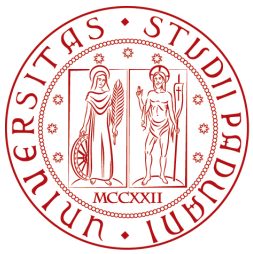
Universitá degli Studi di Padova

Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in

Statistica e Gestione delle Imprese



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

RELAZIONE FINALE

A STATISTICAL APPROACH TO IMPROVE THE RELIABILITY OF FUNCTIONAL MAGNETIC RESONANCE IMAGES

Relatore: Prof. Livio Finos

Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Padova, Italia

Correlatore: Dott. Francesco de Bertoldi

Dipartimento di Neuroscienze e di Salute Mentale, UOC Psichiatria, Ospedale Maggiore Policlinico, Universitá di Milano, Milano, Italia

Correlatore: Prof. Gianmarco Altoé

Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Padova, Italia

Laureando: Marco Zanetti

Matricola N. 1051328

Anno Accademico 2015/2016

Acknowledgements

Now that I have reached the end of this important course of study, I want to thank all the people who helped me to face the difficulties encountered in the thesis, in the exams and in the life, to their go my gratitude, even if to me it is the responsibility for any error contained in this work.

I would like to thank my supervisor Livio Finos, for giving me the opportunity to work with him in the research of fMRI, making me grow in the professional and personal skills. He has never made me weigh the mistakes I made in these months of work, I'm deeply indebted.

I thank my assistant supervisor Francesco de Bertoldi because he was been very helpful and he shared with us useful data for the thesis.

I thank my assistant supervisor Gianmarco Altoé because he creates a pleasant work environment during the stage.

I thank the Department of Statistics and all the teachers, because I was fully satisfied for all their lessons, for the seriousness and the excellent organization of the offices of St. Catherine, if I go back I certainly would choose this course of study again.

I thank all my statistic friends met at the University, for helping me during exams and for making happier the study days in their company.

I thank Marta because she has been a reference point with which I could constantly confront in the past three academic years.

I thank Gaspare, for the bond of true friendship that is created between us from the first day of lesson.

I thank my parents, Luana and Sandro, because they grew up me with healthy values, for taught me what it means the sacrifice and for supporting me morally and

economically in all the choices I made in the almost 22 years.
I thank my brothers Luca and Giacomo beacuse despite the misunderstandings, we learned to share everything and to help us.
I thank Davide, Desiré, Elisa, Enrico, Federico, Giorgia, Guido, Isacco, Marco, Matilde, Mirko and all my friends, for the time that they have spent with me and for the wonderful experiences that we have shared over the years.
Finally, I want to dedicate this work to my grandparents because I feel me really lucky to still receive their advice and their love.

Aim of the thesis

The studies of functional magnetic resonance images (fMRI) are very diffuse and important because the researchers study the neuronal activity of the brain through this technique. In this studies there are many problem of reliability to understand which are the areas in the brain that are activated at the stimulus, because the results change from subject to subject and also between a test and another. In a previous article (*"Improving the reliability of single-subject fMRI by weighting intra-run variability"*, *Bertoldi et al., 2015*), the authors have proposed a solution to alleviate this problem using an index to penalized the p-value of those voxels (volumetric pixels in the brain) that have an instable signal during the stimulus but they have left open a problem. In this thesis we propose a possible solution to control the 1-st tipe error of the weighted p-value proposed by *Bertoldi et al.* using some properties of F distribution.

Contents

Chapter 1

fMRI: data acquisition and preprocessing

Most of the material in this chapter has been taken from *"The statistical analysis of fMRI data", Lindquist, 2008*. This thesis work was limited to part of these aspects but however they can help to better understand the statistical method which will be discussed in the following chapters.

1.1 Introduction

Functional Magnetic Resonance Imaging (fMRI) is a noninvasive technique for studying brain activity. During the course of an fMRI experiment, a series of brain images are acquired while the subject performs a set of tasks. Changes in the measured signal between individual images are used to make inferences regarding task-related activations in the brain. The data comprise a sequence of magnetic resonance images (MRI), each consisting of a number of uniformly spaced volume elements, or voxels, that partition the brain into equally sized boxes. Changes in brain hemodynamics, in reaction to neuronal activity, impact the local intensity of the MR signal, and therefore changes in voxel intensity across time can be used to infer when and where activity is taking place.

Though a good number of these voxels consist solely of background noise, and can be excluded from further analysis, the total amount of data that needs to be analyzed

is staggering. In addition, the data exhibit a complicated temporal and spatial noise structure with a relatively weak signal. A full spatiotemporal model of the data is generally not considered feasible and a number of short cuts are taken throughout the course of the analysis. Statisticians play an important role in determining which short cuts are appropriate in the various stages of the analysis, and determining their effects on the validity and power of the statistical analysis.

1.2 Acquiring fMRI Data

The data collected during an fMRI experiment consists of a sequence of individual magnetic resonance images, acquired in a manner that allows one to study oxygenation patterns in the brain with an MRI scanner. Therefore, to understand the nature of fMRI data and how these images are used to infer neuronal activity, one must first study the acquisition of individual MR images. The standard approach toward 3D imaging is to acquire a stack of adjacent slices (e.g., 20 - 30) in quick succession.

An fMRI study consists of a series of brain volumes collected in quick succession. The temporal resolution of the acquired data will depend on the time between acquisitions of each individual volume; once the k-space has been sampled, the procedure is ready to be repeated and a new volume can be acquired. This is one reason why efficient sampling of k-space is important. Typically, brain volumes of dimensions 64 64 30 (i.e. 122,880 voxels) are collected at separate time points throughout the course of an experiment, where varies between 100 - 2000. Hence, the resulting data consists of roughly 100,000 time series of length . On top of this, the experiment is often repeated for M subjects, where M usually varies between 10 and 40. It quickly becomes clear that fMRI data analysis is a time series analysis problem of massive proportions.

1.3 Understanding fMRI Data

Functional magnetic resonance imaging is most commonly performed using blood oxygenation level dependent (BOLD) contrast (Ogawa et al., 1992) to study local changes in deoxyhemoglobin concentration in the brain. BOLD imaging takes advantage of inherent differences between oxygenated and deoxygenated hemoglobin. Each of these states has different magnetic properties, diamagnetic and paramagnetic respectively, and produces different local magnetic fields. Due to its paramagnetic properties, deoxy-hemoglobin has the effect of suppressing the MR signal, while oxy-hemoglobin does not. The cerebral blood flow refreshes areas of the brain that are active during the execution of a mental task with oxygenated blood, thereby changing the local magnetic susceptibility and the measured MR signal in active brain regions (see Figure 1.1). A series of properly acquired MR images can therefore be used to study changes in blood oxygenation which, in turn, can be used to infer brain activity. The underlying evoked hemodynamic response to a neural event is typically referred to as the hemodynamic response function (HRF, see Figure 1.2). The increased metabolic demands due to neuronal activity lead to an increase in the inflow of oxygenated blood to active regions of the brain. Since more oxygen is supplied than actually consumed, this leads to a decrease in the concentration of deoxy-hemoglobin which, in turn, leads to an increase in signal. This positive rise in signal has an onset approximately 2 seconds after the onset of neural activity and peaks 5-8 seconds after that neural activity has peaked (Aguirre et al., 1998). After reaching its peak level, the BOLD signal decreases to a below baseline level which is sustained for roughly 10 seconds. This effect, known as the post-stimulus undershoot, is due to the fact that blood flow decreases more rapidly than blood volume, thereby allowing for a greater concentration of deoxy-hemoglobin in previously active brain regions. The measured fMRI signal is also corrupted by random noise and various nuisance components that arise due both to hardware reasons and the subjects themselves. For instance, fluctuations in the MR signal intensity caused by thermal motion of electrons within the subject and the scanner gives rise to noise that tends to be highly random and independent of the experimental task. Another

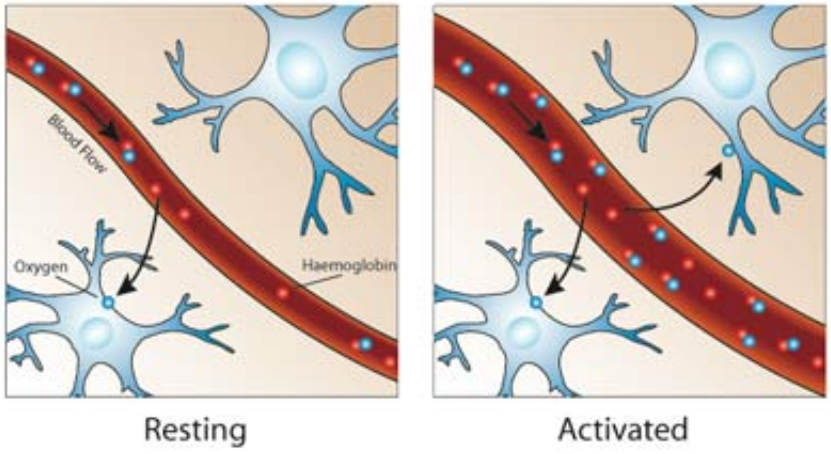


Figure 1.1: The comparison in the BOLD effect between task and rest.

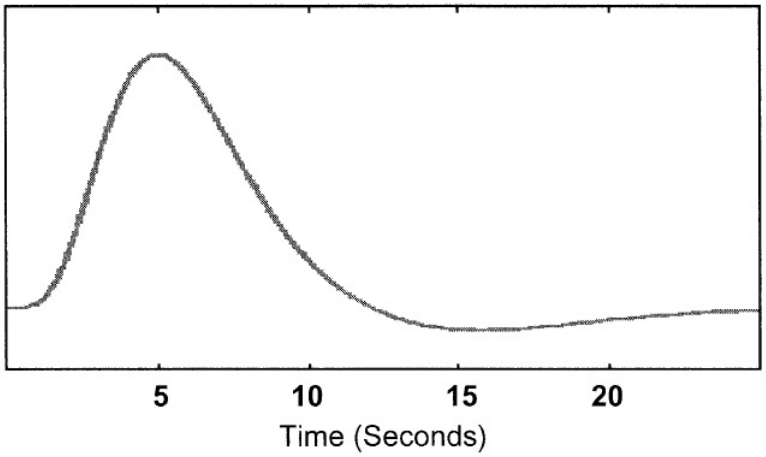


Figure 1.2: The standard canonical model for the HRF used in fMRI data analysis illustrates the main features of the response.

source of variability in the signal is due to scanner drift, caused by scanner instabilities, which result in slow changes in voxel intensity over time (low-frequency noise). Finally, physiological noise due to patient motion, respiration and heartbeat cause fluctuations in signal across both space and time. Physiological noise can often be modeled and the worst of its effects removed.

The temporal resolution of an fMRI study depends on the time between acquisition of each individual image, or the repetition time (RT). Heart-rate and respiration gives rise to periodic fluctuations that are difficult to model. In this situation the periodic fluctuations will be distributed throughout the time course giving rise to temporal autocorrelation. If these terms are properly removed, there is evidence that the resulting error term corresponds to white noise (Lund et al., 2006). Note that for high temporal resolution studies, heart-rate and respiration can be estimated and included in the model, or alternatively removed through application of a band-pass filter, a passive device that allows the passage of frequencies within a given interval (the bandwidth) and attenuates frequencies outside of it.

1.4 Preprocessing

Prior to statistical analysis, fMRI data typically undergoes a series of preprocessing steps aimed at removing artifacts and validating model assumptions. The main goals are to minimize the influence of data acquisition and physiological artifacts, to validate statistical assumptions and to standardize the locations of brain regions across subjects in order to achieve increased validity and sensitivity in group analysis. When analyzing fMRI data it is typically assumed that all of the voxels in a particular brain volume were acquired simultaneously. Further, it is assumed that each data point in a specific voxel's time series only consists of a signal from that voxel (i.e., that the participant did not move in between measurements). Finally, when performing group analysis and making population inference, all individual brains are assumed to be registered, so that each voxel is located in the same anatomical region for all subjects. Without preprocessing the data prior to analysis, none of these assumptions would hold and the resulting statistical analysis would be invalid. The

major steps involved in fMRI preprocessing are slice timing correction, realignment, coregistration of structural and functional images, normalization and smoothing.
When analyzing 3D fMRI data it is typically assumed that the whole brain is measured simultaneously. In reality, because the brain volume consists of multiple slices that are sampled sequentially, and therefore at different time points, similar time courses from different slices will be temporally shifted relative to one another. Slice timing correction involves shifting each voxel's time course so that one can assume they were measured simultaneously.
An important issue involved in any fMRI study is the proper handling of any subject movement that may have taken place during data acquisition. Even small amounts of head motion during the course of an experiment can be a major source of error if not treated correctly. When movement occurs, the signal from a specific voxel will be contaminated by the signal from neighboring voxels and the resulting data can be rendered useless. Therefore, it is of great importance to accurately estimate the amount of motion and to use this information to correct the images. If the amount of motion is deemed too severe, it may result in the subject being removed completely from the study.
Functional MRI data is typically of low spatial resolution and provides relatively little anatomical detail. Therefore, it is common to map the results obtained from functional data onto a high resolution structural MR image for presentation purposes. Is called coregistration the process of aligning structural and functional images. For group analysis, it is important that each voxel lie within the same brain structure for each individual subject. Of course individual brains have different shapes and features, but normalization attempts to register each subjects anatomy to a template brain. The main benefits of normalizing data are that spatial locations can be reported and interpreted in a consistent manner, results can be generalized to a larger population and results can be compared across studies and subjects. It is common practice to spatially smooth fMRI data prior to analysis. The smooth function is used as a final step in spatial pre-processing to blur the functional images. The reason to do this is to correct for slight remaining functional/anatomical difference between subjects. The trade-off, homever, is that you lose resolution by

smoothing.

1.5 Data Analysis

There are several common objectives in the analysis of fMRI data. These include localizing regions of the brain activated by a certain task, determining distributed networks that correspond to brain function and making predictions about psychological or disease states. All of these objectives are related to understanding how the application of certain stimuli leads to changes in neuronal activity. In fMRI the signal typically drifts slowly over time due to scanner instabilities. Therefore, most of the power lies in the low-frequency portion of the signal. To remove the effects of drift, it is common to remove fluctuations below a specified frequency cut-off using a high-pass filter. The most important issue when using a high-pass filter is to ensure that the fluctuations induced by the task design are not in the range of frequencies removed by the filter, as we do not want to throw out the signal of interest. Physiological noise can in certain circumstances be directly estimated from the data (Lindquist et al., 2008), or it can be removed using a properly designed band-pass filter. Some groups have therefore begun directly measuring heart beat and respiration during scanning and using this information to remove signal related to physiological fluctuations from the data (Glover et al., 2000). In standard time series analysis, model identification techniques are used to determine the appropriate type and order of a noise process. In fMRI data analysis this approach is not feasible due to the large number of time series being analyzed, and noise models are specified a priori.

Statistical parametric maps (SPM) are used to present the results of the statistical analysis. Voxelwhose-values are below a certain threshold are color-coded to signify that they contain significant task-related signal. The results are superimposed onto a high-resolution anatomical image for presentation purposes (see figure 1.3).

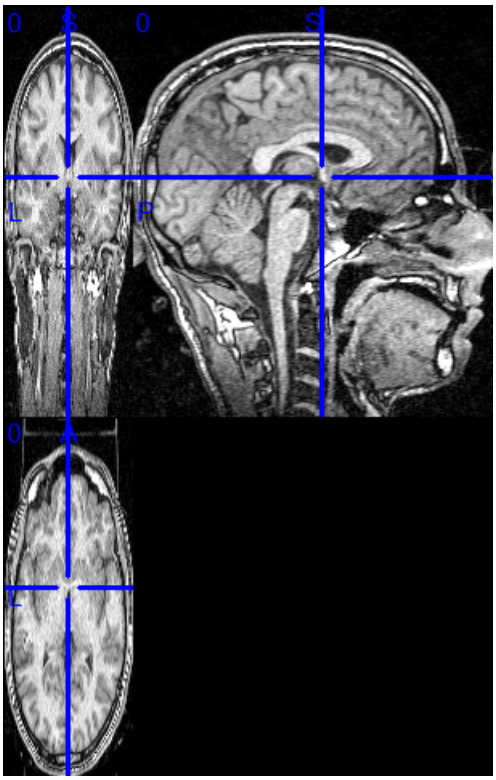


Figure 1.3: An anatomical image in a 2D map open with MRIcron.

Chapter 2

The experiment

Most of the material in this chapter has been taken from *"Improving the reliability of single-subject fMRI by weighting intra-run variability", Bertoldi et al., 2015.*

2.1 Subjects

Seven healthy subjects (4 males, 3 females, aged 19 - 24), all previously fMRI naive, were scanned while performing 4 identical runs with a block design (15/15 seconds of task/rest conditions, with 45 volumes recorded; 4 blocks for each run). The subjects performed a self-paced finger-tapping task with their right hand (all the subjects were right-handed). To minimize subject error, all underwent specific training before the start of the experiment. Four runs were acquired in 2 sessions of 2 runs each, with a 15 min gap between one session and the next. All subjects gave their informed consent and the experimental protocol was approved by the Udine Hospital Ethics Committee. In this thesis we analyze only the 4 runs of the subject 3.

2.2 MRI acquisition and SPM8 preprocessing

The MR images were acquired using a Siemens 1.5 T MRI whole-body scanner (Siemens Avanto, Erlangen, Germany, see Figure 2.1), a 12-channel matrix head coil, and a custom-built head restrainer to minimize head movements. Both structural and functional images were recorded during the MRI sessions. The subjects wore

special MR-compatible glasses while receiving instructions.

Figure 2.1: A Siemens 1.5 T MRI whole-body scanner used to record the functional images during the MRI session.

We have preprocessed data using Spm8 (Friston et al., 2007, see Figure 2.2), a open source package that runs within MATLAB (MathWorks Inc., 2012). In every subject, for each run, at the start of the process we have a number N of raw brain images (f-data with ".img" extension) that we need to preprocess, where N is the number of scan for every run, in our case 45. So, the first step is to realing these images because we do delete the error of head motion produced during the test. This step create 45 realigned images (rf-data) that we have to coregistrate using (in the settings of SPM8) the anatomical image as reference image and the mean image as source image. The third step is to smooth the same raw images and set full width half maximum (FWHM) values with double ratio of voxel dimensions. With this step the software create the final images which we use to studing the appropriate model (srf-data).

2.3 General Linear Model

To estimate the correct model we have used a general linear model (GLM). GLM is an equation $Y = X * \beta + \epsilon$ (see Figure 2.3) that expresses the observed response

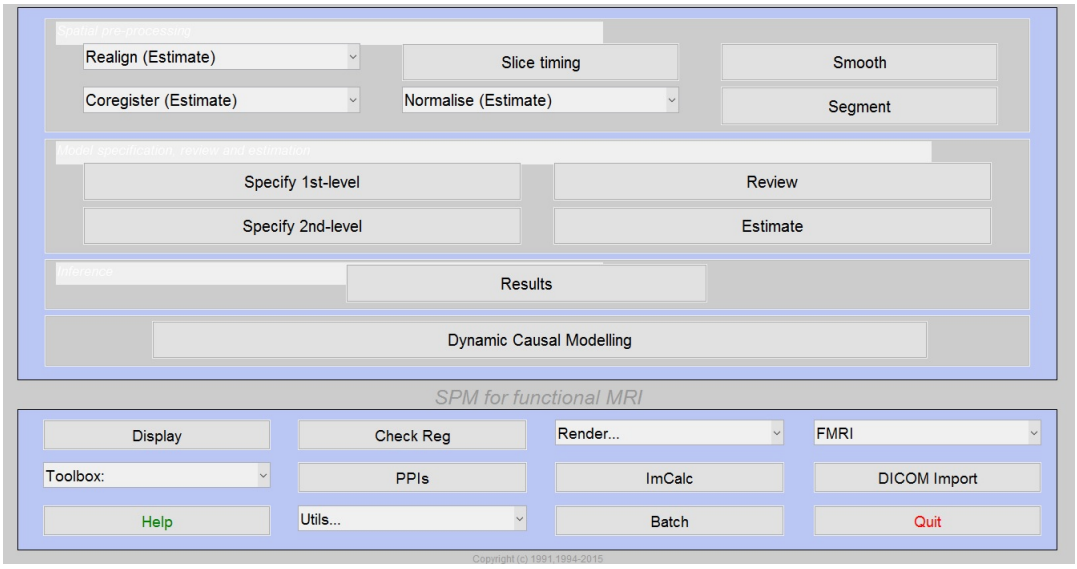


Figure 2.2: The initial settings of the software SPM8.

variable Y in terms of a linear combination of explanatory variables X plus an error term (Friston et al., 1995). The GLM is variously known as "analysis of covariance" or "multiple regression analysis" and subsumes simpler variants, like the "t test" for a difference in means. The matrix X that contains the explanatory variables is called the design matrix and each column of the design matrix corresponds to some effect one has built into the experiment. The effects on the response variable by certain conditions are modeled in terms of functions of the presence of these conditions and constitute the columns of the design matrix. The relative contribution of each of these columns is assessed using standard least squares and inferences about these contributions are made using t or F statistics. In our case Y is the BOLD signal at various time points at a single voxel, X is the design matrix with several components which explain the observed data, i.e. the BOLD time series for the voxel, β define the contribution of each component of the design matrix to the value of Y and estimated so as to minimise the error, ϵ, i.e. least sums of squares, and ϵ is the difference between the observed data, Y, and that predicted by the model, X*β. Normally, the degrees of freedom in a general linear model are calculated as the number of observations minus the number of regressors in the model, whereas the GLM in this case has been extended (Worsley and Friston, 1995) to incorporate intrinsic non-sphericity, or correlations among the error terms. This generalization

brings with it the notion of effective degrees of freedom, which are less than under i.i.d. assumptions. They are smaller because the temporal correlations reduce the effective number of independent observations (Friston, 2003).

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_j \\ \vdots \\ Y_J \end{bmatrix} = \begin{bmatrix} X_{11} & \dots & X_{1l} & \dots & X_{1L} \\ \vdots & & \vdots & & \vdots \\ X_{j1} & \dots & X_{jl} & \dots & X_{jL} \\ \vdots & & \vdots & & \vdots \\ X_{J1} & \dots & X_{Jl} & \dots & X_{JL} \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_l \\ \vdots \\ \beta_L \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_j \\ \vdots \\ \epsilon_J \end{bmatrix}$$

Figure 2.3: The matricial form of general linear model: $Y_{Jx1} = X_{JxL} * \beta_{Lx1} + \epsilon_{Jx1}$.

2.4 Experimental Design

A good experimental design attempts to maximize both statistical power and psychological validity. The statistical performance can be characterized by its estimation efficiency (i.e., the ability to estimate the HRF) and its detection power (i.e., the ability to detect significant activation). The psychological validity is often measured by the randomness of the stimulus presentation, as this helps control for issues related to anticipation, habituation and boredom. When designing an experiment there is inherent trade-offs between estimation efficiency, detection power and randomness. The optimal balance between the three ultimately depends on the goals of the experiment and the combination of conditions one is interested in studying. For example, a design used to localize areas of brain activation stresses high detection power at the expensive of estimation efficiency and randomness.

In a block design the different experimental conditions are separated into extended time intervals, or blocks. For example, one might repeat the process of interest (e.g., finger tapping) during an experimental block and have the subject rest during a control block. The A-B comparison can than be used to compare differences in signal between the conditions (see figure 2.4). In general, increasing the length of each block will lead to a larger evoked response during the task. This increases the separation in signal between blocks, which, in turn, leads to higher detection power.

In addition, it is important that the same mental processes are evoked throughout each block. If block lengths are too long, this assumption may be violated due to the effects of fatigue and/or boredom. The main advantages to using a block design are that they offer high statistical power to detect activation and are robust to uncertainties in the shape of the HRF. The latter advantage is due to the fact that the predicted response depends on the total activation caused by a series of stimuli, which makes it less sensitive to variations in the shape of responses to individual stimulus. The flip side is that block designs provide imprecise information about the particular processes that activated a brain region and cannot be used to directly estimate important features of the HRF (e.g., onset or width).

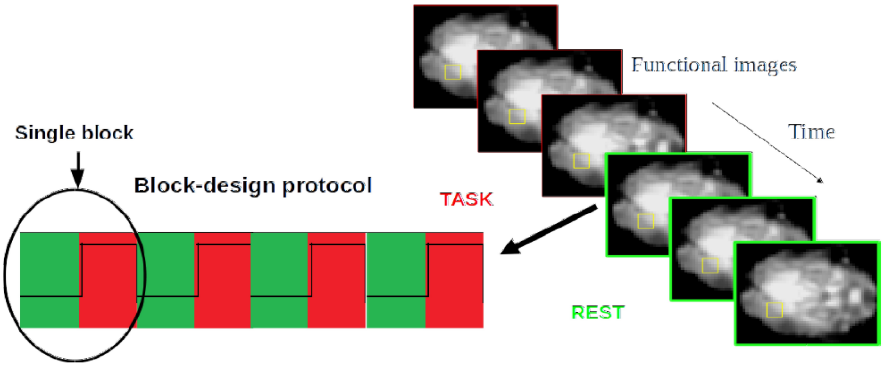


Figure 2.4: A simple graphic of block design.

On SPM8 for estimate the model we have to set the interval interscan (the difference in seconds between two successive scan), all conditions (onsets and durations of every task) and the high-pass filter (1,5 multiplied for the duration in second of one task/rest block, in our case 30 seconds).

2.5 Simple and Complex Model

The method presented here is for a simple task/rest design (see Figure 2.5), called for convenience Model 0. The prototypical model for the task/rest design can be described as follows:

$$Y = \alpha_0 + \beta_0 D + \epsilon \tag{1}$$

where D is a 0/1 indicator function which is 0 in the scans of rest and which is 1 in

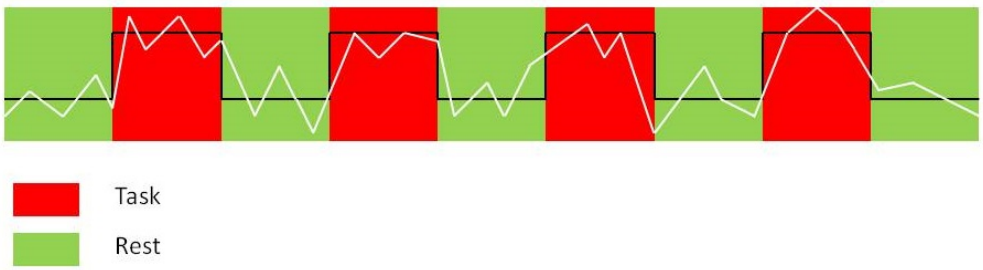


Figure 2.5: The simple model of task/rest design.

the scans under task condition, α_0 is the average effect at rest scan, β_0 is the extra activation induced by the stimulus and ϵ is the error term (i.e., the homoschedastic white noise with null mean and variance σ^2). The matricial design of the simple model is represented in Figure 2.6.

$$\mathrm{Y}=\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ \vdots \\ Y_{10} \\ Y_{11} \\ \vdots \\ Y_{15} \\ Y_{16} \\ \vdots \\ Y_{45} \end{bmatrix} \quad \mathrm{X}=\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix} \quad \beta=\begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} \quad \epsilon=\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \vdots \\ \epsilon_{10} \\ \epsilon_{11} \\ \vdots \\ \epsilon_{15} \\ \epsilon_{16} \\ \vdots \\ \epsilon_{45} \end{bmatrix}$$

Figure 2.6: The matricial design of the simple model: $Y_{45x1} = X_{45x2} * \beta_{2x1} + \epsilon_{45x1}$

The present study assumes that the effect (i.e. the true coefficient of the model)

of the task may change over time (i.e., between the 4 blocks). Thus, we defined a slightly more complex model, called for convenience Model 1:

$$Y = \alpha_0 + \beta_0 D + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \alpha_4 Z_4 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \epsilon \quad (2)$$

where Z_i has a value of 1 in the block i and 0 otherwise, and D_i has a value of 1

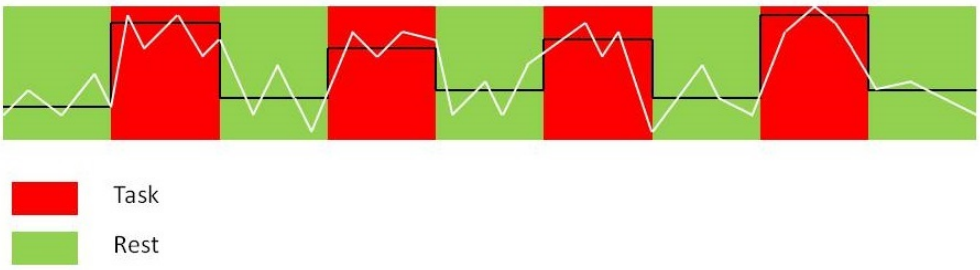


Figure 2.7: The complex model of task/rest design.

under the stimulus in the block i and 0 otherwise (see Figure 2.7). Likewise, the a_i coefficients represent the effects at rest, and b_i denotes the effect of the stimulus in trial i, except β_0 that represents the value of the stimulus of the block 4 to avoid the dummy's trap. The white noise ϵ is assumed homoschedastic, but it can be trivially extended to noise which has equal variance only within the block (i.e., heteroschedastic errors). The matricial design of the simple model is represented in Figure 2.8.

After fitting the models, 2 quantities were computed for each model: the explained deviance by the model (ESS) and the residual deviance (RSS). The F statistic, which is typically used to infer the effect of a stimulus, was derived from the comparison of RSS of Model (1) and Model (2):

$$\mathrm{F} = \frac{(RSS_0 - RSS_1)/RDF_{0-1}}{RSS_1/RDF_1} = F_{RDF_{0-1};RDF_1}$$

where we define:

RSS_0 =Residuals Sum of Square of the simple model,

RSS_1 = Residuals Sum of Square of the complex model,

RDF_0 = Residuals Degrees of Freedom of the simple model,

RDF_1 = Residuals Degrees of Freedom of the complex model,

RDF_{0-1} = Difference of residuals degrees of freedom between simple and complex

$$\mathrm{Y}=\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ \vdots \\ Y_{10} \\ Y_{11} \\ \vdots \\ Y_{15} \\ Y_{16} \\ \vdots \\ Y_{45} \end{bmatrix} \quad \mathrm{X}=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \beta=\begin{bmatrix} \alpha_0 \\ \beta_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \quad \epsilon=\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \vdots \\ \epsilon_{10} \\ \epsilon_{11} \\ \vdots \\ \epsilon_{15} \\ \epsilon_{16} \\ \vdots \\ \epsilon_{45} \end{bmatrix}$$

Figure 2.8: The matricial design of the complex model: $Y_{45x1} = X_{45x9} * \beta_{9x1} + \epsilon_{45x1}$

model.

2.6 Test F with SPM8

An alternative and very common way to calculate the Test F is through the contrasts. This is the prefered method in SPM8 which provides a convenient Gui for this (see Figure 2.9). In this case we have four regressors for the task, the columns from second to fourth estimate respectively the value of task of the first, second and third block while the first column estimate the value of task of the fourth block (so we avoid the dummy's trap). The columns from fifth to eighth estimate the value of all the four blocks while the last column estimate the value of the firsts five scans of rest (see Figure 2.10 and Figure 2.11). We call the values of F Fblock. We can calculate the expected value of the Fblock in our case as:

$Fblock \sim F_{7,30} \rightarrow \frac{1}{Fblock} \sim F_{30,7} \rightarrow E[F_{30,7}] = 7/(7-2) = 7/5 = 1.4$

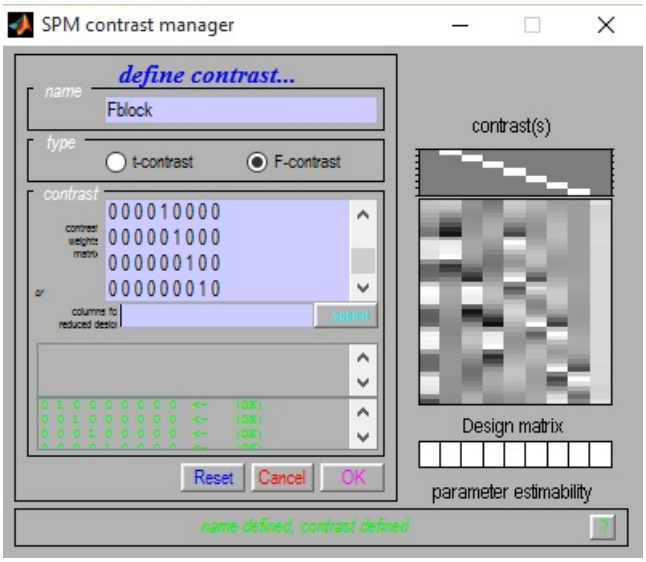


Figure 2.9: Contrasts setting in Spm8.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Figure 2.10: The matricial design of the contrasts. The two colums with only zero are those of the simple model.

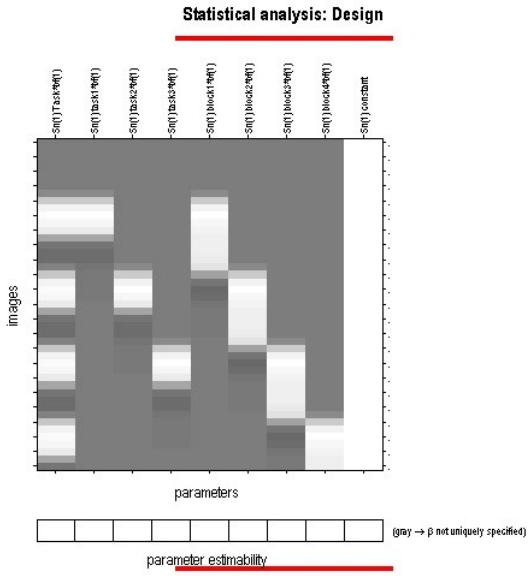


Figure 2.11: The design of the model to calculate F.

Chapter 3

How improve reliability

3.1 IRV index used in the paper

Bertoldi et al., 2015 defined ESS_1 as the explained deviance of Model 1, ESS_0 as the explained deviance of Model 0 and ESS_{1-0} as the difference between the two models. It can be interpreted as the gain in the explained deviance when moving from Model 0 to Model 1 (i.e., moving from a common stimulus effect among the trials to a different effect in each trial). If the effect of the stimulus in each block was constant, then ESS_{1-0} would be (about) equal to zero.

So we can formulated two Hypothesis Test for every voxel:

the first is $\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}$

where the null hypothesis tests if there is a difference in signal between the task/rest scans.

The second is $\begin{cases} S_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \beta_1 = \beta_2 = \beta_3 = 0 \\ S_1 : \bar{S_0} \end{cases}$

where the null hypothesis tests if there is a difference in signal among blocks.

Normalizing ESS_{1-0} in a zero-one interval, we obtained an initial raw indicator of the IRV as follows:

$$\text{IRV} = \frac{ESS_{1-0}}{ESS_{1-0}+RSS_1} = \frac{ESS_{1-0}}{RSS_0} = \frac{RSS_0-RSS_1}{RSS_0}$$

with $RSS_1 = RSS_0 - ESS_{1-0}$ being the residual deviance of the Model 1.

The IRV index resembled the definition of the R^2 index of regression analysis. Indeed, IRV index can be interpreted as the proportion of residual deviance due to the

difference in the estimated effects between blocks. Small IRV values (i.e., close to 0) indicate that the coefficients vary little between the blocks, whereas IRV values close to one suggest that there is large variation between the blocks.

We can compare two voxels with similar t-value but very different values of IRV (see Figure 3.1 and Figure 3.2). We can see that IRV index is able to understand how much a model can correctly provide the signal of a voxel. So, although both voxels have an high t-value only the first voxel is significant at a low threshold α, i.e. the voxel with stable signal of the stimulus among blocks (see Figure 3.3).

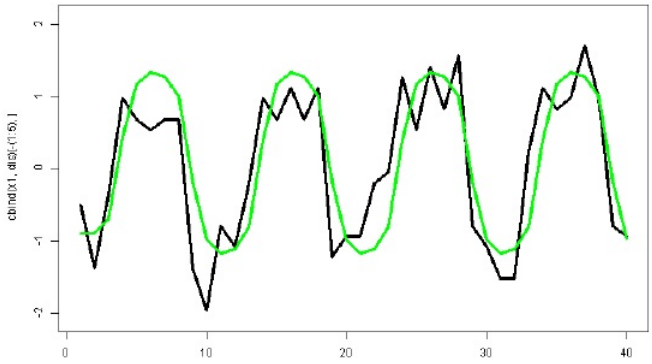


Figure 3.1: Voxel 1: t=2.28 and IRV=0.02

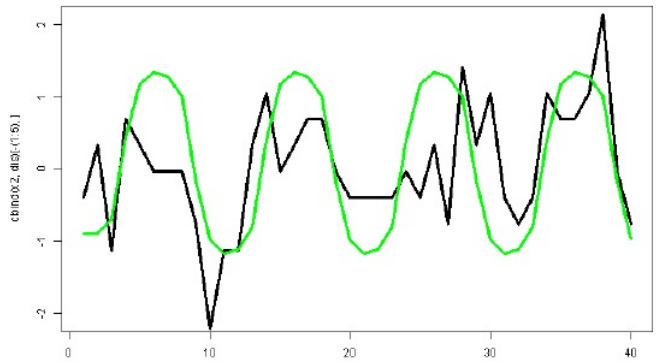


Figure 3.2: Voxel 2: t=2.36 and IRV=0.52

Until now we have concentrated our analysis on an only run but now we need all runs of the subject. The measure of reliability of a given voxel was defined as the

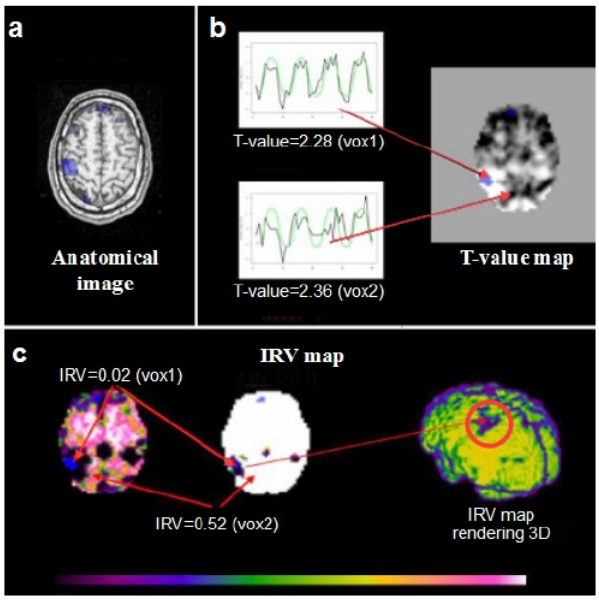


Figure 3.3: In the center IRV map we can see like voxel 1 with small IRV is significant while we can't see voxel 2 because it has high IRV.

overlap score, which denoted the proportion of runs in which the significance of a given voxel was $p<0.05$ (the index assumes values 0, 0.25, 0.5, 0.75, or 1 when there are four valid runs and 0, 0.33, .66, or 1 when there are only three). Then in the article they explored the relationship between IRV and the reliability of the overlap score (i.e., inter-runs). For each map of the seven subjects, they computed the correlation between the overlap score and the IRV. For each t-map, they calculated a correspondent IRV-weighted map in order to assess the potential improvement of the standard GLM analysis after the introduction of the IRV parameter (see Figure 3.4). The values of the computed IRV maps were standardized using the overall mean for each subject:

$$w_i = \frac{1-IRV_i}{\sum_{i=1}^{m} \frac{1-IRV_i}{m}}$$

where $\sum_{i=1}^{m} \frac{1-IRV_i}{m}$ means the average of (1-IRV) among all the m voxels. Subsequently, the t-maps were converted to p-maps (i.e., maps of p values), and the value of each voxel was divided by the correspondent weighted value. In this way, the voxels with stable signal among blocks (i.e. small IRV and high w) are favored, while voxels signal changing over blocks (i.e. high IRV and small w) are penalized. The resulting IRV-weighted p-map was then thresholded with the same threshold

used in the other steps of the analysis (p<0.05 or p<0.01).

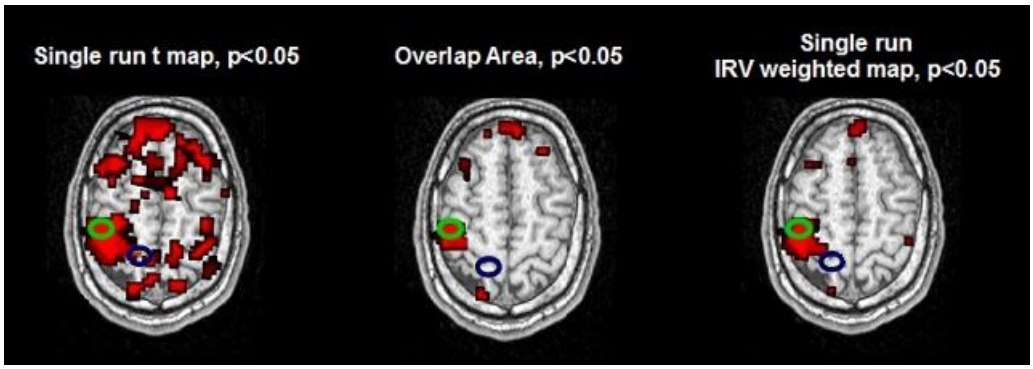


Figure 3.4: These are the maps of p-values, Overlap Area and IRV weighted p-values

3.2 Reliability of this IRV method and its problems

The Theorem 1 of the article Bertoldi et al., 2015 is the following.

Theorem 1

Let be p_i, i=1,...,m the p-values derived from the test statistics, that is, testing the null hypothesis H_0 of coefficient $b_0 = 0$ for the voxel i.

i) Define w_i for each voxel and $p_i^w = \frac{p_i}{w_i}$. The average type I error among all m tests is bounded by α for each α in (0,1], that is:

$$\sum_{i=1}^{m} \frac{P(p_i^w \leq \alpha | H_0)}{m} \leq \alpha$$

ii) More, generally, for any choice of w_i subject to a) the weights w_i are a function of the observed data only trough IRV_i and b) $\sum_{i=1,...,m} w_i = m$. (i.e., their sum equals m) the same property holds.

Theorem 1 states that the Type I error (i.e., 1-Specificity) is controlled at level α on average among all voxels. This mean that, considering all voxels together, the proportion of false positives is the same as in the standard analysis. We can make additional comments about any single voxel. The condition $p_i^w = \frac{p_i}{w_i} \leq \alpha$ used to select the active voxels can be restated as $p_i \leq \alpha w_i$. This means that weighting the p-values is equivalent to thresholding the (unweighted) p-values at different levels,

i.e. αw_i , which depends on the IRV_i of the single voxel.

The problems of this method is that is not true to control the threshold on average of the voxels under the S_0 hipotesis because:

$$\begin{aligned} E[IRV|S_0] &\leq E[IRV|S_1] \\ E[IRV|S_0] &\leq \frac{m_0 * E[IRV|S_0] + m_1 * E[IRV|S_1]}{m} \\ E[IRV|S_0] &\leq E[IRV] \\ E[1 - IRV|S_0] &\geq E[1 - IRV] \\ 1 = \frac{E[1 - IRV|H_0]}{E[1 - IRV|S_0]} &\leq \frac{E[1 - IRV|H_0]}{E[1 - IRV]} \\ 1 = E[w|S_0] &\leq E[w] \end{aligned}$$

where m_j is the sum of all voxel under the hipotesis S_j and $m = \sum m_j$, with j = 0,1. So, if the weights of all the voxels are on average bigger than 1 we threshold the voxels under S_0 on average with a value bigger than α:

$$\alpha \leq \alpha E[w]$$

i.e.

$$\sum_{i=1}^{m_0} \frac{P(p_i^w \leq \alpha|H_0)}{m_0} \leq \alpha E[w] \neq \alpha$$

while we want

$$\sum_{i=1}^{m_0} \frac{P(p_i^w \leq \alpha|H_0)}{m_0} \leq \alpha E[w|S_0] = \alpha$$

3.3 Some possible solutions

To solve this problem we want to find $E[w|S_0]$, starting from the real distribution of the voxels under the S_0 hipotesis ($E[1 - IRV|S_0]$), like a relationship between IRV and the statistic F distribution. We use F in this way because we know its distribution and this means that we know its expected value.

The Test F is: $\frac{(RSS_0 - RSS_1)/RDF_{0-1}}{RSS_1/RDF_1} = F_{RDF_{0-1};RDF_1}$

where:

RDF_0=37

RDF_1=30

RDF_{0-1}=7

There are two properties of distribution F that will be useluf subsequently:

- $1/F_{a;b} = F_{b;a}$

- $E[F_{a;b}] = \frac{b}{b-2}$

We know that $IRV = \frac{RSS_0 - RSS_1}{RSS_0}$ so, starting by this formule, we can find $E[(1 - IRV)|S_0]$ and $E[(1/IRV)|S_0]$:

$$\mathbf{E[(1 - IRV)|S_0]}$$

$$\begin{aligned}
E[(1 - IRV)|S_0] &= 1 - \frac{RSS_0 - RSS_1}{RSS_0} \\
&= \frac{RSS_0 - RSS_0 + RSS_1}{RSS_0} \\
&= \frac{RSS_1}{RSS_0} \\
&= (\frac{RSS_0}{RSS_1})^{-1} \\
&= (\frac{RSS_0}{RSS_1} - 1 + 1)^{-1} \\
&= (\frac{RSS_0 - RSS_1}{RSS_1} + 1)^{-1} \\
&= (\frac{RSS_0 - RSS_1}{RSS_1} : \frac{RDF_{0-1}}{RDF_1} * \frac{RDF_{0-1}}{RDF_1} + 1)^{-1} \\
&= (F_{RDF_{0-1};RDF_1} * \frac{RDF_{0-1}}{RDF_1} + 1)^{-1}
\end{aligned}$$

which substituting with the data of ours experiment we have:

$$\begin{aligned}
E[(1 - IRV)|S_0] &= E[(F_{7;30} * \frac{7}{30}) + 1)^{-1}] \\
&= mean((F_{7;30} * \frac{7}{30} + 1)^{-1})
\end{aligned}$$

where $F_{7;30}$ is a vector with 100000 casual value Fisher's F with 30 and 7 degrees of freedom. $E[(1 - IRV)|S_0]$ result approximately 0.81 (calculated with Rstudio).

$$\mathbf{E[(1/IRV)|S_0]}$$

$$\begin{aligned}
E[(1/IRV)|S_0] &= \frac{RSS_0}{RSS_0 - RSS_1} \\
&= \frac{RSS_0}{RSS_0 - RSS_1} - 1 + 1 \\
&= \frac{RSS_0 - RSS_0 + RSS_1}{RSS_0 - RSS_1} + 1 \\
&= \frac{RSS_1}{RSS_0 - RSS_1} + 1 \\
&= \frac{RSS_1}{RSS_0 - RSS_1} : \frac{RDF_1}{RDF_{0-1}} * \frac{RDF_1}{RDF_{0-1}} + 1 \\
&= F_{RDF_1;RDF_{0-1}} * \frac{RDF_1}{RDF_{0-1}} + 1
\end{aligned}$$

which substituting with the data of ours experiment we have:

$$\begin{aligned}
E[(1/IRV)|S_0] &= E[F_{RDF_1;RDF_{0-1}} * \frac{RDF_1}{RDF_{0-1}} + 1] \\
&= E[F_{RDF_1;RDF_{0-1}}] * \frac{RDF_1}{RDF_{0-1}} + 1 \\
&= \frac{RDF_{0-1}}{RDF_{0-1} - 2} * \frac{RDF_1}{RDF_{0-1}} + 1 \\
&= \frac{RDF_1}{RDF_{0-1} - 2} + 1 \\
&= \frac{30}{7 - 2} + 1 \\
&= \frac{30}{5} + 1 \\
&= 7
\end{aligned}$$

The new weights become:

I) $w_1 = \frac{1-IRV}{E[(1-IRV)|S_0]}$

II) $w_2 = \frac{1/IRV}{E[(1/IRV)|S_0]}$

Chapter 4

Results

4.1 P value of Fblock

We can see in Figure 4.1 the p-value distribution of F of all runs compared with a $F_{7;30}$. This p-values are very near to zero in all runs, specially in Run 2 and Run 3, it means that there are high values of F (specially in these runs), that is, signal instability between the blocks. Under the S_0 hipotesis the p-value distribution should be an uniform distribution.

4.2 Reliability Index

From the single-run t-maps of each subject, the index of reliability (Irel) between all possible pairs of runs (e.g., run 1 vs. run 2, run 1 vs. run 3, etc.,) is computed using the overlap method (Rombouts et al., 1997). More specifically, for each pair of runs, we considered the size of the activated area in the first (Va) and second (Vb) runs at the 2 chosen thresholds ($p<0.01$ and $p<0.05$). In Figure 4.2 we can see an example of a map of an overlap area. Next, as a measure of reproducibility, we computed the size of the areas activated in both runs (V overlap). Finally, the reliability of each pair of runs is defined as $\frac{2*Voverlap}{Va+Vb}$. To obtain a single value of Irel for each subject, we averaged the various indices calculated for the different pairs of runs. The Irel ranged between zero and one.

At both the thresholds adopted ($p<0.01$ and $p<0.05$), we calculated the Irel index

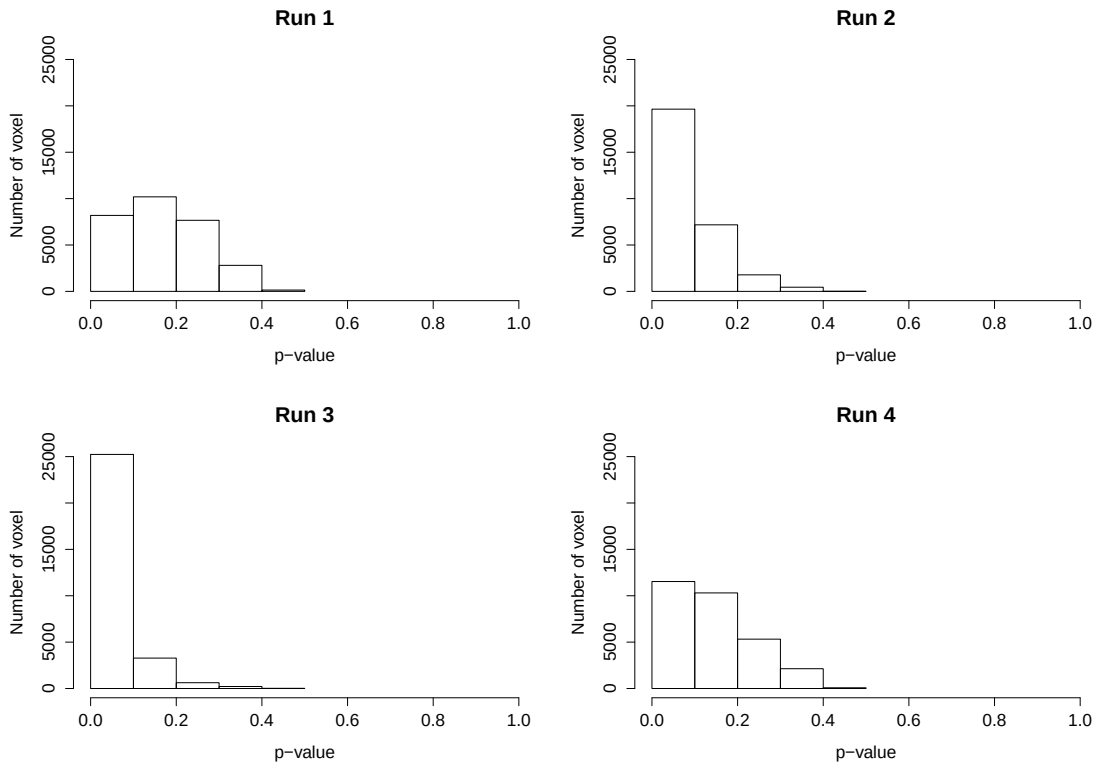


Figure 4.1: Histograms of the distribution of p-value of F taken from Rstudio.

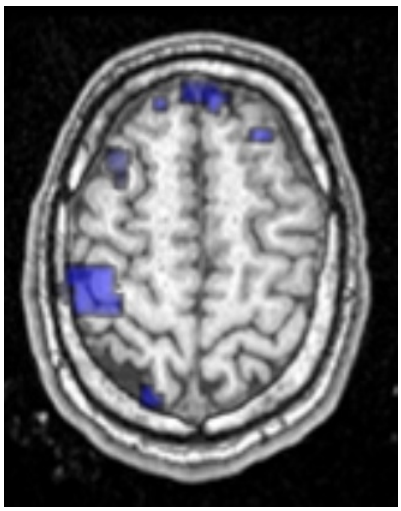


Figure 4.2: An example: a map of reliability of overlap areas.

for the standard t-maps (Irel std) and the IRV-weighted t-maps (Irel wgt). Subsequently, for each subject, we quantified the gain in reliability reached with the proposed IRV-weighting method (gain Irel), as follows:

$$\text{gain Irel} = \frac{Irelwgt}{Irelstd} - 1$$

The results of reliability of the methods used here are listed in the tabels of Figure 4.3 and 4.4. We provide a graphical rapresentation overlaying ours maps of thresholded voxels to the reference brain image using MRIcron (http://www.mricro.com by C. Rorden, see Figure 4.5).

Method	Overlap Area	Mean Areas	Reliability
$\frac{1-IRV}{\sum(1-IRV)/m}$	243	1289,75	0,2108
$\frac{1/IRV}{\sum(1/IRV)/m}$	243,1667	1280	0,2127
$\frac{1-IRV}{E[(1-IRV)\|S_0]}$	203,5	1150,75	0,1998
$\frac{1/IRV}{E[(1/IRV)\|S_0]}$	118,8333	814,75	0,1718

Figure 4.3: Results with the threshold 0.01

Method	Overlap Area	Mean Areas	Reliability
$\frac{1-IRV}{\sum(1-IRV)/m}$	1027,6667	3356	0,3178
$\frac{1/IRV}{\sum(1/IRV)/m}$	1038,1667	3352,75	0,3216
$\frac{1-IRV}{E[(1-IRV)\|S_0]}$	866,6667	2995,25	0,3023
$\frac{1/IRV}{E[(1/IRV)\|S_0]}$	504,8333	2090,25	0,2576

Figure 4.4: Results with the threshold 0.05

In this case we had no earnings, then it is possible that the method is less powerful than the standard method, but it will need to make further investigations (in other subjects and studies), because there are many aspects that could be discussed in future.

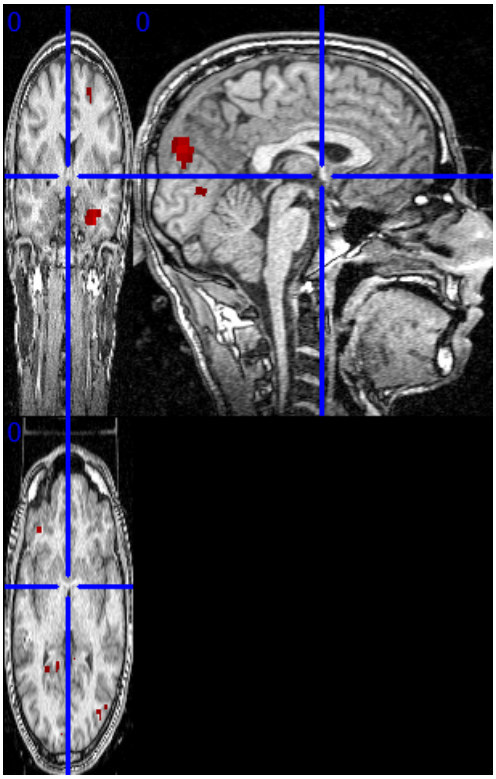


Figure 4.5: An anatomical image (open with MRIcron) where the red voxels are those that are actives with the thresold α.

Bibliography

Aguirre, G. K., E. Zarahn, and M. D'Esposito (1998). "The Variability of Human, BOLD Hemodynamic Responses". In: *NeuroImage* 8.4, pp. 360–369.

Bertoldi, F. de, L. Finos, M. Maieron, L. Weis, M. Campanella, T. Ius, and L. Fadiga (2015). "Improving the reliability of single-subject fMRI by weighting intra-run variability". In: *NeuroImage* 114, pp. 287–293.

Friston, K.J. (2003). "Introduction: Experimental design and statistical parametric mapping". In: *Human Brain Function*. Ed. by R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny. 2nd. Academic Press.

Friston, K.J., A.P. Holmes, K.J. Worsley, J.B. Poline, C. Frith, and R.S.J. Frackowiak (1995). "Statistical Parametric Maps in Functional Imaging: A General Linear Approach". In: *Human Brain Mapping* 2, pp. 189–210.

Friston, K.J., J. Ashburner, S.J. Kiebel, T.E. Nichols, and W.D. Penny, eds. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.

Glover, G. H., T. Q. Li, and D. Ress (2000). "Image-based method for retrospective correction of physiological motion effects in fMRI: Retroicor". In: *Magnetic resonance in medicine* 44.1, pp. 162–167.

Lindquist, M. A. (2008). "The statistical analysis of fMRI data". In: *Statistical Science* 23.4, pp. 439–464.

Lindquist, M. A., C. Zhang, G. Glover, and L. Sheep (2008). "Acquisition and statistical analysis of rapid 3d fmri data". In: *Statist. Sinica* 18, pp. 1395–1419.

Lund, T., K. H. Madsen, K. Sidaros, W. L. Luo, and T. E. Nichols (2006). "Non-white noise in fmri: Does modelling have an impact?" In: *NeuroImage* 29, pp. 54–66.

MathWorks Inc., Matrix Laboratory 7.6 (2012).

Ogawa, S., D. Tank, R. Menon, J. Ellerman, S. Kim, H. Merkle, and K. Ugurbil (1992). "Intrinsic signal changes accompanying sensory simulation: Functional brain mapping and magnetic resonance imaging". In: *Proc. Nat. Acad. Sci.* 89, pp. 5951–5955.

Rombouts, S. A., F. Barkhof, F. G. Hoogenraad, M. Sprenger, J. Valk, and P. Scheltens (1997). "Testretest analysis with functional MR of the activated area in the human visual cortex". In: *AJNR Am. J. Neuroradiol* 18, pp. 1317–1322.

Worsley, K.J. and K.J. Friston (1995). "Analysis of fMRI Time-Series Revisited - Again". In: *NeuroImage* 2, pp. 173–181.