UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea Triennale in Fisica

Tesi di Laurea

Connettività dinamica in un modello computazionale del cervello affetto da ictus

Dynamical connectivity in a whole-brain computational model of stroke

Relatore

Dr. Michele Allegra

Laureando

Silvia Collicelli

Anno Accademico 2022/2023

Contents

Abstract

In this thesis, we test the ability of a mathematical model of brain activity to capture dynamic functional connectivity. We will initially describe the meaning of dynamic functional connectivity: after briefly describing the typical data analysed in our work, fMRI time series, we will define functional connectivity between different regions of the brain and discuss the meaning of the "functional connectivity speed". We will then introduce a computational model that simulates brain regions' activity, discussing the choice of parameters, and how to infer them from empirical data. Finally, we will fit the model on empirical data to simulate the system for both subjects in good health and patients affected by stroke. We will compare the results obtained from simulations with those obtained directly from empirical data, in order to determine whether the model reproduces well the observed behaviour of dynamic functional connectivity and observes differences between healthy and unhealthy subjects.

Nella seguente tesi viene testata la capacità matematica per un modello computazionale di attività cerebrale di riprodurre la connettività funzionale dinamica. Inizialmente viene descritto il significato della dynamic functional connectivity: dopo aver brevemente illustrato il tipo di dato utilizzato per l'analisi, ovvero serie temporali fMRI, viene definito il concetto di functional connectivity tra differenti regioni del cervello e il significato della sua velocità. Viene successivamente introdotto un modello computazionale che simula l'attività cerebrale delle diverse regioni, discutendo i vari parametri di tale modello e esponendo un metodo di interferenza per trovarli partendo dai dati empirici. Viene infine fittato tale modello e simulato il sistema sia per soggetti sani, sia per pazienti affetti da ictus. Vengono confrontati i risultati ottenuti dalle simulazioni con quelli ottenuti direttamente dai dati empirici, così da determinare se il modello utilizzato è in grado di riprodurre fedelmente il comportamento osservato per la dynamic functional connectivity e di osservare differenze tra soggetti sani e malati.

Chapter 1

Introduction

In the last decades, studies on brain cognition have progressively expanded as many non-invasive recording techniques have been developed, first of them functional magnetic resonance imaging (fMRI), which measures the blood-oxygen-level dependent (BOLD) signal, a proxy of neuronal population activity. fMRI is now commonly used to investigate brain activity both during active cognition and in resting state. A property that has been thoroughly analyzed is functional connectivity (FC), which is specific of the entire brain system, as it depends on how different regions interact with each other and can be easily calculated from the correlation between two BOLD signals. The FC matrix, whose entries contain the correlation between all pairs of regions in the brain, yields a map of the large-scale network of the brain, showing how and how strongly one region is correlated with the others.
The prominence of FC in brain studies is also due to the fact that this property changes depending on different neurological and psychiatric conditions, as shown by many authors. In particular, there is evidence that specific brain diseases are linked to specific alterations in FC. This phenomenon emerges both in diseases classically regarded as network pathologies such as Alzheimer, epilepsy, schizophrenia or Autism spectrum disorders, but also in pathologies characterized by focal or multifocal lesions like stroke, tumors and multiple sclerosis [4].
Actually, disease-related anomalies are much more evident in what is called dynamic functional connectivity (dFC) and in its speed. As a matter of fact, FC was traditionally analyzed as a static property of the brain, measured as a temporal average over an entire recording session; but in the last years much more interest has been directed to analyze temporal changes of this observable across short time windows. *Dynamic* FC is an index of the temporal fluctuations of FC which reflect the dynamic changes in the brain network. Visually, we can think of dFC as a random walk between transient FC states; the mean and the type of distribution of its speed are a first quantitative characterization of this walk [7]. The most common way to evaluate FC is by taking the so-called "sliding-windows" approach. Within this approach, we consider a time-ordered sequence of (short) time windows and calculate a FC matrix for each window. The correlation between successive FC matrices defines the dFC "speed", to be considered as the speed of a random walk in FC space. By creating appropriate surrogates, we can test specific properties of the random walk. Switching the matrices of the sequence, we can test the null hypothesis of complete randomness (lack of consequential correlation). By looking at phase-randomized BOLD time series, we can test the null hypothesis of stationarity (order). It has been shown that the empirical distribution of the dFC speed deviates both from an ordered and a randomized scenario, meaning that dFC is a complex random walk [1]. The distribution of the dFC speed can depend on the psychiatric and pathological conditions of the subject; it has been shown that with aging the speed tends to have slower values, meaning that the random walk tends to be less complex. There is also a connection between dynamic FC and sleep deprivation, as dFC speed shows evidence of slowing down after 24 hours of uninterrupted wake [8].
In this thesis we will verify if there are differences in dFC speed between healthy subjects and stroke patients, comparing the distributions obtained in the two different samples. Following the studies mentioned before, we expect to notice some dissimilarities between the two; in particular we are going to test the hypothesis that slower speeds are associated to patients with stroke. Furthermore, we

are going to verify if a simple mathematical model of brain activity (the *MOU* model [3]) is able to reproduce the empirical distribution of the dFC speed along with the differences between patients and healthy subjects (if any exist). The hypothesis we are going to test in this case is that a simple linear model may not exactly reproduce a desired dynamic property like the dFC speed.
In the first section of this work, we describe the methods used for the analysis. Starting from the type of data we consider, we define the properties we focus on, like the FC matrix and the dFC speed, giving also an insight on the meaning of "time window" and defining the ranges of window sizes used for the analysis. Then, we describe the model used for simulations, mentioning the parameters of the corresponding stochastic differential equation. Finally, we show an inference method to estimate those parameters from empirical data, called Lyapunov Optimization.
In the second section, we show the results found in our analysis. First we detail the parameters used for the simulation and the procedure followed for the analysis. Then, we show an example of the static FC matrix for each considered sample, differentiating healthy subjects from patients with stroke and empirical data from simulated ones. Lastly, we do the same for the distributions of the dFC speed and we use the Kolmogorov-Smirnov test to verify if some differences can be detected between the two clinical groups.
Finally, in Discussion and Conclusions we resume the results and we discuss the validity of our original hypothesis. In conclusion we raise some questions for possible additional future work, explaining which improvements could be made to obtain better and more precise results.

Chapter 2

Methods

In order to completely understand the work done for this thesis project, first we are going to give a brief explanation of the data used for the analysis (which are fMRI time series), of the model implemented for the simulation and the properties that can be derived from these data.

2.1 fMRI time series

The data analyzed in this project are time series representing the neural activity of different brain regions in the course of time. The method used to detect neural activity is called Functional Magnetic Resonance Imaging (fMRI) and is based on the fact that neuronal activation is connected to an increase of blood flow towards the interested region, and consequently an increase in the fraction of oxygenated blood in the region. A specific technique to detect this phenomenon is the blood-oxygen-level dependent (BOLD) contrast, which is a brain scan used to map neural activity using magnetic resonance to measure the presence of oxygen in the blood. An important feature of this type of data is that it only represents an indirect measure of brain activity and it can be significantly corrupted by noise, meaning that it is necessary to use statistical methods for the analysis.

2.2 FC matrix

Once the activity of each brain region is known, it is possible to calculate different metrics summarizing its spatiotemporal behavior and in particular have an estimation of the *functional connectivity (FC)*. FC emerges when distant cortical areas exhibit a strong correlated neural activity, which typically occurs even during resting state (rs). The degree of correlation is inherently dynamic, as it changes in time and it is important to study its fluctuations, which are referred to as *dynamic Functional Connectivity (dFC)*. Studies show that these fluctuations may have different trends in elderly people or in people with certain pathological issues, so studying them can improve our understanding of these diseases too, and possibly assist us in detecting and preventing them.

Let us now give a proper definition of the aforementioned properties. The functional connectivity between the i-th region and the j-th region at a fixed time t is defined as the Pearson correlation between the neural activity of the two regions, computed over a window of duration of length τ. In order to calculate the functional connectivity with this type of data we first computed the sample covariance:

$$cov(x_i, x_j) = \frac{1}{\tau - 1} \sum_t (x_i(t)) - \hat{x_i})(x_j(t)) - \hat{x_j})$$

where in our case x_i refers to the neural activity of region i, and $\hat{x}_i$ is the averaged activity.
Then we calculated the correlation for the functional connectivity as

$$FC_{ij} = \frac{cov(x_i, x_j)}{\sqrt{cov(x_i, x_i) cov(x_j, x_j)}}$$

If we calculate this property for every pair of regions, we produce a matrix of the functional connectivity of the brain.

In figure 2.1 we can see an example of how a typical static FC matrix may look like. It is a squared matrix where the entries refer to a specific brain region, for this particular case we used a set of data that we are going to use for the following analysis and we consider a number of 119 different brain regions. Here the colour indicates the strength of the correlation and in this example the yellow colour states for strong correlation, whereas red is for not correlated areas and black for regions with inverse correlation (coefficient equals to -1). As expected we can see that the diagonal has a perfect correlation, with a Pearson coefficient equal to one.

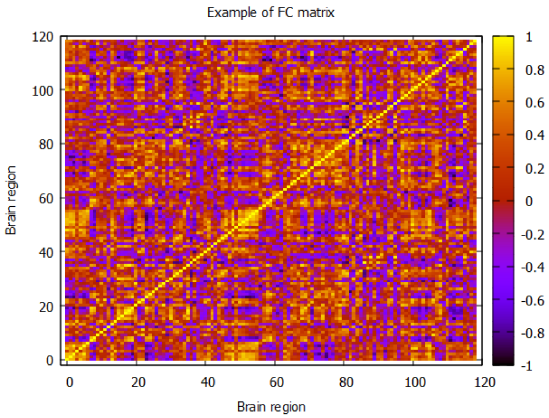

Figure 2.1: Example of a typical FC matrix

The FC matrix depends on the specific time window used. Let us denote by $FC_{ij}(t)$ the matrix computed in the time window $[t - \tau/2, t + \tau/2]$. If we put many FC matrices evaluated at different times in a time-ordered sequence, we obtain the so-called dynamic Functional Connectivity (dFC) stream, which shows how Functional Connectivity changes in time. As an example we can see in figure 2.2 a time series of brain activity of a certain region that has a duration of $300s$ and a window size of $50s$, meaning that can be calculated six FC matrices. Starting from the BOLD time series, we divide it in many parts with a duration of τ each, so that in the dFC stream, every FC matrix is evaluated at a distance of one τ with respect to the previous one.

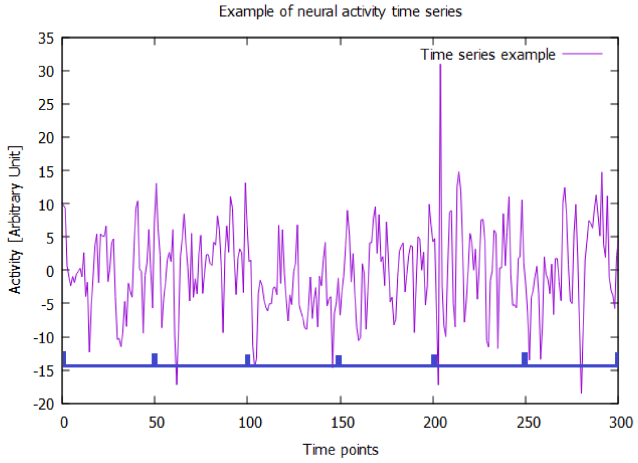

Figure 2.2: Example of temporal windows in a time series

2.3 dFC speed

We are going to focus our study on dFC and on its speed. Considering dFC as a "smooth flow across continually morphing connectivity configurations" [1], we can then define a speed V_{dFC} and analyze its distribution. We estimate the dFC matrix from the FC matrices of the dFC stream, measuring the similarity between two FC matrices at time t_1 and t_2 through the Pearson correlation:

$$dFC(t_1, t_2) = corr[FC(t_1), FC(t_2)]$$

In order to calculate the correlation between two matrices, we consider each entry of the matrix as a component of the sample, meaning that each sample has $n \times n$ elements where n is the number of

regions. We estimate then the sample covariance and the correlation as seen before for the FC. We define the dFC speed as the variation between two FC matrices evaluated at time t and $t + \tau$, where τ is the window size. In particular we define V_{dFC} as:

$$V_{dFC,\tau}(t) = 1 - dFC(t, t + \tau) = 1 - corr[FC(t), FC(t + \tau)]$$

The distribution of this quantity depends on the window size τ. In order to see this dependency, we divide the window sizes in three ranges depending on their duration, in particular we have a *short* window-size range, where we consider $6s < \tau < 15s$ and a number of 3-8 time points of the time series, as in the time series each point is taken every two seconds; a *medium* window-size range with $15s < \tau < 60s$ with 9-30 time points, and a *long* window-size range with $60s < \tau < 210s$ with 33-105 time points.
Studies show that the distribution of this speed tends to presents a peak on a certain value of typical V_{dFC}, but deviates from Gaussianity and distances itself both from an ordered and a completely randomized situation [1]. In particular, Battaglia et al. tested two null hypotheses: the first one for stationarity, indicating an ordered scenario and the second one for time uncorrelation, which is related to randomness. The results showed that the dFC speed cannot be considered a stationary property nor it has a randomized distribution; it lies between these two scenarios, being thus complex. Furthermore, Battaglia et al. showed an interesting feature of the dFC speed, namely that it decreases with age, as the peak of its distributions results shifted towards lower values. This means that there is a (negative) correlation between aging (in this case we talk about healthy aging, as none of the tested patients were diagnosed with some issues) and the speed of dynamic functional connectivity. In elderly people the change between two different FC matrices is slower so we can say that speed tends to be more trivial and less complex.
In this research we estimated V_{dFC} for each tested patient and created a unique final graph with the information obtained for all subjects, keeping separated the three window-size ranges. In the latter part of this thesis we are going to test if the speed distributions have a meaningful correlation also with pathological issues, analyzing if it presents different distributions particularly on patients affected with stroke.

2.4 Model for the simulation

For this thesis project we are interested in a comparison of V_{dFC} in two different cases: in the first one we estimate those features directly from the empirical data of time series, in the second analysis we use a specific model for our system and generate simulated time series. Once we have the time series from this second method, we can give another estimation of the dFC speed and compare the two results in order to evaluate the accuracy of the chosen model.

Models for these time series data generally consist in stochastic differential equation models. Such models can have many forms and represent different stochastic processes, in particular they can be divided into two different categories: linear and non-linear models. For fMRI time series it has been demonstrated that the best models are linear ones, although the brain is a typical non-linear system. This is because the nonlinearities at the level of large-scale brain activity, which is the scope in which we are interested, are very little, meaning that a linear approximation yields good results.

2.5 Multivariate Ornstein-Uhlenbeck process

The simulation for our system follows a Multivariate Ornstein-Uhlenbeck (MOU) process, which is a stationary (as its distribution does not change in time) stochastic Gaussian process that corresponds to a network with linear feedback. This linear feedback is given by a weight for each interaction between two brain regions. The stochastic component of the system follows a normal distribution, so it has a specific mean and variance. In particular the model is defined by the following stochastic

linear differential equation:

$$dx_i = \left[-\frac{x_i}{\tau_x} + \sum_{j \neq i} C_{ij} x_j \right] dt + dB_i$$

This evolution represents an exponential decay for the activity of each i-th region with a time constant equal to τ_x [1]. This activity is excited by the activity of the other areas each of them multiplied by their respective recurrent weight C_{ij} that represents the *effective connectivity* of the network; dB_i is a Gaussian noise with variance σ_i^2 that adds fluctuations to the system. The effective connectivity represents interconnections between different brain regions, so that each region receives inputs from other regions to which it is connected.
This process is commonly used in neuroscience to model fluctuating activity and results very efficient in representing the main features of neural activity. In particular this model reproduces a whole-brain system, which is necessary in order to estimate the functional connectivity, that is a property that blends all together all the regions of the brain. Furthermore this model captures both spatial and temporal changes, as the differential equation is valid for each region. In addition to that, the C matrix of the recurrent weights reproduces the causality given by the interactions between different regions, meaning that strong C weights correspond to strong causal relationships [3].
To use this model to produce simulated time series, we need to know the parameters. In particular it is very important to have an estimation of the effective connectivity C_{ij}, as it represents how and how strong different areas are connected to one another, giving the possibility for example to differentiate receptors if the area exhibits strong incoming connections or feeders if the outgoing connections are more significant. In Ref. [2], Gilson et al. focus on testing an inference method for the estimation of the C_{ij} weights and the variances of each region gathered together in a unique matrix Σ, defined as

$$\Sigma dt = \mathbb{E}[dB_i dB_j]$$

Starting from the time series of the brain activity for each region, they compute the (shifted) covariances, which are a sufficient statistic for the model, as they conatain all the information needed to estimate the parameters. The covariances are calculated as

$$\hat{Q}_{ij}^{\hat{\tau}} = \frac{1}{T} \sum_{0 \leq t \leq T} \left(x_i^t - \hat{x}_i \right) \left(x_j^{t+\hat{\tau}} - \hat{x}_j \right)$$

where T is the duration of the simulation (in this case $T = 1000$ s); $\hat{x}_i^n$ is the averaged activity of i-th region. This definition of covariance matrices considers a time shift of a duration of $\hat{\tau}$. In the article it is showed that the best estimation is obtained by using simultaneously the non-shifted covariance matrix $\hat{Q}_{ij}^0$ and the shifted $\hat{Q}_{ij}^{\hat{\tau}}$ with a time shift $\hat{\tau} \approx 2s$.

2.5.1 Lyapunov Optimization

Gilson et al. manage to estimate C_{ij} and dB_i by implementing a method called Lyapunov Optimization, where C_{ij} is tuned iteratively in order to reproduce theoretical covariances as more similar as possible to the empirical ones. They use a gradient descent approach and start by considering the following Lyapunov function

$$L(Q^0(\Sigma, C), Q^\tau(\Sigma, C)) = \sum_{m,n} (Q_{m,n}^0 - \hat{Q}_{m,n}^0)^2 + \sum_{m,n} (Q_{m,n}^\tau - \hat{Q}_{m,n}^\tau)^2$$

where the hatted covariances are the empirical ones. The aim of this method is to reduce this function (which is positive and has a minimum when the covariances of the model are equal to the empirical ones) in order to find the best estimation for C.

Firstly one defines the Jacobian matrix

$$J_{ij} = -\frac{\delta_{ij}}{\tau_x} + C_{ij} \tag{2.1}$$

[1] The time constant of the process τ_x has a different meaning than the aforementioned window-size τ

where δ_{ij} is a Kronecker delta, meaning that the decay affects only diagonal elements. Then by deriving the covariances with respect to t and using the so-called Ito's formula [5], it is possible to derive an equation (called Lyapunov equation):

$$JQ^0 + Q^0 J^T + \Sigma = 0 \tag{2.2}$$

where J^T indicates the transposed matrix of J. Next one derives the covariances with respect to τ in order to find a relation between Q^τ and Q^0:

$$Q^\tau = Q^0 \mathbf{expm}[J^\dagger \tau] \tag{2.3}$$

where **expm** is the matrix exponential.

Once we have all the needed elements and equations we can start the iterative process to find the model parameters. In this treatment we consider Σ to be known, but in his article Gilson demonstrates that with a minor modification the method can be used to find the noise covariance Σ along with the effective connectivity C. The process begins with zero weights ($C_{ij} = 0$), then it calculates the Jacobian J using (2.1) and finally Q^0 and Q^τ through equations (2.2) and (2.3). The purpose is to update C so as to obtain the following desired changes for the covariances:

$$\Delta Q^0_{mn} = \epsilon\big(\hat{Q}^0_{mn} - Q^0_{mn}\big)$$

$$\Delta Q^\tau_{mn} = \epsilon\big(\hat{Q}^\tau_{mn} - Q^\tau_{mn}\big)$$

where ϵ is a small chosen parameter. In order to calculate the variation of J, ΔJ, that causes ΔQ^0 and ΔQ^τ first one finds a relation between ΔJ and ΔQ^0, ΔQ^τ. By inverting (2.3) we have:

$$J = \frac{1}{\tau}\left\{\mathbf{logm}\Big[\big(\hat{Q}^0\big)^{-1}\hat{Q}^\tau\Big]\right\}^\dagger \tag{2.4}$$

One differentiates (2.4) with respect to $X = (Q^0)^{-1}Q^\tau$ and finds

$$\begin{aligned}
\Delta J &= \frac{1}{\tau}(\Delta X X^{-1})^\dagger \\
&= \frac{1}{\tau}\left\{\Big[(Q^0)^{-1}\Delta Q^0 (Q^0)^{-1}Q^\tau + (Q^0)^{-1}\Delta Q^\tau\Big]\Big[(Q^0)^{-1}Q^\tau\Big]^{-1}\right\}^\dagger \\
&= \frac{1}{\tau}\Big[\big(Q^0\big)^{-1}\big(\Delta Q^0 + \Delta Q^\tau \mathbf{expm}(-J^\dagger\tau)\big)\Big]^\dagger
\end{aligned} \tag{2.5}$$

The connectivity update is then

$$\Delta C_{ij} = \begin{cases} \Delta J_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

To have an estimation of Σ as well, we can tune it iteratively in parallel with C. This is possible if we consider Σ as a diagonal matrix, meaning that each entry on the diagonal of the matrix correspond to the variance of the respective brain region: $\Sigma_{ii} = (\sigma_i)^2$.
For this optimization we have considered the time constant of the system (τ) known and equal to $\tau = -\frac{1}{C_{00}}$.
The parameters used for the simulations in this thesis are derived using the Lyapunov Optimization just described.

Chapter 3

Implementation and Results

The data used come from two different sessions for healthy participants and one unique session for unhealthy ones. In particular we have 31 recordings for the first session of healthy subjects, 26 for the second session and 127 for subjects with illness. Actually, for the last collection of samples from people with stroke, 10 samples out of 127 turn out to be corrupted, as they contain NaN data or time series identically equal to 0, which cannot be considered realistic. As a consequence, these corrupted files are omitted from the analysis. For each patient we have three different files: the first one with the corresponding time series, needed for the direct estimation of the measures of interest; the second and the third containing the C and the Σ matrices needed to generate simulated time series (both are calculated using the aforementioned Lyapunov Optimization, with Σ being diagonal).

The simulation is implemented using the Euler method to solve the differential equation. This method consists in dividing the total duration of the simulation into N small timesteps of duration dt; one starts from a given initial value and then updates the system to the value of $x(t + dt)$, which is calculated by using the equation evaluated in $x(t)$ and multiplied by the duration of the timestep dt. In our analysis the initial values for the simulation are chosen to be equal to 0 for each brain region and for each sample. The other parameters are collected in the following table:

Number of brain regions	n=119
Simulation duration	T=1000 s
Simulation timestep	dt=$\tau \times 0.005$
Steps of the simulation	$N = T/dt$
Time constant	$\tau = -1/C_{0,0}$

Table 3.1: Parameters for the simulation

For the simulation, as the differential equation is a stochastic one, we use an algorithm to generate random numbers that follow a Gaussian distribution with mean equal to 0 and variance taken from the Σ matrix. With the simulation we obtain the time series of the system, that we use to estimate the *Functional Connectivity* matrix first and the *Functional Connectivity Speed* later. These two estimations depend on the chosen window size. After the integration of the differential equation, we calculate the FC matrices for any window size, starting from 3 time points and reaching 105 time points. We divide the results into three categories in order to gather together estimations from similar window sizes.

3.1 FC matrices

We present now the results we obtained from the analysis, starting from the FC matrices. As representatives, we show a matrix from a simulated time series of a healthy subject, one from the simulation of an unhealthy patient, and the other two corresponding to estimations from the direct time series.

Every matrix that we show as an example is evaluated with a window size of 20 time points, which is from the *medium* window-size range.

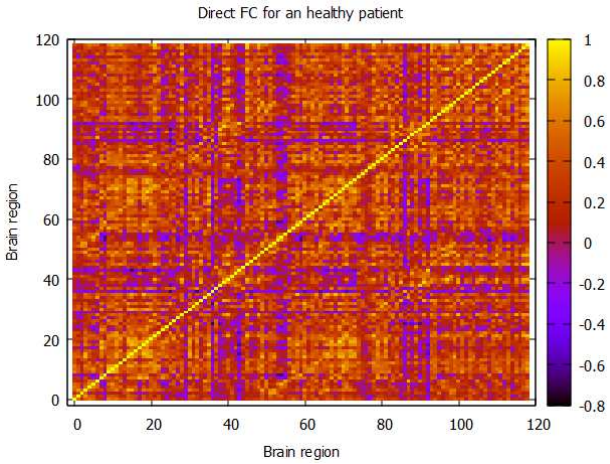

Figure 3.1: FC for dh

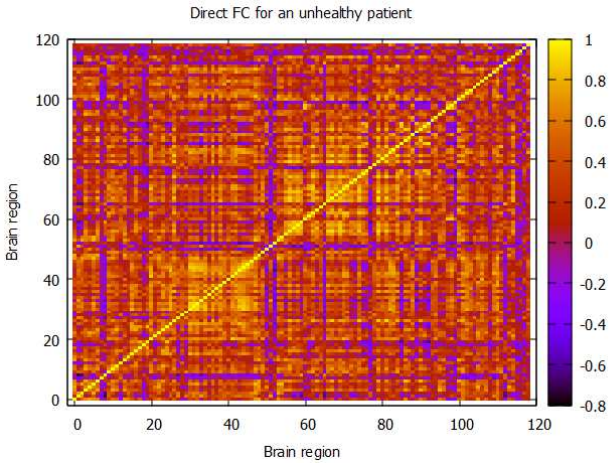

Figure 3.2: FC for dp

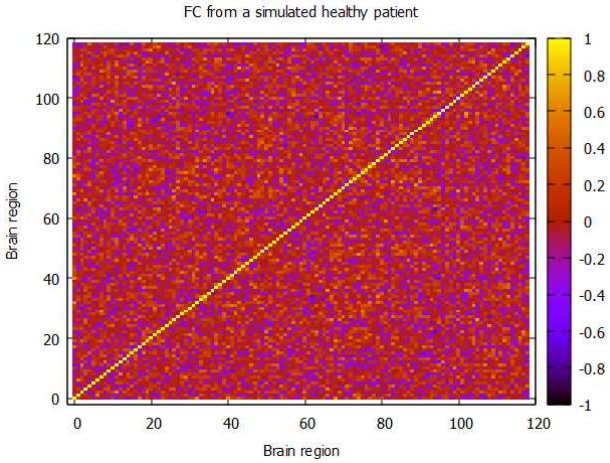

Figure 3.3: FC for sh

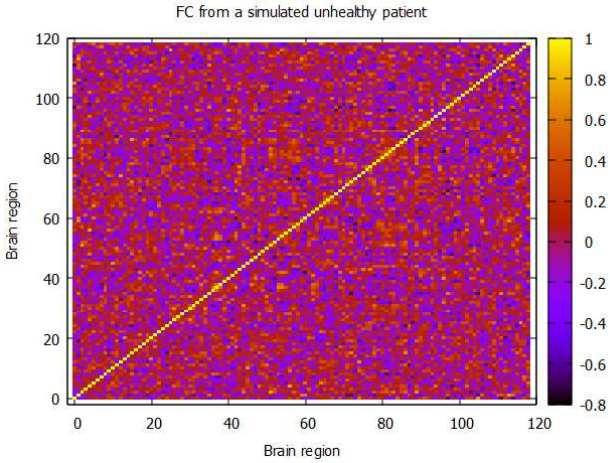

Figure 3.4: FC for sp

The initials used in the captions of the just showed figures stand for:

dh	estimations corresponding to direct data of healthy patients
dp	estimations corresponding to direct data of unhealthy patients
sh	estimations corresponding to simulated data of healthy patients
sp	estimations corresponding to simulated data of unhealthy patients

Table 3.2: Meaning of the initials used in the analysis

From the obtained figures one can see immediately that the simulated matrices do not reproduce the typical "checkerboard" scheme that is easy to see in the direct matrices, both for healthy and unhealthy patients. Furthermore, visually one can notice that the correlation values of simulated matrices have a general shift toward lower values, as the colours of the matrices tend to the blue instead of the yellow. These results will be discussed in the last part of this thesis.

3.2 dFC Speed

Once we have calculated the Functional Connectivity for each patient and for each session, we focus on the *speed* of the dynamic Functional Connectivity. We collect in three different groups all the speed values of all patients and divide them with respect to the window-size range. We can then plot in a unique graph the distributions for the three ranges, in order to see the differences given by the different considered window-sizes. We repeat this procedure for the 4 different kind of samples, as done for the FC matrices. In the following results we see that the graphs corresponding to the unhealthy subjects have a much bigger number of counts, as we have more availability of files corresponding to this type of patients; also for the estimations from the simulated time series, the number of counts is much bigger with respect to the estimations from direct data, that is because the duration of the implemented simulation is longer than the available direct time series.
We report here the four final graphs, two of them corresponding to the estimations from direct time series, and the other two to the simulated data.

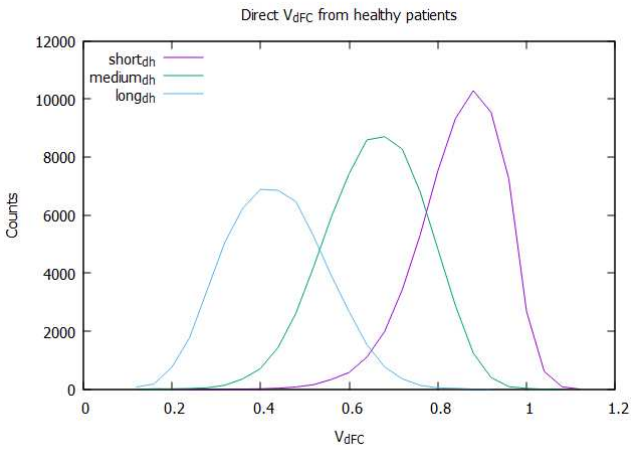

Figure 3.5: V_{dFC} for *dh*

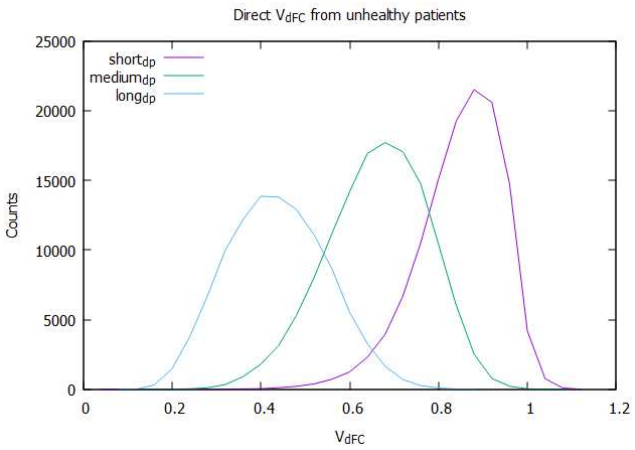

Figure 3.6: V_{dFC} for *dp*

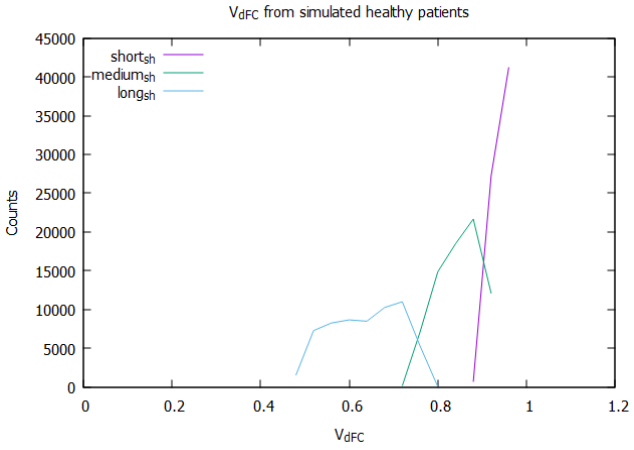

Figure 3.7: V_{dFC} for *sh*

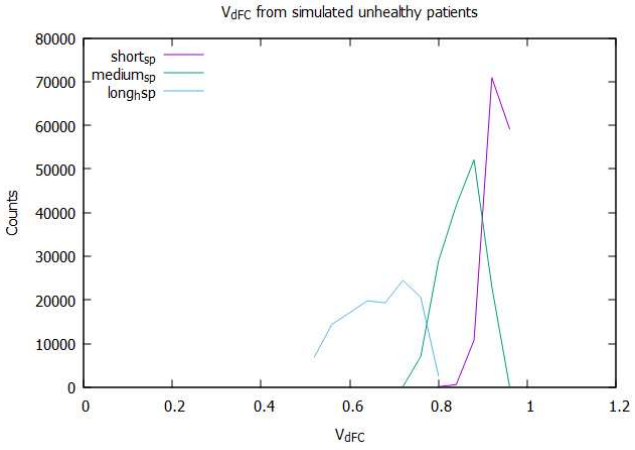

Figure 3.8: V_{dFC} for *sp*

One immediately notices that the distributions of the V_{dFC} from the direct time series are much smoother with respect to the simulated ones, that result in being very fragmented, both for the patients and the healthy subjects. The tendency for the simulated estimations seems to be shifted towards higher values of the speed, in particular for the *long* and *medium* window-size ranges. In addition to that, we can observe that there are no values bigger than 1 for the *short* window-size of simulated time series, causing a distribution with no peak, at least for the sample corresponding to

healthy subjects.
Another characteristic that one can see from the graphs is that there is no immediate evidence of a difference between the distributions of healthy subjects and of patients affected with stroke, particularly for the distributions from direct time series.

3.2.1 Kolmogorov-Smirnov Test

In order to understand whether there is no actual difference in the distributions for the *dFC Speed* or if the dissimilarities are too little to be noticed visually, we used the Kolmogorov-Smirnov test to compare the distributions of the two different samples. This test allows verifying if two different samples come from the same distribution. The null hypothesis H_0 states that both samples come from the same distribution[1]. The procedure starts by calculating the cumulative function for both samples, which we call F(x) and G(x) and the difference between the two. We consider then the maximum of this difference:

$$D_{m,n} = \max_x |F(x) - G(x)| \tag{3.1}$$

where m and n are the dimensions of the two samples. We then compare the value of (3.1) with

$$D_{m,n,\alpha} = c(\alpha)\sqrt{\frac{m+n}{m \times n}} \tag{3.2}$$

where $c(\alpha)$ is the inverse of Kolmogorov distribution and can be found in statistical tables, depending on the value of the significance level α [6]. If $D_{m,n} > D_{m,n,\alpha}$ the null hypothesis is rejected with a significance level equal to α. We summarize on the table below the parameters and the results obtained from this test for each group of two samples, separating them with respect to the window-size range. The significance level we consider is $\alpha = 0.001$.

	window-size range	$D_{m,n}$	$D_{m,n,\alpha}$	H_0
Direct data	*long*	0.012	0.010	False
	medium	0.016	0.009	False
	short	0.016	0.010	False
Simulated time series	*long*	0.12	0.01	False
	medium	0.071	0.009	False
	short	0.184	0.009	False

Table 3.3: Kolmogorov-Smirnov Test

From the results of the test we can say that the distributions for each window-size range and for both direct estimations and estimations from simulated time series belong to different distributions and do not represent the same observable. As a matter of fact, this means that healthy subjects and patients affected with stroke can be distinguished by analyzing their *dynamic Functional Connectivity Speed*.

[1] It is important to underline that this test does not give any information on which is the distribution of the samples.

Chapter 4

Discussion and Conclusions

In this thesis we explored functional connectivity in large-scale brain activity, with a specific focus on its temporal fluctuations represented by dFC speed. We considered two different samples of fMRI time series corresponding to a group of healthy subjects and a group of patients affected by stroke. For these two kinds of clinical conditions we calculated the respective functional connectivity and speed for each subject. We gathered together the values found for the speed and created some graphs showing the distributions corresponding to each sample and to each considered range of window-size. We followed the same procedure for two other samples corresponding to time series which have been simulated following the MOU model of stochastic differential equation (both for healthy subjects and for patients with illness). In the end we exposed the results we found showing one graph for each sample in order to compare them and detect any difference, which we verified also using the Kolmogorov-Smirnov test.
In the introduction we set out two main purposes for this work, each of them followed by a hypothesis we aimed to test. Firstly, we wanted to see if we could detect some differences in the dFC speed distributions of healthy subjects with respect to the ones belonging to patients affected by stroke, supposing to find slower speed values corresponding to unhealthy subjects. The results we found show that at a first sight there is no evidence of any difference between the two different clinical conditions. Using a statistical test to confront them, we found very weak evidence that the two types of tested data correspond to different distributions.
The other purpose we had for this thesis was to test the effectiveness of the MOU model of stochastic differential equations in describing and simulating the regarded system, the brain at a large scale. Starting from the FC matrices exposed in section 3.1, we can already notice some important differences between the matrices corresponding to empirical time series with respect to the ones corresponding to simulated time series: the first type of matrices reproduces a "checkerboard" scheme which is typical of functional connectivity. On the contrary, for the simulated matrices this type of scheme is completely obscured and in addition to that, the correlation entries tend to have slower values. In section 3.2 we found other strong differences between the distributions of speed obtained by empirical time series and the ones from simulated fMRI data. The first type of sample has "smooth" distributions that at least for the long and medium window-size ranges resemble Gaussian distributions. On the contrary, the simulated dFC speeds result in being completely fragmented and without a clear peak. Moreover, all three distributions corresponding to each window-size range are biased towards higher values of speed. Despite all these dissimilarities, the order of the distributions of the different window-size ranges is maintained, meaning that for both samples the long range corresponds to slower values of speed, the medium range has the middle values and the short range is connected to the highest speed values.
The great differences in the results obtained from simulated data with respect to the ones found from empirical time series can be attributed to the fact that a linear model, as the one we used in this work, may not represent in an efficient way the properties we wanted to test. A model like this one can still give some reliable information, for example regarding the speeds of the different ranges of window-sizes, but cannot be considered a good and complete simulation for the system. If we want a more accurate description of both functional connectivity and its fluctuations, a more complex model

is needed, probably non linear and with more parameters.
Giving some examples on how the results we found could be extended to additional future work, first of all a possibility is to investigate also patients with specific lesions and not just to compare patients and healthy subjects. One could differentiate many types of diseases and list for each of them the respective characteristics in functional connectivity and speed distributions, trying to correlate FC speed with distinct markers of structural lesions. Another possible path, as we said in the previous paragraph, is trying to use more advanced models to simulate large-scale brain activity, in order to test if the results are more accurate in reproducing FC speed distributions.

Bibliography

[1] D. Battaglia, T. Boudou, E. C. Hansen, D. Lombardo, S. Chettouf, A. Daffertshofer, A. R. McIntosh, J. Zimmermann, P. Ritter, and V. Jirsa, *Dynamic functional connectivity between order and randomness and its evolution across the human adult lifespan*, NeuroImage, 222 (2020), p. 117156.

[2] M. Gilson, R. Moreno-Bote, A. Ponce-Alvarez, P. Ritter, and G. Deco, *Estimation of directed effective connectivity from fmri functional connectivity hints at asymmetries of cortical connectome*, PLoS computational biology, 12 (2016), p. e1004762.

[3] M. Gilson, G. Zamora-López, V. Pallarés, M. H. Adhikari, M. Senden, A. T. Campo, D. Mantini, M. Corbetta, G. Deco, and A. Insabato, *Mou-ec: model-based whole-brain effective connectivity to extract biomarkers for brain dynamics from fmri data and study distributed cognition*, BioRxiv, (2019), p. 531830.

[4] M. Guye, G. Bettus, F. Bartolomei, and P. J. Cozzone, *Graph theoretical analysis of structural and functional connectivity mri in normal and pathological brain networks*, Magnetic Resonance Materials in Physics, Biology and Medicine, 23 (2010), pp. 409–421.

[5] K. ITO, *109. stochastic integral*, Proceedings of the Imperial Academy, 20 (1944), pp. 519–524.

[6] D. E. Knuth, *Art of computer programming, volume 2: Seminumerical algorithms*, Addison-Wesley Professional, 2014.

[7] R. Liégeois, T. O. Laumann, A. Z. Snyder, J. Zhou, and B. T. Yeo, *Interpreting temporal fluctuations in resting-state functional connectivity mri*, NeuroImage, 163 (2017), pp. 437–455.

[8] D. Lombardo, C. Cassé-Perrot, J.-P. Ranjeva, A. Le Troter, M. Guye, J. Wirsich, P. Payoux, D. Bartrés-Faz, R. Bordet, J. C. Richardson, O. Felician, V. Jirsa, O. Blin, M. Didic, and D. Battaglia, *Modular slowing of resting-state dynamic functional connectivity as a marker of cognitive dysfunction induced by sleep deprivation*, NeuroImage, 222 (2020), p. 117155.