001) developed a correction for some forms of dependency) or has to rely oRFT to make cluster-based inference (Chumbley & Friston, 2009).

With respect to the inference methods, we saw that, confirming the results of Polzehl et al. (2010), adaptive smoothing combined with multiscale tests (i.e. adaptive segmentation) results in higher sensitivity levels compared to using RFT based inference. We focused on voxelwise inference using RFT corrected thresholds but we expect that the results for other inference methods, such as cluster-based techniques, would be similar because they would suffer from the same blurring effect induced by non-adaptive smoothing. Similarly, other adaptive smoothing techniques that are not considered here (e.g. Lu et al., 2003; Harrison et al., 2008; Yue et al., 2010), but also make use of the data to adaptively smooth the images, are expected to behave in the

same way. By using knowledge on shape and extent of the activated region that is present in the data, these more advanced techniques will provide more reliable results with respect to the obtained specificity.

By specifically focusing on two cases applicable to single-subject fMRI, we left the question of the impact of spatial smoothing on higher-level analyses open for the time being. We used spatial smoothing in the first place as a method to increase the SNR of the data, but in higher-level analyses spatial smoothing is also used to make intersubject comparison more feasible. It is possible that some degree of specificity loss is necessary to enable enough overlap between different subjects but this will be a subject for future research.

In summary, more advanced adaptive smoothing procedures can be used as an inference strategy to obtain more specificity for example with unequally sized or neighbouring regions. Compared to the widely used non-adaptive spatial smoothing, adaptive smoothing has the advantage of controlling the number of false positives while increasing the power. In addition, using adaptive segmentation methods always includes a built-in multiple comparisons correction based on multiscale tests.

References

Adler, R. (2000). On excursion sets, tube formulae, and maxima of random fields (special invited paper). *Annals of Applied Probability*, *10*, 1–74.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, *57*, 289–300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*, 1165–1188.

Carp, J. (2012). The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage*, *63*, 289–300.

Chumbley, J., & Friston, K. (2009). False discovery rate revisited: Fdr and topological inference using gaussian random fields. *NeuroImage*, *44*, 62–70.

Descombes, X., Kruggel, F., & Cramon, D. von. (1998). Spatio-temporal fMRI analysis using markov random fields. *IEEE Transactions on Medical Imaging*, *17*, 1028–29.

Dümbgen, L., & Spokoiny, V. (2001). Multiscale testing of qualitative hypotheses. *Annals of Statistics*, *29*, 124–152.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., & Penny, W. (Eds.). (2007). *Statistical parametric mapping: The analysis of functional brain images*. Massachussets, USA: Academic Press.

Friston, K., Worsley, K., Frackowiak, R., Mazziotta, J., & Evans, A. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, *1*, 214–220.

Harrison, L., Penny, W., Daunizeau, J., & Friston, K. (2008). Diffusion-based spatial priors for functional magnetic resonance images. *NeuroImage*, *41*, 408–423.

Logan, B., & Rowe, D. (2004). An evaluation of thresholding techniques in fMRI analysis. *NeuroImage*, *22*, 95–108.

Lu, Y., Jiang, T., & Zang, Y. (2003). Region growing method for the analysis of functional mri data. *NeuroImage*, *20*, 455–465.

Poldrack, R., Mumford, J., & Nichols, T. (2011). *Handbook of functional* MRI *data analysis.* New York, USA: Cambridge University Press.

Poline, J., & Mazoyer, B. (1994). Enhanced detection in brain activation maps using a multifiltering approach. *Journal of Cerebral Blood Flow Metabolism*, *14*, 639–642.

Polzehl, J., & Spokoiny, V. (2006). Propagation-separation approach for local likelihood estimation. *Probability Theory and Relative Fields*, *135*, 335–362.

Polzehl, J., Voss, H., & Tabelow, K. (2010). Structural adaptive segmentation for statistical parametric mapping. *NeuroImage*, *52*, 515–523.

R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org` (ISBN 3-900051-07-0)

Rosenfeld, A., & Kak, A. (1982). *Digital picture processing 2.* Orlando, Florida: Academic Press.

Seurinck, R., Lange, F. de, Achten, E., & Vingerhoets, G. (2011). Mental rotation meets the motion aftereffect: The role of hV5/MT+ in visual mental imagery. *Journal of Cognitive Neuroscience*, *23*, 1395–1404.

Tabelow, K., & Polzehl, J. (2011). Statistical parametric maps for functional MRI experiments in r: The package fmri. *Journal of Statistical Software*, *44*, 1–21.

Tabelow, K., Polzehl, J., Voss, H., & Spokoiny, V. (2006). Analyzing fMRI experiments with structural adaptive smoothing procedures. *NeuroImage*, *33*, 55–62.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., & Rosseel, Y. (2011). neuRosim: An R package for generating fMRI data. *Journal of Statistical Software*, *44*, 1–18.

Welvaert, M., & Rosseel, Y. (2012). How ignoring physiological noise can bias the conclusions from fMRI simulation studies. *Journal of Neuroscience Methods*, *211*, 125–132.

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

Worsley, K. (2003). Detecting activation in fmri data. *Statistical Methods in Medical Research*, *12*, 401–418.

Yue, Y., Loh, J., & Lindquist, M. (2010). Adaptive spatial smoothing of fmri images. *Statistics and Its Interface*, *3*, 3–13.

7 | Discussion and conclusions

In this dissertation, we focused on the simulation of fMRI data within a statistical validation framework. Within this framework, the most important aspect of the data generation is to capture all relevant components that are present in the data and will most likely have an effect on the results of the statistical analysis of these data. Mainly driven by the lack of consensus on how fMRI data should be generated from a statistical perspective, several studies were conducted in order to (1) compile knowledge on the fMRI data generation methods applied until now (Chapter 2 and 3), (2) propose new tools in order to use these techniques (Chapter 4), (3) deliver proof of the need for more advanced simulation techniques (Chapter 5), and (4) demonstrate how the proposed data generation method can be applied in a statistical validation context (Chapter 6). In this general discussion, we will present an overview of the main findings and their implications for the field, followed by a number of suggestions for future research.

7.1 Overview of the main findings and implications

Review of fMRI simulation studies

This dissertation started with a review of fMRI simulation studies (**Chapter 2**). In this review, an fMRI simulation database was compiled consisting of representative papers. The contents of this database were analysed with respect to the goal and the experimental design of the simulation study, and the data generating process that was used to simulate the fMRI data. The most crucial finding in this review was the discrepancy between the simulated components of fMRI data and the components known to be present in real data. In particular the noise model that was adopted in the simulation studies captured only a fraction of real fMRI noise. Another surprising result was that many current fMRI simulation studies lack thorough experimental manipulation. Based on the oversimplification of the noise model and the limited parameter variation of these simulation studies, the validity of the

conclusions in these studies can be questioned. In order to improve the validity of future simulation studies, guidelines were presented that encourage designing thorough experimental manipulations and using more complex data generation models.

A common definition of signal-/contrast-to-noise ratio

Chapter 3 focused on the definition of SNR and CNR for fMRI data. In the fMRI literature, multiple definitions of these quantities exist and the specific properties of these definitions are unknown. We provided an overview of the most current definitions and discussed their advantages and drawbacks in particular with respect to fMRI data. Further, we explored the linear relationship between these definitions and the power to detect activation. Unfortunately, we had to conclude that there was no solid evidence to promote one of these definitions as a reasonable candidate for a common SNR/CNR definition.

However, the main contributions of this chapter are the reference tables. These tables allow for easy comparison of SNR/CNR values and provide an estimate of the maximal power that can be expected to detect activation in data with a given SNR/CNR value. First, these tables facilitate comparison between experiments. For example, when an fMRI study fails to replicate previous findings, assessing the data quality using an SNR or CNR measurement of the data could provide an explanation on why the conclusions do not converge. For group studies, it can also be important to compare the data quality in terms of SNR/CNR between subjects. Second, a better understanding of the SNR/CNR values might encourage fMRI researchers to report these measurements in a more systematic way.

Software for simulation of fMRI data

In **Chapter 4** the R package **neuRosim** was presented that allows for fast and flexible simulation of fMRI data (Welvaert et al., 2011). Based on the additive data generation model (Bellec et al., 2009), three components have to be modelled in order to simulate fMRI data: (1) a baseline image, (2) BOLD

activation and (3) fMRI noise. In **neuRosim** a flexible environment was created to use customised baseline images. BOLD activation can be modelled using, for example, a gamma function (Boynton et al., 1996; Cohen, 1997), a (canonical) double-gamma function (Friston et al., 1998; Glover, 1999) or the Balloon model (Buxton et al., 1998, 2004). To model fMRI noise, the following noise components are implemented: white system noise, temporally correlated noise, low-frequency drift, physiological noise, task-related noise and spatially correlated noise. All these noise sources can be combined using a weighting function. In this chapter, the specifics of each function were described and illustrated with several examples.

Around the same time, another simulation package, **simTB**, has been released to the community (Erhardt et al., 2012). This Matlab-toolbox was written using the same philosophy as **neuRosim**, namely, allowing for fast and flexible fMRI data generation using complex noise models. However, while **neuRosim** was developed under a data-driven perspective, **simTB** originated from an ICA validation context. Therefore, the data generation model in this toolbox is partially model-driven. It is not entirely clear how this ICA data generation model can be applied to validate other statistical methods for fMRI data. However, the independent release of two software packages for the simulation of fMRI data is proof that, until then, there was a great need for common simulation protocols. Making the software available to the neuroscience community was definitely a first step in an attempt to achieve consensus on the fMRI data generating process.

The role of physiological noise

Chapter 5 focuses on the discrepancy between the noise model used in fMRI simulation studies and the knowledge we have about the noise sources in real fMRI data. Simulation studies that generate purely artificial data generally include random Gaussian noise in their data generating process. This noise component only accounts for the white noise that is present in real data. Next to containing white noise, fMRI data are known to suffer from low-frequency drift, physiological noise, motion noise and task-related noise. In three simu-

lation studies, we focused on the impact of physiological noise. The first two simulation studies demonstrated that, compared to white noise data, adding physiological noise to the data generating process resulted in a dramatic decrease of the power and increase of the false positive rate in a GLM analysis of a block and event-related fMRI experiment. The drop of the ROC curve was mainly attributed to the mismatch in the data generation model and the analysis model. Indeed, due to the violation of the assumptions of the GLM model, less sensitivity and specificity could be expected. However, the third simulation study delivered proof that this decrease does not always have to be present. More robust analysis methods did not suffer from the presence of physiological noise. Consequently, when comparing robust methods with the GLM model using a simple noise model, the wrong model could be preferred as the better one. These results deliver explicit evidence that the data generating process for fMRI simulation studies should be constructed with utmost care. Especially with regard to the simulated noise, it is important to take into account the multiple noise sources found in fMRI data in order to avoid biased conclusions.

Simulation application: Validation and comparison of spatial smoothing techniques

In **Chapter 6** an application was presented to demonstrate how simulations can be applied to validate statistical methods. In this particular example, the effect of spatial smoothing on the sensitivity and specificity of an activation detection analysis was investigated. The main goal of the validation study was to consider alternative smoothing techniques to avoid specificity loss in two situations, namely, fMRI data with multiple regions that differ in size and fMRI data with neighbouring regions that are activated by different tasks. Standard Gaussian smoothing is known to have the benefit of increasing the SNR/CNR, but this sensitivity increase comes with a cost of losing specificity (i.e. information on the shape of the activated region is partially lost). Two alternatives to this Gaussian smoothing were systematically evaluated: (1) False Discovery Rate (FDR) control of unsmoothed

data (Benjamini & Hochberg, 1995), and (2) adaptive smoothing (Tabelow et al., 2006; Polzehl et al., 2010). The difference between adaptive smoothing and the standard (non-adaptive) smoothing is the use of adaptive weights that are determined based on the functional activity. Simulation studies in which two scenarios, i.e. unequally sized or neighbouring regions, were reconstructed clearly showed the superiority of adaptive smoothing with respect to maintaining specificity during activation detection (i.e. less false positives compared to Gaussian smoothing) while the power was higher compared to the results of FDR control on unsmoothed data. Applying the techniques on a real data example, further demonstrated the benefits of using adaptive spatial smoothing instead of standard Gaussian smoothing. Based on these results, we made a call to apply more advanced analysis strategies.

This simulation application is of interest to both methodological and applied researchers. Methodological researchers can rely on this study as an example on how different analysis methods are compared in a systematic evaluation. In particular the design of the simulation studies should be of great interest. The simulation studies were constructed in order to keep as close as possible to real fMRI data experiments by choosing scenarios that were representative for fMRI research and a data generation model that contained the different sources of fMRI data. As such, the generalisability of the simulation studies will be assured. The application of the investigated techniques on real data provided additional evidence that will substantiate the conclusions.

For applied researchers this study provides insight in the effects of frequently used techniques with respect to the results of their studies. Next to demonstrating that the current procedures are not ideal in some cases, the benefits of an alternative were presented. Studies like this application can serve as a guidance for applied researchers and are most likely to be highly welcomed because they satisfy a need for making informed decisions on the fMRI analysis pipeline.

7.2 Future research

In this dissertation, the topic of research was simulation of fMRI data with a specific focus on (1) activation detection within single-subject fMRI data based on experimental tasks and (2) the noise model in the data generating process. Consequently, there are still some topics related to the simulation of fMRI data that were outside the scope of this dissertation but are definitely of interest for future research projects.

Validation of analysis methods

First of all, the simulation strategies discussed in this dissertation can be applied to validate statistical methods for fMRI data. In the General Introduction some questions with regard to the fMRI data analysis pipeline were posed. For example, is it useful to add the estimated movement parameters in the design matrix; is it better to control the number of false negatives or rather the number of false positives? In Chapter 6 we addressed one of these questions: How should we choose the width of the smoothing kernel? In the same manner and using the tools provided in this dissertation, the other questions can be answered and the insights gathered from these validation studies will further increase the quality of the analysis of fMRI data. Further, any shortcomings of current analysis strategies, which would come up during the validation process, could lead to the development of new, more suitable and robust analysis methods for fMRI data.

Simulation of multi-subject studies

Single-subject validation studies should also be extended to multi-subject studies. Especially, the second-level analysis (i.e. combined analysis of multiple subjects) is still largely unexplored terrain. Although widely applied, it is not clear at all how the analysis procedure behaves in practice. For example, what is the effect of the parameter choices that are made on the single-subject level? How are the results effected by *influential* subjects? Multi-subject datasets can easily be simulated using **neuRosim** and the prop-

erties of specific dynamics and their effects can be systematically evaluated. Again, thorough validation of this analysis procedure will be highly beneficial to those researchers who analyse their data, often blind-sighted to the specific effects of the choices made during the analysis process.

Resting-state fMRI data

fMRI is not only used to image the brain in function but also at rest. The number of resting-state fMRI studies has increased over the last years (see Biswal, 2012, for a review). Together with the measurement of resting-state data, the analysis techniques for this type of data have emerged. For example, a GLM analysis is not suitable for resting-state data since the model relies on a design matrix. Instead, model-free procedures like voxel-wise cross-correlation and ICA have been used to distinguish the resting-state activation from the noise. Validation studies based on simulations are highly necessary and a sound simulation method is required. Since physiological noise is present in fMRI data around the same frequency range as the resting-state activation (see Birn, 2012, for a discussion on the role of physiological noise in resting-state data), a data generating process that includes this type of noise can play a major role.

Brain connectivity

A final extension to the research that has been conducted in this dissertation is the simulation of brain networks. Simulating a brain network, either task-induced or a resting-state network, imposes additional challenges on the data generating process. The main difficulty is defining a data generation model that is not influenced by an analysis technique. For example, although biophysically inspired, networks simulated using the DCM simulator (Friston et al., 2003) cannot be considered as model-free simulations. Therefore, they are not suitable for the validation of the retrieval of brain networks using DCM. What a suitable data generation model should look like is still open for debate. At the time of writing, there is not yet common ground on how to model directed links between regions in the *ground truth network*.

7.3 Final conclusion

Simulation studies are an excellent method to validate statistical methods under the condition that the data generating process is representative for the data on which the statistical methods are applied. However, with regard to fMRI data, the data generating process is more often *model-based* instead of *data-driven*, e.g. important sources that are present in the data are not modelled during the data generation because they are not adequately represented in the analysis technique. Therefore, these model-based simulations can be questioned in terms of reliability and generalisability. As a solution, we presented a data-driven simulation method for fMRI data. A validation study established that our technique can indeed alter the conclusions of a simulation study. Further, the applicability of our approach to real-life questions has been demonstrated.

References

Bellec, P., Perlbarg, V., & Evans, A. (2009). Bootstrap generation and evaluation of an fMRI simulation database. *Magnetic Resonance Imaging*, *27*, 1382–1396.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, *57*, 289–300.

Birn, R. (2012). The role of physiological noise in resting-state functional connectivity. *NeuroImage*, *62*, 864 – 870.

Biswal, B. (2012). Resting state fMRI: A personal history. *NeuroImage*, *62*, 938 – 944.

Boynton, G., Engel, S., Glover, G., & Heeger, D. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *The Journal of Neuroscience*, *16*(13), 4207–4221.

Buxton, R., Uludağ, K., Dubowitz, D., & Liu, T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, *23*, S220–S233.

Buxton, R., Wong, E., & Frank, L. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine*, *39*, 855–864.

Cohen, M. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage*, *6*, 93–103.

Erhardt, E., Allen, E., Wei, Y., Eichele, T., & Calhoun, V. (2012). SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability. *NeuroImage*, *59*, 4160–4167.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*, 30–40.

Friston, K., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.

Glover, G. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*, 416–429.

Polzehl, J., Voss, H., & Tabelow, K. (2010). Structural adaptive segmentation for statistical parametric mapping. *NeuroImage*, *52*, 515–523.

Tabelow, K., Polzehl, J., Voss, H., & Spokoiny, V. (2006). Analyzing fMRI experiments with structural adaptive smoothing procedures. *NeuroImage*, *33*, 55–62.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., & Rosseel, Y. (2011). neuRosim: An R package for generating fMRI data. *Journal of Statistical Software*, *44*, 1–18.

Nederlandstalige samenvatting
Summary in Dutch

Functionele MRI (fMRI) wordt vaak toegepast om de functies van de hersenen in beeld te brengen. De data die het resultaat zijn van deze beeldvormingstechniek staan bekend om hun complexe structuur doordat imperfecties tijdens het fysische meetproces interageren met de fysiologie en psychologie van de proefpersonen. Gezien de complexiteit van de data is in de fMRI literatuur een veelheid van analysetechnieken voorhanden. Deze technieken dienen gevalideerd te worden alvorens ze met voldoende betrouwbaarheid en validiteit toegepast kunnen worden op reële datasets. De validatie van analysetechnieken voor fMRI data wordt echter gehypothekeerd doordat de werkelijke structuur van de data niet bekend is, wat een noodzakelijke voorwaarde is voor een gegronde validatiestudie. Een oplossing is om de fMRI data artificieel te simuleren en de validatie te baseren op de gesimuleerde data waarvan exact geweten is hoe ze tot stand is gekomen. Dit doctoraatsproefschrift neemt de huidige methodologie voor het genereren van fMRI data onder de loep en stelt, waar nodig, alternatieven voor die de betrouwbaarheid van simulatiestudies kunnen verhogen.

In hoofdstuk 2 werd een database samengesteld die een representatieve steekproef bevat van fMRI simulatiestudies. Op basis van de informatie in deze database werden de simulatiestudies geëvalueerd op grond van het simulatiedesign en het model dat werd gebruikt voor het genereren van de fMRI data. Het eerste opvallende resultaat was dat de meeste simulatiestudies niet vaak gebaseerd zijn op een doordacht experimenteel design. Parameters in de simulatiestudie worden bijvoorbeeld niet gevarieerd. De tweede conclusie in deze review was de discrepantie tussen het model dat artificiële fMRI noise genereert en de kennis over de bronnen van fMRI noise. Dikwijls blijkt de noise structuur in veel modellen sterk vereenvoudigd. Bijgevolg kan de betrouwbaarheid en generaliseerbaarheid van deze simulatiestudies in vraag gesteld worden. Opdat deze betrouwbaarheid en generaliseerbaarheid in de toekomst gegarandeerd zou kunnen worden, werden enkele richtlijnen opgesteld voor onderzoekers die fMRI simulatiestudies uitvoeren.

Een bijkomend aspect, naast de betrouwbaarheid en generaliseerbaarheid, is de transparantie en vergelijkbaarheid van simulatiestudies. In hoofdstuk 2 ligt de focus op de definitie van SNR en CNR voor fMRI data. Studie van

de literatuur leert dat er een enorme waaier aan SNR/CNR waarden terug te vinden zijn, dit wordt veroorzaakt door het gebruik van verschillende definities. In dit hoofdstuk werden de meest gangbare definities besproken en hun voor- en nadelen met betrekking tot hun gebruik voor fMRI data toegelicht. Kortweg hadden alle definities zowel voor- als nadelen en afhankelijk van de specifieke context kan de ene dan wel de andere definitie verkozen worden. Bijkomend werden enkele referentietabellen opgesteld die toelaten om eenvoudige vergelijkingen te maken tussen de waarden van de verschillende definities. Deze tabellen kunnen aangewend worden om de vergelijking tussen simulatiestudies transparanter te maken.

Om tegemoet te komen aan het gebrek aan software voor het simuleren van fMRI data werd een R pakket ontwikkeld dat toelaat om op een snelle en flexibele manier fMRI data te simuleren (hoofdstuk 4). Bij het implementeren van de verschillende functies werd bijzondere aandacht besteed aan de verschillende componenten van fMRI data. **neuRosim** maakt het onder andere mogelijk om complexe noise structuren te genereren die representatief zijn voor fMRI data. De functies in het *open-source* software pakket zijn uitermate geschikt om in een simulatiescript ingebed te worden. **neuRosim** werd dan ook gebruikt bij alle simulatiestudies in dit proefschrift.

In hoofdstuk 2 werd reeds aangehaald dat de meeste simulatiestudies een vereenvoudigde noise structuur van de fMRI data hanteren. Om een beter inzicht te verwerven op de exacte invloed van deze vereenvoudiging werden drie simulatiestudies uitgevoerd, hierin werden de resultaten op basis van verschillende noise modellen met elkaar vergeleken (hoofdstuk 5). In het bijzonder werden data met een eenvoudige noise structuur, random *witte* noise, vergeleken met data waarin fysiologisch gerelateerde noise werd toegevoegd. De introductie van fysiologische noise in de data resulteerde in een drastische daling van de sensitiviteit bij een standaard GLM-analyse. Daarnaast werd aangetoond dat, wanneer twee analysetechnieken met elkaar vergeleken worden, de conclusies van simulatiestudies vertekend kunnen zijn indien enkel een eenvoudig noise model wordt gebruikt in de simulatiestudies.

Tenslotte werd in hoofdstuk 6 een toepassing gepresenteerd waarin simulatiestudies werden aangewend om analysetechnieken voor fMRI data te

evalueren en te vergelijken. In dit specifiek voorbeeld werden courante spatiale smoothing procedures vergeleken met alternatieve technieken. Spatiale smoothing creëert een uitvlakking van de data waardoor de signaal-ruisverhouding verbetert. Als vertrekpunt werd het probleem vooropgesteld dat wanneer fMRI data gesmoothed worden, de informatie over de precieze vorm van een actieve regio verloren gaat. De daling in specificiteit dient dus te worden afgewogen ten opzichte van de hoeveelheid van de winst in SNR die door het smoothing proces ontstaat. Dit geldt in het bijzonder voor fMRI data waarin meerdere regio's actief zijn die een verschillende grootte hebben of voor data waarin actieve gebieden naast elkaar gelokaliseerd zijn. Als alternatief kan gebruik gemaakt worden van adaptieve smoothing. In deze smoothing procedure worden lokale gewichten afgeleid op basis van de functionele activatie in de data met als gevolg dat niet over de grenzen van actieve regio's gesmoothed wordt. Zowel simulatiestudies als voorbeeldanalyses van experimentele data toonden aan dat de methodes die adaptieve smoothing gebruiken superieur zijn in vergelijking met de courante smoothing procedures in termen van specificiteit, maar met behoud van sensitiviteit. Met andere woorden: de adaptieve technieken slaagden erin om een even hoge SNR te bekomen als de standaard techniek, maar bovendien was er een behoud van de vorm van de actieve gebieden. Dit resulteerde bijvoorbeeld niet tot een overlap tussen de nabijgelegen regio's terwijl de standaard smoothing procedure dit wel deed.

Tot slot kon er op basis van de fMRI simulatieliteratuur worden geconcludeerd dat de betrouwbaarheid en generaliseerbaarheid van huidige simulatiestudies in vraag gesteld kunnen worden, dit voornamelijk met betrekking tot de modellering van de noise. Deze simulatiestudies werden benoemd als modelgebaseerde simulaties, waarbij het analysemodel dat gevalideerd wordt ook als basis dient voor het genereren van de gesimuleerde data. In dit proefschrift werden verscheidene instrumenten gepresenteerd om een data-gebaseerde simulatiestudie uit te voeren. Daarnaast werd tijdens het onderzoek duidelijk dat de data-gebaseerde benadering de betrouwbaarheid en generaliseerbaarheid van simulatiestudies, die analysetechnieken voor fMRI data valideren, kan verhogen.

Bibliography of academic output

Peer-reviewed publications

Welvaert, M., Farioli, F. and Grainger, J. (2008). Graded effects of number of inserted letters in superset priming. *Experimental Psychology, 55(1),* pp. 54–63.

Van Assche, E., Drieghe, D., Duyck, W., **Welvaert, M.** and Hartsuiker R. (2011). The influence of semantic contraints on bilingual word recognition during sentence reading. *Journal of Memory and Language, 64(1),* pp. 88–107.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G. and Rosseel, Y. (2011). neuRosim: An R package for generating fMRI data. *Journal of Statistical Software, 44(10),* pp. 1–18.

Welvaert, M. and Rosseel, Y. (2012). How ignoring physiological noise can bias the conclusions from fMRI simulation results. *Journal of Neuroscience Methods, 211(1),* pp. 125–132.

Submitted manuscripts

Welvaert, M., Tabelow, K., Seurinck, R. and Rosseel, Y. (2013). *Adaptive smoothing as inference strategy: more specificity for unequally sized or neighbouring regions.* Manuscript submitted to *Neuroinformatics*.

Welvaert, M. and Rosseel, Y. (2013). *A review of fMRI simulation studies.* Manuscript submitted to *NeuroImage*.

Welvaert, M. and Rosseel, Y. (2013). *On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data.* Manuscript submitted to *PLoS One*.

Conference contributions

Welvaert, M. and Rosseel, Y. (2009). *Neural activity with spatial and temporal correlations as a basis to simulate fMRI data.* Poster presented

at the 15th annual meeting of the Organization for Human Brain Mapping, June 18–23, San Francisco, USA.

Welvaert, M. and Rosseel, Y. (2010). *Comparison of noise models through ROC analyses of simulated data.* Poster presented at the 16th annual meeting of the Organization for Human Brain Mapping, June 06–10, Barcelone, Spain.

Welvaert, M. and Rosseel, Y. (2011). *Simulating fMRI data with realistic noise using neuRosim.* Poster presented at the 17th annual meeting of the Organization for Human Brain Mapping, June 26–30, Quebec, Canada.

Welvaert, M. and Rosseel, Y. (2011). *neuRosim: an R package for simulation of fMRI magnitude data with realistic noise.* Contributed talk presented at UseR! The R user conference, August 15–18, Warwick, UK.

Welvaert, M. and Rosseel, Y. (2011). *Using neuRosim to simulate magnitude fMRI data containing physiological noise.* Poster and demo session presented at the 4th INCF Congress of NeuroInformatics, September 4–6, Boston, USA.

Welvaert, M. and Rosseel, Y. (2011). *Simulating fMRI data: the R package neuRosim.* Contributed talk presented at the Workshop on Statistics and Neuroimaging, Berlin, Germany.

Welvaert, M., Rosseel, Y. and Tabelow, K. (2012). *How to smooth your fMRI data? A comparison between Gaussian and adaptive smoothing.* Poster presented at the 18th annual meeting of the Organization for Human Brain Mapping, June 10–14, Beijing, China.

Welvaert, M., Tabelow, K., Seurinck, R. and Rosseel, Y. (2013). *Defining ROIs based on localizer studies: more specific localization using adaptive smoothing.* Poster to be presented at the 19th annual meeting of the Organization for Human Brain Mapping, June 16–20, Seattle, USA.

Now this is not the end.
It is not even the beginning of the end.
But it is, perhaps, the end of the beginning.

Winston Churchill