NeuroImage 127 (2016) 242–256

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI?

R. Hindriks [a,*], M.H. Adhikari [a], Y. Murayama [d], M. Ganzetti [b,c], D. Mantini [b,c], N.K. Logothetis [d], G. Deco [a,e]

[a] Center for Brain and Cognition, Computational Neuroscience Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
[b] Department of Health Sciences and Technology, ETH Zurich, Switzerland
[c] Department of Experimental Psychology, University of Oxford, United Kingdom
[d] Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[e] Instituci Catalana de la Recerca i Estudis Avanats (ICREA), Universitat Pompeu Fabra, Barcelona, Spain

ARTICLE INFO

*Article history:*
Received 16 June 2015
Accepted 23 November 2015
Available online 26 November 2015

*Keywords:*
Resting state
Functional MRI
Dynamic functional connectivity
Surrogate data

ABSTRACT

During the last several years, the focus of research on resting-state functional magnetic resonance imaging (fMRI) has shifted from the analysis of functional connectivity averaged over the duration of scanning sessions to the analysis of changes of functional connectivity within sessions. Although several studies have reported the presence of dynamic functional connectivity (dFC), statistical assessment of the results is not always carried out in a sound way and, in some studies, is even omitted. In this study, we explain why appropriate statistical tests are needed to detect dFC, we describe how they can be carried out and how to assess the performance of dFC measures, and we illustrate the methodology using spontaneous blood-oxygen level-dependent (BOLD) fMRI recordings of macaque monkeys under general anesthesia and in human subjects under resting-state conditions. We mainly focus on sliding-window correlations since these are most widely used in assessing dFC, but also consider a recently proposed non-linear measure. The simulations and methodology, however, are general and can be applied to any measure. The results are twofold. First, through simulations, we show that in typical resting-state sessions of 10 min, it is almost impossible to detect dFC using sliding-window correlations. This prediction is validated by both the macaque and the human data: in none of the individual recording sessions was evidence for dFC found. Second, detection power can be considerably increased by session- or subject-averaging of the measures. In doing so, we found that most of the functional connections are in fact dynamic. With this study, we hope to raise awareness of the statistical pitfalls in the assessment of dFC and how they can be avoided by using appropriate statistical methods.



## Introduction

Resting-state blood-oxygen level-dependent (BOLD) functional magnetic resonance imaging (fMRI) studies have traditionally investigated patterns of functional connectivity (FC) that are static within the scanning period. More recently, attention shifted towards temporal fluctuations in FC within sessions. The latter is referred to as *dynamic functional connectivity* (dFC), as opposed to the former, which is referred to as *static functional connectivity* (sFC). The progress made in the study of dFC has recently been reviewed in Hutchison et al. (2013a). The most common and straightforward way to investigate dFC is using windowed FC, which consists of calculating a given FC measure, for example, the Pearson correlation coefficient or phase-locking factor (Pereda et al., 2005), over consecutive windowed segments of the data. This gives a time series of FC values, which can subsequently be used to assess fluctuations in FC within sessions (Chang and Glover, 2010; Hutchison et al., 2013b; Handwerker et al., 2012; Keilholz et al., 2013; Tagliazucchi et al., 2012; Jones et al., 2012; Allen et al., 2012; Zalesky et al., 2014; Barttfeld et al., 2015). Although such an analysis seems straightforward, there are two pitfalls that have not always been recognized in previous studies.

The first pitfall is to identify an observed value of a test statistic with its true underlying value. This means that the mere presence of fluctuations in an observed FC time series is taken as evidence for the presence of dFC. The pitfall is that of overlooking the fact that the observed FC values are *estimates* of the true (and unobservable) values, and hence, are subject to statistical uncertainty. As an analogue, consider repeated measurements of a physical quantity, say the speed of an approaching car, by using a laser gun. While the car is approaching, multiple

* Corresponding author at: Center for Brain and Cognition, Computational Neuroscience Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Carrer Tanger 122-140, 08018 Barcelona, Spain.
*E-mail address:* rikkert.hindriks@upf.edu (R. Hindriks).

http://dx.doi.org/10.1016/j.neuroimage.2015.11.055

measurements are made, which, due to the imperfections in the instrument and ambient noise, produces a time series of fluctuating values. Although the fluctuations are real, they are due to noise, and do not necessarily reflect fluctuations in the car's speed, which could be constant. In the same way, observed FC values can be viewed as measurements of a quantity, namely, the true (and unobservable) FC. In classical statistical terms, one needs to distinguish between the *sample* FC, which is an *estimator* of the *population* FC.

Thus, to decide whether fluctuations in an observed FC time series are due to statistical uncertainty or reflect true changes in population FC, an appropriate statistical test has to be carried out. This is typically done by calculating a *test statistic* (also called a measure, index, or biomarker) that characterizes the fluctuations in the FC time series and subsequently test if the observed value of the test statistic falls outside the test statistics' null distribution, that is, its distribution *if the correlations would be static*. Several test statistics have been proposed to test for the presence of dFC, including the variance of the FC time series (Sakoglu et al., 2010), test statistics based on the FC time series' Fourier-transform (Handwerker et al., 2012), and non-linear test statistics (Zalesky et al., 2014), among others (Chang and Glover, 2010; Keilholz et al., 2013). Crucially, the null hypothesis under which the distribution of the test statistic is constructed should correspond to the FC being static. This might seem trivial, but the construction of such a distribution is far from trivial and this forms the second pitfall in assessing dFC, which is the use of an inappropriate null-hypothesis.

Since the null distribution cannot be derived mathematically for most dFC measures, it needs to be approximated from the data at hand. Ideally, such *surrogate data* is constructed such that they share all statistical properties with the observed data, except that they lack the property one wants to test for, in this case, dFC (Schreiber and Schmitz, 2000; Pereda et al., 2005). In the literature on dFC, several methods have been proposed to approximate null distributions for dFC. For example, by randomly shuffling the Fourier phases of the BOLD time series (Handwerker et al., 2012; Leonardi et al., 2013) or by randomly selecting BOLD time series from different scanning sessions (Keilholz et al., 2013). The pitfall here is that these two approaches destroy the sFC in the data and hence correspond to a *different* null hypothesis, namely, that of the FC being static *and* equal to zero. Additionally, *a priori* it is unclear how this affects the results of the subsequent statistical testing. A more appropriate way of constructing surrogate data is to fit a time series model to the data and to approximate the null distribution by bootstrapping from the model residuals, as done, for example, in Chang and Glover (2010) and Zalesky et al. (2014). Yet another way, which might be easier to use in practice, is to shuffle the Fourier phases in such a way so that the sFC is preserved (Prichard, 1994). As far as we know, this method has only been applied in Allen et al. (2012). In this study, we focus on the Fourier-based surrogate method.

## Material and methods

### *Statistical assessment of dynamic FC*

Suppose we have recorded resting-state BOLD-fMRI time series from two voxels or regions-of-interest (ROIs) like those displayed in Fig. 1A and we want to decide if the functional connectivity between the two time series is dynamic, that is, if it changes over the duration of the scan. Although the concept of functional connectivity (FC) is wide and includes any kind of statistical relationship between time series (Pereda et al., 2005; Friston, 2011), we focus on the (Pearson) correlation coefficient, which is the most widely used FC measure in resting-state fMRI research (Sakoglu et al., 2010; Chang and Glover, 2010; Hutchison et al., 2013b; Handwerker et al., 2012; Keilholz et al., 2013; Tagliazucchi et al., 2012; Jones et al., 2012; Thompson et al., 2013; Zalesky et al., 2014). The most straightforward way to proceed is to calculate correlation coefficients on overlapping segments of the time series. This results in a time series of correlation values as shown in Fig. 1B. Note that the windowed correlations have different values for different windows. In particular, we observe both negative and positive correlations, the latter are referred to as "hypersynchrony states" in Hutchison et al. (2013b). Although in some studies, the observed fluctuations in FC are taken as evidence for the presence of dynamic FC (dFC), most studies agree that a statistical test is needed to draw this conclusion. Indeed, an appropriate statistical test for dFC answers the question if the observed fluctuations in the correlation time series can be distinguished from those that would be observed *if the correlation were static*, that is, independent of time.

One way to answer this question is to construct confidence intervals around the values in the correlation time series, as done, for example, in Kang et al. (2011) and Hutchison et al. (2013b). If the data is a white-noise Gaussian process, the confidence intervals can even be calculated analytically. Otherwise, they can be approximated by resampling of the windowed time series, a technique referred to as bootstrapping. The confidence intervals in Fig. 1B (dotted lines) were obtained by such a bootstrap procedure. More specifically, for each window, we selected (with replacement) unpaired sample-pairs to build a bootstrapped

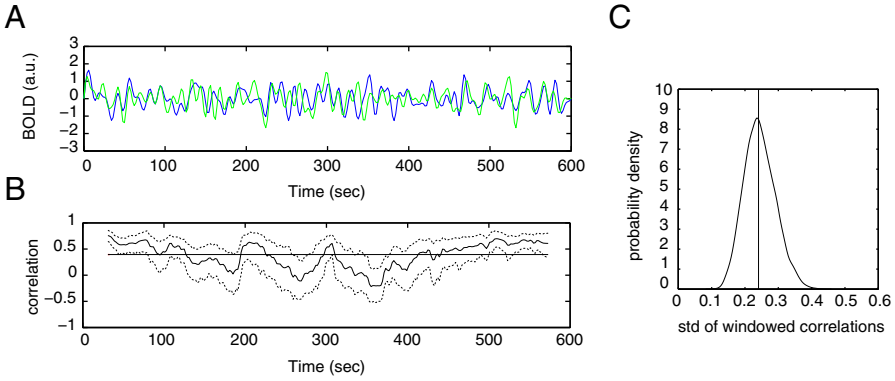


**Fig. 1.** Statistical testing for dynamic FC. A, Simulation of two simultaneously recorded fMRI time series from two different voxels or ROIs. B, Time series of Pearson correlations obtained by calculating the correlation coefficients on successive 60 s segments of the fMRI time series (maximal overlap). The correlation values are plotted as a function of the window-centers. The dotted lines denote the 95% confidence intervals of the correlation values obtained by repeatedly permuting the windowed fMRI time series. The horizontal line denotes the average correlation between the fMRI time series. C, Probability density of the standard deviation of the correlation time series under the null hypothesis. The observed value was 0.24 and is marked by the vertical line.