UNIVERSITY OF SOUTHAMPTON

EEG connectivity measures and their application to assess the depth of anaesthesia and sleep

by

Giulia Lioi

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Engineering and the Environment

Institute of Sound and Vibration Research

January 2018

Declaration of Authorship

I, Giulia Lioi, declare that this thesis titled, 'EEG connectivity measures and their application to assess the depth of anaesthesia and sleep' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

- Parts of this work have been published as:

  Conference Papers

  Lioi G, Bell SL, Smith DC, Simpson DM. *Characterization of functional brain connectivity networks in slow wave sleep.* PGBiomed 2015, Liverpool, Germany.

  Lioi G, Bell S L and Simpson D M 2016. *Changes in Functional Brain Connectivity in the Transition from Wakefulness to Sleep in different EEG bands.* MEDICON 2016 Paphos, Cyprus, 2016. Ed E Kyriacou et al (Berlin: Springer) IFMBE Proceedings, 57,pp 38.

  Lioi G, Bell SL, Smith DC, Simpson DM. *Changes in EEG directional connectivity during a slow induction of propofol anesthesia.* EMBEC and NBC 2017 (IFMBE Conference),Tampere, Finland.(**Abstract**)

  Journal Papers

  Lioi G, Bell SL, Smith DC, Simpson DM. *Directional Connectivity in the EEG is able to discriminate wakefulness from NREM sleep.* Physiol. Meas. 38 (2017) 18021820

  Lioi G, Bell SL, Smith DC, Simpson DM. *Changes in EEG directional connectivity during slow induction of anaesthesia.* British Journal of Anaesthesia (**In Preparation**)

Signed:

Date:

"And the LORD God caused a deep sleep to fall upon Adam, and he slept; and He took one of his ribs, and closed up the flesh in its place"

Genesis (2:21)

"I like the scientific spirit-the holding off, the being sure but not too sure, the willingness to surrender ideas when the evidence is against them: this is ultimately fine-it always keeps the way beyond open-always gives life, thought, affection, the whole man, a chance to try over again after a mistake-after a wrong guess"

Walt Whitman

Abstract

General anaesthesia has been used for more than two centuries to guarantee unconsciousness, analgesia and immobility during surgery, yet our ability to evaluate the level of anaesthesia of the patient remains insufficient. This contributes on one hand to occasional episodes of intraoperative awareness and recall and on the other to 'controlled' drug over-dosage that increases hospital costs and patients recovery times. At present parameters used in clinical practice to monitor anaesthesia are indirect measures of the state of the brain, which is the target organ of anaesthetics. The lack of a reliable monitor of anaesthetic depth has led to considerable effort to develop new monitoring methods based on electrophysiological measurements. This progress has produced a series of depth of anaesthesia monitors based on various features of the electroencephalogram (EEG) signal. Even though these indexes are practically useful, their theoretical and physiological validity is poorly evidenced and they suffer from some practical limitations. As a result, their clinical uptake has been quite low. In recent years increasing attention has been given to brain connectivity as a powerful tool to investigate the complex behaviour of the brain. Theoretical and experimental findings have identified the disruption of brain connectivity as a crucial mechanism of anaesthetic-induced loss of consciousness. In this work a novel index of anaesthetic depth based on brain connectivity estimated from non-invasive scalp recordings (EEG) is proposed. Firstly, robust estimators of directed connectivity were identified in the framework of multivariate autoregressive (MVAR) models. With a series of simulation studies the performances of these methods in estimating causal connections were assessed in particular with respect to the deleterious effects of instantaneous connectivity due to volume conduction. Recently published solutions were also tested (and rejected). From a comparison of connectivity measurements in simulations, MVAR based estimators were most robust to the effects of volume conduction than conventional coherence measurements. Next the performances of directed connectivity estimators were tested in two experimental studies on NREM sleep and on anaesthesia. Features that exhibited the most robust changes with the individual level of consciousness were identified and their performances in discriminating wakefulness from anaesthesia tested on ten patients undergoing a slow induction of propofol anaesthesia. The performance of the proposed method were also compared with established depth of anaesthesia indexes such as Bispectral Index (BIS) or Auditory Evoked Potentials (AEP). Results suggest that EEG connectivity features are sensitive to the anaesthetic induced changes and that they have the potential to be integrated in future monitors of intra-operative awareness and anaesthetic adequacy.

Acknowledgements

Firstly I would like to express my sincere gratitude to my supervisors Steven Bell, David Simpson and David Smith. I thank them in the first place for giving me the opportunity to work in their group and for the constant support and positive presence during these years. They have always encouraged my intellectual freedom, supported me in attending meetings and conferences and stimulated the exchange of ideas. I deeply thank them for their wisdom, generosity and humanity: they have made my PhD a cheerful journey and taught me to be a better scientist and person.

I also would like to thank the members of my internal review panel Professor Paul White and Dr. Thomas Blumensath for their useful comments to my work.

My days in the Signal Processing and Control Group have been joyful also thanks to my fellow labmates: Luigi, Dario, Michele Iodice, Michele Zilletti thanks for all the laughs!

I am very grateful to Alessandro Beda, Nadjia Cristinne Carvalho, Hyorrana and Cristiano to have warmly welcomed me in UFMG and helped me with the experiments: their support and kindness have been very precious for me and never to be forgotten.

My deepest gratitude full of good memories and a little of nostalgia goes to the Shaftesbury 65 crew: Damien, Domenico, Tual, Alessandra, Laura, Jana, Armand, Tobias, Xander and all the persons gravitating towards this unusual collection of people that have been my family here in Southampton and made these years unforgettable.

Finally, I would like to thank my parents and my brother that have always believed in me and supported me with their unlimited love, regardless of the distances that stand between us.

Contents

List of Figures

List of Tables

Abbreviations

ABR	Auditory Brain-stem Response
AEP	Auditory Evoked Potential
CCD	Cortical Current Density (source model)
COH	Coherence
BIS	Bispectral Index
BISP	Bispectrum
DAG	Directed Acyclic Graph
DoA	Depth of Anaesthesia
EEG	Electroencephalogram
ESC	Effect site concentration
ECoG	Electrocorticogram
eMVAR	Extended Multivariate Autoregressive (Model)
FC	Functional Connectivity
$\boldsymbol{F_{sp}}$	F-value at a single point
FPE	Final Prediction Error (Criterion)
fMRI	Functional Magnetic Resonance
GA	General Anaesthesia
HSD	Honest Significance Test (Tukey)
IFT	Isolated Forearm Technique
LBP	Ljung-Box Portmanteau Test
LFM	Lead Field Matrix
LOC	Loss of Consciousness
LS	Least Square
MLS	Maximum Length Sequence
MVAR	Multivariate Autoregressive (Model)

PAM	Post-Auricular Muscle
SBC	Schwarz's Bayesian Criterion
scMVAR	Strictly Causal Multivariate Autoregressive (Model)
SNR	Signal to Noise Ration
SL	Synchronization Likelihood
SWS	Slow-Wave Sleep
PSD	Power Spectral Density
TMS	Transcranial Magnetic Stimulation
TP	Triple Product

Symbols

$\mathbf{R}$	correlation matrix, with generic element $r_{i,j}$
ρ	correlation coefficient
$\mathbf{S}$	spectral densitiy matrix, with generic element $S_{i,j}$
$\mathbf{G}$	inverse spectral matrix, with generic element $g_{i,j}$
Coh	coherence
$PCoh$	partial coherence
A	MVAR model coefficient matrix
ε	vector of innovations (residuals)
$\mathbf{\Sigma_\varepsilon}$	residual covariance matrix, with generic element $\sigma_{i,j}$
$\mathbf{H}$	MVAR model transfer matrix, with generic element $H_{i,j}$
DC	directed coherence
PDC	partial directed coherence (generalized)
DTF	directed transfer function
$orPDC$	partial directed coherence (ordinary)
f_{samp}	sampling frequency [Hz]
p_{opt}	MVAR model optimal order
vec	column stacking operator

Chapter 1

Introduction

More than 250 years have passed since the first successful demonstration of anaesthesia (Mashour 2006) and anaesthetics are at the present day used during surgery in millions of patients per year. However monitoring the anaesthetic effects still represents a challenge and our ability to evaluate the level of the awareness of the patient remains insufficient (Lau et al. 2006). This is due to the intrinsic complexity of anaesthesia, which is a multicomponent process, where different agents, with various neural actions, target different conditions. An adequate level of anaesthesia prevents awareness and recall of the surgical procedure while minimizing the untoward effect of excessively deep sedation. Awareness of the intra-operative events is a relatively rare (0.1 to 0.2% (Sandin et al. 2000, Shepherd et al. 2013)) but a serious complication of general anaesthesia, associated with anxiety, depression and a high rate of post-traumatic stress disorders (Bruchas et al. 2011, Shepherd et al. 2013). The incidence of unexpected awareness may be much higher in procedures where anaesthesia is kept to a minimum like caesarean section and cardiac surgery (Kaul et al. 2002) and it is a significant concern for patients admitted for surgical procedures (McCleane & Cooper 1990). Moreover the use of neuromuscular blocking drugs makes the detection of intra-operative awareness more difficult since it removes both respiration and movement in response to a noxious stimulus, which are valuable indicators of anaesthetic inadequacy. To avoid this, anaesthetists are used to administer a 'controlled overdose' of anaesthetics that increases patient recovery times, hospital costs and the risk of postoperative morbidities (Weiskopf 2000).

The reasons for the lack of a reliable monitor that permits a conclusive statement about anaesthetic depth in individual patients are of different kind. Firstly, mechanisms of how anaesthetics suppress consciousness are still unclear. Secondly, at present the parameters used in the clinical practice to monitor anaesthesia (blood pressure, tear formation and sweating, together with presumed drug pharmacokinetis) are not considered reliable indicators (Lau et al. 2006) primarily because they are indirect measures of the state

of the brain, which is the target organ of anaesthetic drugs. The necessity for a more reliable index of consciousness, able to correspond with neurobiological processes, has increased the interest of both clinicians and researchers in exploring alternative methods to monitor anaesthetic depth. This has resulted in a series of commercially available depth of anaesthesia (DoA) monitors. They are typically based on a series of electroencephalogram (EEG) features that enter a proprietary algorithm whose output is a dimensionless index, usually ranging from 0 (cortical silence) to 100 (full wakefulness). The most widely used is the Bispectral Index (BIS®, Aspect Medical System, USA), whose computation is based on time features (i.e. burst suppression) and frequency domain parameters (i.e. beta activity, bispectrum) of EEG recordings from 2 or more electrodes on the forehead.

Studies comparing the performances of these parameters have reported contrasting results (Marchant et al. 2014) and a technology assessment report commissioned by National Institute of Health Research has concluded that *the available evidence on the impact of these technologies on reducing the likelihood of intraoperative awareness is limited.* (Shepherd et al. 2012). This is probably one reason why the clinical uptake of DoA monitors at the present day is quite low (only 2% of anaesthetists use them routinely (Pandit & Cook 2014)) despite the fact that the use of DoA monitoring is strongly recommended in total intravenous anaesthesia with neuromuscular blockade in the UK anaesthetic standards guidelines (Association of Anaesthetists of Great Britain and Ireland 2016). Moreover the DoA monitors that are commercially available are affected by some limitations and reliability issues and their interpretability as indicators of patient awareness is debated (Pandit & Cook 2014). There is therefore a need for improvement of current DoA monitors, in particular with regard to their ability to detect unexpected awareness (Marchant et al. 2014).

Theories of how anaesthetics suppress consciousness (Alkire et al. 2008, Mashour 2006, John & Prichep 2005) have identified brain functional disintegration as a crucial mechanism: anaesthetics are thought to disrupt the communication of brain regions that would otherwise be coupled during wakefulness. This hypothesis is supported by a series of experimental findings that show impaired cortico-cortical and thalamo-cortical connectivity in general anaesthesia (Ferrarelli et al. 2010, Gómez et al. 2013, Boveroux et al. 2010, Lee et al. 2009) and in other states of suppressed consciousness as NREM sleep (Massimini et al. 2005*a*) and vegetative state (King et al. 2013, Sitt et al. 2014, Boly et al. 2012). In the search for a more reliable index of consciousness, it may be therefore promising to concentrate on measures that more closely capture the neurophysiology of the brain. While many neuroimaging studies have focussed on aspect such as regional baseline metabolism (Alkire et al. 1995) or response to external stimuli (Bell et al. 2006, Thornton et al. 1992) changes during anaesthesia, in recent years an increasing relevance

has been attributed to measures of brain connectivity, as they can grasp the global organized behaviour of neural circuits, beyond the simple mapping of their localized activity (Hudetz 2012, David et al. 2004). Among the various neuroimaging techniques, non-invasive scalp measures (EEG) of brain connectivity seem promising because of their relatively cheap and easy to apply technology (Sitt et al. 2014).

1.1 Aims and Hypothesis of the Research

The aim of this research project is to develop and test a novel indicator of anaesthetic depth based on EEG brain connectivity measures. Based on literature review findings presented in chapters 2 and 3, the research aims can be summarized as follow:

1. From the various estimators proposed to assess brain connectivity, identify those that are more robust and, at the same time, compatible with a possible online application for anaesthesia monitoring (chapters 4 and 5).
2. Characterize the topological features of EEG connectivity patterns during loss of consciousness (i.e. in deep sleep and anaesthesia) as compared to wakefulness (chapters 5 and 6)
3. Identify, among EEG connectivity estimators and features, those that exhibit the most robust changes with loss of consciousness (LOC) and therefore are able to reliably discriminate wakefulness from anaesthesia (chapter 6). From this develop a single index of awareness.
4. Compare the performances of the selected connectivity features with respect to more established DoA methods (chapter 6).
5. Assess the suitability of the proposed index to monitor anaesthesia in individual subjects and in a clinical environment (chapter 6).

As theoretical and experimental findings suggest that brain connectivity is critically related to anaesthetic-induced unconsciousness, we hypothesize that connectivity measures will more efficiently capture changes in the level of consciousness of the subject than measures based on the local neuronal activity. Therefore, brain connectivity features can be expected to have better performances than classical DoA monitors in discriminating wakefulness from anaesthesia. We also speculate that this methodology has the potential to be implemented for routine anaesthetic monitoring, given the non-invasive,

cheap and applicable at bed-side nature of EEG technology.

This work contributes to the development of a new signal processing tool that could be incorporated in future DoA monitors to improve their performances in the detection of intraoperative awareness or adequacy of anaesthesia. Anaesthesia represents also a powerful tool to 'probe' the unconscious brain in the search for neural correlates of consciousness and many scientists argue for an integrated approach to the study of consciousness and anaesthesia (Mashour 2006, Beecher 1947). The results of this study may therefore be of interest not only for clinicians but also for neuroscientist as they are likely to shed new light on mechanisms of anaesthetic actions and on the patterns of reorganization of connectivity networks with loss of consciousness.

1.2 Thesis Overview

In this work we[1] assessed brain connectivity from scalp EEG using spectral estimators derived from the multivariate autoregressive (MVAR) modelling of EEG data. We investigated the rationale behind the development of these measures and carried out an analysis of their performances by means of different simulation studies. Due to the remarkable behavioural and electrophysiological homologies between sleep and anaesthesia we first tested the connectivity estimators in a normative study on sleep, where parameters likely to indicate awareness were identified. We ultimately investigated the performances of the selected connectivity features in an anaesthetic study, in comparison with alternative DoA indexes such as BIS and Middle Latency Response (MLR).

In chapter 2 general mechanisms of anaesthetic action and their effect on electrophysiologic measures are reviewed, with a focus on the latest experimental findings on brain connectivity changes during anaesthesia. A concise review of the different DoA monitors and approaches to EEG analysis of current use in the anaesthesia field is also presented.

Chapter 3 gives an overview of the methods used to estimate linear coupling and causality between time series with a particular focus on spectral measures of causality developed in the framework of MVAR models. A description of the refinement and properties of these estimators will be performed as well by means of an illustrative example. Crucial aspects for accurate model identification will be discussed and statistical tools useful to

[1] In this manuscript the pronoun 'We' is used to indicate the work carried out by the author, unless otherwise stated.

assess the validity of the model described in the last section of the chapter.

A major issue in the estimation of causality among EEG derivations is the adverse effect of instantaneous correlations among time series, which is, in turn, related to the problem of volume conduction. The performances of different methodologies with respect to this issue is investigated in different simulation studies in chapter 4.

In chapter 5, the spectral connectivity estimators described and tested in previous chapters are applied to the analysis of sleep EEG. The sleep study was designed as a controlled experiment in a laboratory setting in order to explore the connectivity changes in different states of arousal, with a view of assessing anaesthetic depth in the last stage of the research work. In chapter 5 the rationale, experimental design and methods of the sleep study is outlined and results of the functional connectivity analysis discussed. The journal paper describing results from the sleep study (Lioi et al. 2017) is reported in Appendix.

Chapter 6 presents results from the anaesthetic study in a clinical setting that included ten patients scheduled for surgery undergoing slow induction of propofol anaesthesia. The changes in topological features of EEG connectivity associated with different anaesthetic depths will be described both at cohort and individual level and their physiological interpretation discussed. We also compared the performances of the connectivity features with more established DoA indicators (BIS and MLR) in discriminating 'awake' from 'anaesthesia' at different levels using both a linear and a non-linear classification approach. Results are outlined in the last section of chapter 6 while a draft version of the paper on the anesthetic study is reported in Appendix.

Finally, chapter 7 draws overall conclusions and discuss future work.

1.3 Original Contributions

- **Testing of the extended MVAR approach (proposed to account for instantaneous connectivity) on simulated and recorded EEG.** In EEG signals, instantaneous connectivity is expected due to strong volume conduction effects and this has the potential to confound the analysis of causal connectivity estimated with the classical MVAR approach. The extended MVAR (eMVAR)framework has been introduced to deal with this issue and has been tested

in previous works on simple connectivity models where it was shown to effectively eliminate spurious causality introduced by zero-lag effects. In this work we tested the eMVAR approach on simulations that more realistically model EEG signals and on EEG recordings. Our results showed that in this case the eMVAR algorithm fails to reliably estimate instantaneous effects. Our findings suggest that the characterization of instantaneous causality between EEG time-series remains a challenging issue and that caution is required when interpreting results from the estimation of EEG connectivity.

- **Characterization of instantaneous volume conduction effects on scalp connectivity by means of a realistic head model.** The effects of volume conduction on the estimation of scalp connectivity with MVAR estimators are not clear. By means of a series of simulation studies using a realistic head model we characterized volume-conducted, non-physiological connectivity as a function of interelectrode distance and reference choice: our preliminary results show that MVAR estimators considerably reduce the effects of volume conduction if compared with classical Coherence and Partial Coherence, in particular for interelectrode distances larger than 10 cm.

- **Correlation of directed connectivity performance with the individual level of consciousness during (Non Rapid Eye Movement) NREM sleep.** We have assessed the performance of different EEG indexes (EEG directed connectivity and normalized spectra in different frequency bands) in relation to sleep stages in a NREM sleep study. Our results show that among all of the EEG measures tested, a proposed index of the direction of connectivity on the rostro-caudal axis performed well at a group level and gave the highest correlation with individuals sleep stage and hence level of consciousness.

- **Assessment of EEG directed connectivity changes during a slow induction of propofol anaesthesia**. We assessed changes in multivariate EEG connectivity during a target-controlled slow induction of propofol anaesthesia with a view to proposing a connectivity-based measure of depth of anaesthesia. We observed an inversion of directed connectivity from posterio-frontal in wakefulness to fronto-posterior in anaesthesia. We have identified a step change of connectivity features with the onset of anaesthesia (in contrast with a more gradual trend with increasing propofol effect site concentration-ESC observed in BIS and MLR) that is broadly consistent at individual level and is relevant in terms of a physiological interpretation of anaesthetic-induced LOC.

- **Proposal of a novel and promising index of anaesthetic depth based on EEG directed connectivity features and assessment of its performances**

in comparison with BIS and MLR. We proposed a novel DoA index based on EEG connectivity features. We assessed the novel index performance in discriminating wakefulness from anaesthesia in a clinical setting and compared it with MLR and BIS features extracted at stable anaesthetic brain concentrations. We showed that directed connectivity features have the best performances in discriminating wakefulness from anaesthesia, as compared with MLR and BIS, with an average accuracy of 96% and results robust across subjects. Our results indicate the potential for directed connectivity to be integrated into future DoA monitors (possibly in combination with other EEG features) to improve the detection of intraoperative awareness.

Chapter 2

Mechanisms and Electrophysiological signatures of Anaesthesia

In this chapter general principles of anaesthetic mechanisms will be discussed, together with theoretical proposals and some experimental findings describing brain activity changes associated with anaesthetic induced unconsciousness. An overview of the principal commercially available monitoring techniques based on electrophysiological recordings is also given in this chapter. The methods developed to date fall into two classes: Pattern recognition analysis of spontaneous cerebral activity, and physiological responses evoked by sensory stimulation. A novel approach for depth of anaesthesia monitoring is to measure changes in brain connectivity. This approach aligns with the hypothesis that anaesthesia alters information processing patterns in the brain from long-range complex connections to localised, simple, stereotyped activity. Relevant theoretical and experimental findings describing brain connectivity changes associated with anaesthetic-induced LOC are reviewed in the last section of this chapter. Due to the complexity and the extent of the topic, only a brief overview is given, since a comprehensive dissertation is beyond the scope of this work.

2.1 Component and Mechanisms of Anaesthesia

Anaesthetic drugs comprise a wide range of molecules that usually fall in two classes: Intravenous agents (i.e. Propofol, Ketamine, etc.) and inhalational anaesthetics (i.e. Nitrous Oxide, Sevoflurane, etc.). Usually a cocktail of drugs is used to achieve a series

of clinical endpoints: Unconsciousness (hypnosis), amnesia, analgesia (nociceptive blockade) and immobility, together with stable cardiovascular and respiratory conditions.
Anaesthetics act by altering the neurotrasmission at multiple sites of the brain (Brown et al. 2010). At a molecular level anaesthetics are thought to interact with neuron's ion channels and their regulatory mechanisms and in particular to hyperpolarize neurons by enhancing inhibitory (gamma aminobutyric acid-GABA) receptors and therefore inhibiting excitatory (N-metyl D-aspartate-NMDA) receptors (John & Prichep 2005). Because a small number of inhibitory interneurons regulates the activity of large populations of excitatory pyramidal neurons anaesthesia can effectively inactivate broad areas of the brain. As a consequence of these cellular actions general anaesthesia leads ultimately to a synchronized firing pattern of action potentials consisting of a hyperpolarized silent state and a depolarized firing state that is reduced to short bursts of action potentials as the drugs doses increase.

2.1.1 Clinical Signs and EEG patterns of General Anaesthesia and their relation to Sleep

The molecular and cellular modifications induced by anaesthetic administration result in a series of macroscopic EEG patterns (associated with specific clinical signs) that vary in relation to the anaesthetic phase (Brown et al. 2010). During induction the patient enters a state of paradoxical excitation (euphoria, delirious speech) characterized by an increase in EEG beta (β) activity (13-25 Hz) that resembles the 'β arousal' observed in REM sleep (also known as 'paradoxical sleep')(Sanders et al. 2012) (figure 2.1). During general anaesthesia maintenance period the EEG patterns undergo different phases: in light anaesthesia β activity decreases and alpha (α, 8-13 Hz) and delta ($\delta < 4$ Hz) power increases. The intermediate state (phase 2) shows a prevalence of frontal δ and α activity ('anteriorization') that is similarly found in NREM sleep. The α and δ rhythms are thought to arise from the pyramidal neurons hyperpolarization due to the physiological (sleep) or drug induced (anaesthesia) reduction of excitatory inputs (Brown et al. 2010). This phase of anaesthesia, where surgery is usually performed, corresponds to Slow-Wave Sleep (SWS) where pain perception is significantly reduced and a strong stimulation is required to produce arousal. As anaesthesia deepens the EEG alternates periods of inactivity to burst of α and β waves known as 'burst suppression' (figure 2.1). Similarly a switch of the thalamus from its tonic to bursting mode is also observed in Slow-wave sleep. Bursts of activity become more rare as anaesthesia deepens.
Together with behavioural (hypnosis, amnesia, immobility) and electrophysiological traits, sleep and anesthesia share specific neurophysiologic mechanisms: A series of sleep-wake nuclei in the brainstem and thalamus are involved in anaesthetic mechanisms

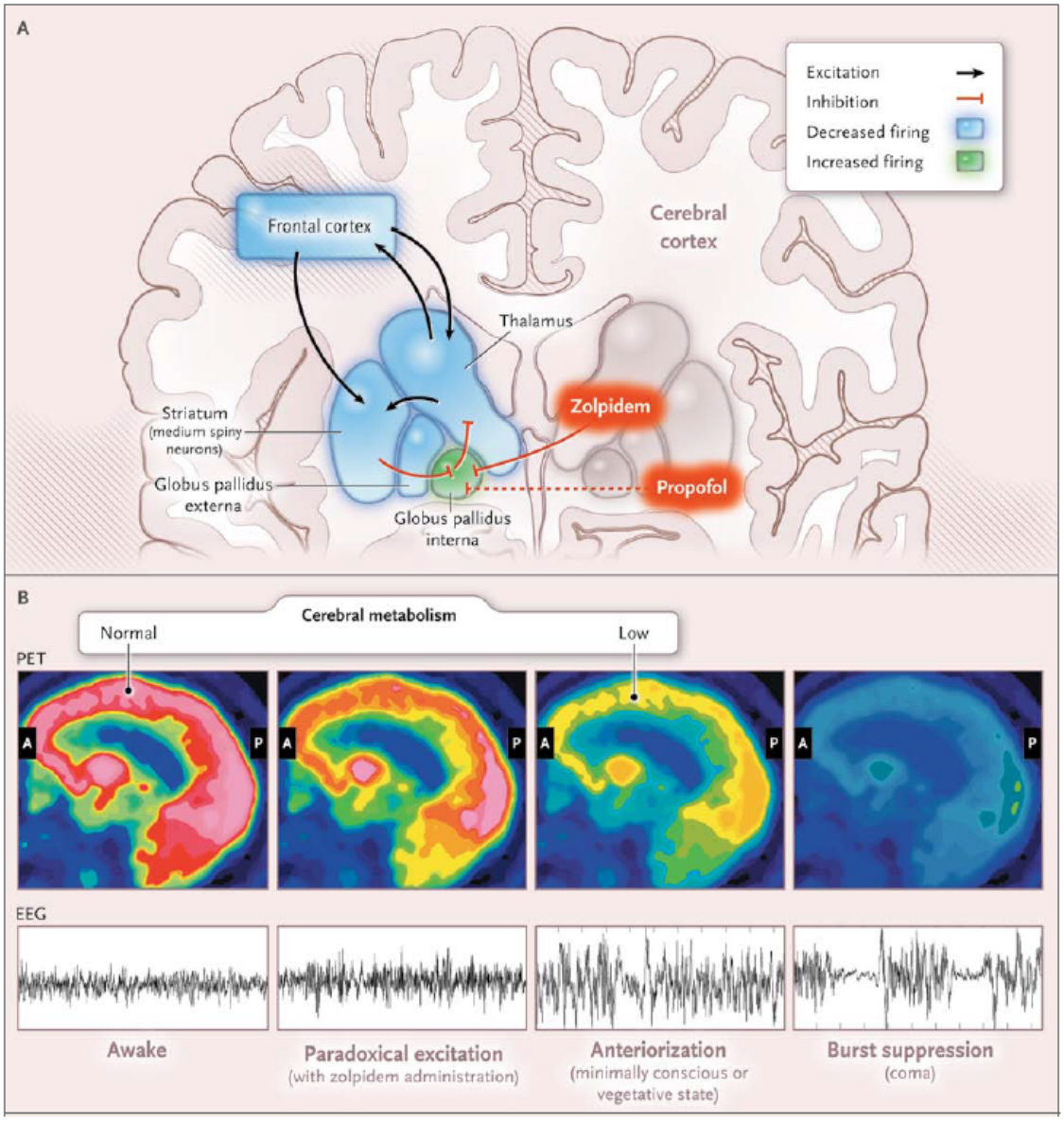

FIGURE 2.1: Mechanisms of anaesthetic actions and EEG patterns in anaesthesia and vegetative states. Anaesthetic drug such as propofol have an inhibitory action on the thalamus and decrease cortical activity. This correspond to a reduction in cerebral activity, as measured by positron emission tomography (PET) and to dramatic changes in EEG patterns. Example of EEG traces in Paradoxical excitation in anaesthesia and Anteriorization and Burst Suppression in vegetative state and coma are given in the bottom plots. General anaesthesia is characterized by similar EEG patterns. From Brown et al. (2010)

(Mashour 2011, 2010). This suggests that, despite the important differences characterizing these two states (Bonhomme et al. 2011), investigating the relationship between sleep and anesthesia could provide important insight in the mechanisms of anaesthetic action (Brown et al. 2010, Mashour 2010).

2.1.2 Models of anaesthetic action

Anaesthetic induced loss of consciousness (LOC) does not arise from a general depression of neuronal activity but from the modulation of specific brain areas activity. The thalamus has been identified as an important target of anaesthetics due to the consistent reduction of thalamic metabolism and blood flow observed in PET studies (Alkire et al. 2000, 1995): although the anaesthetic state was found to be associated with a global decrease in brain metabolism, this global deactivation was not uniform but involved primarily the thalamus. These observations lead to the hypothesis that the thalamus serves as a switch for consciousness. There is supporting evidence in the literature that anaesthetic (and slow-wave sleep) induced LOC is related to the switch of thalamic cells from a tonic-firing pattern, characteristic of vigilance, to a burst-firing behaviour (Llinás & Steriade 2006). This switch in the thalamic neurons occurs coincident with a change in EEG patterns from low voltage high frequency 'Activated EEG' to slow waves activity and it is caused by an hyperpolarization of thalamic cells that block the transmission of sensory input to the cortex (Alkire et al. 2000). Whether this thalamic switch is a direct effect of anaesthetics or it is mediated by cortical activity is an open question (Alkire et al. 2008): it is hypothesized that primary effects of anaesthesia occur at the cortical level (with higher order cortical areas more sensitive than the lower order ones) and then secondary effects occur in subcortical areas (Velly et al. 2007), however the exact sequence of events remains to be defined (Marchant et al. 2014). These investigations suggest that thalamus may be a 'marker' of cortical activity, i.e. its deactivation could be an indirect outcome of the cortical effects of anaesthetics (Alkire et al. 2008).

Current theories of general anaesthesia have therefore attempted to describe how anaesthetics induce LOC by focussing on cortical cognitive processes. The 'unified narcosis theory' proposed by Alkire (Alkire et al. 2000, 2008) suggest that anaesthetics induce unconsciousness by disrupting thalamo-cortical circuits. This hypothesis is coherent with the 'cognitive unbinding' theory proposed by Mashour (Mashour 2004): This framework identifies the mechanism that mediates the unity of conscious experiences ('cognitive binding') as central to sustain awareness and speculates that anaesthetics act by impairing cognitive binding processes at different levels of brain architecture. At global level synthetic processes may be interrupted by the uncoupling of brain structures that would otherwise be synchronized in wakefulness (Mashour 2006). John and colleagues (John & Prichep 2005) have attempted to describe the successive steps by which anaesthetics induce unconsciousness ('anaesthetic cascade'): while the first actions take place at molecular level the final step of the 'cascade' is a functional cortico-cortical uncoupling that ultimately results in suppressed awareness. All these theoretical frameworks suggest that the ultimate effect of anaesthetic is a disruption of cognitive processing that is

INDEX	MANUFACTURER	PARAMETERS	VALUES
BIS Bispectral Index	Aspect Medical Systems, USA	Beta Activity SynchFastSlow Burst Suppression	0-100
RE-SE Response Entropy, State Entropy	GE Heathcare, Finland	Power Spectrum Shannon Entropy	RE 0-100 SE 0-91
PSI Patient State Index	SEDline, CA	Power Spectrum Inter-hemispheric Coherence Power Asymmetries	0-100
CSI Cerebral State Index	Danmeter-Goalwick, Denmark	Alpha Beta Ratio Alpha Beta Difference Burst Suppression	0-100
Narcotrend	Arbeitsgruppe Informatik/Biometrie, Germany	Power Spectrum Entropy Autoregressive	0-100 Letters
WAVcns Wavelet-based Anesthetic Value	NeuroWave Systems Inc, OH	Wavelet Coefficients	0-100
aepEX Auditory Evoked Potentials Index	Medical Device Management Ltd, UK	MLR latency and amplitude	0-99

TABLE 2.1: Principal commercially available monitors of the depth of anesthesia, with the EEG features entering their algorithm, and the indexes reference values (adapted from Marchant et al. (2014))

associated with thalamo-cortical and cortico-cortical uncoupling. As we will describe in section 2.3 these hypotheses are supported by a wide scope of experimental studies investigating brain connectivity in states where consciousness is diminished or suppressed and suggest that, in the search for a more reliable monitor of DoA, it may be promising to concentrate on measures that more closely capture the neurophysiology of the brain. In the next section we will describe the commercially available DoA monitors extracted from the EEG signal and discuss some of their limitations.

2.2 Electrophysiological Measures for Anaesthesia Monitoring

The interest of clinicians and researchers in exploring alternative methods to monitor anaesthetic depth has resulted in a series of commercially available depth of anaesthesia (DoA) monitors. These indicators are typically extracted from a wide range of EEG features that enter a proprietary algorithm whose output is a dimensionless index usually ranging from 0 (cortical silence) to 100 (subject awake and orientated). At present at least seven market solutions are available for routine DoA monitoring: we have summarized the main features of the different indexes in Table 2.1.

The most widely used is the Bispectral Index (BIS®, Aspect Medical System, USA), whose computation is based on time features and frequency domain parameters of EEG recordings from two or more electrodes on the forehead.
Another classical approach, which has resulted in the commercial device aepEX (Medical Device Management Ltd, Essex, UK), is based on auditory evoked potentials (brain

responses elicited by the presentation of an auditory stimulus), and in particular on the middle latency response (MLR) features. Other approaches are grounded on the computation of one channel EEG entropy (Bruhn et al. 2000, Olofsen et al. 2008). The rationale behind entropy indexes is found on the reduced 'disorder' and increased predictability of the EEG signal in anaesthesia (Marchant et al. 2014). Other indexes (i.e. the Patient State Index (Prichep et al. 2004)) result from a combination of 'mixed' EEG features as power in specific bands, burst-suppression, spectral asymmetries, etc. Several studies have investigated the performances of different DoA indexes in predicting the hypnotic effect (Bell et al. 2006, Schneider et al. 2003, Gajraj et al. 1999, Loveman et al. 2001), the anaesthetics concentration (Irwin et al. 2002, Thornton et al. 1992) or the response to surgical stimulation (Schneider et al. 2002, Myles et al. 2004): Results are however not conclusive. So far no superiority of one type of index over others has been established (Marchant et al. 2014) and the correlation between EEG indexes and anaesthetic effects depends on the drug and the induction protocol used (Olejarczyk et al. 2017). In this section we will focus on the description of the BIS and the MLR indexes primarily because they were historically developed first and they represent the most studied and tested DoA indexes; secondly because we have compared their performances with the proposed index based on brain connectivity measures in the last stage of this work (chapter 6).

2.2.1 Bispectral Index

The BIS index monitor (Aspect Medical Systems, Norwood, MA, US) is based on spectral analysis of scalp recordings from two or more electrodes (in addition to the ground electrode) on the forehead and is one of the most used monitors of anaesthesia in clinical practice. It was demonstrated to reduce the incidence of intraoperative awareness (Myles et al. 2004) and to predict movement in response to skin incision (indicative of inadequate anaesthetic depth) in patients anaesthetized but not paralysed (Kearse et al. 1994). The BIS combines, using a proprietary algorithm (Chamoun et al. 1995), several variables extracted from the EEG in a multivariate index scaled to a range between 0 and 100. The variables used in the computation of the BIS index are both time domain features of the EEG traces (burst-suppression ratio, BSR) and frequency domain measures (power spectrum, bispectrum and higher order spectral indexes). The power spectrum is used for the computation of the beta ratio parameter (β_R), while the bispectrum is used in the computation of the SynchFastSlow ($SynFS$). The proprietary algorithm combines the sub-parameters with clinically predetermined coefficients extracted from a model based on a database of EEG recordings matched to corresponding hypnotic levels. The weights are assigned using a non-linear function and they vary

depending on the anaesthetic stage: the β_R is dominant in light sedation, the $SynFS$ during EEG activation and surgical level of hypnosis while the BSR detects deep anaesthesia (Rampil 1998). While the computation of the spectrum measures the dramatic changes in EEG power distribution as anaesthesia deepens, bispectral analysis quantifies quadratic phase coupling between different frequency components. During continuous monitoring the BIS index is updated through the preprocessing and analysis of multiple epochs overlapping in a 60 s time window. The bispectrum ($BISP$) is the average of the triple products (TP) calculated for each epoch into which the original EEG signal $x(t)$ was segmented, as follow:

$$TP_j\left(f_1, f_2\right) = X_j\left(f_1\right) X_j\left(f_2\right) X^*\left(f_1 + f_2\right) \tag{2.1}$$

$$BISP(f_1, f_2) = \sqrt{\frac{1}{M} \sum_j TP_j\left(f_1, f_2\right)} \tag{2.2}$$

where $BISP(f_1, f_2)$ represents the bispectrum computed for two frequencies f_1 and f_2, M is the number of epochs in the time window and $X_j(f)$ is the Fourier Transform component at frequency f and for epoch j.

Whether the computation of the phase coupling information importantly contributes in monitoring anaesthesia is a debated issue. Results from a study involving 58 subjects who underwent a range of anaesthetic levels between light sedation and deep anaesthesia showed that using a combination of spectral and bispectral features increased the 'responsive' vs 'anaesthetized' classification performances, if compared with spectrum or bispectrum alone (Holt et al. 1998). On the other hand, other works reported contrasting results: A study on 39 subjects undergoing elective surgery Miller and colleagues (Miller et al. 2004) compared the performance of the bispectrum computed by emulating the proprietary algorithm and an equivalent parameter depending only on the spectrum; the results showed that the two indexes closely tracked each other, suggesting that almost all the changes in BIS index during anaesthesia is explained by the decrease in the high frequency EEG spectrum. Moreover the same index computed using bicoherence, a measure of phase coupling that is independent of spectral amplitude and it is defined as the normalized degree of phase coupling, showed negligible changes with the induction of anaesthesia, suggesting that the application of bispectral analysis does not add clinically useful information that could not be obtained from the power spectrum. Similarly, another study (Hagihira et al. 2001) found modest changes in low frequency bicoherence as the concentration of anaesthetics was increased and demonstrated that to accurately estimate the bispectrum at least three minutes of EEG monitoring are required, while

the BIS monitor uses 60 s EEG segments to update the index value. Moreover the bispectrum computation involves the combination of three estimators (equation 2.1), hence its variability could represent an issue.

Despite being a well-established method, the BIS monitor suffers from several limitations. BIS values show a high variability during stable physiological conditions and have been demonstrated to be unable to detect the transition from unconsciousness to the recovery of awareness (Gajraj et al. 1999). The ability of BIS in measuring the hypnotic level was also questioned in a study on 20 patients monitored with the isolated forearm technique (IFT) (Schneider et al. 2002) where the BIS index could not reliably distinguish between responsive and unresponsive patients. In a similar study the relationship between BIS and postoperative recall was investigated in 56 patients and similar BIS values were found in subjects with and without conscious recall (Kerssens et al. 2003). Moreover BIS shows a gradual increase after the termination of anaesthesia, thus suggesting it may actually measure the anaesthetic induced suppression of EEG activity, thus being a monitor of the clearance of drugs instead of the state of arousal of the brain, which is the result of the complex balance between analgesic and hypnotic levels and the surgical stimulation effects (Gajraj et al. 1999). Some authors suggest that BIS value could drop during the anaesthetic procedure as a result of neuromuscular blockade and consequent decrease of scalp electromyographic (EMG) activity or that, equally, EMG artifacts may falsely elevate the BIS value (Bard 2001). In addiction the BIS monitor is 'blind' to some anaesthetics as Nitrous Oxide (Barr et al. 1999).

2.2.2 Auditory Evoked Potentials

Auditory evoked potentials (AEPs) are brain responses elicited by the presentation of an auditory stimulus. The electrical activity is recorded via scalp electrodes at specific locations on the head. The characteristic features (amplitude and latency of peaks) of AEPs have been extensively investigated and represent the activity of neural generators along the auditory pathway[1]. Usually AEPs are classified on the basis of the latency (the timing of the responses relative to the stimulus onset) into auditory brainstem responses (ABR) -within 15 ms after stimulation; mid-latency responses (MLR) -which occurs in the interval of 15-50 ms after stimulation; and Late cortical responses (ALR). The MLR is characterized by a series of positive ("P" waves) and negative ("N" waves) waves; The first MLR negative wave was called Na, followed by the positive wave Pa and by the Nb dip and Pb peak. While the waves of early evoked responses are stable with the level of arousal, later components, that represent activity going from the thalamus to the

[1] In view of anaesthetic monitoring, it is useful to point out that AEPs may give information about different anaesthetic endpoints as compared with EEG as the first involve the processing of sensory information, while spontaneous EEG does not.

primary and association auditory cortices, undergo significant changes during sleep and anaesthesia and are also modulated by the level of attention. In particular, the MLR seems to be the most promising for discriminating conscious from unresponsive patients during anaesthesia. Literature results report evidence of an increase in the latency and a decrease in the amplitude of MLR peaks with anaesthesia. These properties are exploited in a commercial device, the aepEX monitor (Medical Device Management Ltd, Essex, UK), that extracts the MLR evoked by click stimuli to produce a diagnostic index of DoA. A correlation between Nb negative wave latency and response to command during anaesthesia induced with different agents was demonstrated (Thornton & Sharpe 1998, Loveman et al. 2001, Tooley et al. 1996, 2004). However the large inter-subject variability in Nb latency (some patients are unresponsive at latencies where others are sufficiently awake to respond to verbal commands) gives rise to the problem of specifying a general cut-off point for awareness. A more recent study by Bell and colleagues (Bell et al. 2006) investigated MLR trends associated with changes in responsiveness to command and reported a switch in MLR power with the onset of anaesthetic induced unresponsiveness but they did not report any latency shift. Also in this study the identification of a threshold across subjects that indicates conscious awareness was complicated by the individual variability. Moreover the large variation in predictive threshold with different drug combinations suggest that MLR is agent specific and that, more specifically, it is not a pure hypnotic parameter (Tooley et al. 2004). AEP is also a difficult signal to record, since it is embedded in EEG background activity, myogenic and electrical artifacts, leading to signal to noise ratio (SNR) often less than -20 dB and issues of data quality may not have been well addressed in early studies. The standard technique to improve SNR is to register multiple responses to repeated stimulation and coherently average them and several methods have been proposed in order to acquire a large number of responses within a relatively short recording time. For instance the use of Maximum Length Sequences (MLS) (Eysholdt & Schreiner 1982), which allows a higher stimulation rate to be used by overlapping the multiple responses in time, was demonstrated to improve the detectability of MLR waves and hence the feasibility of using MLR in clinical monitoring (Bell et al. 2002). However the detection of significant differences between anaesthetic infusion rates remains challenging, due to the extremely small differences in amplitude (typical variations are a fraction of 1 μV) (Bell et al. 2006). These factors, together with worsening SNR with increasing hearing impairment, limit the usefulness of MLR in clinical practice for anaesthetic monitoring (Loveman et al. 2001).

2.2.3 Comparison of DoA monitors

Despite the large number of studies on commercial monitors, comparing the performances and effectiveness of different DoA indexes is challenging due to the heterogeneity of anaestehtic protocols and patients populations investigated (Bruhn et al. 2006). In a study comparing AEP, BIS and processed EEG performances including 90 patients (**?**) the AEP was found most sensitive to surgical stimulation, while the BIS had the better signal quality. BIS and entropy (RE,SE) showed similar accuracy in detecting loss of response to verbal command. As previously discussed, BIS is also thought to be more affected by EMG artifacts than other monitors as Narcotrend or Entropy. However, in terms of algorithm complexity (or,equivalently, performances in monitoring rapid changes in sedation level) Narcotrend (together with PSI) are slower if compared to BIS or Entropy monitors (**?**)

In a recent update of UK guidelines for minimal standards of anaesthetic monitoring the use of DoA monitors is strongly recommended ('if not essential') in total intravenous anaesthesia with neuromuscular blockade (Association of Anaesthetists of Great Britain and Ireland 2016). These guidelines are based on a technology assessment report commissioned by National Institute of Health Research that examined 22 randomized control trials and compared BIS, Entropy and Narcotrend with standard monitoring in terms of risk of intraoperative awareness, patient outcomes (recovery times, comorbidities), drugs administration and cost-effectiveness of DoA monitoring. The main conclusions of this detailed assessment are that the evidence for preferring DoA to standard monitoring is stronger for BIS-guided anaesthesia than for Entropy or Narcotrend monitoring. However the impact of BIS monitoring in reducing the incidence of intraoperative awareness is limited (and significant only in patients at higher risk of awareness). All the monitors included in the analysis are associated with reduced anaesthetic consumption and patient recovery times however the relative cost savings are counterbalanced by the additional costs of DoA monitoring. The ability of current DoA monitors to reduce the risk of intra-operative awareness is questioned (Pandit & Cook 2014) and, as we have discussed, although the methods described above correlate well with the delivered anaesthetic concentration, they suffers from some limitations. Together with practical limitations, a primary issue in the use of these DoA monitors is that they do not provide direct information about the state of the brain, which is the target of anaesthetics; in other words they lack a theoretical and physiological basis (Massimini et al. 2010). Moreover current DoA monitors measure only the hypnotic component of general anaesthesia disregarding, for instance, the anti-nociception level, which is a very important indicator of the surgical stress of the patient. These are all prossible reasons why at the

present their uptake is quite low (only 2% of anaesthetists use them for routine monitoring (Pandit & Cook 2014)). There is therefore a need for improvement of current DoA monitors, in particular with regard to their ability to detect unexpected awareness. In the search for a more reliable index, it may be fruitful to concentrate on correlation with neurobiological processes involved in the conscious state, such as cortico-thalamic information flow. In this regard theoretical and experimental observations indicate that electrophysiological measures of brain connectivity are promising.

In the next section a review of studies investigating the relation between brain connectivity and consciousness will be given together with the rationale and methods to be explored in developing a novel indicator of depth of anaesthesia based on cerebral connectivity.

2.3 Anaesthesia and Brain Connectivity

As seen in section 2.1, loss of consciousness induced in anaesthesia may be a result of changes in the information flow between cortical and subcortical areas. According to the integrated information theory of consciousness introduced by Tononi (2008), large scale coupling that integrates information from a complex of brain areas is essential to generate a conscious experience. In this framework a possible measure of consciousness was proposed, the Integrated Information, defined as *the amount of information generated by a complex of elements, above and beyond the information generated by its parts*; Tononi and colleagues also suggested that network patterns characterized both by large activation and high differentiation could integrate the ensemble of stimuli characteristic of the conscious experience.

Brain connectivity reflects statistical or causal dependencies between brain regions (Friston 2011) and it can be used to describe functional networks activated in some specific behavioural states. Its potential to provide new insight into the neural mechanisms underlying LOC (Nallasamy & Tsao 2011) has given rise to a series of works that have investigated the anaesthetic modulation of brain connectivity using a heterogeneity of methodologies and anaesthetic protocols. These studies have suggested that anaesthetic-induced LOC is associated with widespread changes in brain connectivity. Convincing evidence of a localized and stereotypic brain connectivity pattern in anaesthetic LOC has been given in Ferrarelli et al. (2010). Ferrarelli and colleagues used a pertubational approach involving Transcranical Magnetic Stimulation (TMS) to induce electrical current inside the head by magnetic pulses, and EEG recording to investigate how triggered neural activity spread from the stimulation site. Whilst during wakefulness the TMS

elicited complex patterns of scalp waves and current spreading in distant cortical areas, during midazolam induced LOC, the TMS evoked response faded shortly after the stimulation and was reduced to a stereotypical and local response, thus indicating a 'breakdown of effective connectivity' (figure 2.2). Interestingly enough, similar find-

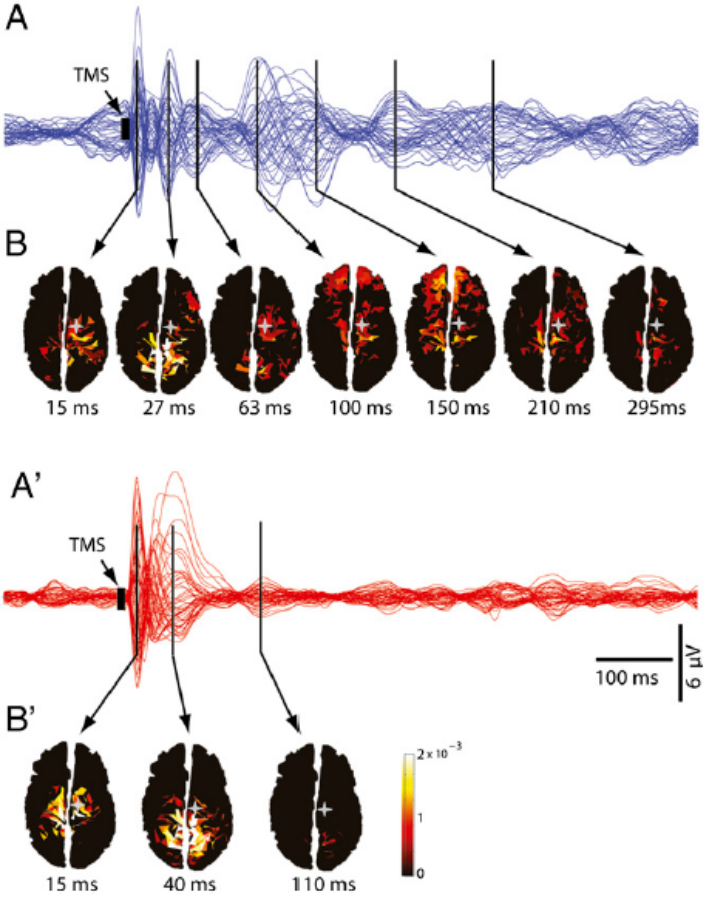

FIGURE 2.2: TMS-evoked response in wakefulness (TOP) and midzodam LOC (BOTTOM). A and A' represent the averaged TMS-evoked potentials for all electrodes. B and B are maps of estimated cortical currents. The grey cross indicates the site of TMS stimulation (premotor cortex). From Ferrarelli et al. (2010). During wakefulness the TMS stimulus triggers a potential (in blue in the top panel) that shows complex patterns and spreads to different cortical locations. On the other end the potential evoked in anaesthesia (in red in the bottom panel) decays shortly after the stimulation and remains localized at the stimulation site.

ings were observed in a TMS/EMG investigation of NREM sleep by the same group (Massimini et al. 2005*a*). A general impairment of brain networks integration (with fronto-parietal connectivity particularly affected) has also been reported in several functional magnetic resonance (fMRI) studies of propofol anaesthesia (Schrouff et al. 2011, Boveroux et al. 2010, Gómez et al. 2013). In a study involving 164 patients anaesthetized with a variety of agents, changes in EEG coherence in different frequency bands were investigated (John & Prichep 2005). Coherence (Coh) measures the proportion of linear dependency between pairs of signal and will be discussed in more detail in Section 3.1.

Results showed that coherence between frontal and occipital electrodes in the gamma (γ) band (25-50 Hz) decreases significantly shortly after LOC and during maintenance of anaesthesia and returns to 'baseline' levels at recovery of consciousness. These findings indicate a functional uncoupling of frontal and posterior brain areas during anaesthetic LOC: this uncoupling is more marked for the γ rhythm, known to arise from the cortico-thalamo-cortical activity regulating the processing of complex environmental stimuli.

Changes in EEG connectivity during general anaesthesia have also been investigated in a few recent studies. These studies have reported significant changes in the direction and functional organization of fronto-parietal networks, using directed connectivity estimators based on EEG phase synchronization (Lee et al. 2009), symbolic transfer entropy (Ku et al. 2011) and Granger Causality (time domain) (Nicolaou & Georgiou 2014). While these works identify a general impairment of frontoparietal connectivity in anaesthetic-induced LOC, results regarding the changes in the direction of coupling are not conclusive and are sometimes contrasting, probably as a result of applying different methodologies for connectivity analysis. The activity in the fronto-parietal associative network has been shown to be systematically altered also in other states of diminished or suppressed consciousness, like NREM sleep, vegetative states or coma (Massimini et al. 2005*a*, Spoormaker et al. 2010, King et al. 2013, Sitt et al. 2014, Boly et al. 2011). A significant impairment of medium and long-range information sharing was demonstrated to occur in patients in vegetative state (King et al. 2013); a weighted symbolic mutual information (wSMI) analysis of EEG auditory stimuli was applied to evaluate information sharing in different disorders of consciousness. To compute wSMI the EEG traces were transformed in discrete symbols and then the joint probability for each pair of symbols was estimated. The changes in mutual information sharing at different distances between electrodes were investigated revealing a significant increase of long range wSMI with the level of consciousness as showed in figure 2.3;

Taken together, these findings provide evidence for the important role of the fronto-parietal association cortices in the maintenance of consciousness (Boly et al. 2008) and the hypothesis that information flow break-down may affect signalling between sensory posterior areas and associative frontal cortices that is essential for a conscious experience (Boveroux et al. 2010). Moreover these studies indicate that anaesthetic-induced unconsciousness is associated with widespread changes in brain connectivity and therefore suggest that empirical measures of information sharing between brain regions represent a useful tool to investigate the basis of anaesthetic modulation of consciousness. Differently from the black-box approach of the current DoA monitors (that obscures the underlying neural mechanisms in an a-dimensional index) brain connectivity measures

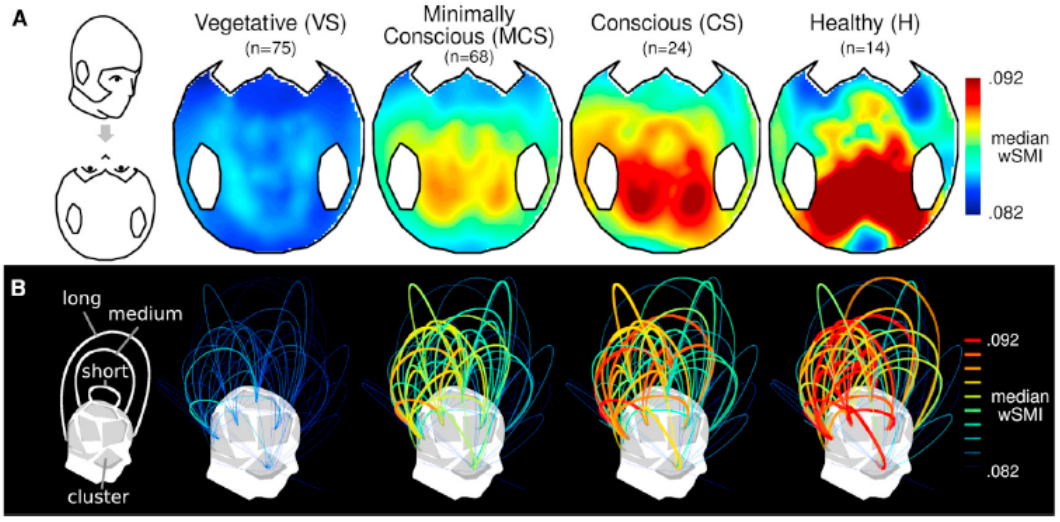

FIGURE 2.3: Weighted symbolic mutual information (wSMI) maps computed for different levels of consciousness (indicate by different head plots). A. Median wSMI computed for each channel with respect to all other channels. B. Mean wSMI computed among 16 clusters (groups of adjacent electrodes) and represented by arcs whose height is proportional to the distance in 3D space between clusters. From King et al. (2013). This indicates that long range connectivity is affected by the level of consciousness of the subjects.

may provide more information about the physiological correlates of consciousness.

Among the various neuroimaging techniques mentioned, non-invasive recordings from surface scalp electrodes (EEG) are an interesting method, due to their high temporal resolution, practicability and limited costs (Sitt et al. 2014). EEG-based systems can be used in routine clinical work in the home or ward, as well as in intensive care units or operating theatres where assessment of level of consciousness may be carried out. Due to its practical advantages, in this work we have focused on connectivity estimated from EEG recordings. Different methodologies have been proposed to assess EEG connectivity, based on various theoretical frameworks (phase dynamics, information theory, Granger causality, etc.). Among them, spectral estimators developed in the framework of Granger causality represent well established, computationally convenient methods, that allow a straightforward interpretation in terms of power content. In the next chapter, a detailed overview of the rationale and properties of these estimators will be outlined.

Chapter 3

Measuring coupling and causality with Multivariate Autoregressive modeling

In the last decades a growing body of neuroimaging studies has investigated neuronal activity during loss of consciousness in sleep, anaesthesia and vegetative states. These studies have analysed different aspect of brain activity: for instance the baseline activity as reflected by cerebral metabolic rate, or the blood oxygen level dependent signal as well as the response of brain networks to external stimuli. In recent years an increasing relevance has been attributed to the concept of brain connectivity as a way to understand the organized behaviour of cortical regions, beyond the simple mapping of their activity (David et al. 2004); currently the functional connectivity of specific large-scale networks is in the forefront of interest (Hudetz 2012) and it is also the most relevant for the current study. As described in section 2.3, different imaging techniques have been used (PET, fMRI, TMS and EEG) with different measures of connectivity (i.e. Coherence, Phase Synchronization, Mutual Information). While brain imaging and invasive recordings are powerful tools to investigate brain activity, including at subcortical level, they suffer from some practical limitations that affect their applicability for routing monitoring: these technologies are often too cumbersome and expensive to be easily applied for clinical monitoring. On the other hand macroscopic, non-invasive scalp recordings such as EEG, due to their practicability (EEG can be applied easily at bed-side) and limited costs, represent a suitable technique for clinical monitoring. In this work we will therefore focus on measures of connectivity extracted from EEG signals that will be described more formally in this chapter.

Functional connectivity is defined as the statistical dependence among measurements of neural activity (Friston 2011) and it is usually inferred through temporal correlations among different neurophysiological events. These correlations may be, for example, a result of stimulus triggered oscillations evoked by a common input, but do not necessarily refer to direct coupling among neural systems mediated by anatomical connections. The estimation of Functional Connectivity is quantifiable with measures of statistical dependencies among recorded data, such as correlation or coherence, and it does not require a model of how the neural systems under investigation are connected, thus being particularly suited for the analysis of large-scale complex networks.

A considerable number of approaches (linear and non-linear) have been proposed in the literature for estimation of Functional Connectivity from scalp recorded EEG signals, however of particular interest are frequency-based methods since it is well known that important information in the EEG is coded in this domain and frequency specific patterns are dominant in some behavioural states, e.g. sleep stages, performances of cognitive tasks (Klimesch 1999). The approach employed in this work makes extensive use of multivariate autoregressive models (MVAR) models to compute linear causality estimator of FC in the frequency domain.

The aim of this chapter is to give an overview of the methods proposed in the literature to measure linear coupling and causality between time series. In particular, the focus will be on measures developed in the framework of MVAR. This approach remains of great interest in the study of physiological signals because it can be strictly connected to the frequency domain. The rationale for the development of these measures will be given together with a description of their properties. In the last section some practical issues that are essential for an accurate model estimation will be addressed (i.e choice of the algorithm, optimum model order selection, appropriate time window length, etc.). To continue, the validity of the estimated MVAR model will be examined and different statistical tools recommended to check for the model validity discussed.

3.1 Non-directed measures of coupling: Coherence and Partial Coherence

Measures of interactions among observed time series of a multivariate process have generally relied on estimates of their correlation matrix. The cross-correlation function $r_{i,j}(k)$ at time lag k of two signals $x_i(n)$ and $x_j(n)$ is defined as the average of the product of $x_i(n)$ and $x_j(n-k)$. For a $M \times 1$ vector of M observed zero-mean time series at time lag n

$$\mathbf{x}(n) = [x_1(n)...x_i(n)...x_M(n)]^T \text{ [1]} \tag{3.1}$$

the correlation matrix at time lag k is defined by the $M \times M$ matrix $\mathbf{R}(k)$

$$\mathbf{R}(k) = E\left[\mathbf{x}(n)\mathbf{x}^T(n-k)\right] \tag{3.2}$$

One common way to quantify the correlation in time between two signals is the correlation coefficient, which is a normalised measure of linear dependencies, as expressed below

$$\rho_{i,j}(k) = \frac{r_{i,j}(k)}{\sqrt{r_{i,i}(0)r_{j,j}(0)}} \tag{3.3}$$

where $r_{i,j}(k)$ is the i,j element of the correlation matrix $\mathbf{R}(k)$ and $r_{i,i}(0)$ represent the variance of the signal x_i.

As previously mentioned, in EEG analysis it is convenient to have information about the coupling in the frequency domain. The Fourier Transform of the correlation matrix is called the spectral density matrix $\mathbf{S}(f)$.

$$\mathbf{S}(f) = \begin{bmatrix} S_{1,1}(f) & S_{1,2}(f) & \cdots & S_{1,M}(f) \\ S_{2,1}(f) & \cdots & \cdots & S_{2,M}(f) \\ S_{M,1}(f) & \cdots & \cdots & S_{M,M}(f) \end{bmatrix} \tag{3.4}$$

In analogy to the definitions given in the time domain, a normalised spectral measure of coupling is given by the Coherence (Coh) that is the ratio of the cross-spectral density function to the product of the autospectral densities.

$$Coh_{i,j}(f) = \frac{S_{i,j}(f)}{\sqrt{S_{i,i}(f)S_{j,j}(f)}} \tag{3.5}$$

The $Coh_{i,j}(f)$ is complex-valued and its squared modulus $|Coh_{i,j}(f)|^2$ (called the *Coherence function* (Bendat & Piersol 2000)) measures the strength of linear interactions between $x_i(n)$ and $x_j(n)$ at frequency f, i.e. the extent to which the signal $x_j(n)$ may be predicted by the signal $x_i(n)$ by an optimum linear least square relationship. It can assume values between 0 and 1 and $|Coh_{i,j}(f)|^2 \neq 0$ when any linear relationship (direct

[1] Vector and matrices will be indicated with bold symbols.

and indirect) between the signals $x_i(n)$ and $x_j(n)$ exists. Coherence has been ubiquitously used to investigate frequency-specific synchronization in EEG and MEG studies (Gross et al. 2001).

In order to quantify *direct* linear interactions between pairs of signals the concept of Partial Coherence was introduced. The Partial Coherence ($PCoh$) is a measure of the dependence between two time series after removing the linear effect of the other series of the dataset (Bendat & Piersol 2000). It plays the same role as ordinary Coh except that it applies to partial (or conditioned) spectral density functions. The partial cross-spectral density function of $x_i(n)$ and $x_j(n)$ is defined as:

$$S_{ij|(X/ij)}(f) = S_{ij}(f) - \mathbf{S}_{i(X/ij)}(f)\mathbf{S}^{-1}_{(X/ij)(X/ij)}(f)\mathbf{S}_{(X/ij)j}(f) \tag{3.6}$$

where X/ij means all the series of the dataset $\mathbf{x}(n)$ except the series $x_i(n)$ and $x_j(n)$ and $\mathbf{S}^{-1}_{(X/ij)(X/ij)}(f)$ is the inverse of the spectral matrix $\mathbf{S}(f)$ remaining when the i_{th} and j_{th} row and column have been removed. The partial cross-spectral density represents the cross spectrum between $x_i(n)$ and $x_j(n)$ once the linear effects of X/ij on $x_i(n)$ and $x_j(n)$ have been removed. The second term in the right side of equation 3.6 represents the linear least square prediction of $x_i(n)$ from X/ij [2] (Brillinger 1981).

The $PCoh$ between $x_i(n)$ and $x_j(n)$ is defined as follows:

$$PCoh_{i,j}(f) = \frac{S_{ij|(X/ij)}(f)}{\sqrt{S_{ii|(X/ij)}(f)S_{jj|(X/ij)}(f)}} \tag{3.7}$$

As for the ordinary Coh, the $PCoh$ is complex, therefore its squared modulus $|PCoh_{i,j}(f)|^2$ (called the *Partial Coherence function*) is used to quantify the strength of interaction between $x_i(n)$ and $x_j(n)$ after subtracting the linear effect of the remaining signals of the process.

It has been demonstrated (Dahlhaus 2000) (see appendix B for the demonstration) that the Partial Coherence can be estimated as a function of the inverse spectral matrix

$$\mathbf{G}(f) = \mathbf{S}^{-1}(f) \tag{3.8}$$

as follows

[2] If the correlated effects of X/ij are removed by $x_i(n)$ it is not necessary to remove them also from $x_j(n)$ to compute the partial cross-spectral density between $x_i(n)$ and $x_j(n)$. It can be demonstrated in fact that the cross spectrum between $x_{i|(X/ij)}(n)$ and $x_{j|(X/ij)}(n)$ must be the same as the cross-spectrum between $x_{i|(X/ij)}(n)$ and $x_j(n)$ (or $x_{j|(X/ij)}(n)$ and $x_i(n)$) (Bendat & Piersol 2000)

$$PCoh_{i,j}(f) = \frac{g_{i,j}(f)}{\sqrt{g_{i,i}(f)g_{j,j}(f)}} \tag{3.9}$$

This approach, that uses the inverse spectral matrix instead of the partial spectral density functions as in equation 3.7, is more convenient as it avoids the computation of all the M(M-1)/2 partial coherences (Dahlhaus & Eichler 1997). It is also recommended in Eichler et al. (2003) for neural spike trains analysis and in Medkour et al. (2009) for EEG application as *an efficient computation of all frequency domain PCoh statistics.* In conclusion, the ordinary Coh quantifies the linear interactions between signals while the $PCoh$ measures the linear relationships between two signals of a multivariate process once the least square dependence from the other signals of the dataset has been removed, therefore it quantifies only direct dependencies between signals. It can be a useful measure to identify the direct influence between two signals where the ordinary Coh can be erroneously high (or low) because of the influence a third (or more) signal has on the considered ones. An illustrative example of how, if the effect of other variables is not considered properly, the ordinary coherence alone can lead to erroneous conclusions is showed in Figure 3.1. Assume that the Coh function computed between $x1$ and $x2$ has a value close to the unity. This would lead to believe that a linear relation is directly relating these two variables. However this result might be only a consequence of the presence of a third signal $x3$ that is highly correlated with both $x1$ and $x2$. In reality there may be no direct relationship between $x1$ and $x2$ at all. In this case, differently from the $Coh_{1,2}$, the $PCoh_{1,2}$ would be zero thus giving an appropriate indication of the degree of linear dependence between the two signals.

The use of $PCoh$ to identify direct interactions presents some limitations however: Firstly, neural signals cannot be completely described by linear dynamics. In this case one must bear in mind that $PCoh$ is able to partial out only linear indirect correlations: it can therefore be different from zero if the signals are driven by some indirect non-linear dynamics. $PCoh$ may also 'falsely' indicate a direct connection in the typical case of 'marrying parents of a joint child' (Dahlhaus & Eichler 1997). This occurs when two processes $x1$ and $x2$ are independent but they both influence a third signal $x3$ (i.e. $x3 = x1 + x2 + \epsilon$): in this case $x1$ and $x2$ are not independent conditional to $x3$ and the $PCoh_{1,2}(f)$ will be different from zero.

Moreover, due to the symmetric properties of the cross-spectral matrices, Coh and $PCoh$ do not provide information regarding the *direction* of the interactions (i.e. $Coh_{i,j}(f) = Coh^*_{j,i}(f)$ and $|Coh_{i,j}(f)|^2 = |Coh_{j,i}(f)|^2$ and the same holds for the $PCoh$) and do not allow inference on the causal structure of the process. In order to be able to quantify directional information, measures derived from a factorization of Coh and $PCoh$ have been developed, primarily in the framework of the multivariate autoregressive models.

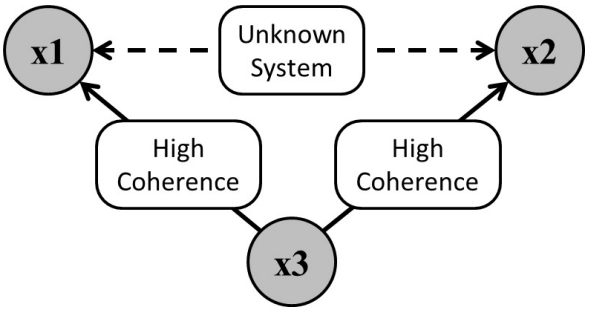

FIGURE 3.1: Illustrative example of erroneous high coherence

3.2 Directed measures of causality derived from MVAR models

In this section directed measures of interactions between time series (i.e. measures able to quantify the strength of coupling from one signal $x_i(n)$ to another $x_j(n)$ as opposite to the feedback interaction from $x_j(n)$ to $x_i(n)$) will be introduced. As previously mentioned these measures make extensive use of the MVAR model framework. A MVAR process is a linear model of the temporal dynamics and interactions among time series. The general expression of a MVAR process is given in equation 3.10 (Lütkepohl 2005).

$$\mathbf{x}(n) = \sum_{l=1}^{p} \mathbf{A}(l)\mathbf{x}(n-l) + \boldsymbol{\varepsilon}(n), n = 1, \cdots, N \tag{3.10}$$

$\mathbf{x}(n)$ represents the M dimensional vector of the time-series at time lag n[3], N is the number of time samples, $\mathbf{A}(l)$ is the coefficients matrix describing the linear interactions among observed time series at lag l, p is the number of time lags used to model interactions, i.e. the model order, and $\boldsymbol{\varepsilon}(n)$ is the vector of white innovations, with non-singular residual covariance matrix

$$\boldsymbol{\Sigma}_{\varepsilon} = \left\{\sigma_{ij}^2\right\} \tag{3.11}$$

and

$$E[\boldsymbol{\varepsilon}(n)\boldsymbol{\varepsilon}(m)] = 0, n \neq m \tag{3.12}$$

An MVAR model describes each signal as a weighted combination of its own past values and the past values of other signals in the model plus an error term. The weights relating

[3]For EEG analysis, M represent the number of channels considered and $\mathbf{x}(n)$ is the vector of EEG recordings at time lag n

the present of one signal to the past of another capture the causal or directed influence between signals.

The expression for a MVAR model in the frequency domain is

$$\mathbf{X}(f) = \mathbf{A}(f)\mathbf{X}(f) + \mathbf{E}(f) \tag{3.13}$$

where $\mathbf{A}(f)$ is the Fourier Transform of the matrix of coefficients $\mathbf{A}(l)$.

$$\mathbf{A}(f) = \sum_{l=1}^{p} \mathbf{A}(l) e^{\frac{-2\pi f l}{f_{samp}}} \tag{3.14}$$

Equation 3.13 can be expressed in an equivalent way as a linear causal filter, defining the transfer matrix $\mathbf{H}(f)$ as indicated in equation 3.15:

$$\begin{aligned} \mathbf{X}(f) &= \mathbf{H}(f)\mathbf{E}(f) \\ \mathbf{H}(f) = [\mathbf{I} - \mathbf{A}(f)]^{-1} &= \overline{\mathbf{A}}(f)^{-1} \end{aligned} \tag{3.15}$$

Taking in account equation 3.15, the spectral matrix and the inverse spectral matrices of a MVAR process $\mathbf{X}(f)$ can therefore be factorized as follows:

$$\begin{aligned} \mathbf{S}(f) &= \mathbf{H}(f)\mathbf{\Sigma}_{\varepsilon}\mathbf{H}^{*}(f) \\ \mathbf{G}(f) &= \overline{\mathbf{A}}(f)^{*}\mathbf{\Sigma}_{\varepsilon}^{-1}\overline{\mathbf{A}}(f) \end{aligned} \tag{3.16}$$

where the superscript $*$ stands for the conjugate transposed. The spectral factorization in equation 3.16 allows the unique decomposition of $Coh_{i,j}(f)$ into two *Directed Coherence* (DC) terms one quantifying the strength of coupling from $x_i(n)$ to $x_j(n)$, the other representing the feedback interaction from $x_j(n)$ to $x_i(n)$. Under the assumption that the residuals are uncorrelated and therefore $\mathbf{\Sigma}_{\varepsilon} = diag\left\{\sigma_{ii}^2\right\}$ we can express the general element of the spectral matrix $S_{i,j}(f)$ as follows:

$$S_{i,j}(f) = \sum_{m=1}^{M} \sigma_{m,m}^2 H_{i,m}(f) H_{j,m}^{*}(f) \tag{3.17}$$

Therefore the Coh can be factorised as (Baccalá & Sameshima 2001):

$$Coh_{i,j}(f) = \frac{\sum_{m=1}^{M} \sigma_{m,m} H_{i,m}(f)}{\sqrt{\sum_{m=1}^{M} \sigma_{m,m}^2 |H_{i,m}(f)|^2}} \frac{\sum_{m=1}^{M} \sigma_{m,m} H_{j,m}^*(f)}{\sqrt{\sum_{m=1}^{M} \sigma_{m,m}^2 |H_{j,m}(f)|^2}} = \sum_{m=1}^{M} DC_{i,m} DC_{j,m}^*(f) \tag{3.18}$$

The Coh is expressed in equation 3.18 as a sum of terms DC called Directed Coherence

$$DC_{i,j}(f) = \frac{\sigma_{j,j} |H_{i,j}(f)|}{\sqrt{\sum_{j=1}^{M} \sigma_{j,j}^2 |H_{i,j}(f)|^2}} \tag{3.19}$$

The $DC_{i,j}(f)$ can be interpreted as a measure of the coupling from $x_j(n)$ to $x_i(n)$ as opposite to $DC_{j,i}(f)$ that quantifies the feedback interaction from $x_i(n)$ to $x_j(n)$. In particular, the squared DC measures the coupling from $x_j(n)$ to $x_i(n)$ as the normalized portion of $S_{i,i}(f)$ due to $x_j(n)$ (or transferred from $x_j(n)$ via the transfer function $H_{i,j}(f)$ to $x_i(n)$). It is demonstrated, in fact, that (Faes & Nollo 2011)

$$\begin{aligned} S_{i|j}(f) &= |DC_{i,j}(f)|^2 S_{i,i}(f) \\ S_{i,i}(f) &= \sum_{j=1}^{M} S_{i|j}(f) \end{aligned} \tag{3.20}$$

where $S_{i|j}(f)$ is the part of the spectrum of $x_i(n)$ due to the signal $x_j(n)$. $|DC_{i,j}(f)|^2$ quantifies the amount of spectrum at frequency f transferred from $x_j(n)$ to $x_i(n)$ normalized by the spectrum of the signal $x_i(n)$.

When the residuals variances are excluded from the definition of DC, or when, equivalently, all the input variances are equal, the definition of DC in equation 3.19 coincides with the Directed Transfer Function (DTF) (Kamiński et al. 1997)

$$DTF_{i,j}(f) = \frac{H_{i,j}(f)}{\sqrt{\sum_{j=1}^{M} |H_{i,j}(f)|^2}} \tag{3.21}$$

The useful equation 3.20 is not valid for the DTF, that, disregarding the signals variances $\sigma_{j,j}$, loses interpretability in the sense of power transfer and spectral causality.

Starting from equation 3.16 and with analogous steps (and assumptions) the $PCoh$ can be factorized as follows [4]

[4] the order of the indexes i and m in this equation is different from equation 3.18 (i.e. $\overline{A_{m,i}^*}(f)$ vs $H_{i,m}(f)$) because of the conjugate transposed matrix $\overline{\mathbf{A}}(f)^*$ in the factorization expressed in equation 3.16.

$$PCoh_{i,j}(f) = \frac{\sum_{m=1}^{M} \frac{1}{\sigma_{m,m}} \overline{A}^{*}_{m,i}(f)}{\sqrt{\sum_{m=1}^{M} \frac{1}{\sigma^2_{m,m}} |\overline{A}_{m,i}(f)|^2}} \frac{\sum_{m=1}^{M} \frac{1}{\sigma_{m,m}} \overline{A}_{m,j}(f)}{\sqrt{\sum_{m=1}^{M} \frac{1}{\sigma^2_{m,m}} |\overline{A}_{m,j}(f)|^2}} = \sum_{m=1}^{M} PDC^{*}_{m,i} PDC_{m,j}(f) \tag{3.22}$$

In the last expression the $PCoh$ has been expressed as a sum of Partial Directed Coherence (PDC) terms (Baccalá & Sameshima 2001, Baccalà & Sameshima 2007)

$$PDC_{i,j}(f) = \frac{\frac{1}{\sigma_i} |\overline{A_{i,j}}(f)|}{\sqrt{\sum_{i=1}^{M} \frac{1}{\sigma^2_i} |\overline{A_{i,j}}(f)|^2}} \tag{3.23}$$

The $PDC_{i,j}(f)$ quantifies the normalized proportion of the inverse spectrum $S^{-1}_{j,j}(f)$ that is transferred from $x_j(n)$ to $x_i(n)$ through the function $|\overline{A^{*}_{i,j}}(f)|$. In fact it can be demonstrated that (Faes & Nollo 2011):

$$g_{i|j}(f) = |PDC_{i,j}(f)|^2 g_{j,j}(f) \tag{3.24}$$

where $g_{*}(f)$ represents the generic element of the inverse spectral matrix $\mathbf{G}(f)$ (see equation 3.8). Being the $PCoh$ a measure of direct coupling, the $PDC_{i,j}(f)$ measures the strength of the direct coupling from $x_j(n)$ to $x_i(n)$, when the linear influences from all the other signals of the MVAR process have been excluded. The quantity defined in equation 3.23 is the so-called *generalized* PDC (Baccalà & Sameshima 2007) that was introduced in order to give the PDC the property of scale-invariance with respect to the signals amplitudes, differently from its original formulation (ordinary PDC, equation 3.25) (Baccalá & Sameshima 2001)

$$orPDC_{i,j}(f) = \frac{|\overline{A_{i,j}}(f)|}{\sqrt{\sum_{i=1}^{M} |\overline{A_{i,j}}(f)|^2}} \tag{3.25}$$

where the weighting by the residual variances is lacking. The rationale of the upgrade from $orPDC$ to PDC [5] was to improve the performances of the estimator with respect to the time series scaling; the same property of scale invariance holds for the DC that, differently from DTF, includes the normalization by the variances of processes. A description of the effects of time series amplitude on the estimation of causality is detailed in Baccalà & Sameshima (2007).

[5] For the sake of simplicity we will indicated the generalized PDC with the symbol PDC in this report

3.2.1 Interpretation in the sense of Granger Causality

The PDC and DC can be interpreted as measures of Granger causality. The notion of Granger Causality was introduced in Econometrics by Granger (1969): by definition a signal $x(n)$ is said to Granger cause a signal $y(n)$ if *the knowledge of the $x(n)$'s past significantly improves the prediction of $y(n)$* (Lütkepohl 2005). This causal relationship is not reciprocal and therefore allows the inference on the direction of information flow between structures. The key concept in Granger causality is the exclusive reference to past samples in the prediction of the time series (the cause must precede the effect). PDC and DC (or DTF) provide an estimation of the coupling that takes into account only the effect of the past of the time series on the current sample. For instance the PDC is a function of the spectral coefficients $A_{i,j}(f)$ that are the Fourier Transform of the coefficients modelling the linear lagged interactions (equation 3.10). Statistical tests of Granger causality can be performed by direct examination of the coefficients of the MVAR model in the time domain; for instance a signal $x_j(n)$ Granger causes a signal to $x_i(n)$ if at least one of the coefficients $A_{i,j}(l), l = 1, \cdots, p$ is significantly different from zero. For this reason PDC and DC are considered frequency domain descriptors of Granger causality[6].

3.2.2 Differences between PDC and DC

There are important differences between PDC and DC.

- Normalization: because of their mathematical derivation, the $DC_{i,j}$ is normalized with respect to the effects produced by all the other signals of the M-variate dataset on the receiving signal $x_i(n)$ (normalization with respect to the receiving structure), while the $PDC_{i,j}$ is normalized with respect to all the contributions of $x_j(n)$ to the other time series of the dataset (normalization with respect to the source).

- DC and PDC are factors in the decomposition of Coh and $PCoh$, respectively, therefore the DC measures linear interactions, while PDC direct linear interactions between signals. This can be seen also looking at their expressions: the $DC_{i,j}$ is function of the transfer matrix $H_{i,j}(f) = \overline{A}_{i,j}^{-1}(f)$ that contains a sum of terms each one related to a possible path from $x_j(n)$ to $x_i(n)$. Hence DC (or DTF) is different from zero whenever any interaction from $x_j(n)$ to $x_i(n)$ is significant. On the contrary PDC is a direct function of the coefficient of the spectral matrix of parameters $A_{i,j}(f)$, therefore it is different from zero only when a direct connection

[6] For the sake of simplicity, from now on the term *causality* will be used to indicate Granger causality.

is present.[7]

These two measures highlight different aspects of coupling between time series: $DC_{i,j}(f)$ measures causality as the amount of information flow from $x_j(n)$ to $x_i(n)$ through all the possible paths, therefore is not a measure of direct coupling; however it has an easier interpretation since it measures the normalized amount of spectrum transferred from one signal to another. This is particular relevant in the analysis of EEG time series that are characterized by specific brain rhythms (frequency bands). On the other hand, the interpretation of PDC in the frequency domain is not straightforward because it is function of the inverse spectra of signals. However PDC is able to resolve the causality structure of the multivariate process, since it measures the amount of information flow from $x_j(n)$ to $x_i(n)$ through the direct path. PDC aims to quantify the direct interactions between time series that cannot be attributed to the other simultaneously observed time series of the process; therefore it represents a useful tool for inferring functional structure. The properties of the spectral causality estimators defined in this section will be clarified with an illustrative example in the next section.

3.2.3 An Illustrative Example

In order to compare the performances of the different measures previously introduced and to give general guidelines for the interpretation of the graphical results, a theoretical example will be considered in this section. The model proposed is adapted from Faes et al. (2001). This choice was made with the additional purpose of comparing the results and validate the implemented software. A simple MVAR model of order $p = 2$ involving $M = 3$ processes is shown in Figure 3.2 and generated by the following equations:

$$\begin{aligned} x_1(n) &= 1.34x_1(n-1) - 0.81x_1(n-2) + \varepsilon_1(n) \\ x_2(n) &= x_1(n-1) + 0.5x_3(n-1) + \varepsilon_2(n) \\ x_3(n) &= 0.5x_2(n-1) + 0.5x_2(n-2) - 0.54x_3(n-1) - 0.81x_3(n-2) + \varepsilon_3(n) \\ \mathbf{\Sigma}_\varepsilon &= diag\{1, 9, 1\} \end{aligned} \tag{3.26}$$

.

The matrix of parameters at the two lags $l = 1, 2$ are therefore

[7] as for the $PCoh$ the PDC is a measure of linear relationships, therefore it can be different from zero also when indirect non-linear dynamics affects the signals considered.

$$\mathbf{A}(1) = \begin{bmatrix} 1.34 & 0 & 0 \\ 1 & 0 & 0.5 \\ 0 & 0.5 & -0.54 \end{bmatrix} \tag{3.27}$$

$$\mathbf{A}(2) = \begin{bmatrix} -0.81 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0.5 & -0.81 \end{bmatrix} \tag{3.28}$$

.

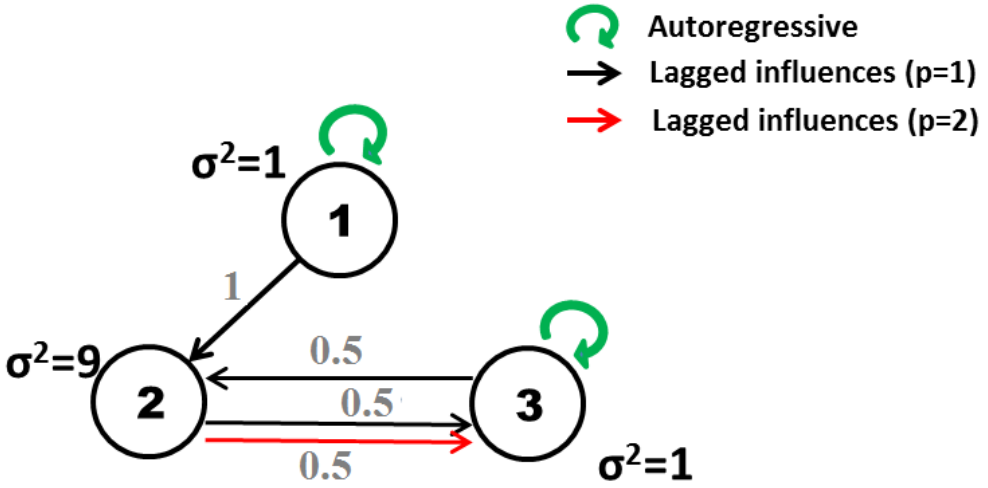

FIGURE 3.2: Imposed model involving three interacting processes. The arrows represent linear direct dependencies among time series at time lag p=1 (black) and p=2 (red); the values of the MVAR coefficients $a_{i,j}(l), i \neq j$ are shown in gray for each connection. The processes variances are indicated by σ^2.

The coupling among the processes is introduced imposing MVAR coefficients $a_{i,j}(l), i \neq j, l = 1, 2$ different from zero. The diagonal elements of the matrix of parameters $\mathbf{A}(l)$ on the other hand determine the autoregressive contribution, i.e. the intrinsic oscillations of the processes. The simulated processes were generated applying the MVAR filter expressed in equation 3.26 to the white Gaussian random processes whose variances were imposed respectively equal to 1, 9 and 1, in order to show the performances of the estimators with respect to the unbalanced variance of signals. The sampling frequency f_{samp} was set to 250 Hz and the length of the simulated time series set to 4 s (1000 samples). The coupling measures were then estimated on simulated data after an MVAR model was fit on the dataset. First the optimum model order was estimated applying the Schwarzs Bayesian Criterion (Schwarz, 1978) order selection criterion. Secondly a Least Square Algorithm was used to estimate the matrix of coefficients (ARfit, Matlab, Schneider and Neumaier, 2001) (details about the estimation of MVAR models from a

given time series set will be given in the section 3.3.1 of this chapter). The generation-estimation procedure was repeated for 100 iterations as for previous simulation studies (Astolfi, Cincotti, Mattia, Marciani, Baccala, de Vico Fallani, Salinari, Ursino, Zavaglia, Ding, Edgar, Miller, He & Babiloni 2007, Faes et al. 2013).

First the time domain measures were computed on simulated data. Figure 3.3 represents the absolute value of the correlation coefficient $\rho_{i,j}(k)$ computed for all pairs of signals and for $k = 1, \cdots, 25$. The cross-correlations coefficients (off-diagonal elements) show that there are linear dependencies among all the processes for the first 10 time lags, with peaks of correlations for time lags 1 and 2, on the other hand the autocorrelations coefficient (diagonal elements) peaks are aligned with zero. The values of cross-correlation are different from zero also for processes 1 and 3 that are not directly dependent and different from zero also when a directed connection is absent, for instance $\rho_{1,2}(k) \neq 0$, (first row, second column of figure 3.3). For these reasons it is not possible to determine the coupling structure of the model from the cross-correlation coefficients estimation.

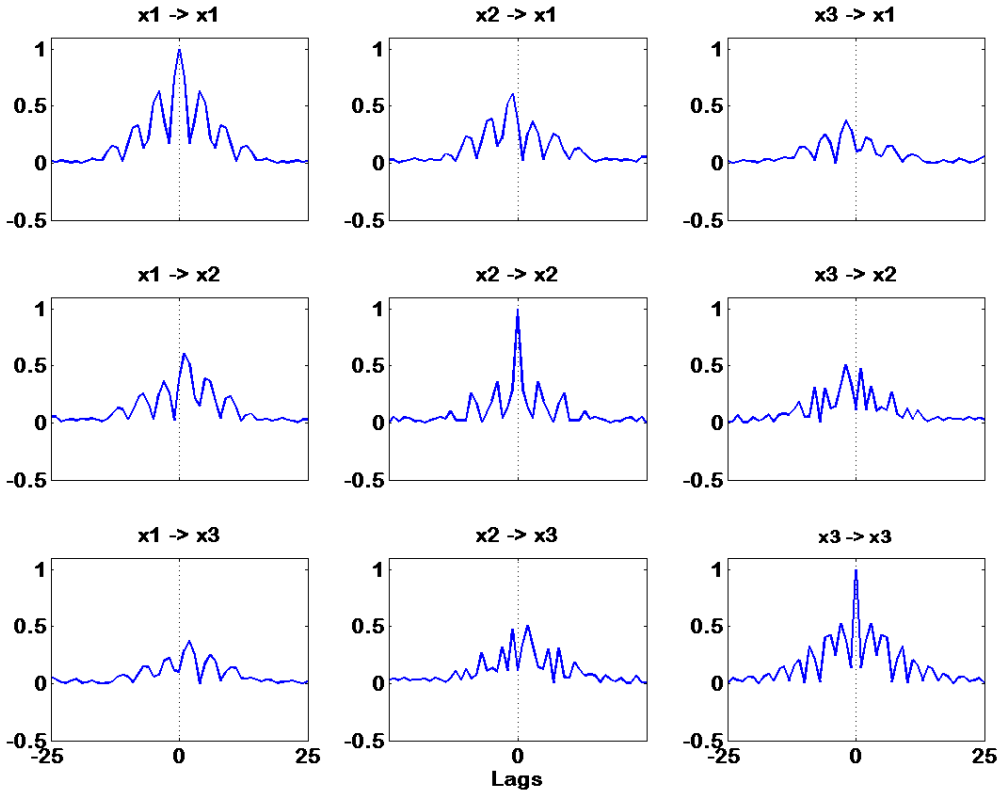

FIGURE 3.3: Matrix plot showing the correlation coefficients values $|\rho_{i,j}(k)|$ computed for all pairs of signals and for time lags $k = 1, \cdots, 25$. From an inspection of the cross-correlation coefficient values it is not possible to determine the coupling structure of the model represented in figure 3.2: for instance $|\rho_{1,3}|$ (first row, last column) and $|\rho_{3,1}|$ (first column, last row) show values different from zero for some time lags even if a connection between node 1 and node 3 in figure 3.2 is absent.

The spectral measures of coupling were computed by first estimating an MVAR model from the simulated processes, then estimating the matrix of parameters $\mathbf{A}(l)$ and its Fourier Transform $\mathbf{A}(f)$. In order to evaluate the accuracy of the estimation for each measure the correspondent theoretical value was computed from the imposed matrix of parameters given in equation 3.27 and 3.28.

The diagonal elements in figure 3.4 show the spectra of the three processes. The peaks of the spectra of x_1 and x_3, at $f_1 = 0.11 f_{samp}\,(27.5Hz)$ and $f_3 = 0.3 f_{samp}\,(75Hz)$ are a result of the oscillations imposed with the autoregressive parameters that introduce complex conjugate poles in the transfer function of the processes. As a result of the imposed linear coupling the peaks are transferred on the process x_2. The Coh is reported in the off-diagonal elements of Figure 3.4 as a function of the frequency: the plot on the i_{th} row and j_{th} column represents the Coh between signal x_j and x_i.

Figure 3.5 shows the inverse measures. The diagonal elements represent the inverse spectra while the off-diagonal elements show the $PCoh$ as a function of the frequency. The Coh measures both the linear direct and indirect coupling, being significantly different from zero for each pair of signals, also for processes not directly dependent (i.e. $Coh_{3,1}(f) \neq 0$, third row, first column in figure 3.4), while the $PCoh$ measures the direct coupling, being close to zero in absence of directed linear interactions (i.e. $PCoh_{3,1}(f)$, third row, first column in figure 3.5). Both measures are symmetric hence they cannot give information regarding the direction of coupling; therefore a full picture of the network structure cannot be inferred from the analysis of Coh or $PCoh$.

Following the same order of the methods section, the measures representing the factorisation of Coh and $PCoh$ are shown. Figure 3.6 is a matrix plot of the DTF estimated on simulated data and the theoretic DTF computed from the imposed matrix of parameters. The generic off-diagonal element $DTF_{i,j}(f)$ represents the normalized transfer of spectrum from signal x_j to signal x_i. For instance $DTF_{2,1}(f)$ (second row, first column) indicates that the spectrum of x_1 is transferred to x_2 mainly for frequency up to 65 Hz, with a peak at the frequency of oscillation f_1, while the higher part of the spectrum of x_2 is due to x_3 (i.e. $DTF_{2,3}(f)$, second row, third column, has a peak at the frequency f_3). The generic diagonal element $DTF_{i,i}$ represents the part of the spectrum of the process x_i that is not due to other time series (i.e. x_1 is not influenced by the other processes, therefore its autospectrum is equal to 1 for all the frequencies considered). As can be observed in the DTF is not symmetric thus allowing to identify the direction of the linear coupling; for instance $DTF_{1,2}(f) = 0$ while $DTF_{2,1}(f)$ is nonzero, indicating that the direction of information flow is $x_1 \rightarrow x_2$ (as in figure 3.2). On the other hand the DTF quantifies all instances of coupling, also indirect ones; for example $DTF_{3,1}(f)$ is non zero because of the indirect coupling path $x_1 \rightarrow x_2 \rightarrow x_3$. This does not hold

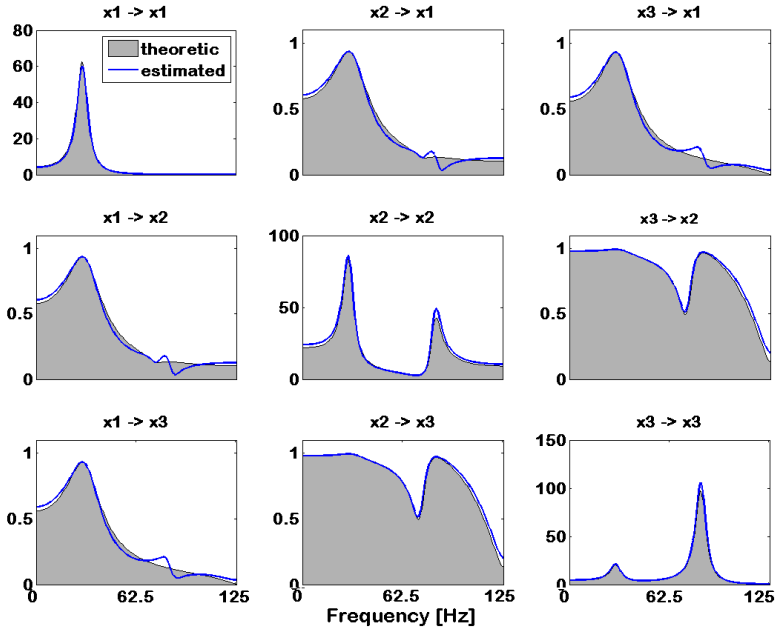

FIGURE 3.4: Diagonal elements: spectra (power spectral density, $[\mu V^2 Hz^{-1}]$) of the processes . Off-diagonal elements: Coh. The Grey area represents the theoretical measure while the estimated measures are plotted in blue. This image shows how the Coh is a symmetrical measure, hence it is not able to convey information about the direction of connections: i.e. $|Coh|1,2 = |Coh|_{2,1}$ therefore it is not possible to establish that the direction of the connection is from x1 to x2 as showed in figure 3.2. Moreover the value of Coh are different from zero even if a direct connection is absent: $|Coh|_{1,3} = |Coh|_{3,1}$ are different from zero even if a direct link between x1 and x3 is absent (figure 3.2).

for the $orPDC$, illustrated in figure 3.7, that is uniformly zero when no direct connection is present in the imposed model (i.e. $orPDC_{3,1}(f) = 0$). The $orPDC$ represents the normalized portion of the inverse spectrum that is transferred from one process to another. The structure of the original network can be inferred examining the $orPDC$ trends (the nonzero off diagonal elements in figure 3.7 corresponds to the imposed direct connections on the model in figure 3.2). Differently from what is observed for the DTF, the elements of the $orPDC$ do not have a straightforward interpretation in terms of spectral content, as they are function of the inverse spectrum.

As a final step in the refinement of the causality estimators the effect of signal scaling has been considered and the DC and the generalised PDC introduced. Results showing the trends of these measures are given in figures 3.8 and 3.9. DC and PDC show trends similar to, respectively, DTF and $orPDC$, but with different scaling. In particular it can be observed that the $DTF_{2,3}(f)$ (second row, third column, figure 3.6) exhibits higher values than the $DTF_{3,2}(f)$ (third row, second column) while one would expect higher values of coupling from x_2 to x_3 rather than in the opposite direction, because of the double arrow in figure 3.2 (see also equation 3.26). The explanation to these results is that the coefficients of the MVAR model are affected by the amplitude of signals. In

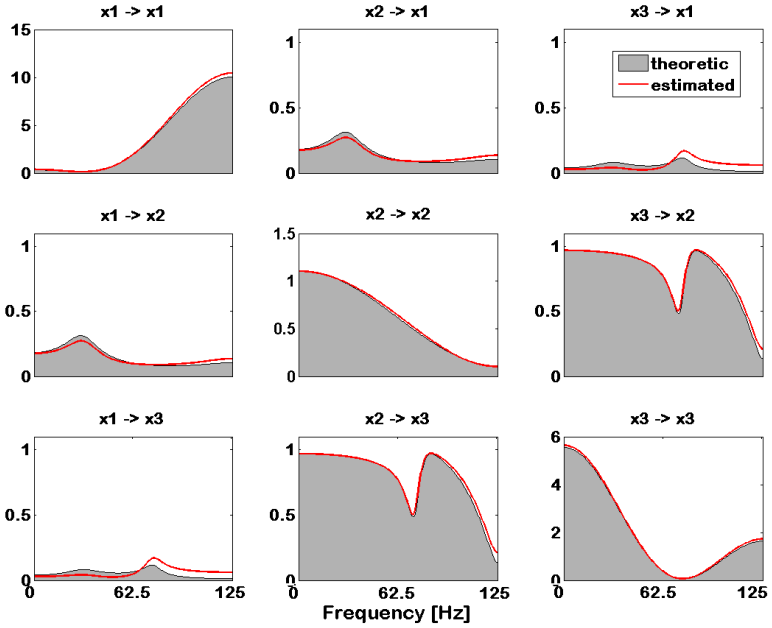

FIGURE 3.5: Diagonal elements: Inverse spectra of the processes. Off-diagonal elements: $PCoh$. The Grey area represents the theoretical measure while the estimated measures are plotted in red. This image shows how the $PCoh$ is a symmetrical measure, hence it is not able to convey information about the direction of connections: i.e. $|PCoh|_{1,2} = |PCoh|_{2,1}$ therefore it is not possible to establish that the direction of the connection is from x1 to x2 as showed in figure 3.2.

particular the variance of the process x_2 is 9 times higher than the variance of the other two processes. This results in the observation that the coefficient quantifying the coupling from x_2 to x_1 and x_3 are scaled by the signal amplitude and thus underestimated. The scaling is corrected introducing the normalisation by the variances of the processes as indicated in equation 3.19: in fact one can observe in figure 3.8 that $DC_{3,2}(f)$ exhibits higher values than $DC_{2,3}(f)$, as expected. The same holds for the PDC that correctly estimated a higher causality in the direction $x_2 \rightarrow x_3$ than in the direction $x_3 \rightarrow x_2$, differently from the analogous $orPDC$.

As this example and the example in Baccalà & Sameshima (2007) (where the PDC was originally introduced) show, if the source signal variance is high with respect to the other signals (i.e. in this case $\sigma_2^2 = 9$ vs $\sigma_3^2 = 1$), the gains of the connections will be 'erroneously' reduced if estimated with the DTF and $orPDC$. For this reason, in order to appropriately estimate the connection strength, it is important to include the variances of the signals to compensate for the scaling (as in DC and PDC). This may be very important to properly estimate the functional structure in the case of highly unbalanced signals amplitudes. For this reason, and because of their straightforward interpretation in terms of power (or inverse power) transfer [8], in this study we will

[8] the decompositions 3.20 and 3.24, very useful for the interpretation in term of power content, hold only for DC and PDC respectively, not for DTF and $orPDC$

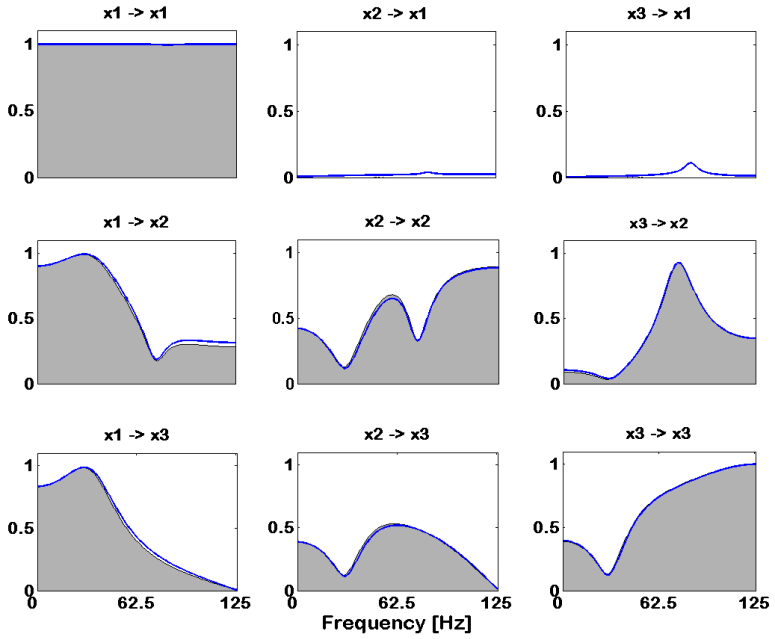

FIGURE 3.6: Directed transfer Function DTF as a function of frequency. The Grey area represents the theoretical DTF, in blue is shown the estimated DTF. Differently from what observed for the Coh and $PCoh$ in the previous figures, the values of the DTF are not symmetrical hence enabling the determination of the coupling direction: i.e. $|DTF|_{1,2}$ (first row, second column) is close to zero while $|DTF|_{2,1}$ (second row, first column) is different from zero thus indicating that a link from x1 to x2 exists, but not from x2 to x1 (as showed in the connectivity model in figure 3.2).The DTF is however different from zero for indirect connections, i.e. x1 $\rightarrow$ x2 $\rightarrow$ x3 results in the profile of $|DTF|_{3,1}$ (last row, first column) being different from zero.

consider the DC and PDC only.

Figure 3.10 provides a graphical representation of how the spectrum of x_2 can be decomposed into power contributions from all the other processes weighted by the respective DC values (squared). As previously mentioned the power spectrum of each signal can be decomposed into a sum of contributions from the spectra of all the time series weighted by the square of the respective $DC_{i,j}$, as expressed in equation 3.20. In the example given in figure 3.10 we observe that the power peak of $S_{2,2}(f)$ at f_1 is entirely due to x_1, while the second peak at f_3 is only in part due to x_3, because of the bidirectional interaction between x_2 and x_3: the oscillation at f_3 is generated in x_3, then transmitted to x_2 and backward to x_3. This example clarifies the interpretation of $DC_{i,j}$ as the normalised portion of the spectrum of x_j that is transferred to x_i.

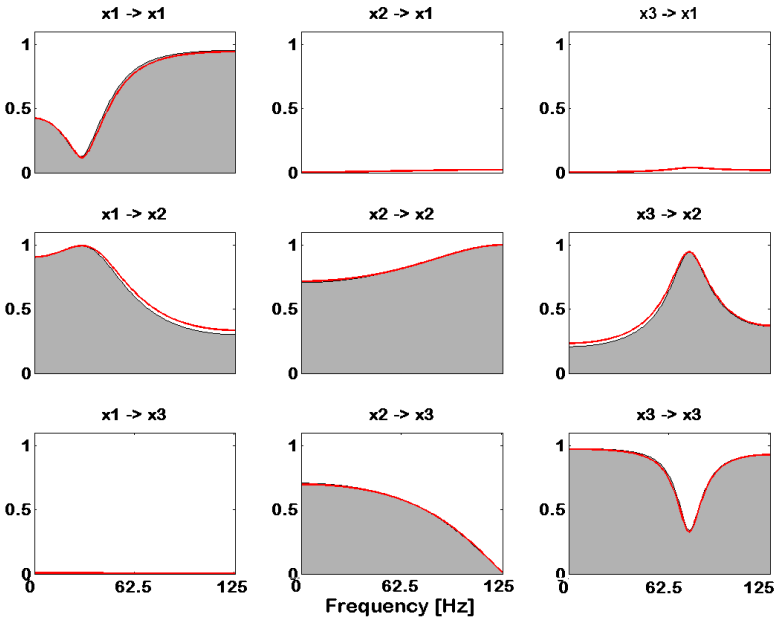

FIGURE 3.7: Ordinary Partial directed Coherence $orPDC$ as a function of frequency. The Grey area represents the theoretical $orPDC$, in red is shown the estimated $orPDC$. An inspection of the $orPDC$ profiles allows to identify the directed causal structure of the imposed connectivity model: $orPDC$ is different from zero only for the imposed connections showed in figure 3.2.

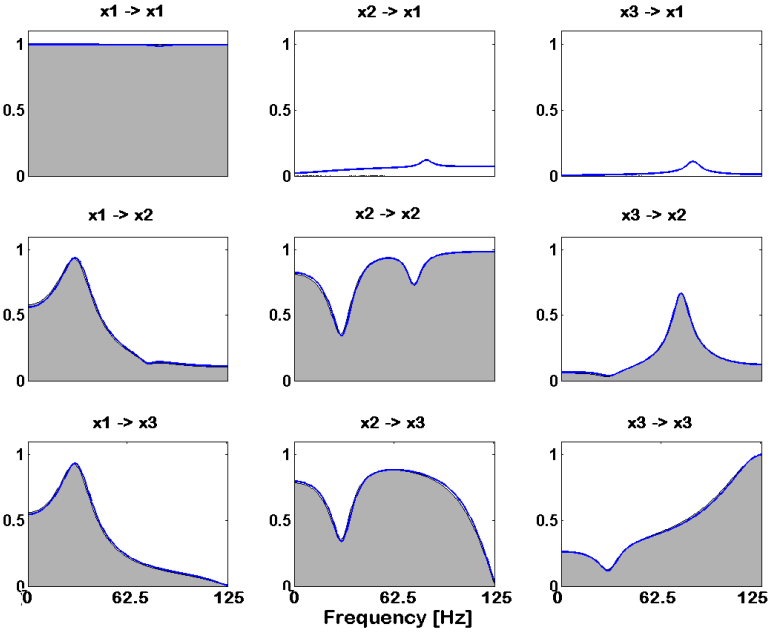

FIGURE 3.8: Directed Coherence DC as a function of frequency. The DC differs from the DTF because it includes the variance of the residuals (compare equations 3.19 and 3.21) and it therefore takes in account the effect of signal scaling in the computation of directed causality. The Grey area represents the theoretical DC, in blue is shown the estimated DC. DC profiles are slightly different from those observed in figure 3.6 for the DTF (i.e. compare $|DC|_{3,2}$ and $|DTF|_{3,2}$): due to the normalization by signal variance, DC more correctly estimate (with respect to DTF) the strength of imposed connections (see figure 3.2).

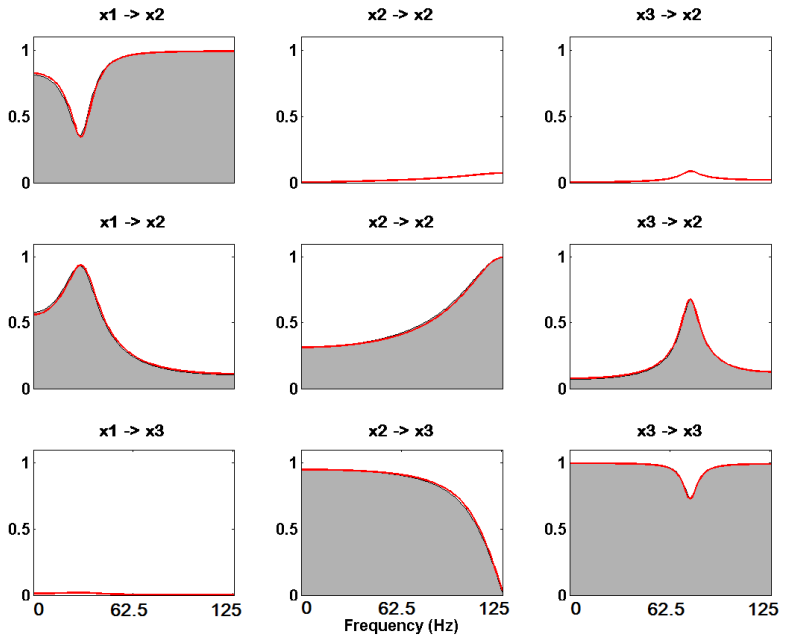

FIGURE 3.9: Generalized Partial directed Coherence PDC as a function of frequency. The PDC differs from the $orPDC$ because it includes the variance of the residuals (equations 3.23 and 3.25) and it therefore takes in account the effect of signal scaling in the computation of directed causality. The Grey area represents the theoretical PDC, in red is shown the estimated PDC. PDC profiles are slightly different from those observed in figure 3.7 for the $orPDC$ (i.e. compare $|PDC|_{3,2}$ and $|orPDC|_{3,2}$): due to the normalization by signal variance, PDC more correctly estimate (with respect to $orPDC$) the strength of imposed connections (see figure 3.2).

3.3 MVAR Model Identification and Validation

3.3.1 MVAR model estimation

As discussed in the previous sections, one necessary precondition for an accurate estimation of a MVAR model from the observed dataset is that the latter is a stationary, stable process (Lütkepohl 2005). This is commonly valid for records of resting EEG data, but caution is required in considering extended segment of EEG data[9] and the transition between different EEG patterns. On the other hand there is a concern that the number of samples considered for the analysis is sufficient to accurately fit the model. Given a M-variate dataset and a model of order p it is necessary to estimate M^2p number of free parameters for the model fitting, therefore a minimum of M^2p data point is required. However, in practice it is recommended to consider a number of data points larger than about $10M^2p$ for an accurate model estimation (Schlögl & Supp 2006). In this work the length of time series was chosen so as to include a number of data points sufficient to estimate the free parameters and at the same time to analyse the shortest EEG segment

[9]The approximate time during which an EEG signal can be considered stationary varies with the condition the EEG was recorded in; EEG epochs of 1-5 s length were found approximately stationary (Sugimoto et al. 1978, Gath & Inbar 1996)

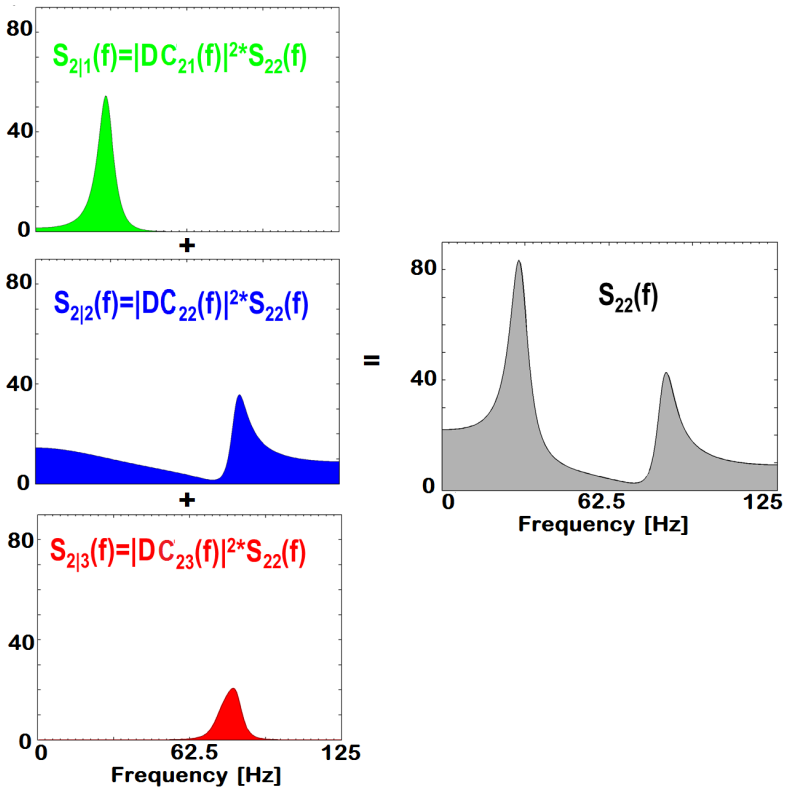

FIGURE 3.10: Example of spectral decomposition. The Spectrum of x_2 $S_{2,2}(f)$ (grey area) is expressed as a sum of 3 contributes from the spectrum of x_1 (green), x_2 itself (blue) and x_3 (red) weighted by the respective DC squared functions. This figure shows how $DC^2_{i,j}$ can be interpreted as the normalized spectrum of x_i due to the signal x_j.

in order to approximate the stationary condition. Short data windows are also desirable in monitoring applications, in order to quickly respond to changes in the signals.

3.3.1.1 MVAR estimation algorithm

Different approaches can be used to estimate a MVAR model from the observed time series, for instance multivariate Least Square (LS) algorithms, lattice algorithms or approaches grounded on the Kalman filter (Lütkepohl 1993). Since all these approaches are based on the minimization of the residuals (prediction error) there are small differences in their performances, especially when a sufficient number of data points are considered for the estimation (Schlögl 2006). However it has been shown that the Nuttal-Strand (1976) and the ARfit (Schneider & Neumaier 2001) algorithms yield the best results, giving a smaller prediction error and describing more accurately the properties of data. In this work the multivariate LS estimator implemented in the ARfit package (Matlab) will be used: it applies LS algorithm to high-dimensional EEG data in order to calculate the matrix of parameters $\mathbf{A}(l)$ and to give an estimate of the residual noise

covariance $\mathbf{\Sigma}_\varepsilon$. A detailed derivation of the LS estimators and their asymptotic is found in Lütkepohl (1993), Chapter 3; we will refer widely to this extensive work.

Given the M-variate process:

$$\mathbf{x}(n) = \sum_{l=1}^{p} \mathbf{A}(l)\mathbf{x}(n-l) + \varepsilon(n), n = 1, \cdots, N \tag{3.29}$$

we will call the time series of p previous samples (with respect to the actual sample n) the *predictor* vector $\mathbf{u}$

$$\mathbf{u}_n = \begin{bmatrix} \mathbf{x}_{n-1} \\ \mathbf{x}_{n-2} \\ \cdots \\ \mathbf{x}_{n-p} \end{bmatrix} \tag{3.30}$$

and cast the M-channels MVAR model in the form of a regression model as follows

$$\mathbf{x}_n = \mathbf{A}\mathbf{u}_n + \boldsymbol{\varepsilon}_n, n = 1, ..., N \tag{3.31}$$

Using the following notation is useful to express the MVAR model in more compact ways: we therefore define

$$\begin{aligned} \boldsymbol{X} &:= (\boldsymbol{x_1}, ..., \boldsymbol{x_N}) && [M \times N] \\ \boldsymbol{A} &:= (\boldsymbol{A_1}, ..., \boldsymbol{A_p}) && [M \times Mp] \\ \boldsymbol{U} &:= (\boldsymbol{u_1}, ..., \boldsymbol{u_N}) && [Mp \times N] \\ \boldsymbol{E} &:= (\boldsymbol{\varepsilon_1}, ..., \boldsymbol{\varepsilon_2}) && [M \times N] \\ \boldsymbol{x} &:= vec(\boldsymbol{X}) && [MN \times 1] \\ \boldsymbol{\alpha} &:= vec(\boldsymbol{A}) && [M^2p \times 1] \\ \boldsymbol{\varepsilon} &:= vec(\boldsymbol{E}) && [MN \times 1] \end{aligned} \tag{3.32}$$

where $\boldsymbol{x}_n$, $\boldsymbol{u}_n$ and $\boldsymbol{\varepsilon}$ are the M-variate vectors of the time series, the predictor vector and the innovation respectively and *vec* is the column stacking operator[10]. We can therefore

[10]Given a $[m \times n]$ matrix $\boldsymbol{A} = (\boldsymbol{a_1}, ..., \boldsymbol{a_n})$ the 'vec operator' transforms $\boldsymbol{A}$ in a $[mn \times 1]$ vector by stacking its column, i.e.

$$vec(A) = \begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \cdots \\ \mathbf{a_n} \\ \end{bmatrix} \tag{3.33}$$

write the MVAR model expression in compact form

$$\mathbf{X} = \mathbf{AU} + \boldsymbol{E} \tag{3.35}$$

or in the column stack form:

$$\mathbf{x} = vec(\boldsymbol{A}\mathbf{U}) + \boldsymbol{\varepsilon} \tag{3.36}$$

.

The multivariate LS algorithm chooses an estimator that minimizes the squared estimation error, i.e. the following cost function CF:

$$CF(\boldsymbol{\alpha}) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \tag{3.37}$$

that is equivalent, for the properties of the *vec* operator (see footnote), to

$$CF(\boldsymbol{\alpha}) = tr[(\boldsymbol{X} - \boldsymbol{AU})^T (\boldsymbol{X} - \boldsymbol{AU})] \tag{3.38}$$

It is demonstrated that the multivariate LS estimator is identical to the ordinary LS estimator (OLS) applied for each of the M equations of the M-variate model in 3.29 (Lütkepohl 1993). This result is due to Zellner (1962) who showed that, if the regressors in all equations are the same, the multivariate LS minimizes the sum of squared estimation errors and also the single (for each time series) errors separately.

The solution (that minimizes the CF)is given by (Lütkepohl 1993)

$$\hat{\mathbf{A}} = \mathbf{X}\mathbf{U}^T \left(\mathbf{U}\mathbf{U}^T\right)^{-1}. \tag{3.39}$$

In terms of the moment matrices

. The vec operator has the following useful property

$$vec(B^T)^T vec(A) = vec(A^T)^T vec(B) = tr(AB) = tr(BA) \tag{3.34}$$

$$\begin{aligned} \boldsymbol{\Upsilon} &= \sum_{n=1}^{N} \mathbf{u}_n \mathbf{u}_n^T \\ \mathbf{V} &= \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T \\ \boldsymbol{\Omega} &= \sum_{n=1}^{N} \mathbf{x}_n \mathbf{u}_n^T \end{aligned} \tag{3.40}$$

The estimated matrix of parameters can be expressed as

$$\hat{\boldsymbol{A}} = \boldsymbol{\Omega}\boldsymbol{\Upsilon}^{-1} \tag{3.41}$$

and the residual covariance matrix is

$$\hat{\boldsymbol{\Sigma}}_{\varepsilon} = \frac{1}{N - pM} \sum_{n=1}^{N} \hat{\boldsymbol{\varepsilon}_n} \hat{\boldsymbol{\varepsilon}}_n^T = \frac{1}{N - pM} \left(\mathbf{V} - \boldsymbol{\Omega}\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Omega}^T \right). \tag{3.42}$$

Under the assumption of a standard white noise process

$$\begin{aligned} E(\boldsymbol{\varepsilon_n}) &= 0; \\ \boldsymbol{\Sigma}_{\varepsilon} \; &non \; singular; \\ E[\boldsymbol{\varepsilon_n}\boldsymbol{\varepsilon_m}] &= 0, n \neq m. \end{aligned} \tag{3.43}$$

it is demonstrated that (Lütkepohl 1993) that the LS estimator of the MVAR model parameter is consistent [11] and asymptotic normal.

3.3.1.2 Model Order Selection

A critical issue in the estimation of an MVAR model is the selection of the model order. For this purpose it is useful to make use of criteria for an adequate selection. The most common approach consists in minimizing an information criterion over a range of model orders. Commonly used criteria are functions of the residual covariance matrix $\hat{\boldsymbol{\Sigma}}_{\varepsilon}(p)$ estimated for a model order p (equation 3.42) and a second term that is a function of the number of free parameters to estimate, which increases with increasing model order.

[11] An estimator of a quantity $\mathbf{A}$ is called *consistent* if

$$lim_{N \to \infty} Pr(\hat{\mathbf{A}} = \mathbf{A}) = 1 \tag{3.44}$$

By minimizing both terms the forecast precision error is minimized and the degrees of freedom of the model are adjusted to avoid data over-fitting with too many parameters. Akaike (1969,1971) introduced a criterion called final prediction error (FPE) that is:

$$FPE(p) = \left[\frac{T + Mp + 1}{T - Mp - 1}\right]^{M} det\hat{\mathbf{\Sigma}}_{\varepsilon}(p) \tag{3.45}$$

and its logarithm

$$ln(FPE(p)) = ln\left[\frac{T + Mp + 1}{T - Mp - 1}\right]^{M} + ln(det\hat{\mathbf{\Sigma}}_{\varepsilon}(p)) \tag{3.46}$$

where $\hat{\Sigma}_{\varepsilon}(p)$ is an estimator of the residuals covariance matrix and T is the number of time samples.

Another commonly used criterion is the Schwarz's Bayesian Criterion (SBC) (Schwarz, 1978)

$$SBC(p) = ln|\hat{\mathbf{\Sigma}}_{\varepsilon}(p)| + \frac{lnT}{T} pM^2 \tag{3.47}$$

For a given criterion the optimum model order is the one that minimizes the information criterion. The FPE criterion (and the similar Akaike Information Criterion, AIC) tends to overestimate the true order of the MVAR model and thus the model order estimate obtained using FPE is not consistent. On the other hand, the SBC criterion severely penalizes large model orders and its estimate of the optimal order is consistent for any dimension of the model (Lütkepohl 1993). Moreover, as observed in EEG applications, usually the FPE criterion does not show a distinct global minimum over a reasonable model orders range, while it is common for the SBC to show a clearer minimum. For these reasons, even if the FPE (and AIC) criterion may have better properties for small samples and for forecasting problems (Lütkepohl 1993), in this work the optimal model order p_{opt} will be chosen on the basis of the SBC criterion, since the aim is to accurately estimate the MVAR model.

3.3.2 Model Validation

As stated before, it is recommended, once estimated the MVAR model, to perform model validation steps to insure a correct interpretation of the obtained estimators. The model validation consists in making use of a range of different tools to check if the MVAR model is adequately estimated. Although these tools have been rarely applied in the literature

and in applications of causality estimators (DC or PDC) for the analysis of EEG data, they represent an important protective measure against making erroneous inferences from misleading results. A detailed description of the validation criteria can be found in Lütkepohl (2005) while in this context we will focus on two critical conditions: the whiteness and independence of residuals.

A simple approach used to validate a MVAR model is to compare the spectra estimated with the MVAR parametric method with another non-parametric approach (Mullen 2010). Even if the assumption that the non-parametric spectra are the optimal estimates may not always be justified (Burg, 1975), this represents a useful validation procedure.

An MVAR model fitting a multiple time series can be regarded as a filter that transforms white noise into the given structured data series (figure 3.11). If we have adequately modelled the full causal structure of the data, the residuals should be white. In this regard a validation criterion is to check whether the residuals are close enough to white noise. If the residuals are correlated in some extent some correlation structure in the data has not been described.

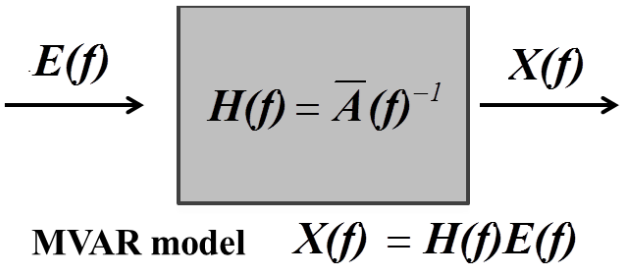

FIGURE 3.11: Schematic representation of a MVAR model as a linear filter having in input a white process $E(f)$ and in output the multivariate dataset $X(f)$: if the filter transfer function $H(f)$ is estimated correctly, then it models all the causal structure of the process $X(f)$ and the residual $E(f)$ are white.

Testing the whiteness of residuals typically involves the analysis of the autocorrelation matrix coefficients up to a fixed lag $h > 0$. A simple test for checking the null hypothesis that a given multiple time series is generated by a white process is based on the asymptotic distributions of white noise autocorrelations and it is called the *Autocorrelation Function Test*.

Given a M-dimensional vector of identically independently distributed (i.i.d.) time series, the covariance matrices at different time lags k are estimated as:

$$\mathbf{\Sigma}_{\varepsilon,k} = \frac{1}{N}\sum_{t=k+1}^{N} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-k}, i = 1, ..., h \tag{3.48}$$

where N is the number of time samples. The correlation coefficient matrix at lag k $\boldsymbol{P}_k$ is obtained from $\mathbf{\Sigma}_{\varepsilon,i}$ as follows

$$\boldsymbol{P}_k = \boldsymbol{D}_{\varepsilon,0}^{-1} \mathbf{\Sigma}_{\varepsilon,k} \boldsymbol{D}_{\varepsilon,0}^{-1} \tag{3.49}$$

where $\boldsymbol{D}_{\varepsilon,0}$ is a diagonal matrix with elements equals to the square roots of the elements of $\mathbf{\Sigma}_{\varepsilon,0}$, i.e. the covariance matrix of residuals at lag zero. The matrix $\boldsymbol{P}_k$ is an estimator of the true correlation matrix for $k \neq 0$. A typical element of $\boldsymbol{P}_k$ is $\rho_{ij,k}$:

$$\rho_{ij,k} = \frac{\sigma_{ij,k}}{\sqrt{\sigma_{ii,0}}\sqrt{\sigma_{jj,0}}} \tag{3.50}$$

where $\sigma_{ij,k}$ is a generic element of the covariance matrix $\mathbf{\Sigma}_{\varepsilon,k}$ at time lag k.

The matrix of autocorrelations and its vector (obtained stacking its columns) up to a time lag h are

$$\begin{aligned} \boldsymbol{P}_h &= \big(\boldsymbol{P}_1 \cdots \boldsymbol{P}_h\big) \\ \boldsymbol{\rho}_h &= vec\big(\boldsymbol{P}_h\big) \end{aligned} \tag{3.51}$$

It is demonstrated that the vector $\sqrt{N}\boldsymbol{\rho}_h$ for $h \geq 1$ converges in distribution to a multivariate normal distribution, therefore the variances of the asymptotic distributions of the elements of $\sqrt{N}\boldsymbol{\rho}_h$ are unity. The null hypothesis of autocorrelation equal to zero is accepted at the 5% level if

$$|\sqrt{N}\boldsymbol{\rho}_h| < 1.96 \tag{3.52}$$

The statistic of the Autocorrelation Function test is a function of the number of autocorrelation coefficients for whom the condition in equation 3.52 is valid, with respect to the total number of autocorrelation coefficients considered:

$$\zeta = \frac{num[|\sqrt{N}\boldsymbol{\rho}_h| < 1.96]}{num[\boldsymbol{\rho}_h]} \tag{3.53}$$

If $\zeta > 0.95$ then the null hypothesis of white residuals is accepted at 5% level of significance. When using this test it is important to consider that for small samples the test may be conservative in the sense that it may tend to accept the null hypothesis of whiteness more often than indicated by the significance level, because $\frac{1}{\sqrt{N}}$ may considerably differ from the true variance of the correlation coefficients. Moreover this test is specifically designed for individual coefficients and, if applied in the multivariate case, it assumes that the elements of the vector $\boldsymbol{\rho}_h$ are independent: this may be not necessarily true for the elements of $\boldsymbol{P}_h$.

The Portmanteau tests (Lütkepohl 2005) are on contrast specifically designed to test the null hypothesis of white residuals of a multivariate process. This class of tests is based on the fact that, under the null hypothesis, the statistic

$$\boldsymbol{Q}_h = N \sum_{i=1}^{h} tr\left(\boldsymbol{\Sigma}_{\varepsilon,i}^T \boldsymbol{\Sigma}_{\varepsilon,0}^{-1} \boldsymbol{\Sigma}_{\varepsilon,i} \boldsymbol{\Sigma}_{\varepsilon,0}^{-1}\right) \tag{3.54}$$

approximates the asymptotic χ^2 distribution with $M^2(h-p)$ degrees of freedom (d.o.f.) (M is the number of time series, h the maximum time lag the covariance is computed for, p the order of the MVAR model). A statistic test can hence be performed comparing the test statistic with the χ^2 distribution. A p-value can be obtained comparing $\boldsymbol{Q}_h$ with the cumulative distribution function (cdf) of the χ^2 distribution. If the p-value is greater than the significance level $\alpha = 0.05$ the null hypothesis of whiteness cannot be rejected (residuals are white). In order to improve the performances of the test for small time samples, the statistic in equation 3.54 has been modified in what is called the Ljung-Box Portmanteau test (LBP) (Lütkepohl 2005).

Testing the whiteness of residuals represents an important tool to check the validity of MVAR estimation. More specifically it represents a way to check if the lagged correlation structure of time series has been adequately estimated (only lagged correlations in fact are considered in whiteness tests). On the other hand the identification of a MVAR model should results in residuals uncorrelated also at lag $h = 0$. When this condition is violated there are significant zero-lag effects that have not been modeled by the MVAR model (Faes et al. 2001). The significance of instantaneous causality between time series can be tested checking the diagonality of the residual covariance matrix $\boldsymbol{\Sigma}_{\varepsilon,0}$. In fact statistical test of lack of instantaneous causality reduce to demonstrate that $\boldsymbol{\Sigma}_{\varepsilon,0}$ is diagonal. As will be detailed in the next chapter, instantaneous effects may seriously affect the estimation of causality, in particular from EEG multichannel data. In this regard checking the significance of instantaneous causality represents a relevant validation step for the correct interpretation of results. Leaving the description of the

impact of instantaneous effects on the estimation of causality to the next chapter, in this paragraph we will demonstrate that the non-diagonality of the residuals covariance matrix $\mathbf{\Sigma}_{\varepsilon,0}$ gives an indication that instantaneous correlations among time series exist.

Given a MVAR model of order p of an M-variate process as expressed in equation 3.55

$$\boldsymbol{x}(n) = \sum_{l=1}^{p} \boldsymbol{A}(l)\boldsymbol{x}(n-l) + \boldsymbol{\varepsilon}(n), n = 1, \cdots, N \tag{3.55}$$

with the covariance matrix of white residuals

$$\mathbf{\Sigma}_{\varepsilon,0} = \left\{\sigma_{ij}^2\right\}, \tag{3.56}$$

this model can be can be expressed in such a way that residuals are uncorrelated. For this purpose the covariance matrix can be decomposed as follows, from the Choleski decomposition[12]

$$\mathbf{\Sigma}_{\varepsilon,0} = \boldsymbol{W}\mathbf{\Sigma}_{\omega,\mathbf{0}}\mathbf{W}^{\mathbf{T}} \tag{3.57}$$

where $\boldsymbol{W}$ is a lower triangular matrix with unit diagonal and $\mathbf{\Sigma}_{\omega,0}$ is a diagonal matrix having the same diagonal elements of $\mathbf{\Sigma}_{\varepsilon,0}$.

Premultiplying equation 3.55 by $\boldsymbol{W}^{-1}$ gives

$$\boldsymbol{W}^{-1}\boldsymbol{x}(n) = \sum_{l=1}^{p} \boldsymbol{A}^{\bullet}(l)\boldsymbol{x}(n-l) + \boldsymbol{\omega}(n), n = 1, \cdots, N \tag{3.58}$$

with

$$\begin{aligned} \boldsymbol{A}^{\bullet}(l) &= \boldsymbol{W}^{-1}\boldsymbol{A}(l) \\ \boldsymbol{\omega}(n) &= \boldsymbol{W}^{-1}\boldsymbol{\varepsilon}(n). \end{aligned} \tag{3.59}$$

Equation 3.58 can be written as

$$\boldsymbol{x}(n) = \boldsymbol{A}^{\bullet}(0)\boldsymbol{x}(n) + \sum_{l=1}^{p} \boldsymbol{A}^{\bullet}(l)\boldsymbol{x}(n-l) + \boldsymbol{\omega}(n), n = 1, \cdots, N \tag{3.60}$$

[12] If a matrix M is positive definite, then it exists a triangular matrix T such that $M = TT'$. The decomposition where T is lower triangular with positive main diagonal is called *Choleski decomposition*(Lütkepohl 1993).

where

$$\boldsymbol{A}^{\bullet}(0) = \boldsymbol{I} - \boldsymbol{W}^{-1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ a_{2,1}(0) & 0 & 0 & 0 \\ \cdots & \cdots & 0 & 0 \\ a_{M,1}(0) & \cdots & a_{M,M-1}(0) & 0 \end{bmatrix} \tag{3.61}$$

The matrix $\boldsymbol{A}^{\bullet}(0)$ represents the matrix of instantaneous effects (zero-lag dependencies) that, given the hypothesis (W lower triangular with unitary diagonal), is a lower triangular matrix with zero diagonal. The elements of the matrix of instantaneous effects are a function of the covariance of the residuals (equation 3.57). If the residuals are uncorrelated the matrix $\boldsymbol{W}^{-1}$ coincides with the identity matrix and as a result the matrix of instantaneous effects is null (equation 3.61). Based in this result, a statistical test of lack of instantaneous causality consists in demonstrating that $\boldsymbol{\Sigma}_{\varepsilon,0}$ is diagonal. In this work the covariance matrix of residuals will be inspected and its non-diagonality considered as an indication of the significance of zero-lag correlations among time series. Instantaneous correlations among residuals indicate that significant zero-lags interactions among time series have not been adequately modeled: In this case the issue of misleading results caused by instantaneous causality must be considered. In the next chapter a throughout description of the instantaneous connectivity issue and of the approaches proposed to model it will be given, together with an analysis of the reliability of these approaches and their impact on EEG causality estimation.

3.4 Summary

In this section a rationale for the development of spectral causality measures based on the factorisation of the Coherence and the Partial Coherence was given. A description of what the different measures aim to quantify has been provided with the help of an illustrative example. In this example the behaviour of all the estimators on data generated imposing a simple connectivity scheme was examined.

PDC and DC provide a frequency domain representation of multivariate time series grounded on the concept of Granger causality. The DC was shown to quantify the direct and indirect causality in the frequency domain as the normalised portion of the spectrum that is transferred from one process to another. The PDC provides a clearer representation of the structural interaction, with PDC equal to zero corresponding to the lack of direct causality. Both measures are able to identify the direction of information flow, something that cannot be inferred from the analysis of other spectral estimators such as

Coh, *PCoh* and their equivalent in the time domain. When $orPDC$ and DTF are normalised with respect to the variances of the processes (thus giving PDC and DC), they have the important property of scale invariance (Baccalà & Sameshima 2007); therefore PDC and DC may be less confounding as causality estimators in the case of time series with highly unbalanced variances. In previous work PDC and DC have been applied[13] to the simultaneous analysis of multichannel EEG in various physiological conditions and behavioral tasks (Astolfi et al. 2006, Astolfi, Cincotti, Mattia, Marciani, Baccala, de Vico Fallani, Salinari, Ursino, Zavaglia, Ding, Edgar, Miller, He & Babiloni 2007, Baccalá & Sameshima 2001, Kamiński et al. 1997, Kuś et al. 2008, Brzezicka et al. 2011) and were shown to be able to effectively estimate functional connectivity patterns, with a remarkable agreement with anatomical and neuroimaging based evidences (Kaminski & Blinowska 2014). They were shown to represent a useful tool to investigate the macroscopic network structure generating the EEG signal, whose information is mainly coded in frequency, and perform well also in case of non-linear and non-stationary interactions (Winterhalder et al. 2005).

PDC and DC are efficient estimators of directed connectivity provided that the conditions that guarantee a reliable fitting of the MVAR model (i.e. adequate signal length, appropriate model order choice etc.) are fulfilled. In the last section of this chapter we have more formally looked at crucial aspects for an accurate MVAR model fitting describing the ARfit algorithm for the estimation of MVAR coefficients and the FPE and SBC criteria to determine the optimal model order. Once a MVAR is estimated from a multivariate dataset, it is recommended to check for the model validity to avoid misleading interpretation of results. As described in the last paragraph a series of statistical tools are available to check if the MVAR model is adequately estimated: in particular we have shown that the conditions of whiteness and independence of residual are important indicators that the lagged and zero-lag causality structure of the dataset are correctly modelled. The critical issue of instantaneous causality and its effects on the estimation of EEG connectivity has been often ignored in the literature: it will be the focus of the next chapter where methods to deal with its adverse effects will be tested and its impact on lagged DC and PDC characterized.

[13] mainly in their non-normalised version $orPDC$ and DTF

Chapter 4

The Impact of Instantaneous Effects on the Estimation of Scalp Connectivity

In this chapter the issue of instantaneous causality and its impact on DC and PDC estimation will be addressed. Firstly we will review the work carried out in literature and formally outline an approach introduced to deal with the adverse effects of zero-lag correlation on lagged causality: the extended MVAR model (eMVAR). We will test the reliability of the eMVAR approach with a particular emphasis on EEG applications. Finally, in section 4.2, the instantaneous mixing of sources (volume conduction) effects on the estimation of scalp connectivity will be more thoroughly characterized by means of a realistic model of EEG signal generation.

4.1 The impact of instantaneous causality on the estimation of DC and PDC between scalp channels

As described in section 3.2.1, the key concept in Granger causality is the exclusive reference to past samples in the prediction of the time series and hence it is fairly simple to deal with in the context of MVAR models. The term *causality* suggests a cause-effect (past-present) relationship between time series. When also the influence of present samples is considered then it is usual to speak of instantaneous causality (Lütkepohl 1993): it refers to the zero-lag effect of one signal upon another signal. Instantaneous causality does not say anything about the cause-effect relationship because the direction of causation cannot be determined in this case. The classical MVAR framework does not

include zero-lag dependencies therefore it does not model instantaneous causality. We can only argue that if the residuals are correlated (at lag zero) then some instantaneous correlation structure exists in the data that has not been described by the estimated MVAR model and the assessment of lagged causality may be erroneous. The problem of zero-lag interactions is of particular relevance for EEG signal analysis because it has been suggested that in EEG data instantaneous causality may be related mainly to non-physiological phenomena such as volume conduction (Nolte et al. 2008): deep cortical sources of EEG activity that spread through the tissues of the head on the scalp could result in instantaneous dependencies among surface recordings .
The hypothesis that DC/PDC are affected by instantaneous causality is debated. Some authors claim that these estimators are only sensitive to phase shifts therefore they are not affected by zero-lag effects (Kaminski & Blinowska 2014). In contrast, various studies (Faes et al. 2013, Faes & Nollo 2010, Billinger et al. 2016) demonstrated both analytically and by means of simulations on simple MVAR models that zero-lag interactions adversely influence lagged causality estimated with DC/PDC. These works have shown in fact that when significant instantaneous dependencies among signals are present, the omission of adequate modelling of zero-lag effects (as in the classical MVAR modeling) can lead to misleading results; in particular instantaneous effects can change the values of time-lagged coefficients and introduce spurious connections (Faes & Nollo 2010). Billinger and colleagues (Billinger et al. 2016) have shown both analytically and by a numerical simulation involving 3 cortical sources and 3 surface signals that volume conduction generates spurious (non physiological) DTF scalp connections.

As a possible solution an *extended* MVAR (eMVAR) approach modelling the full causal structure (instantaneous and lagged) of time series has been proposed by Faes & Nollo (2010). The eMVAR approach was tested on simple MVAR processes showing to reduce estimation errors with respect to the classical MVAR model, provided that the assumptions for its identification were respected. The eMVAR framework was also tested on two dataset recorded from, respectively, the cardiovascular and neurophysiologic systems (Faes et al. 2013) showing promising results, but an analysis of the reliability of the eMVAR algorithm applied for EEG causality estimation is missing. Moreover a characterization of the instantaneous causality generated in EEG time series by volume conduction and its impact on the estimation of Granger causality is lacking, because this issue has been ignored in the majority of practical applications.

In the following sections we will first describe the rationale and assumptions of the eMVAR framework and then outline the objectives of this study.

4.1.1 The extended MVAR model

As showed in the previous chapter, the traditional MVAR model (which in this chapter will now be referred to as a strictly causal MVAR model scMVAR- from now on) used for the estimation of causality from EEG recordings, takes into account only the linear influences of past samples on present samples .

$$\boldsymbol{x}(n) = \sum_{l=1}^{p} \boldsymbol{A}(l)\boldsymbol{x}(n-l) + \boldsymbol{\varepsilon}(n), n = 1, \cdots, N \tag{4.1}$$

The scMVAR does not model instantaneous dependencies ($l = 0$) among time series, therefore any significant zero-lag influence among signals is transferred in the model residuals generating a correlation structure within them and making the innovation covariance matrix not diagonal (Lütkepohl 2005). This results in two practical issues:

- the model is not estimating the causal structure correctly and the performances of PDC and DC in estimating lagged causality are potentially degraded (Faes & Nollo 2010, Faes et al. 2013), as stated in the introduction.
- The factorizations expressed in equations 3.18 and 3.22 are not valid, because they are based on the assumption of uncorrelated residuals; therefore the interpretation of the estimators as normalized portions of spectrum (or inverse spectrum) transferred between signals is not correct, and the interpretation of results may be misleading.

In order to integrate both instantaneous and lagged effects in multivariate time series modelling, an alternative framework was introduced (Shimizu et al. 2006, Hyvarinen et al. 2010, Faes & Nollo 2010) and defined by the following expression (note the lag index l now ranges from 0 to p):

$$\boldsymbol{x}(n) = \sum_{l=0}^{p} \boldsymbol{B}(l)\boldsymbol{x}(n-l) + \boldsymbol{\omega}(n), n = 1, \cdots, N \tag{4.2}$$

with residual covariance matrix

$$\boldsymbol{\Sigma}_{\omega} = diag\left\{\sigma_{ii}^2\right\} \tag{4.3}$$

As previously indicated, this new approach is named extended MVAR (eMVAR) to distinguish it from the traditional multivariate modelling. The hypothesis for the scMVAR

model is that the residuals (or innovation processes) are white, while for a model that includes instantaneous effect (eMVAR) the assumption is that residuals are white and uncorrelated (Lütkepohl 1993)[1]. In order to find the relation between the scMVAR and eMVAR model we can express the equation 4.2 as follows

$$\boldsymbol{x}(n) = \boldsymbol{B}(0)\boldsymbol{x}(n) + \sum_{l=1}^{p} \boldsymbol{B}(l)\boldsymbol{x}(n-l) + \boldsymbol{\omega}(n), n = 1, \cdots, N \tag{4.5}$$

$$\big(\boldsymbol{I} - \boldsymbol{B}(0)\big)\boldsymbol{x}(n) = \sum_{l=1}^{p} \boldsymbol{B}(l)\boldsymbol{x}(n-l) + \boldsymbol{\omega}(n), n = 1, \cdots, N \tag{4.6}$$

$$\tag{4.7}$$

Therefore the eMVAR model can be expressed in a scMVAR 'form' using the matrix $\boldsymbol{L} = \big[\boldsymbol{I} - \boldsymbol{B}(0)\big]^{-1}$ as follows:

$$\boldsymbol{x}(n) = \sum_{l=1}^{p} \boldsymbol{L}\boldsymbol{B}(l)\boldsymbol{x}(n-l) + \boldsymbol{L}\boldsymbol{\omega}(n), n = 1, \cdots, N \tag{4.8}$$

Hence the relation between the two models, comparing equations 4.1 and 4.8, is

$$\boldsymbol{A}(l) = \boldsymbol{L}\boldsymbol{B}(l) \tag{4.9}$$

$$\boldsymbol{\varepsilon}(n) = \boldsymbol{L}\boldsymbol{\omega}(n) \tag{4.10}$$

$$\boldsymbol{\Sigma}_{\varepsilon} = \boldsymbol{L}\boldsymbol{\Sigma}_{\omega}\boldsymbol{L}^{T} \tag{4.11}$$

From these considerations it is evident that the matrices of lagged effects for the two models are different (because of the multiplication by L) and it is (mathematically) demonstrated that instantaneous interactions affect the values of parameters expressing lagged effects. In the absence of instantaneous effects, the two models coincide ($\boldsymbol{B}(0) = \boldsymbol{0}$, $\boldsymbol{L} = \boldsymbol{I}$). Moreover equations 4.9 and 4.10 show that an eMVAR model can be estimated from the identification of a scMVAR model provided that $\boldsymbol{L}$ (which is in turn function of the matrix of instantaneous effects $\boldsymbol{B}(0)$) is known. If this is the case, once the matrix $\boldsymbol{A}(l)$ and the vector of residuals $\boldsymbol{\varepsilon}(n)$ have been identified from the estimation

[1] Note that in this contest we use the term *white* to refer to time series that have the following property

$$E[\boldsymbol{\varepsilon}(n)\boldsymbol{\varepsilon}(m)] = 0, n \neq m \tag{4.4}$$

and that are therefore white in the temporal dimension. With the term *uncorrelated* we indicate time series that are independent 'spatially', therefore their multivariate covariance matrix is diagonal.

of the scMVAR, the correspondent matrix $\boldsymbol{B}(l)$ and vector of innovations are obtained from $\boldsymbol{L}^{-1} = \boldsymbol{I} - \boldsymbol{B}(0)$.

The estimation of the zero-lag effects matrix however is challenging. While for a lagged influence the direction of causality is from the less recent sample to the most recent one, this directional information is lacking in the case of instantaneous interactions. Therefore, while the knowledge of the covariance matrix is sufficient for the estimation of the matrices of lagged effects, additional information is required to identify the matrix of instantaneous effects. Following the fitting of the correspondent scMVAR, two possible approaches can be adopted to estimate the instantaneous matrix $\boldsymbol{B}(0)$ (and the eMVAR model as a consequence):

1. A priori approach. Imposing a priori the structure of the instantaneous causality, causally ordering the time series, for instance in accordance with the order the signals are recorded, and hence imposing the direction but not the strength of instantaneous effects[2]. This approach, called *a priori* eMVAR (Faes & Nollo 2010), is however not suitable for EEG signals since they are measured simultaneously and the direction and order of zero-lag effects is not known a priori.

2. Non-Gaussian approach. This approach was developed based on arguments similar to those behind the use of non-Gaussianity in finding independent components in independent components analysis (ICA). It exploits the non-Gaussianity of the corresponding scMVAR model residuals to estimate the matrix $\boldsymbol{B}(0)$ of the eMVAR model. As previously stated, differently from what happens with lagged effects, the concept of instantaneous causality does not give information about the cause and effect relation. The direction of instantaneous causality cannot be derived from the covariance matrix of the multivariate process that is symmetrical, therefore further knowledge about the relationship between variables must be included (Lütkepohl 2005). Algorithms based only on second-order statistics (i.e. PCA, based on the covariance matrix) generally fail to estimate the full causal structure. For example, in the case of two variables x_1 and x_2 , such methods cannot prefer the instantaneous model $x_1 \rightarrow x_2$ over $x_2 \rightarrow x_1$, because they have the same value in the covariance matrix, therefore different models are plausible for the same set of data[3]. Other information, for instance obtained from higher order statistics, has to be exploited to identify a unique model underlying the data. Higher order statistics (e.g. skewness, and kurtosis) are not informative when the data are Gaussian.

[2]The amplitude of the coefficient of $\boldsymbol{B}(0)$ can then be estimated from a decomposition of the residual covariance matrix, as showed in equation 4.11

[3]Even if both models ($\boldsymbol{B}(0)$ matrices) fit the data, only one $\boldsymbol{B}(0)$ is 'correct': if the allocation of zero-lag effects is not correct this may lead to misleading values of the matrix of lagged coefficients (see equation 4.9).

However in the case of non-Gaussian residuals, one can exploit the information of higher order statistics and estimate the full structure of instantaneous effects (Dodge & Rousson 2001, Shimizu et al. 2006, 2005).

In equation 4.10 the matrix $\boldsymbol{L}$ can be considered as a mixing matrix that transforms independent time series $\boldsymbol{\omega}$ into correlated time series $\boldsymbol{\varepsilon}$. Equation 4.10 in fact defines the standard linear ICA model. This approach (non-Gaussian eMVAR) has been proposed by Shimizu and colleagues (Shimizu et al. 2006, Hyvarinen et al. 2010) in order to estimate causal influences that occur either instantaneously or with time lags. It is rooted on the following assumptions:

(a) The observed data can be arranged in a causal order, meaning they can be represented by a directed acyclic graph (DAG): A DAG is a graph that contains no cycles (Pearl 1993), i.e. there is no way to start at some vertex and loop again to it through a sequence of directed edges. The assumption of acyclicity insure that no later variable causes an earlier variable. This property is reflected in the matrix $\boldsymbol{B}(0)$ having zero diagonal elements and the existence of a a permutation matrix that makes $\boldsymbol{B}(0)$ lower triangular. Note that this assumption is very strict and may be flawed in the case of EEG signals: it is likely that a cortical source may instantaneously affect two or more scalp signals thus generating between them a bidirectional zero-lag connection. This issue was ignored in previous application of eMVAR but may be central in case of eMVAR application to recorded EEG.

(b) The residuals are independent random variables with a non-Gaussian distribution. This assumption is a fundamental hypothesis for the ability to apply ICA and the accurate estimation of the de-mixing matrix $\boldsymbol{L}^{-1}$.

The de-mixing matrix $\hat{\boldsymbol{W}} = \boldsymbol{L}^{-1}$ that transforms correlated time series $\boldsymbol{\varepsilon}$ into independent components $\boldsymbol{\omega}$ is estimated applying ICA to the residuals of the scMVAR model, $\boldsymbol{\varepsilon}$. $\hat{\boldsymbol{W}}$ is however an unordered and non-normalised version of the true de-mixing matrix $\boldsymbol{L}^{-1}$, since the ICA gives the factors in a random order. Therefore additional steps are required to establish the correct order of components and hence the correct correspondence between residuals $\boldsymbol{\varepsilon}$ and $\boldsymbol{\omega}$. This step is performed applying the constraint that the matrix $\boldsymbol{B}(0)$ has a null diagonal (by definition) hence the matrix $\hat{W} = \boldsymbol{L}^{-1} = \boldsymbol{I} - \boldsymbol{B}(0)$ has diagonal elements equal to 1[4]. More details about the estimation algorithm can be found in Shimizu et al. (2006), Hyvarinen et al. (2010).

[4]It is demonstrated in fact that for a DAG model it exists one and only one permutation matrix $\boldsymbol{P}$ that would give a matrix $\tilde{\boldsymbol{W}} = \boldsymbol{P}\hat{\boldsymbol{W}}\boldsymbol{P}^T$ with no zeros on the diagonal elements. Once $\tilde{\boldsymbol{W}}$ has been obtained, the rows are normalized in order to obtain a diagonal with ones and then $\boldsymbol{B}(0)$ is estimated as $\hat{B}(0) = \boldsymbol{I} - \tilde{W}$

The scMVAR does not model instantaneous dependencies among time series and any significant zero-lag influence may affect the estimation of causality, 'appearing' as lagged. On the contrary, the eMVAR model (provided that the estimation of instantaneous and lagged effects is accurate) opens the possibility to estimate Granger causality considering only pure lagged effect, hence excluding the contribution of the estimated matrix $\boldsymbol{B}(0)$ of zero-lag effects in the computation of spectral causality. Being $\boldsymbol{B}(l), l = 0, ..., p$ the matrix of parameters of the eMVAR model then

$$\boldsymbol{B}_{ext}(f) = \boldsymbol{I} - \sum_{l=0}^{p} \boldsymbol{B}(l) e^{\frac{-2\pi f l}{f_{samp}}} \tag{4.12}$$

is the eMVAR coefficient matrix in the frequency domain. The spectral coefficient matrix that includes only lagged causality is then obtained by removing from $\boldsymbol{B}(f)$ the zero-lag matrix $\boldsymbol{B}(0)$:

$$\boldsymbol{B}_{lag}(f) = \boldsymbol{I} - \sum_{l=1}^{p} \boldsymbol{B}(l) e^{\frac{-2\pi f l}{f_{samp}}} = \boldsymbol{B}_{ext}(f) + \boldsymbol{B}(0) \tag{4.13}$$

The focus in this work will be on PDC and DC estimated from $\boldsymbol{B}_{lag}(f)$. The purpose of the analysis is in fact to estimate causality in the Granger sense, hence considering only the effect of past samples in the estimation of spectral measures of coupling. These estimators were called *lagged* PDC and DC by Faes and colleagues (Faes et al. 2013). From now onward the terms PDC and DC will refer implicitly to lagged PDC and DC. In is important to notice from equation 4.13 that the spectral matrix of eMVAR coefficient $\boldsymbol{B}_{lag}(f)$ is function also of the matrix of instantaneous coefficients $\boldsymbol{B}(0)$: it is crucial therefore that the matrix of zero-lag coefficients is accurately estimated for a correct assessment of lagged connectivity (DC or PDC) from $\boldsymbol{B}_{lag}(f)$.

4.1.2 Objectives

As mentioned in the introduction, the eMVAR approach tested on simple MVAR models seems promising in removing the effect of instantaneous causality (Faes et al. 2013). However the eMVAR algorithm reliability in EEG applications has not been thoroughly tested. The eMVAR approach has been tested (Faes & Nollo 2010, Faes et al. 2013) on simple connectivity models where the direction of zero-lag connection was imposed a priori: this is quite different from how EEG zero-lag connectivity is generated (i.e. by instantaneous mixing of underlying source signals). The assumption of acyclical unidirectional zero-lag connections has been therefore imposed or taken for granted in

previous work. We believe that this hypothesis may be flawed in the case of application to recorded EEG and lead to misleading results. The purpose of this chapter is therefore to compare the performances of eMVAR approach with the classical scMVAR model in estimating lagged connectivity in simulated and recorded EEG data. The analysis aims at determining if in our (and related work) the more complex and less established eMVAR should be used, or if the simpler original algorithms should be recommended, addressing the following research question:

1. Is the eMVAR model a reliable estimator of instantaneous causality in simulated and recorded EEG data?
2. How does the eMVAR approach perform if applied on multichannel EEG recordings, compared to the classical MVAR method? Does it improve the identification of expected physiological patterns?

These issues have been addressed performing a series of simulation studies and applying the methodologies to EEG recordings. The methods and results of these studies will be outlined in the following sections.

4.1.3 Methods

Simulation study I The aim of the first simulation study is to investigate the performances of the causality measures obtained using the scMVAR approach compared to the eMVAR method on a simple connectivity model. For this purpose theoretical values computed from the imposed model and measures estimated from simulated data will be compared. Another objective of the study is to investigate the reliability of the non-Gaussian approach and compare it to the a-priori method for the estimation of the eMVAR model. In this case the a-priori method is applicable because the causal order of instantaneous interactions is imposed and thus known. The current study is similar to that carried out by Faes and colleagues (Faes & Nollo 2010, Faes et al. 2013) that performed an analysis on simulated signals obtained applying an MVAR filter with white noise innovations. The simulation study presented here was designed to model more realistically (compared to the model of Faes and colleagues) EEG time series, generating simulated signals with spectral features typical of EEG data. The imposed connectivity scheme involved four signals having lagged, instantaneous and autoregressive dependencies, as showed in figure 4.1. In order to generate time series with spectral features similar to EEG signals the first process (x_1) was a 4 s EEG trace recorded during rest with closed eyes (sampling frequency $f_{samp} = 250Hz$). Other signals were generated imposing the connectivity scheme, iteratively applying the equation of the eMVAR model

(equation 4.14) and adding white innovations $\boldsymbol{\omega}$ with a uniform amplitude distribution to allow for the eMVAR estimation based on non-Gaussianity.

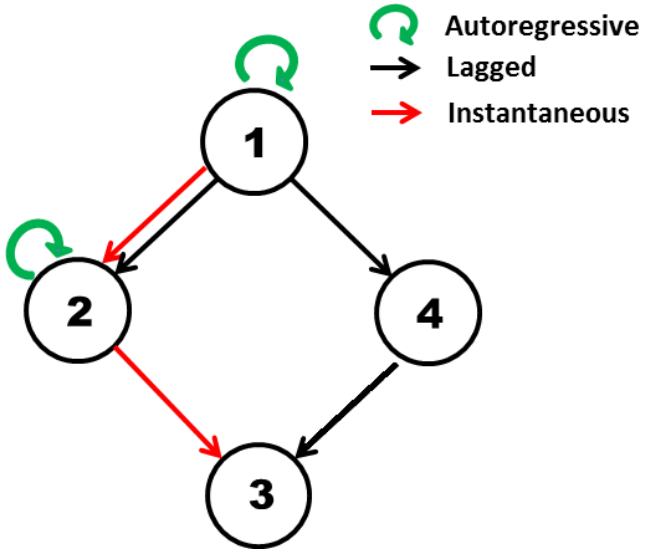

FIGURE 4.1: Imposed connectivity scheme for the Simulation study I. Note that the imposed instantaneous connections are unidirectional, to fulfill the DAG assumption. In case of EEG signal this may not be physically 'true' because the zero-lag dependencies are generated by instantaneous mixing of underlying source signals. This model is similar to the connectivity scheme used by Faes & Nollo (2010), Faes et al. (2013) however we modified it in order to generate time series with spectral features similar to EEG signals by imposing the first process x_1 equal to a EEG recorded signal.

The chosen model was given by the following set of equations:

$$
\begin{aligned}
x_2(n) &= 0.8x_1(n) + 0.4x_1(n-1) - 0.5x_1(n-2) - 0.64x_2(n-1) + 0.2x_2(n-2) + \omega_2(n) \\
x_3(n) &= 0.7x_2(n) + 0.5x_4(n-2) - 0.4x_4(n-3) - 0.3x_4(n-4) + \omega_3(n) \\
x_4(n) &= 0.6x_1(n-1) - 0.3x_1(n-2) - 0.4x_1(n-4) + \omega_4(n)
\end{aligned}
\tag{4.14}
$$

As can be observed from the equations of the eMVAR model (4.14), both instantaneous and lagged effects are present and the maximum lag of interaction is $p = 4$ (imposed model order). The autoregressive coefficients for the signal x_2 were chosen in order to generate a peak in the spectrum of x_2 distinct from the alpha peak (8-12 Hz) in the EEG signal x_1, while the off-diagonal coefficients were chosen in a range of realistic values as observed in previous studies testing the performances of connectivity estimators (Astolfi, Cincotti, Mattia, Marciani, Baccala, de Vico Fallani, Salinari, Ursino, Zavaglia, Ding, Edgar, Miller, He & Babiloni 2007, Baccalá & Sameshima 2001). Both direct (i.e. $1 \rightarrow 2, 1 \rightarrow 4$) and indirect (i.e. $1 \rightarrow 3$) connections were included in the model.

	scMVAR	**eMVAR a-priori**	**eMVAR non-Gaussian**
PDC	scPDC	apPDC	ngPDC
DC	scDC	apDC	ngDC

s

TABLE 4.1: Table of symbols indicating the acronyms for the connectivity estimators PDC and DC estimated with the classical scMVAR approach (scDC, scPDC), the a-priori eMVAR framework (apDC, apPDC) and the non-Gaussian eMVAR approach (ngDC,ngPDC).

A scMVAR model was then fitted on simulated time series. First the optimum model order was estimated applying the SBC criterion. Secondly the LS algorithm (ARfit, (Schneider & Neumaier 2001)) was used to estimate the matrix of coefficients. From the estimated spectral matrix of parameters $\boldsymbol{A}(f)$ and transfer function $\boldsymbol{H}(f)$, PDC and DC were computed. Following the estimation of the scMVAR, a corresponding eMVAR model was fitted on the simulated data using both the a-priori and the non-Gaussian approaches (the non-Normality of residuals was always true, because the white innovations $\boldsymbol{\omega}$ were imposed with a uniform amplitude distribution). In the first case (a-priori) the (known) direction of zero-lag interactions was imposed on the matrix $\boldsymbol{B}(0)$. In the second case (non-Gaussian) ICA was applied to the residuals of the scMVAR model in order to estimate the matrix of instantaneous effects as described in section 4.1.1. The eMVAR model was estimated using the eMVAR toolbox implemented by the original authors in Matlab (Faes et al. 2013). Lagged DC and PDC were then computed as a function of the spectral coefficient matrix of the extended model $\boldsymbol{B}_{lag}(f)$ applying equations 3.19 and 3.23.

The generation-estimation procedure was performed with 100 repetitions in order to examine the repeatability of the estimators. The number of repeats was chosen equal or larger than the number of realizations used in previous simulation studies (Astolfi, Cincotti, Mattia, Marciani, Baccala, de Vico Fallani, Salinari, Ursino, Zavaglia, Ding, Edgar, Miller, He & Babiloni 2007, Faes et al. 2013). In order to distinguish among the spectral causality estimators derived from the three approaches (scMVAR, a-priori eMVAR, non-Gaussian eMVAR) the different measures were named as indicated in table 4.1. This nomenclature holds for all the studies described in this chapter.

Simulation study II In this second simulation study the instantaneous interactions among EEG channels were more realistically modelled as mixing of source signals. Differently from the previous simulation model where a connectivity scheme between EEG signal was imposed, here a more realistic EEG generation model involving 3 sources and 3 signals simulating scalp EEG was designed, as showed in figure 4.2. EEG signals are in fact measured at the scalp, thus being a result of the volume conduction through

the tissues of the head of cortical signals. The aim of this model is to represent EEG signals as a mixture of source signals and noise and to investigate the impact of instantaneous source mixing on the lagged causality measured 'at the scalp'. The model in figure 4.2 was designed in order to investigate how the instantaneous mixing of sources affects lagged causality among scalp signals (estimated with eMVAR and scMVAR) and to compare it with the imposed causality among source signals. The diagonal elements of the matrix $\boldsymbol{A}$ chosen produce distinct spectral peaks in each source when they are driven with white noise. This was chosen to simulate sources $s_i, i = 1, 2, 3$ with different frequencies ($f_1 = 30Hz$, $f_2 = 50Hz$, $f_3 = 75Hz$). The off-diagonal elements of $\boldsymbol{A}$ were set different from zero, as shown in equations 4.15 and 4.16, in order to introduce some linear dependencies among sources. The sampling frequency was set to $f_{samp} = 250Hz$.

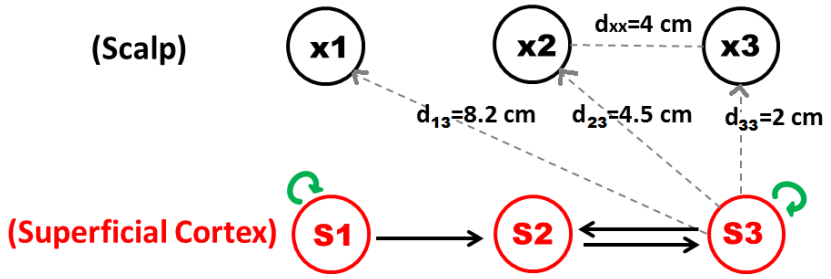

FIGURE 4.2: Scheme of the imposed source model. Note that linear lagged interactions have been imposed among the sources (black arrows). For the sake of clarity of the representation, we have indicated only the S3 signal 'spreading on the scalp' with grey arrows, but this holds also for the other sources S1 and S2. The expression of scalp signals $x_i, i = 1, 2, 3$ as a function of source signals $s_j, j = 1, 2, 3$ is and the distances $d_{i,j}$ is given in equation 4.18

.

$$\boldsymbol{A}(1) = \begin{bmatrix} 1.34 & 0 & 0 \\ 1 & 0 & 0.5 \\ 0 & 0.5 & -0.54 \end{bmatrix} \tag{4.15}$$

$$\boldsymbol{A}(2) = \begin{bmatrix} -0.81 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0.5 & -0.81 \end{bmatrix} \tag{4.16}$$

As innovations vector $\boldsymbol{\varepsilon}(n)$ a non-Gaussian white noise was considered in order to respect the assumption of the non-Gaussian approach[5]. The simulated scalp signals $x_i, i = 1, 2, 3$

[5]The non-Gaussian noise was generated applying to a Gaussian white noise vector $\boldsymbol{\eta}(n)$ the nonlinear transformation, as in Faes et al. (2013):

$$\varepsilon_i(n) = sign\big(\eta_i(n)\big)|\eta_i(n)|^a, i = 1, 2, 3 \tag{4.17}$$

The exponent a was chosen equal to 0.7 to generate a sub-Gaussian distribution.

were obtained applying the following equations:

$$\begin{aligned} x_1(n) &= \frac{1}{d_{11}^2}s_1(n) + \frac{1}{d_{12}^2}s_2(n) + \frac{1}{d_{13}^2}s_3(n) + \varepsilon_1(n) \\ x_2(n) &= \frac{1}{d_{21}^2}s_1(n) + \frac{1}{d_{22}^2}s_2(n) + \frac{1}{d_{23}^2}s_3(n) + \varepsilon_2(n) \\ x_3(n) &= \frac{1}{d_{31}^2}s_1(n) + \frac{1}{d_{32}^2}s_2(n) + \frac{1}{d_{33}^2}s_3(n) + \varepsilon_3(n) \end{aligned} \tag{4.18}$$

$s_i(n), i = 1, 2, 3$ being the source processes at time lag n, d_{ij} the imposed distance between source i and scalp electrode j and $\varepsilon_i(n), i = 1, 2, 3$ the non-Gaussian noise. As expressed in equation 4.18, the source signals were attenuated with the square of their distances from the hypothetical scalp locations, considering that cortical sources contributions to the scalp potential decay with the square of the distance between source location and scalp electrode (Buzsáki et al. 2012). The objective of this simulation was to investigate instantaneous and lagged causality on simulated EEG signals resulting from a mixture of sources signals. At this stage, a realistic model of signal attenuation and distortion through the soft and hard head tissues was beyond the purpose of the study. A more realistic model of the source signal propagation through the tissues of the head will be proposed in section 4.2. In this context by 'volume conduction' we will indicate the weighted mixing of sources as expressed in equation 4.18. The variance of the noise $\varepsilon_i(n)$ was set in order to have a SNR=10 (power of sources before attenuation/power of noise $\varepsilon(n)$). The distance among simulated scalp signals was set equal to 4 cm, as it is a reasonable estimate of the average distance between two electrodes in the International 10-20 system.

Also for this second study, simulated signals were generated with 100 repetitions each of $N = 1000$ samples in order to test the reliability of estimators. For each repetition a MVAR autoregressive model was estimated on the simulated data x_i using a LS Algorithm and the optimum model order selected with the SBC criterion. As for the Simulation Study I, two different approaches were applied for the fitting of the MVAR model, the scMVAR and the non-Gaussian eMVAR approach, where the latter is used to model instantaneous effects. In this case the a priori eMVAR approach could not be applied since the direction of instantaneous effects is not known. In this simulation zero-lag dependencies among scalp channels are generated by instantaneous mixing of source signals; as a consequence the hypothesis that zero-lag connections are unidirectional and acyclical may not be respected (because not physiologically 'true' when considering the effects of the spread of electric fields). The aim of this second simulation study is to test the performance of the non-Gaussian eMVAR approach in estimating instantaneous

causality in this case, i.e. when the direction of zero-lag connectivity is not imposed a-priori.

Application to recorded EEG data As a final step of the work carried out in this chapter to assess the performances of the eMVAR approach in estimating lagged connectivity, the spectral measures previously tested on simulated EEG data were applied to multichannel EEG recordings from 2 representative subjects in resting with eyes closed. The purpose of this study was to check the validity of the ngDC and ngPDC estimators (i.e. their ability to identify expected physiological patterns) and to compare their performances with those of the classical MVAR approach in estimating connectivity on recorded EEG data.

Fifteen 2 s epochs of resting state EEG with closed eyes sampled at $f_{samp} = 250Hz$ were considered in the analysis. In order to reduce the computational effort only 6 well separated electrodes were considered (two frontal, two central, two occipital) as the interest was focussed on long range connectivity. The length of EEG traces was chosen so as to include a number of data points sufficient to estimate the free parameters and at the same time to analyse a short EEG segment in order to better approximate the stationary conditions (see 3.3.1). Estimation of connectivity (PDC, DC) was performed fitting EEG data first using the scMVAR (ignoring the modelling of instantaneous effects) and then the extended model using the eMVAR toolbox (Faes et al. 2013). In this case, differently from the simulation studies, information about the theoretical causality patterns among signals was not available. However the validity of connectivity estimation could be examined on the basis of physiological considerations about the expected spectral coupling among brain areas and by comparison with results in the literature. Moreover in order to examine the performances of the two approaches (scMVAR vs eMVAR) with respect to volume conduction artifacts, the estimated PDC and DC between EEG channels were also compared with connectivity estimated between the underlying source signals estimated solving the so-called *Inverse problem*, i.e. the problem of finding the putative electric sources in the cortex that generate the given set of scalp recordings. The source time series were estimated making use of a distributed source model (Dale & Sereno 1993) and a realistic head model (Montreal Neurological Institute brain, Collins, 1994) implemented in the software eConnectome (Bin et al. 2011). The intensity of the current dipoles distributed on the cortex was then estimated applying a weighted minimum norm algorithm (Fuchs et al. 1999) with anatomical constraints. The solution of the inverse problem gives an estimate of cortical signals, thus removing the volume conduction artifacts that affects signals recorded on the scalp. Once estimated, in the current work the cortical dipole signals were then averaged in six regions

of interest (ROIs), corresponding to the six electrodes considered in Figure B.1. More details about the model and algorithms used to solve the Inverse Problem are given in Appendix B.

4.1.4 Results

In this section the results of the simulation studies and the application on EEG recordings are presented. The nomenclature given in table 4.1 will be used to indicate all the results of the study.

Simulation Study I The following figures show the results relative to the first simulation study, whose imposed model is showed again in 4.4. An analysis of the quality of MVAR estimation showed that for all the 100 realisations, the optimum model order estimated with the SBC criterion coincided with the imposed one. An example of the information criteria' trend for one repetition is given in figure 4.3: the SBC curve presents a clear minimum for $p = 4$, while the FPE curve presents a shallower profile. A test for whiteness of residuals was performed for each repetition: for 99 of 100 repetitions the residuals resulted white when tested with the Ljung-Box Portmanteau test ($p < 0.05$), therefore the lagged correlation structure of the time series was adequately estimated for all the repetitions.

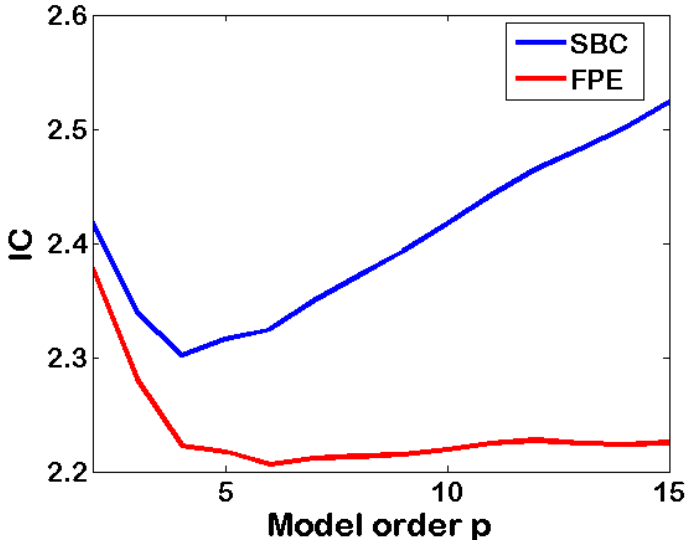

FIGURE 4.3: Example of model order selection for one iteration: SBC and FPE criteria values as a function of model order. The model order used in generating signals was 4. This is a representative example of the SBC and FPE criteria trends and shows that, while the SBC presents a clear minimum for the optimal model order, the FPE has a shallower, less interpretable profile.

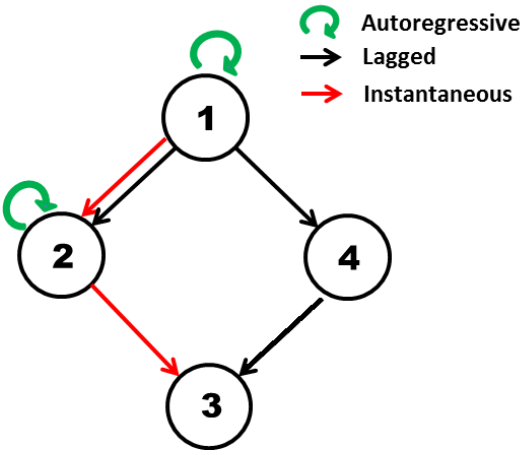

FIGURE 4.4: Imposed connectivity scheme for the Simulation Study I, repeated from figure 4.1 in order to compare the effect of instantaneous causality (red) with the off-diagonal element of the covariance matrix-next figure 4.5. Note how the instantaneous connections between 1 and 2 and 2 and 3 correspond to the elements of the covariance matrix different from zero.

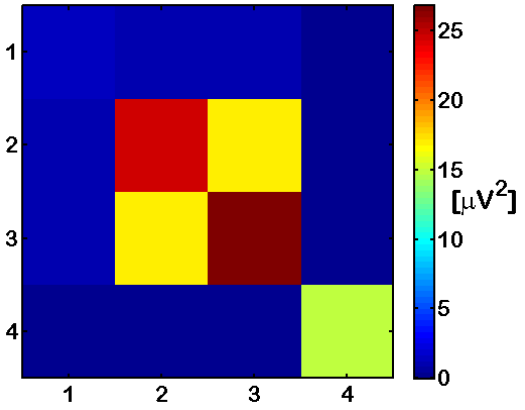

FIGURE 4.5: Covariance matrix $\mathbf{\Sigma}_{\varepsilon,0}$ of the scMVAR residuals $\boldsymbol{\varepsilon}$ averaged for the 100 iterations. The diagonal elements σ_{ii} represent the variance of residuals. The off-diagonal elements σ_{ij}^2 represent the cross-covariance between residuals and indicate the presence of zero-lag effects.

An inspection of the covariance matrix of residuals (figure 4.5) reveals that they are correlated at lag-zero, as some off-diagonal elements are different from zero; therefore there are instantaneous dependencies among signals that have not been modelled with the classical scMVAR. It is useful to remember here that the matrix of instantaneous effects is a function of the residual covariance matrix (as demonstrated in section 3.3.2) and that the non-diagonality of the residual covariance matrix indicates the presence of instantaneous causality. The results showed in figure 4.5 were therefore expected since, when significant zero-lag effects are present among time series (in this case they were imposed in the generating model) the scMVAR residuals are correlated. The covariances (off-diagonal elements) different from zero correspond in fact to the zero-lag interactions imposed on the model (i.e. $1 \rightarrow 2$, $2 \rightarrow 3$, as well as those in the opposite direction $2 \rightarrow 1$ $3 \rightarrow 2$ since the covariance matrix is symmetric).

In order to check the validity of the non-Gaussian hypothesis for the application of the eMVAR method, the non-normality of residuals was checked for each repetition with the Jarque-Bera Test[6]: for all the repetitions, residuals were found to be non-Gaussian, and therefore the necessary assumption for the application of the non-Gaussian algorithm was justified.

[6]The Jarque-Bera test compares the skewness and kurtosis of a given process with the theoretical distribution of a Gaussian process (Lütkepohl 2005).

The following plots represent the coupling measures derived from the estimation of scMVAR and eMVAR on simulated data.

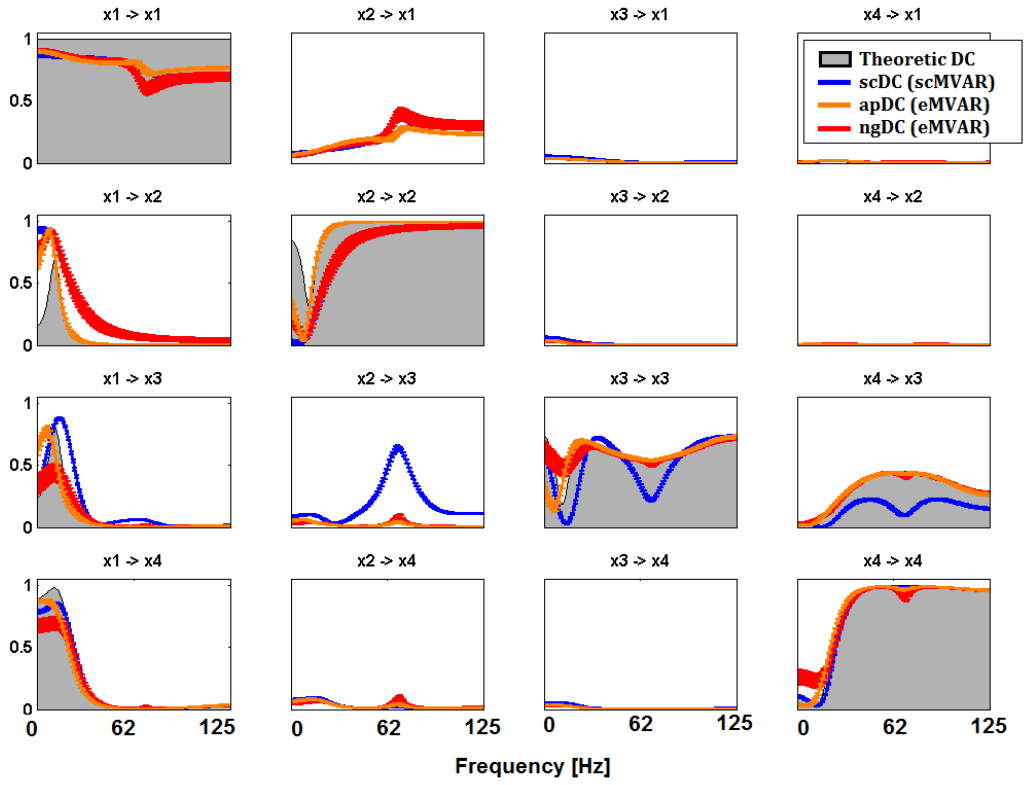

FIGURE 4.6: DC (squared) estimated with the three methods: scMVAR-blue, eMVAR non-Gaussian Approach-red, eMVAR a priori approach-orange and compared with the theoretical values-shaded gray area. The errorbars represent the 99% confidence interval for the 100 repetitions.

Figure 4.6 is a matrix plot of the DC estimated with the classical scMVAR and the eMVAR compared with theoretical values. Theoretical values (in grey in figure 4.6 and following) were computed on the basis of the imposed matrix of parameters[7] and show that, as expected, none of the other signals contribute to the spectrum of x_1 (the off diagonal theoretical values of the first row are all zero while the autoregressive component of x_1 is equal to 1 across all frequencies). The peak at α frequency (10 Hz) of the signal x_1 is transferred to x_2, x_3 and x_4 as the $DC_{i,1}(f), i = 2, 3, 4$ (first column) has a peak at the α frequency. Since in the computation of DC only lagged coefficients are considered, the theoretical values are different from zero only when lagged connections are present in the original model. By contrast the DC estimated on simulated data shows a non-zero profile also in absence of lagged interactions (causality), for instance the profile of the

[7]Being the signal x_1 a EEG recording from which all the others signals $x_i (i = 2, 3, 4)$ were generated, the theoretical values of the first row of the matrix MVAR parameters were not known a priori. For this reason the autoregressive coefficients of x_1 were estimated considering x_1 as an univariate autoregressive process with the LS algorithm (ARfit, MATLAB), imposing the model order equal to the theoretic one ($p = 4$).

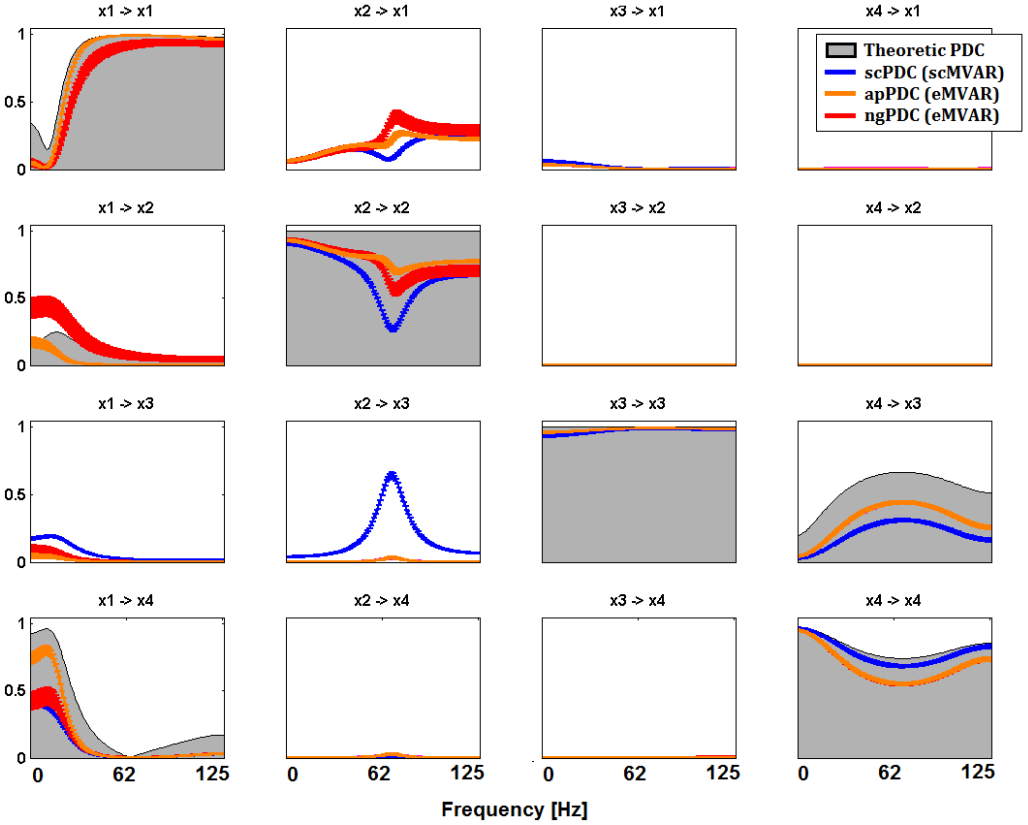

FIGURE 4.7: PDC (squared) estimated with the three methods: scMVAR-blue, eMVAR non-Gaussian Approach-red, eMVAR a Priori approach-orange and compared with the theoretical values-shaded gray area. The errorbars represent the 99% confidence interval for the 100 repetitions.

DC is non-zero for $DC_{1,2}(f)$ (first row, second column) and $DC_{3,2}(f)$ (third row, second column) as a result of the zero-lag links $x_1 \leftrightarrow x_2$ [8] and $x_2 \leftrightarrow x_3$ but in the absence of a lagged correlation. The apDC however shows the best performance, closely matching the theoretical results (except for $apDC_{1,2}(f)$). The non-Gaussian method (ngDC) is affected by a larger variance across repetitions and has poorer performances (for instance note the bias for $ngDC_{3,1}(f)$, third row, first column). However ngDC is close to zero when lagged effects are absent but there are instantaneous effects (e.g. $x_2 \rightarrow x_3$), differently from scDC estimated with the classical scMVAR model, that indicates as lagged the dependence $x_2 \rightarrow x_3$, where only instantaneous coupling is present (third row, second column).

Similar considerations are valid for the estimated squared PDC, showed in Figure 4.7. Theoretical PDC shows values different from zero only when lagged *direct* effects are present, because the PDC measures only direct lagged coupling. In this case apPDC and ngPDC are able to identify only direct effects, with values close to zero when

[8] Note that we have used a bidirectional arrow to indicate instantaneous effects: Even if in the imposed connectivity model zero-lag connections are unidirectional (i.e.1 $\rightarrow$ 2), their effects on estimated connectivity is bidirectional (i.e. $DC_{1,2}(f)$ is different from 0)

instantaneous ($x_2 \rightarrow x_3$) or indirect ($x_1 \rightarrow x_3$) effects are present; scPDC cannot do that. For instance $scPDC_{3,2}(f)$ (third row and second column of figure 4.7) clearly presents a peak at f_2 quantifying as lagged what is in fact an instantaneous connection from $x_2 \rightarrow x_3$. Therefore the lagged connectivity structure of the imposed model can be inferred only from the observation of PDC estimated from the eMVAR models (ngPDC and apPDC). However the performances of PDC estimated with the non-Gaussian approach are affected by a larger variance across repetitions and larger bias and, as for the DC in figure 4.6, ngPDC values for the link $x_2 \rightarrow x_1$ are not ideal.

In conclusion, this first simulation study used an imposed connectivity scheme to simulate signals with spectral features similar to an EEG recording. Simulated signals were generated from the imposed (instantaneous and lagged) connectivity scheme. The aim of the study was to examine the performances of the eMVAR approach in estimating lagged connectivity in comparison with the classical scMVAR. The model produces the intended connectivity patterns (as calculated from theoretical DC and PDC), but all the estimates fail to reproduce this pattern accurately. The apDC (and apPDC) is probably closest to the correct results; the scDC is furthest, with ngDC showing intermediate performance. However the relative performance was frequency dependent, and also varied between channels considered. Furthermore, it should be emphasized that this is only one illustrative case, and while it aims to provide a realistic example, one cannot generalize to all cases.
Results are in line with the finding in the literature (Faes & Nollo 2010, Faes et al. 2013) and suggest that the scMVAR model, in the presence of significant zero-lag dependencies, may lead to erroneous estimates of lagged causality. Both scPDC and scDC estimated with the classical scMVAR model indicate as lagged the zero-lag interaction from x_2 to x_3 (i.e. the zero-lag connectivity has affected the lagged connectivity). An analysis of the consistency of the estimators for all the connections was not previously performed: results indicate that the non-Gaussian approach for the estimation of the eMVAR model presents a larger variance across repetitions and a bias in estimating certain connections as compared with the a-priori method.
In this simulation study the imposed model included significant instantaneous dependencies among signals, but all these zero-lag dependencies were unidirectional, which is also the assumption made for an accurate eMVAR model estimation. This is unrealistic in the case of EEG signals, where volume conduction effects spread all signals widely. In the next section this will be included in the simulation with zero-lag effects generated by instantaneous mixing of sources at the scalp-level.

Simulation Study II In the following, results obtained from the source simulation study are shown. In order to make the interpretations of findings easier, the scheme of

the imposed model is shown again in Figure 4.8 (see also figure 4.2). As for the previous study, the results of the different tests to check the model validity will be commented first.

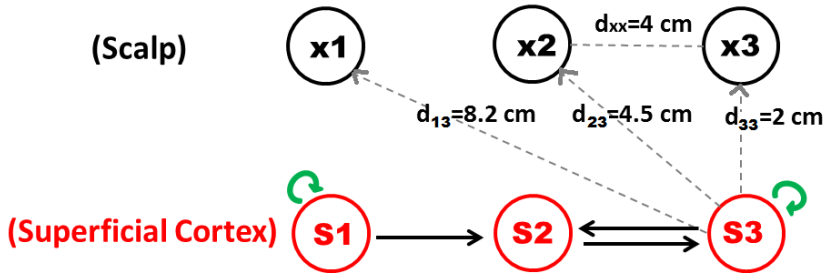

FIGURE 4.8: Scheme of the imposed model for Simulation II repeated from figure 4.2 in order to compare results. The green arrows represent the autoregressive contributions to the source signals while the black arrows are linear interactions among sources. The scalp signals x are the results of the instantaneous mixing of the three sources s.

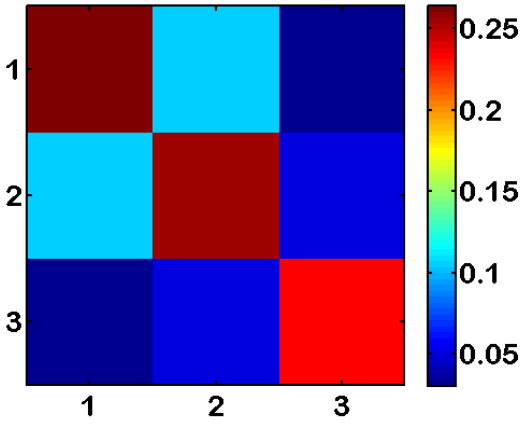

FIGURE 4.9: scMVAR residual covariance matrix averaged across repetitions. The diagonal elements σ_{ii} represent the variance of residuals. The off diagonal elements σ_{ij} represent the covariance between residuals.

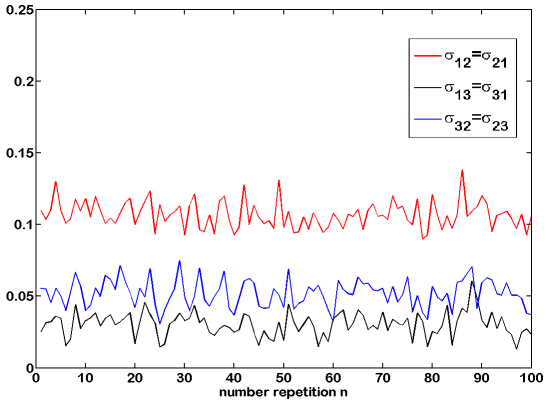

FIGURE 4.10: Off-diagonal elements of the residual covariance matrix σ_{ij} as a function of the number of the repetition.

The model validity analysis showed that for 84 out of 100 repetitions the estimated optimum model order was p=3, and p=4 for the rest of the repetitions. Note that the imposed sources model order was $p = 2$. In this case the estimated model of the scalp signals has a higher order than the imposed source model, differently from the previous EEG simulation were for all the repeats the estimated optimum model order was equal to the imposed one. This may be due to the fact that the scalp signals are generated as a mixture of sources and therefore additional correlation among them is introduced. This effect could increase the model order among scalp signals with respect to the source model. The test for whiteness of residuals revealed that for 55 iterations the hypothesis of whiteness of residuals of the scMVAR model was rejected, therefore for half of the

repetitions the full causal structure was not estimated properly. The average covariance matrix of residuals and the profile of the covariance for all the repetitions are shown in figures 4.9 and 4.10. The covariance values are different from zero, thus indicating the presence of significant instantaneous effects that are higher for adjacent electrodes in all the repetitions. In this case the covariance[9] between signals x_1 and x_2 has higher values with respect to $\sigma_{23} = \sigma_{32}$, because the signal x_1 has a larger power than x_2 and x_3 (results not shown for the sake of brevity).

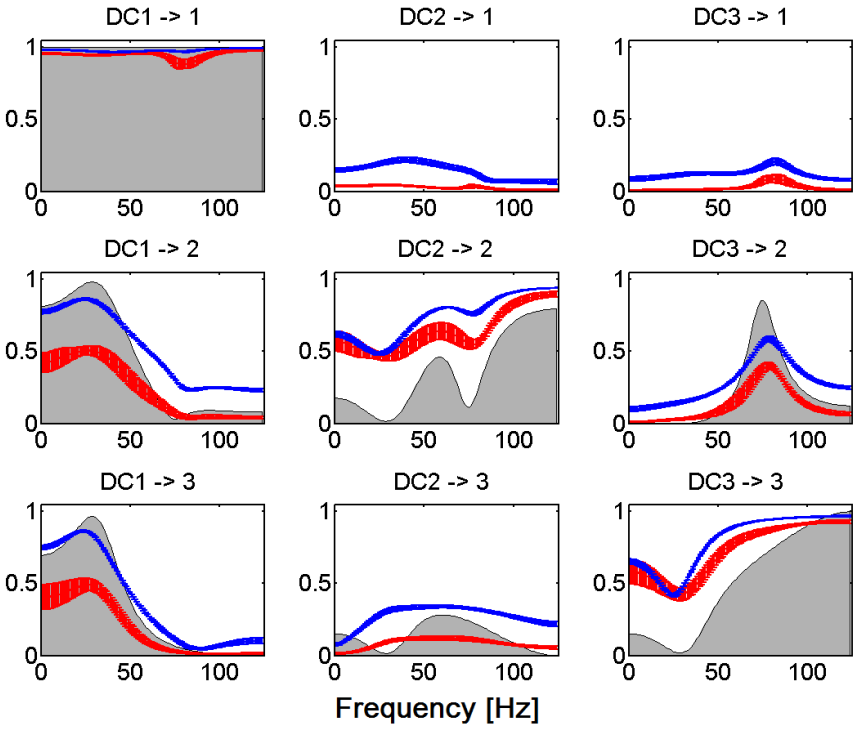

FIGURE 4.11: DC estimated from the scMVAR model (scDC, blue) and with the eMVAR non-Gaussian approach (ngDC, red). The errorbars represent the 99% confidence interval over 100 repetitions. The grey areas represent the theoretical values of DC computed from the source signals, i.e. the theoretic profile of DC obtained from the imposed coefficient matrix of the source model.

Figure 4.11 is a matrix plot of the DC estimated on simulated data and the theoretical values computed from the imposed source model. In this case the a priori eMVAR approach was not applicable as the direction of zero-lag connections is not known a priori. Comparing the theoretical and estimated values two cases can be identified:

1. When theoretical values of $DC_{ij}(f)$ among sources are different from zero (i.e. a connectivity path exists in the original source model) then the theoretical DC shows larger values than the DC estimated on scalp signals but the scDC (scMVAR model) performs better than the ngDC (eMVAR model), showing values of scDC

[9]The scale of all the results of the Simulation Study II is relative to the source signals that were generated numerically with unit variance.

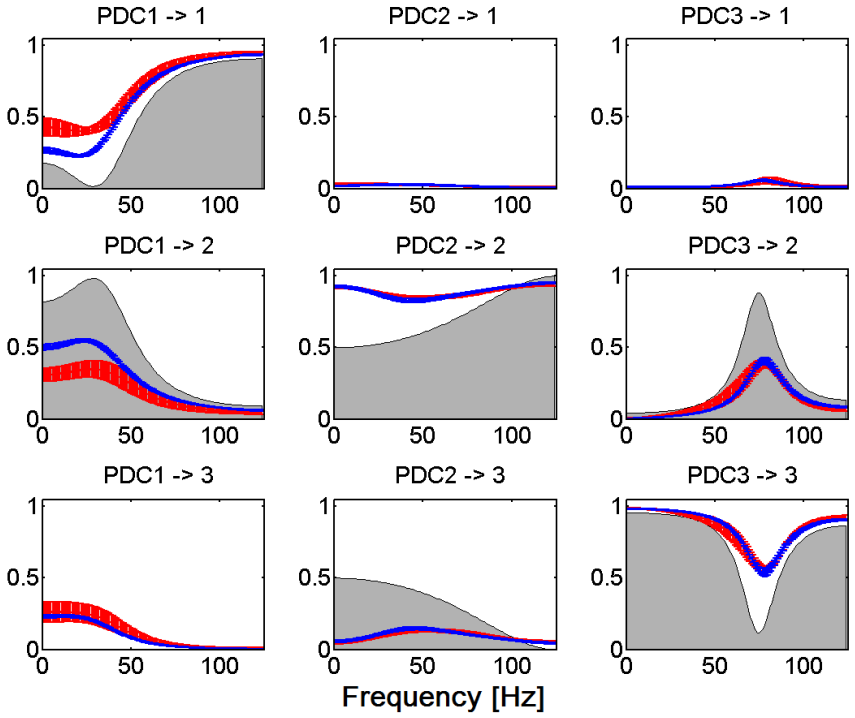

FIGURE 4.12: PDC estimated from the scMVAR model (scPDC, blue) and with the eMVAR non-Gaussian approach (ngPDC, red). The errorbars represent the 99% confidence interval over 100 repetitions. The grey areas represent the theoretical values of PDC computed from the source signals, i.e. the theoretic profile of PDC obtained from the imposed coefficient matrix of the source model.

much closer to the theoretical sources values, for example $DC_{12}(f)$ (first column, second row) or $DC_{13}(f)$ in figure 4.11.

2. When theoretical values of $DC_{ij}(f)$ among sources are equal to zero (i.e. absence of connections among sources) then in the process of mixing sources to generate scalp signals some instantaneous dependencies are introduced in the absence of connections in the source model, here $DC_{21}(f)$ or $DC_{31}(f)$, therefore the estimated values are non-zero while the theoretical values are zero. In this case the eMVAR ngDC performs better than scDC, being closer to zero.

The matrix plot of the PDC (figure 4.12), shows that the PDC underestimates the coupling between scalp signals if compared with theoretical values, but, differently from DC, is close to zero when a lagged connections is absent among corresponding sources (i.e. $PDC_{2\to1}$ or $PDC_{3\to1}$). Moreover, the two methods (scMVAR and eMVAR) gave more similar results then were observed for the DC. An explanation for this has to be found in the properties of the PDC. The PDC is a factorization of the PCoh that quantifies the strength of the direct linear interactions between two signals, after removing the effects of the other time series of the MVAR process, i.e. the effect common to any other signal

combination; it therefore also removes some common effects of the source mixing, even if not included in the original model (scMVAR).

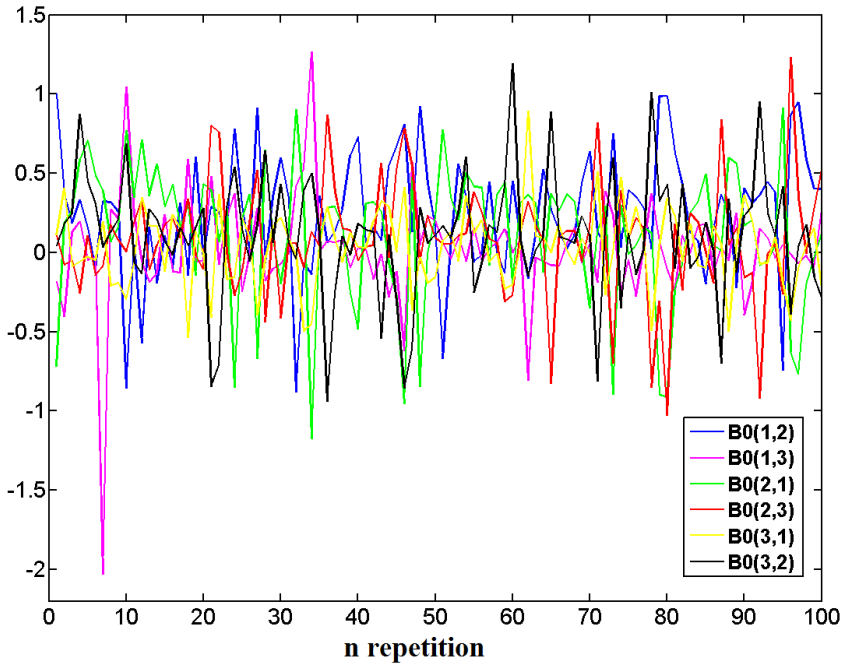

FIGURE 4.13: Coefficients of the estimated instantaneous interactions (i.e. off-diagonal elements of the matrix $\boldsymbol{B}(0)$ estimated with the non-Gaussian eMVAR approach) for all 100 repetitions. Note how the estimated $\boldsymbol{B}(0)$ coefficients present large variability from one repetition to another. This is in contrast to what expected since the process that generates zero-lag connectivity (instantaneous mixing of source by multiplication with a matrix) is constant across repetitions. These results indicate that the estimation of $\boldsymbol{B}(0)$ by means of the non-Gaussian approach is adversely affected by random initial conditions.

A plot of the estimated zero-lag model coefficients (using the non-Gaussian eMVAR approach) for all the 100 simulated signals is presented in figure 4.13. It shows that the values of zero-lag coefficients present a large variability across repetitions. Moreover the zero-lag coefficients for distant scalp signals ($B_{1,3}(0)$-magenta- and $B_{3,1}(0)$-yellow) are larger than the instantaneous coefficients for adjacent signals in some repetitions, differently from what was expected and was observed from the covariance matrix of residuals in figure 4.10. These results are quite unexpected, since if instantaneous interactions among scalp signals are mainly generated by the instantaneous mixing of sources we would expect consistent values of the matrix $\boldsymbol{B}(0)$ across repetitions, being the imposed mixing matrix identical. The residual cross-covariances give information about instantaneous effects among time series, therefore an agreement of results would be expected comparing the residuals covariance matrix (figure 4.10) and the elements of $\boldsymbol{B}(0)$. Contrary to this the profiles are very different; for instance we cannot conclude from estimates that $\boldsymbol{B}(0)$ values (for single iterations, but also on average) for adjacent scalp signals are higher than for distant electrodes.

The non-Gaussian algorithm for the estimation of the instantaneous effects matrix seems to be affected by the random noise $\varepsilon_i(n), i = 1, 2, 3$ in equation 4.18, which is the only component that changes from one repetition to another, and gives a wide scatter in the results.

Application to recorded EEG data The estimators first tested on simulation signals were then applied to EEG recordings of rest with closed eyes. As described in the methods section, the different connectivity estimators were applied on two datasets:

1. Multichannel EEG signals from 6 electrodes (Fp1, Fp2, C3, C4, O1, O2)
2. Putative cortical source signals reconstructed solving the inverse problem for six cortical areas underlying the electrodes Fp1, Fp2, C3, C4, O1, O2 (figure B.1 in Appendix B).

The DC estimated with the classical scMVAR model on the raw EEG signals (blue) and the underlying cortical sources signals (black) is represented in figure 4.14.

As expected, the causal dependence between EEG signals has generally higher values than the coupling among corresponding sources: in fact the dependencies among scalp signals result both from correlation among the underlying cortical sources and the additional correlation of the spreading of sources (volume conduction). It is possible to identify a significant bidirectional coupling between occipital electrodes O1 and O2 with a peak at α frequency (around 10 Hz, $O1 \rightarrow O2$ fourth row, last column and $O2 \rightarrow O1$ last row, fourth column) for both the dataset (EEG estimated sources). The α oscillations are in fact known to originate in the occipital visual cortex due to synchronisation of neurons discharge in the α frequency range (8-12 Hz) during rest with closed eyes (Klimesch 1999). Moreover results show an information flow from posterior channels to frontal ones in the α frequency range both on EEG and source estimated scDC, for instance for $O1 \rightarrow Fp2$, $C4 \rightarrow Fp2$, $O1 \rightarrow C4$, $O2 \rightarrow C4$. The detection of a preferential back-to-front direction for the α EEG activity is in agreement with previous findings, and is explained by the occipital nature of the α oscillations, which are thought to originate in the visual cortex and then spread towards the central and frontal regions in the brain (Kamiński et al. 1997, Faes et al. 2013). As stated previously, the diagonal elements of the matrix plot represent the part of the signals spectrum that is not due to other signals of the multivariate series. An inspection of the diagonal elements of figure 4.14 reveals that most of the spectrum of the EEG channels is due to the autoregressive contribution because the DC is close to 1 for all the frequencies except the ones in which a significant transfer of spectrum occurs. For instance, $scDC_{1,1}$ (first row, first column,

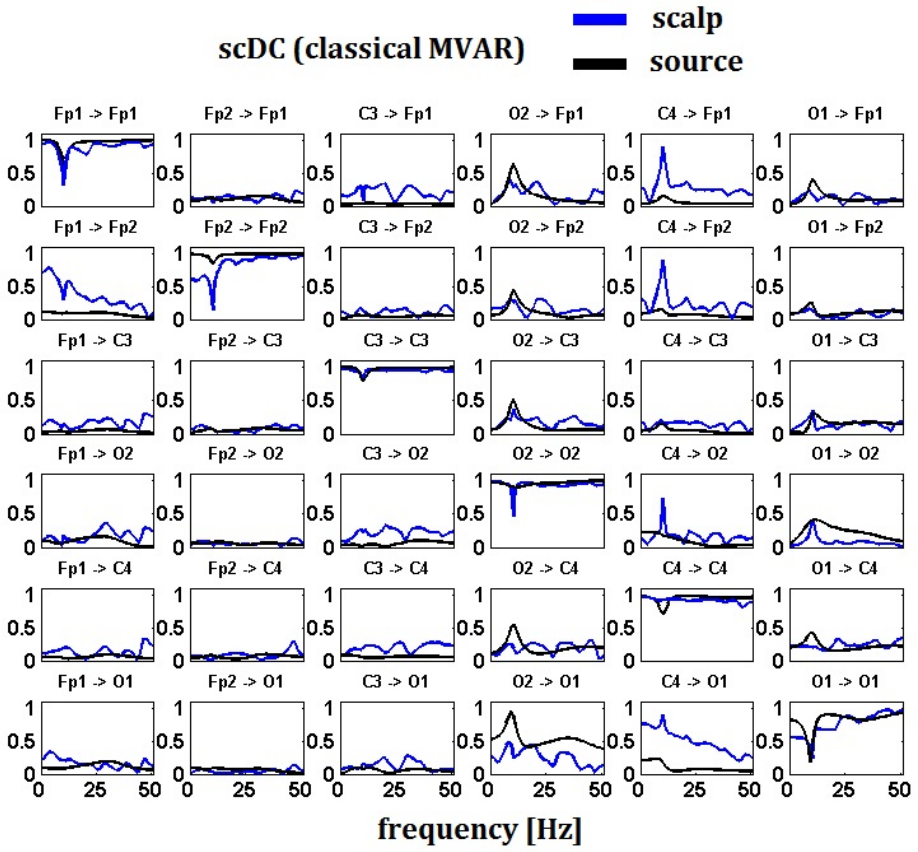

FIGURE 4.14: scDC estimated from the **classical scMVAR** model on raw EEG data (blue) and sources signals (black) as a function of the frequency. The diagonal elements $scDC_{ii}(f)$ represents the autoregressive component of the scDC (i.e. the part of the spectrum of the process x_i that is not due to other time series) while the off-diagonal elements $scDC_{ij}(f)$ represent the normalised spectrum that is transferred from signal x_j to signal x_i

channel Fp1) is close to one for most of the frequencies but presents a sharp dip at the α frequency (8-12 Hz), since in this frequency range there is a significant information flow from O1 to Fp1 (first row, last column).

Results relative to the DC estimated with the eMVAR approach applied on EEG data are showed in figure 4.15: the ngDC estimated with the non-Gaussian eMVAR approach on EEG exhibits a flat behaviour across frequencies for all the connections. Moreover ngDC values largely vary depending on the original dataset (EEG, sources) and it is difficult to interpret the results: for certain connections source values are very close to ngDC estimated on EEG, i.e $O2 \rightarrow C3$, for others the values are very different, i.e.$Fp2 \rightarrow O1$, $O1 \rightarrow C4$.

Results obtained from a second subject with the classical scMVAR approach are similar (results not shown): an information flow from occipital and central electrodes towards

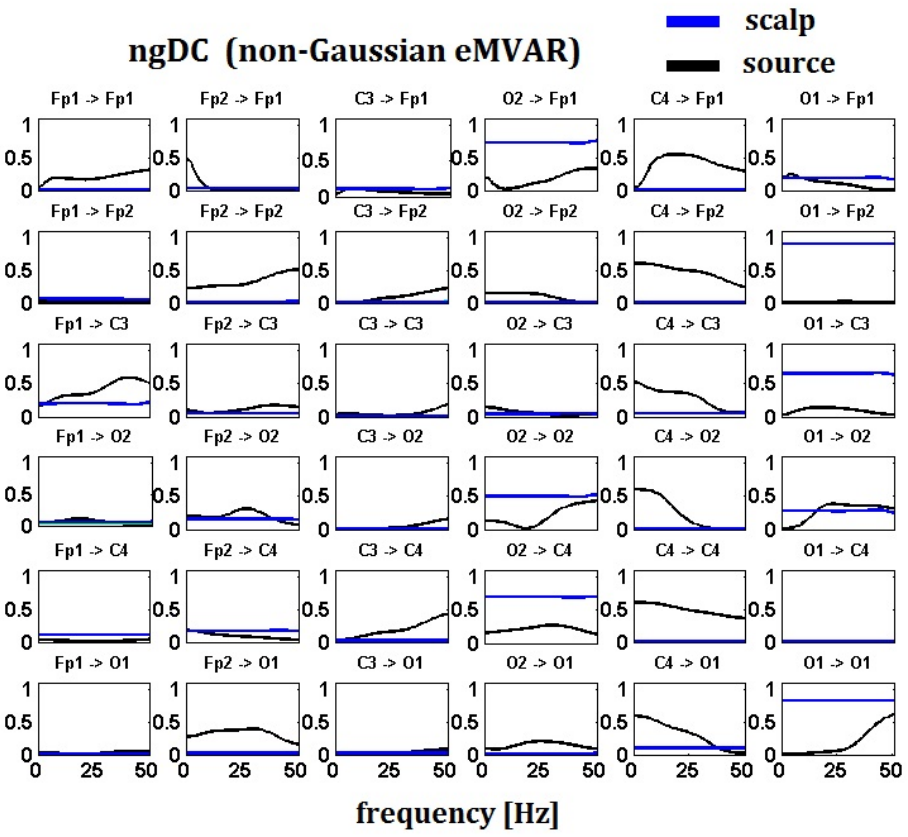

FIGURE 4.15: ngDC estimated with the **non-Gaussian eMVAR** approach on raw EEG data (blue) and sources signals (black) as a function of the frequency. The diagonal elements $ngDC_{ii}(f)$ represents the autoregressive component of the ngDC (i.e. the part of the spectrum of the process x_i that is not due to other time series) while the off-diagonal elements $ngDC_{ij}(f)$ represent the normalised spectrum that is transferred from signal x_j to signal x_i

the frontal ones is observed and connectivity estimated on EEG data and source data show comparable trends, with coupling among sources having lower amplitude. As for the first subject, lagged connectivity estimated with the eMVAR approach (ngDC) shows a flat behaviour in frequency and very different values of scalp and source connectivity, making the interpretation of results quite difficult.

Figure 4.16 shows the covariance matrix of residuals from the scMVAR estimation. As demonstrated in section 3.3.2, instantaneous effects are a function of the covariance of the residuals and when residual covariances are non-zero some instantaneous correlation among time series exists. The variance of residuals (diagonals elements) is extremely small compared to the power of EEG data, which is of the order of $100\mu V^2$, indicating that the prediction error of the model is relatively small. However, the residual covariance matrix is not diagonal, suggesting that some instantaneous effects exist among signals (in particular between the adjacent channels O2-O1) and have not been modelled

by the classical scMVAR model. The matrix of instantaneous effects estimated on the same EEG set is shown in figure 4.17: the highest zero-lag dependences are estimated between C3-Fp1 and C3-Fp2. Similarly to what observed for the Simulation Study II there is poor consistency between the instantaneous effects indicated by the residual covariance matrix and zero-lag coefficients $\boldsymbol{B}(0)$ estimated with the non-Gaussian eMVAR approach.

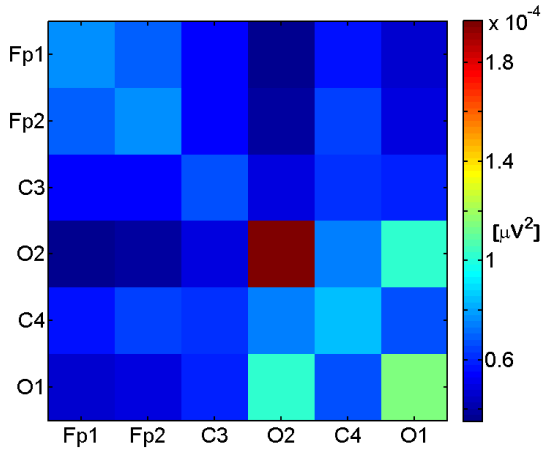

FIGURE 4.16: Residuals covariance matrix of the scMVAR model estimation on EEG signals.The off-diagonal elements $\sigma_{ij}(0)$ indicate instantaneous causality between channels x_j and x_i.

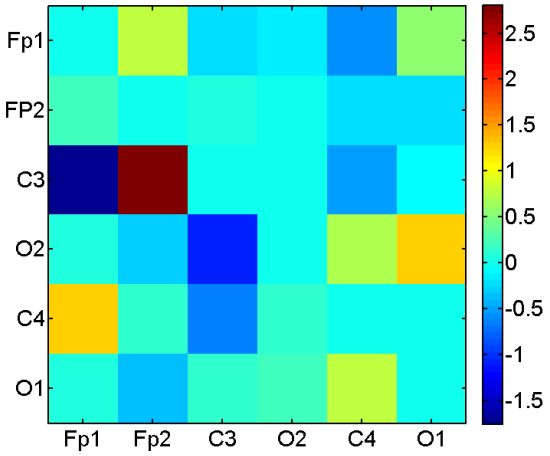

FIGURE 4.17: Matrix $\boldsymbol{B}(0)$ estimated with the eMVAR non Gaussian approach. The generic element $B_{ij}(0)$ represents the zero-lag linear dependency from channel x_j to x_i.

The connectivity measures estimated with the non-Gaussian eMVAR approach exhibit unexpected flat profiles in frequency when estimated on recorded EEG. These flat trends cannot be ascribed to low model order because the eMVAR is estimated starting from the scMVAR fitting and thus the classical and the extended models have the same order (p=18 in this case). These trends were observed also for the respective transfer functions $\boldsymbol{H}(f)$ (scMVAR) and $\boldsymbol{eH}(f)$ (eMVAR) (results not shown). In the computation of the matrix of lagged coefficients for the eMVAR model $\boldsymbol{B}_{lag}(f)$ the constant matrix of instantaneous effects $\boldsymbol{B}(0)$ is added for each frequency (see equation 4.19). An erroneous estimation of the matrix $\boldsymbol{B}(0)$(in particular an overestimation) may be the reason why the profile of $\boldsymbol{B}_{lag}(f)$ (and the spectral connectivity measures estimated from it) is flat in frequency.

$$\boldsymbol{B}_{lag}(f) = \boldsymbol{I} - \sum_{l=1}^{p} \boldsymbol{B}(l) e^{\frac{-2\pi f l}{fsamp}} = \boldsymbol{B}_{ext}(f) + \boldsymbol{B}(0) \tag{4.19}$$

Our examples suggest that, when applied on recorded EEG data, the non-Gaussian framework fails to characterize reliably zero-lag interactions and that caution is required

when interpreting results of this approach. Moreover there is a very poor agreement between the connectivity networks estimated on scalp EEG channels with the eMVAR approach and those on underlying source signals. These results add further uncertainty to the reliability of estimation of instantaneous effects on real EEG data with the non-Gaussian eMVAR algorithm. Scalp connectivity estimated with the classical scMVAR approach (scDC, scPDC), on the other hand, exhibit patterns that are in agreement with previous findings in literature and show similarity with the connectivity patterns estimated between underlying source signals.

4.1.5 Summary

The results of the two simulation studies and the application on recorded EEG data are discussed and summarized in the following points:

1. Neither of the two methods (scMVAR and eMVAR) accurately estimate lagged (theoretical) connectivity. The a-priori eMVAR approach cannot be applied in EEG simulated or recorded data as the direction of zero-lag interactions is not known a priori. The non-Gaussian eMVAR framework performs better than the scMVAR when the zero-lag dependences are imposed as directed and acyclic in the original model (Simulation I). However when the zero-lag effects are more realistically modelled as source mixing (Simulation II) and in application to recorded EEG the performances of the eMVAR are degraded, and not superior to those of the classical scMVAR.

2. Our examples suggest the non-Gaussian algorithm estimates of instantaneous interactions are not consistent across repetitions and the results obtained from the estimation of the non-Gaussian eMVAR model may be misleading. This was in some way expected in simulated and recorded EEG data because the assumption of unidirectional acyclical zero-lag effects is physiologically 'violated' and the correct application of non-Gaussian eMVAR compromised.

3. Compared with the DC, the PDC estimated with the two methods (ngPDC and scPDC) present similar results and is less affected by instantaneous causality (even if the eMVAR methods exhibited larger variance across repetitions). These results are in line with the theoretical properties of this estimator: The PDC is in fact a measure of linear dependencies between two time series once the effect common to all the other signals of the multivariate dataset is excluded. In EEG multichannel analysis, volume conduction effects are common to adjacent electrodes, therefore the use of PDC should reduce the volume conduction artifacts in the computation of directed causality, independently from the algorithm used (eMVAR or

scMVAR). This suggest that the PDC may be the optimal choice for causality structure inference because of its properties of excluding indirect connectivity and reducing instantaneous effects.

These results suggest that the estimation (and removal) of instantaneous effects in the estimation of lagged causality remains a challenge. Given that, the issue of instantaneous causality remain unresolved. In the case of the EEG signal, which is our main focus, the removal of instantaneous causality is beneficial because it represents the non physiological correlation produced by volume conduction of underlying sources. If removing instantaneous connectivity between scalp channels is not possible, then it is useful to characterize to what extent the DC and PDC estimators are affected by the instantaneous effects of volume conduction in order to correctly interpret the significance of estimated connectivity. We will address this issue with a more realistic model of EEG signal generation in the next section.

4.2 The effect of volume conduction on the estimation of DC and PDC between scalp channels: a realistic simulation

Whether PDC and DC estimated from EEG scalp channels are affected by the instantaneous mixing of cortical sources (volume conduction) is a debated issue. Kaminski and Blinowska, who introduced the DTF estimator, claim that DTF is not influenced by volume conduction (Kaminski & Blinowska 2014) because it is nonzero only if there is a phase difference between channels. Since the volume conduction is an instantaneous propagation of electromagnetic field that does not generate phase difference at electrodes, it should not influence connectivity estimators DC/DTF and PDC. To demonstrate that Kaminski performed a simulation adding a sinusoid at 20 Hz to the set of EEG signals (with the same phase for each EEG channel) and showed that the estimated DFT was not affected. Other authors (Schlögl & Supp 2006) justify the independence of PCoh and PDC from the volume conductions effects because of the property of these estimator to remove the components common to other channels from the estimation of linear dependencies between two channels.
However we have shown in the previous section how instantaneous correlation between residuals affects the estimation of lagged causality. These instantaneous effects are likely to represent, in the case of the EEG signal, the effect of the zero-lag mixture and propagation of sources across the head tissues on the scalp. Moreover a recent work (Billinger et al. 2016) has demonstrated, both mathematically and with a simulation, that scalp

connectivity (DTF) is adversely affected by volume conduction. The simple simulation model used involves only 3 scalp electrodes placed directly above 3 coupled sources and is lacking a more realistic characterization of volume conducted sources currents. Similarly in the work by Faes and colleagues (Faes & Nollo 2011, Faes et al. 2013) volume conduction is modelled imposing directed zero-lag dependencies between signals: this, as suggested in the previous section, may not realistically model the effect of source mixing on scalp signals. The work of Faes suggests that PDC and DC reduce the effects of instantaneous correlation if compared to Coh and PCoh, however a comparison of the performances of the different MVAR spectral estimators by means of a realistic model of volume conduction of sources is lacking.

In a series of studies Nunez and colleagues (Nunez et al. 1997, 1999, Srinivasan et al. 1998) have investigated and extensively characterized the effects of volume conduction on scalp Coh estimates as a function of the interelectrode distance and reference choice. Using a multilayer sphere model of the head they have characterized both analytically and with simulations the Random Coh between scalp potentials only due to the effect of instantaneous volume conduction from random uncorrelated cortical sources (called *Random* because it does not arise from actual correlation among underlying cortical sources but exclusively from the spread of source signals on the scalp (Nunez et al. 1997)). In particular analytical and simulations results showed that for an inter-electrode distance of 4 cm the *Random* Coh is in the range 0.4-0.9 (for a linked ears or average reference) (Nunez et al. 1997, 1999, Srinivasan et al. 1998). They also have shown how Coh estimates are affected by the reference choice providing a useful baseline for interpreting the physiological significance of coherence estimates for scalp data. To the best of our knowledge, a similar characterization of volume conduction effects for MVAR derived estimators (DC, DTF, PDC and orPDC) is lacking.

4.2.1 Objectives

In this last section of the chapter we will address the issue of the impact of volume conduction (instantaneous mixing of cortical sources) on EEG lagged causality by means of a series of simulations and using a realistic head model. The aim of this second study is to characterize volume conduction effects on DC, DTF, PDC and orPDC (estimated with the classical scMVAR) as a function of the interelectrode distance and reference electrodes. We believe that these issues are fundamental for the correct physiological interpretation of scalp connectivity (estimated with DC or PDC) as compared to the connectivity between underlying sources.

4.2.2 Methods

Differently from the simulation study II (section 4.1.4) where EEG signals were generated by the instantaneous mixing of sources weighted by the inverse of the squared 'distance' from the scalp, in this study a Lead Field Matrix (LFM) from a realistic head model (Montreal Neurological Institute, Collins et al. (1994)) describing the signal propagation through the tissues of the head was used. The high-resolution LFM was available in the software eConnectome (Bin et al. 2011). The Montreal Neurological Institute brain was defined by using a large series (250) of MRI scans on normal subjects to obtain a 'brain' representative of the population. The head model determines the way the current dipoles at a given cortical location produces the scalp signals and includes electromagnetic and geometric properties of the whole head volume, thus taking in account both superficial and deep sources. Once the properties of the head model are known it is possible to generate a LFM. Each j_{th} column of the LFM describes the potential distribution generated on the scalp electrodes by the j_{th} source current dipole. Since the Montreal Neurological Institute cortex is segmented in 7850 triangles, the high resolution LFM relates the 7850 current dipoles centred in the triangles (source space) and the voltage at the 62 channels on the scalp (sensor space).
In a first simulation scalp potentials for 62 electrodes were generated by mixing uncorrelated uniformly distributed source dipoles. We selected 1600 sources (out of the 7850) evenly distributed on the cortex surface (figure 4.18) [10] and imposed the white noise variance equal to 10^{-2} (mA*mm), that corresponds to the typical dipole strength from a patch of $1cm^2$ of cortex (Hämäläinen & Sarvas 1989).

The sensor measurements $\boldsymbol{x}(n)$ were generated as a linear instantaneous mixture of sources $\boldsymbol{s}(n)$ through the LFM $\boldsymbol{M}, [62 \times 1600]$

$$\boldsymbol{x}(n) = \boldsymbol{M} * \boldsymbol{s}(n), n = 1, \cdots, N \tag{4.20}$$

where N is the number of samples considered (5000).
A MVAR model (scMVAR) was then estimated from the scalp time series and then the MVAR spectral estimators (including Coh and PCoh) computed using the usual equations (3.5,3.19,3.23,3.25). The generation-estimation procedure was repeated 100 times. At each repetition random uncorrelated source signals were generated using the Matlab function rand.m, that generates uniformly distributed random numbers. The scalp potentials were referenced with respect to the average value across all electrodes

[10] Nunez and colleagues used 4200 dipole sources in their simulations. In the current work, we maintained the order of magnitude but cut down the number of sources considered in order to reduce computational cost.

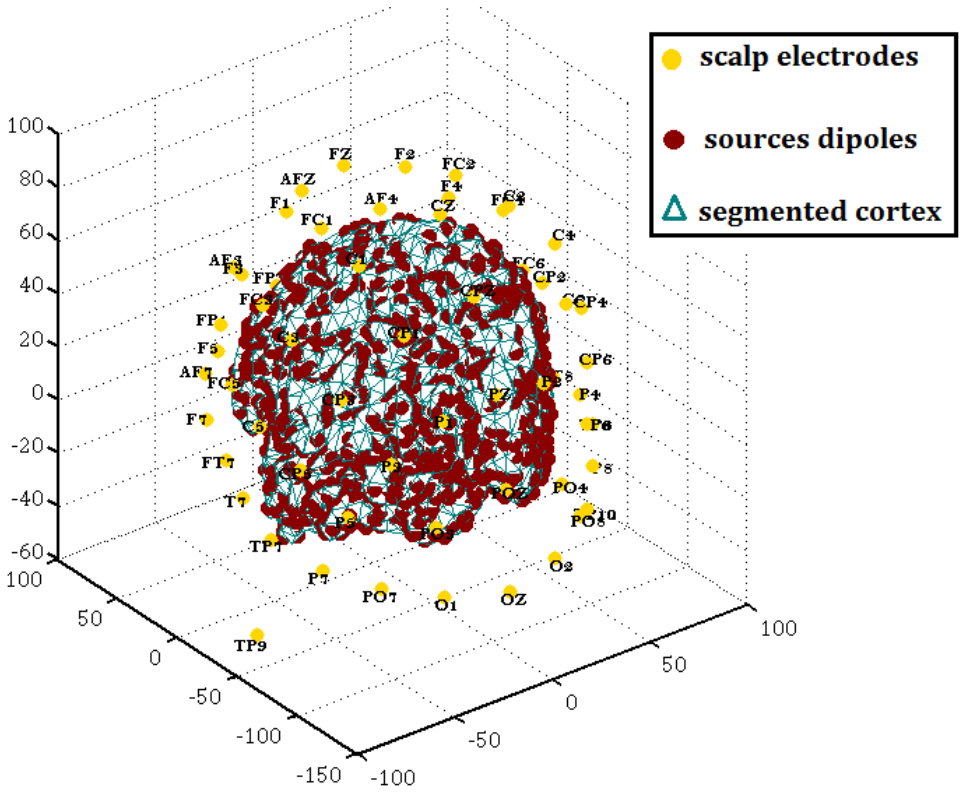

FIGURE 4.18: Locations of the sources dipoles (1600, in red) selected on the cortex surface to generate the voltage distribution on the scalp and selected electrodes locations and labels (62, in yellow). The figure shows how the source dipoles are evenly distributed and cover all the cortex surface.

(Average Reference) and the average of channels T7 and T8 (Linked mastoids Reference) as in Nunez et al. (1997). For each index, the estimated values of 'random' connectivity[11] between channels were plotted as a function of the interelectrode distance and for different frequencies in the interval (1, fs/2), with sampling frequency set equal to 250 Hz.

In order to reduce the computational cost we repeated the same procedure using a reduced number of electrodes (32) and sources (62) (figure 4.19), obtaining very similar results. In this case we also introduced some coloured noise sources: The uniform noise was pass-band filtered with Butterworth filters (order 3) with central frequencies ranging from 1 to 121 Hz and pass band of roughly 4 Hz in order to obtain uncorrelated coloured noise sources. The hypothesis of uncorrelated sources is fundamental here because the aim of the simulation is to quantify the random connectivity generated only by instantaneous mixing of sources, in absence of dependencies between them. The independency of sources was therefore tested by means of a Kendall test for each

[11]The term Random Coh introduced in the previous section to indicate spurious non physiological Coh introduced only by the volume conduction of uncorrelated sources, will be here extended to *Random Connectivity* to indicate the volume conduction induced spurious connectivity assessed with DC/PDC.

repetition. The usual criterions for optimal model order selection (SBC) and model validation (whiteness of residuals) were used.

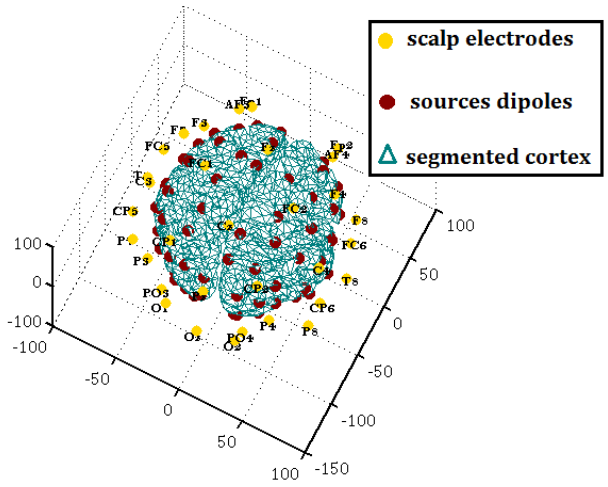

FIGURE 4.19: Locations of the 62 sources dipoles (red) selected in the second simulation study to generate the voltage distribution on the scalp and 32 selected electrodes locations and labels (yellow). The figure shows how the scalp electrodes and the source dipole are evenly distributed on, respectively, the scalp and cortex surface.

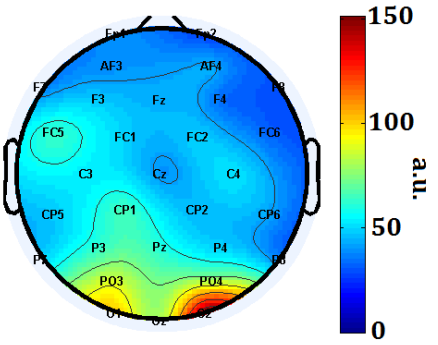

FIGURE 4.20: Example of scalp signals variance distribution for the linked mastoid reference and white sources showing how the signals at electrodes in different locations have different power.

4.2.3 Results

An example of the simulated scalp variances is given in figure 4.20: even if source signals are generated with equal variance, scalp channels have different power. This is because sources contributions are weighted differently depending on their position and the electrical property of the local tissues and correspond with unbalanced power distribution observed in recorded EEG signals (i.e. the EEG power of some channels is higher than for electrodes in other locations).

Figures 4.21 and 4.22 show results for squared Coh, PCoh, DC and PDC estimated respectively from average and linked mastoid referenced scalp signals and for f=10 Hz. The white noise sources are uncorrelated, therefore the scalp electrodes non-zero correlations are only due to the instantaneous volume conduction and reference electrode effects. Results indicate that, as extensively described by Nunez and colleagues for the Coh (Nunez et al. 1997, Srinivasan et al. 1998), the random Coh and PCoh values are significantly different from zero, especially for shorter interelectrode distances, and are affected by the reference choice. The Coh show expected trends (see figure 4.23 from

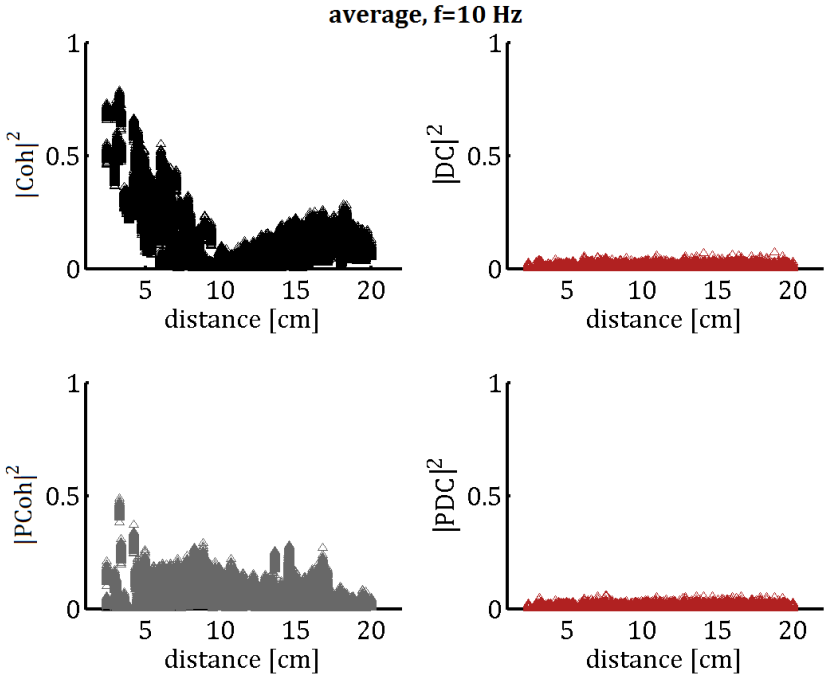

FIGURE 4.21: Simulated squared measures (Coh, PCoh, DC and PDC) of 'random connectivity' estimated from scalp signals at frequency 10 Hz as a function of the interelectrode distance. Sources are 1600 dipoles distributed on the cortex surface and scalp signals are referenced with respect to the average potential. Each plot is showing results for all the possible pairs of 60 electrodes and 100 repetitions (for a total of 354000 points).

Srinivasan et al. (1998) for a comparison): for the average reference, random Coh exhibits values ranging from 0.4 to 0.9 for surface distances shorter than 5 cm. Coh falls off to zero for distances of roughly 10 cm. At larger distances there is a small rise in coherence as the each source contributes a small negative potential due to the known behaviour of superficial dipoles[12]. The simulated Coh values for the mastoid reference show a more gradual decay with the interelectrode distance and long distances volume conduction effects are close to zero. The random PCoh trends are similar to those observed for the Coh, however random PCoh shows lower values for both the reference choices (mostly below 0.5 also for distances lower than 5 cm) and more steeply falls off with distance in the linked mastoid reference case.

Interestingly both random DC and PDC are close to zero for all the interelectrode distances range and are not affected by instantaneous mixing of white random sources. In

[12]The field of each dipole source falls off to zero with the square of distance but then rises again with opposite sign for larger distances (Srinivasan et al. 1998). A dipole in a spherical volume conductor generates a potential (with respect to infinity) whose integral over the surface is zero (Bertrand, 1985). This is also valid for multilayer spherical volume. The positive and negative potentials of each dipole always appear anti-symmetrically on the sphere surface, thus their integral is zero. This behaviour is observed for the reference with respect to infinity and for the average reference, but is not observed for single channel or linked mastoid reference (see figure 4.22). The average potential in fact approximate the potential at infinity (is close to zero) when numerous and widespread channels are considered.

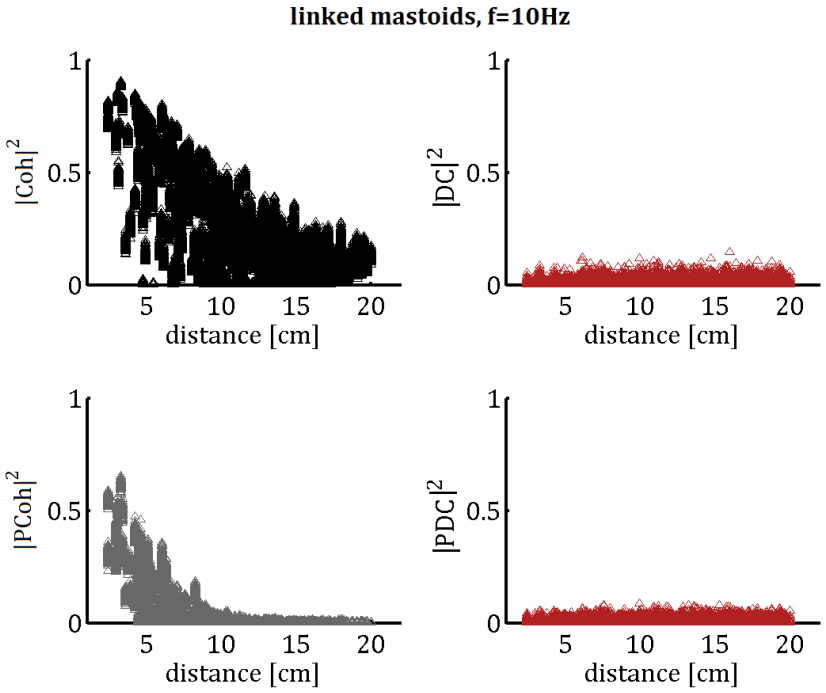

FIGURE 4.22: Simulated squared measures (Coh, PCoh, DC and PDC) of 'random connectivity' estimated from scalp signals at frequency 10 Hz as a function of the interelectrode distance. Sources are 1600 dipoles distributed on the cortex surface and scalp signal are referenced with respect to the (digitally) linked mastoids. Each plot is showing results for all the possible pairs of 60 electrodes and 100 repetitions (for a total of 354000 points).

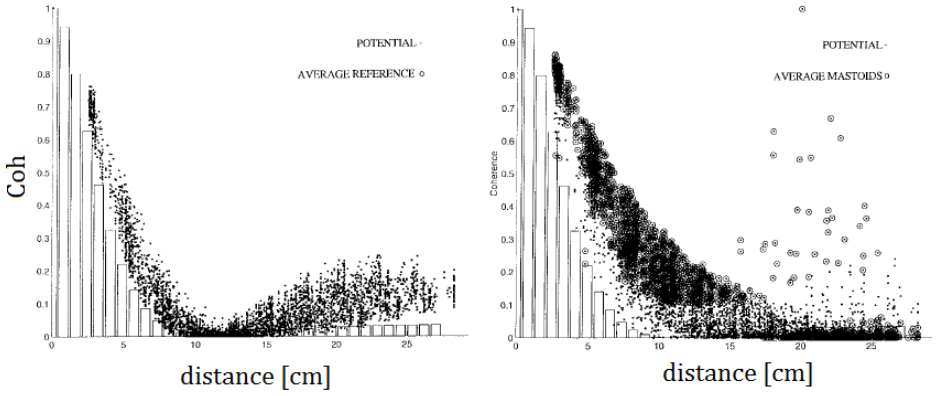

FIGURE 4.23: Results from Srinivasan et al. (1998). Simulated random coherence due to uncorrelated white sources for (a) average reference and (b) linked mastoids. The scalp potentials are simulated by mean of a four spheres head model. The bars indicate the reference-independent analytical solution. The dots represent simulated data for 111 scalp electrodes. Adapted from Srinivasan et al. (1998) and reported here in order to compare with the Coherence results obtained from the present simulation study and showed (black triangles) in figures 4.21 (average reference) and 4.22 (linked mastoid).

figures 4.21 and 4.22 results for one frequency (f=10 Hz) are shown. These results are consistent for all the frequencies analysed (see figure 4.24 where results for the average reference for another frequency, f=1 Hz, are shown to give an example) as expected since

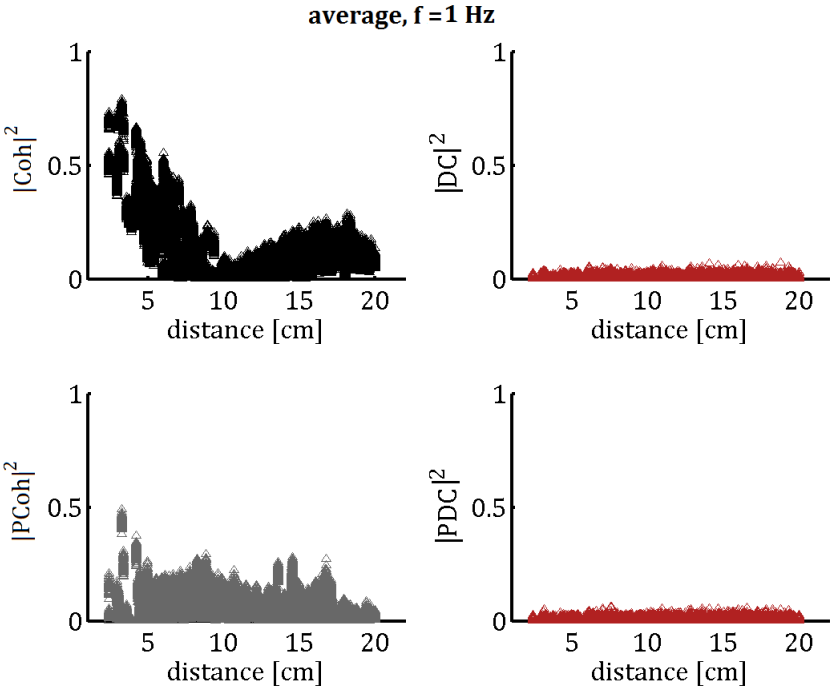

FIGURE 4.24: Simulated squared measures (Coh, PCoh, DC and PDC) of random connectivity estimated from scalp signals at frequency 1 Hz as a function of the inter-electrode distance. Scalp signal are referenced with respect to the average potential. Results are remarkably consistent with those showed in figure 4.21 for f=10 Hz and show how the instantaneous volume conduction of white sources is independent from frequency.

the instantaneous mixing of sources (the passive conduction of dipole currents across the tissues of the head) is independent of frequency[13].

These trends are consistent also if a smaller number of sources and electrode is considered, as for the second simulation study. The Kendall test indicated that source signals were independent for all the iterations. We also performed a test of the whiteness of residuals to check for MVAR estimation accuracy: the whiteness test indicated that simulated scalp potentials were accurately modelled for most of the repetitions (residuals were found white in 89 out of 100 repetitions for the average reference and in 88 repetitions for the linked mastoid reference). The optimal MVAR model order was found equal to 1 for all the repetitions and for both reference choices. In figure 4.25 averaged results across the 100 iterations for the linked mastoid reference are showed for 5 frequencies (1, 10, 20, 50, 100 Hz). Random PDC and DC average values are lower than 0.005 for all frequency surface distances. Results remarkably matches for all the frequencies considered as expected since, as previously mentioned, in the range of frequencies

[13]Over the frequency range usually considered for EEG analysis, tissue resistivities are insensitive to temporal frequency and the spread of cortical current on the scalp can be modelled as linear weighed sum of signals (Nunez 1981).

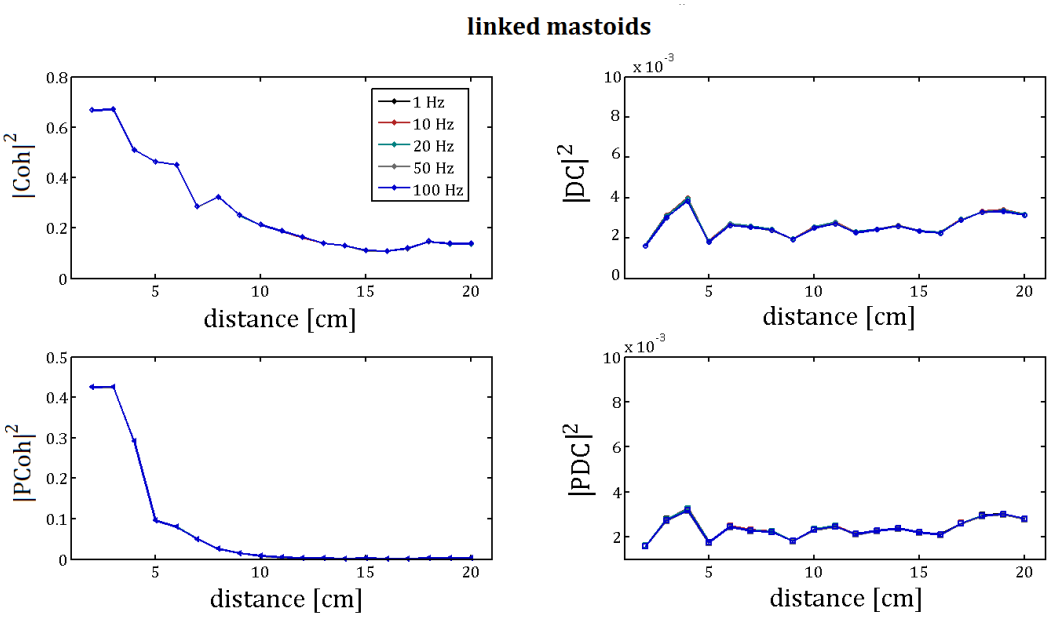

FIGURE 4.25: Average values (across 100 repetitions) of simulated squared measures for the realistic head model showed in figure 4.19 (62 sources, 32 electrodes). Scalp potentials are generated from white uncorrelated sources and referenced with respect to the linked mastoid. Although perfectly overlapping in this case, results for 5 frequencies are showed.

considered for EEG analysis, the volume conduction can be modelled as passive current spread independent from temporal frequency.

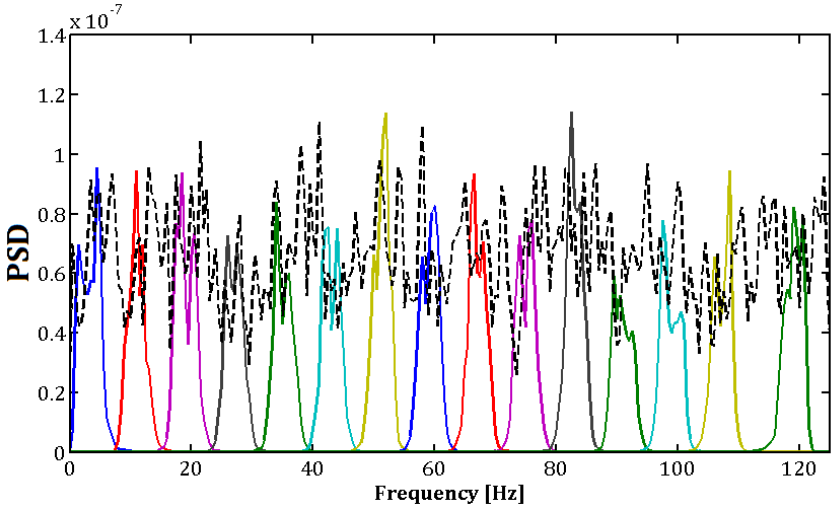

FIGURE 4.26: An example of Power Spectral Density -PSD (computed using the Welch method, Hanning window, 50% overlap) for 15 coloured sources as compared to a white noise dipole signal (dashed black line).

We have also investigated the effect of coloured uncorrelated sources on causality measures. As described in the methods section, 30 of the 62 source signals were generated by band pass filtering the white noise in order to obtain narrow band coloured sources, with

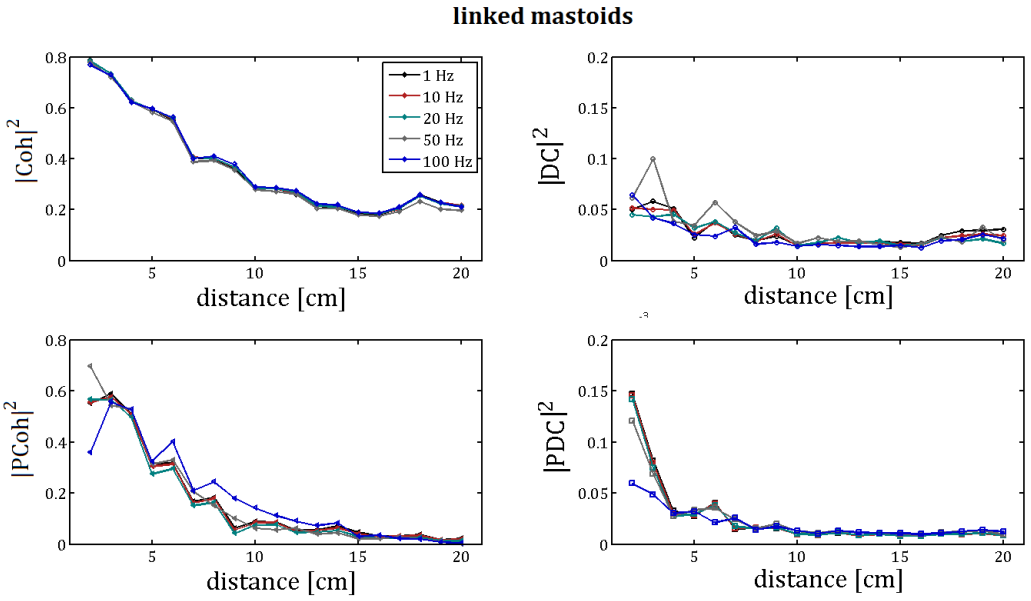

FIGURE 4.27: Average values (across 100 repetitions) of simulated squared measures for the realistic head model showed in figure figure 4.19. Scalp potentials for 32 electrodes are generated from coloured uncorrelated sources (62) and referenced with respect to the linked mastoid.

non-overlapping central frequencies. An example of the PSD for 15 coloured sources in comparison with a white noise dipole PSD is given in figure 4.26. Also in this case sources were found independent for all the repetitions (Kendall test). In this case the optimum MVAR model order was equal to 2 for both the linked mastoid and average references in all the repetitions, suggesting that the instantaneous mixing of coloured signals introduces some lagged correlation between scalp channels.

As figure 4.27 shows, Coh and PCoh exhibit average trends similar to those observed for white noises sources, however random PCoh show slightly higher values for coloured sources across all the distances range. Random DC and PDC are significantly affected by volume conduction of coloured sources and show much higher values than those observed for white uncorrelated sources, in particular for interelectrode distances lower than 10 cm. In the case of coloured sources small differences in estimates for different frequencies can be observed as well, which are expected (Nunez et al. 1997) and caused by the differences in the spectral properties of the sources. Interestingly, the random PDC exhibits higher values for close electrodes than the DC (figure 4.27, right plots). On the other hand random PDC shows a monotonic decrease with interelectrode distance, while random DC exhibit a slight increase for distances larger than 15 cm. In both cases volume conduction induced connectivity is lower than 0.05 for interelectrode distances larger than 10 cm.

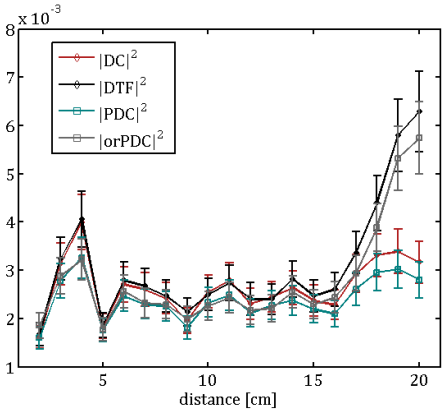

FIGURE 4.28: MVAR estimators for white noise sources: a comparison. Mean and standard errors (across 100 repetitions) values for squared DC (red), DTF (black), PDC (blue) and orPDC (gray) as a function of interelectrode distances. Scalp potentials are generated by volume conduction of 62 white sources. The figure shows how random connectivity generated by white uncorrelated sources and quantified with MVAR estimators is close to zero for all the interelectrode distances considered.

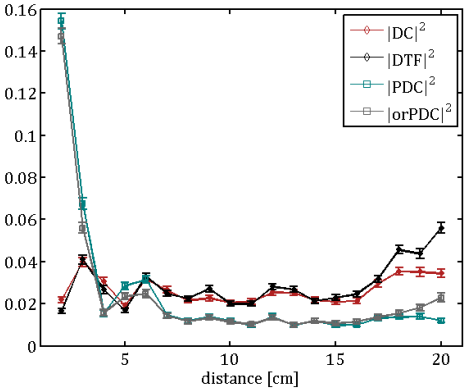

FIGURE 4.29: MVAR estimators for coloured noise sources: a comparison. Mean and standard errors (across 100 repetitions) values for squared DC (red), DTF (black), PDC (blue) and orPDC (gray) as a function of interelectrode distances. Scalp potentials are generated by volume conduction of coloured sources. The figure shows how random connectivity generated by coloured uncorrelated sources and quantified with MVAR estimators is different from zero for interelectrode distance shorter than 10 cm. DTF and orPDC random connectivity for large interelectrode distances ($>$ 15 cm) it is higher than the respective measures DC and PDC.

In figures 4.28 and 4.29 the average trends of the four estimators DC, DTF, PDC and orPDC are compared for white and coloured sources respectively. In both cases DTF and ordinary PDC (orPDC) show higher values for large interelectrode distances if compared the respective estimators normalized by signals variances (DC and PDC). This may be a results of the poor performances of DTF and orPDC (as compared to DC and PDC, respectively) when the time series of the multivariate set have different variances (Baccalà & Sameshima 2007). As shown in figure 4.20, distant electrodes are more likely to have different variances: as a result the unscaled estimators DTF and orPDC erroneously overestimate random connectivity for large interelectode distances, while DC and PDC show more robust results.

4.2.4 Summary

In summary, we have investigated the effect of instantaneous mixing of uncorrelated cortical sources on scalp connectivity using a realistic model of the head (Montreal Neurological Institute brain). In particular we looked at the effects of the type of sources

(white or coloured), the scalp interelectrode distance and the reference choice. The aim of the study was not only to confirm that instantaneous source spreading affects DC and PDC, but to investigate under what conditions lagged connectivity is significantly affected and how this vary as a function of interelectrode distance. Results showed that Coh and PCoh are significantly affected by volume conduction both for white and coloured sources (and roughly to the same extent in the two cases). As expected, by subtracting the common linear effects, the PCoh remove some of the volume conducted random connectivity and shows lower values of random connectivity if compared to Coh. This is, however, not enough to remove all the volume conducted connectivity generated by white or coloured uncorrelated sources. On the other hand, DC and PDC are affected only by the instantaneous mixing of coloured sources, when random connectivity is significantly different from zero especially for close electrodes. A comparison of the different MVAR derived spectral estimators confirmed results in literature by Baccalà & Sameshima (2007): DC and PDC perform better than DTF and orPDC in the case of unbalanced signal variances (i.e. for distant electrodes, in this case). We have also shown that, despite being affected by volume conduction, DC and PDC significantly reduce the spurious effects of uncorrelated sources mixing if compared with Coh and PCoh. In particular our preliminary results show that volume-conducted spurious connectivity estimated with PDC or DC is reduced to almost zero for interelectrode distances larger than 10 cm.

4.3 Discussion

Several methods have been proposed to more reliably estimate multi-channel EEG connectivity by (partially) removing the spurious, non-physiological coherence introduced by volume conduction. One approach reduces the correlation between scalp channels by applying a Laplacian filter to the EEG time series. The Laplacian is the second spatial derivative of the scalp potentials and it was shown to remove almost all volume conduction artifacts for distances larger than few centimetres (Nunez et al. 1997). This approach is based on the assumption that scalp potential are generated mainly by superficial radial dipoles and is insensitive to tangential or deep sources. Moreover it has been shown that the surface Laplacian improves the resolution at intermediate length coherence however it underestimate the value of many coherences at long distances (> 20 cm) because its spatial bandpass characteristic attenuates low spatial frequencies (Srinivasan et al. 1998). Moreover, the application of such a filter is thought to alter the phase distribution of the original time series, eliminating any physiological phase relationships

and therefore is considered an invalid method of calculating coherence and directed coherence (Kaminski & Blinowska 2014, Thatcher 2012) [14].
Another approach uses inverse imaging methods (see Appendix B) to estimate the time course of cortical sources activity in areas of interest (ROI). This approach requires a realistic electric and geometric model of the head and the solution of the inverse problem (the problem of finding the putative cortical sources given the scalp voltage distribution). It involves therefore a considerable computation effort that may not be compatible with real time EEG analyis.

Classical MVAR models do not include zero-lag coefficients, therefore disregarding the modeling of possible volume conduction effects. However the presence of instantaneous effects, often ignored in practical applications, affects the estimation of pure lagged causality and invalidates the assumption in the usual definition of Granger causality, that only the lagged interaction need to be considered in the computation of 'causality'. To deal with the misleading effects of zero-lag interactions on the estimation of time-lagged causality, an eMVAR framework has been introduced (Shimizu et al. 2006, Hyvarinen et al. 2010, Faes & Nollo 2010). The eMVAR approach includes instantaneous effects in the model introducing zero-lag coefficients. The aim of the first series of simulations was to compare the performances of the eMVAR approach and the classical scMVAR method in estimating lagged connectivity from imposed connectivity models of EEG signals and on recorded EEG data. For this purpose the two frameworks were applied in two simulation studies and eventually on EEG recordings (appendix B). In the first study the simulated signals were generated from an imposed connectivity scheme including directed zero-lag interactions. Results showed that, in agreement with previous finding, the eMVAR approach leads to better estimates of pure lagged causality.

In the second simulation study the frameworks were applied to a more realistic EEG model, where instantaneous effects were not imposed a priori but generated as zero-lag mixing of source signals. Results in this case showed that the reliability of the non-Gaussian eMVAR approach was low and the eMVAR based connectivity did not perform notably better than the classical scMVAR in removing the volume conduction effects. In addition, the eMVAR based spectral measures showed a dramatic decrease of spectral resolution when applied on EEG recordings, exhibiting flat profiles in frequency. These findings suggest that, when applied on more realistic simulations or on recorded EEG data, the non-Gaussian approach may fail to characterize reliably zero-lag interactions

[14] *'The process of adding together all of the phase differences from all electrodes destroys the physiologically based time differences that were present in the original time series and replaces the original time series with scrambled phase and thus an inability to accurately relate coherence and phase to an underlying neurophysiology'* (Thatcher 2012).

and that caution is required when interpreting results of this approach. This may be due to both the complexity of this computational method and the underlying assumptions which are incompatible with omnidirectional volume conduction effects. The authors of the non-Gaussian algorithm Shimizu et al. (2005) have in effect expressed their concern about the computational stability of this method, alerting on the fact that it may be affected by the random initial conditions. The assumptions behind the application of this method may represent the fundamental problem. As described in section 4.1.1, the non-Gaussian algorithm is based on the hypothesis that the instantaneous effects can be arranged in a causal order, meaning they can be represented by a directed acyclic graph (DAG). Imposing a sequence to the instantaneous effects either based on a priori assumptions or on the non-Gaussianity may be the main reason why this method fails on more realistic EEG simulations or on recorded EEG data, since from a physical point of view such a restriction may be incorrect. These considerations suggest that estimation and characterization of instantaneous causality in the EEG remains unresolved and challenging issue.

The scMVAR approach applied in realistic simulations and on recorded data is adversely affected by zero-lag effects. However the scMVAR connectivity estimators gave results in line with literature findings and were found to detect expected spectral features such as the spread of α oscillation from occipital to frontal areas, typical of EEG at rest with closed eyes. Of the two estimators considered (scDC and scPDC), scPDC was found to be less affected by instantaneous effects than scDC, indicating that it may represent a better estimator of lagged causality from EEG recordings.
Since our examples showed that the eMVAR approach is not able to remove the effect of instantaneous causality and that in practical EEG applications it may only add computational effort without improving the estimation of lagged causality, in the following we will use the classical scMVAR framework to estimate scalp connectivity because, even if affected by volume conduction, it represents a simpler, more reliable and established method than eMVAR.

In the study of the EEG signal the issue of instantaneous connectivity is mainly related to problems like volume conduction effects and the impact of a reference electrode (Nunez 1981, Nunez et al. 1997, Srinivasan et al. 1998). Given that the estimation and removal of instantaneous effects using the eMVAR remains a challenge, it is important to quantify the impact of volume conduction on scMVAR lagged causality in order to more appropriately interpret the physiological significance of connectivity estimators. We addressed this issue in a second series of simulations that more realistically modelled volume-conducted connectivity. Scalp potentials were generated by mixing random

source dipoles uniformly distributed on the cortex through a realistic model of the head (LFM).

Results in section 4.2 showed that DC/DTF and PDC/orPDC considerably reduce the effects of volume conduction as compared to Coh and PCoh and that for interelectrodes distance larger than 10 cm spurious DC/PDC connectivity is close to zero. In line with previous findings in literature, we showed that Classical Coh (and PCoh) are significantly affected by both white and coloured sources. Going beyond previous work we showed that DC and PDC are affected only by the instantaneous mixing of coloured sources. When sources are white their autocorrelation is positive only at lag zero (i.e. given a source signal, its samples at different time lags are not correlated). In contrast, for coloured sources, the autocorrelation is different from zero also for positive time lags (samples at different lags are correlated) and the instantaneous (zero-phase) mixing of non-white source signals affects scalp lagged causality estimated with DC and PDC.

In other words, we showed that the claim of Kaminski (Kaminski & Blinowska 2014) that the scMVAR derived estimator DTF/DC is not affected by zero phase mixing holds only for white sources. The zero phase mixing of coloured sources (being their lagged correlation different from zero) introduces some phase delay between scalp electrodes, and it therefore generate spurious connectivity.

Moreover, in contrast to the hypothesis that PCoh and PDC partialize out the effects of volume conduction(Schlögl & Supp 2006), we showed that both PCoh and PDC are significantly affected by volume conduction: the partialization of Coh (PCoh) removes some of the volume conduction effects common to other electrodes but does not reduce them to zero.

A comparison of the different scMVAR derived spectral estimators confirmed that DC and PDC perform better than DTF and orPDC in the case of different scalp signal variances. In these study we also showed, in agreement with previous results on Coh (Nunez et al. 1997, Srinivasan et al. 1998), that the linked mastoid reference cancels some of the long distances volume conduction effects that are present if the average reference is used. These findings oriented the choice to use DC and PDC as connectivity estimators (from EEG data referenced with respect to the linked mastoid) in the following studies on sleep and anaesthesia (chapters 5 and 6).

That being said, this study should be considered as a preliminary work since it presents some limitations. Firstly we have investigated scalp voltage generated by a specific source dipole distribution (therefore using only one lead field matrix LFM): future work should investigate results obtained randomly varying the source dipole distribution, as in Srinivasan et al. (1998). A second consideration regards the sources signal choice: we first have investigated effects generated from a mixture of white uncorrelated sources as

in Nunez et al. (1997), and secondly introduced coloured sources by band-pass filtering the white noise to test the hypothesis that lagged connectivity is affected by the volume conduction of uncorrelated coloured sources. In future work realistic cortical dipole spectral features may be used to more accurately model volume conduction effects. For instance we could have estimated more realistic source signals applying the inverse model to EEG recordings. In this case however it would have been more challenging to ensure the independence of source signals, which was the assumption of the simulation study to assess random volume conducted connectivity. We believe however, that despite these limitations, our preliminary result provide novel contribution in the characterization of volume conduction impact on PCoh, PDC and DC, and can represent useful guidelines for the interpretation of physiological significance of PDC and DC for the experimental studies on sleep and anaesthesia that will be outlined in the next chapters. It also opens the way for a new series of simulations studies that could provide considerable insight in the interpretation of EEG MVAR connectivity.

Chapter 5

Functional Connectivity Analysis of Sleep

The spectral connectivity estimators previously described and tested in simulation studies were applied for the analysis of EEG recordings collected during a sleep experiment. Considering the results of the simulations and their properties, the DC and PDC estimators derived from the classical MVAR model were applied in this first experimental study. In this study the changes in functional brain connectivity with level of consciousness were explored comparing wakefulness with different sleep stages. In this chapter the experimental design and methods will be described, then results from the EEG connectivity analysis will be shown and discussed.
Preliminary results from the sleep study were presented at the PGBiomed 2015 conference in Liverpool and at the MEDICON 2016 (XIV Mediterranean Conference on Medical and Biological Engineering and Computing) in Paphos (Cyprus).

Results from this study are published in: Lioi G, Bell SL, Smith DC, Simpson DM. *Directional Connectivity in the EEG is able to discriminate wakefulness from NREM sleep*. Physiol. Meas. 38 (2017) 18021820

5.1 Introduction

Sleep is a naturally-occurring rest condition *'in which the eyes usually close and consciousness is completely or partially lost, so that there is a decrease in bodily movement and responsiveness to external stimuli'* (Stedman 2005). The macrostructure of sleep consists of two different phases: NREM and rapid eye movement (REM) sleep. During a

whole night's sleep there is a periodic alternation of NREM and REM sleep. The NREM phase is in turn divided in different stages classified on the basis of their neurophysiological features: N1, N2 and N3 sleep, the latter also named slow wave sleep (SWS) because the brain rhythms in this stage are characterized by highly synchronized oscillation at frequency $f < 4Hz$ (Iber et al. 2007). SWS represents the deepest stage of sleep, where the subject is not responsive and is unaware of the external environment. During REM sleep, in contrast, reports of dream-like conscious experiences are common. A nap generally includes only NREM sleep (N1, N2 and N3) with sleep N2 representing the largest proportion and the time spent in N3 increasing with the nap-length for naps longer than 10 minutes (Brooks & Lack 2006). Naps, as also used by Massimini and Tononi (Massimini et al. 2005*a*), therefore represent a convenient condition to investigate LOC at the onset of NREM sleep, with the advantage of only requiring a simple experimental setup that does not entail the practical challenges of overnight sleep recordings.

Sleep and anaesthesia are the result of distinct sequences of events (Bonhomme et al. 2011) and have different behavioural endpoints. However the view of general anaesthesia as 'deep sleep' has deep roots, and it is common to hear the expression 'putting someone to sleep' to indicate the administration of general anaesthetics. NREM sleep and anaesthesia share important features such as closing of the eyes, hypnosis, reduced mobility and amnesia. As described in section 2.1.1, beside these behavioural analogies, these two states share common EEG properties (slow waves patterns, sleep spindles and similar scalp distributions of EEG activity (Lydic & Baghdoyan 2005, Murphy et al. 2011)) and important neurophysiology mechanisms as brain networks involved in sleep regulation are modulated by anaesthetics (Mashour 2010). These important homologies represent one of the rationales of this sleep study as a first experimental step, with a view to assessing the depth of anaesthesia in the later stages of this project, and comparing the normative sleep study with the anaesthetic study. In this study, performed in a controlled laboratory setting, brain connectivity features changes with the level of consciousness were investigated on good-quality recordings. This represented a useful step before moving to a clinical setting typically affected by electrical noise.

Changes in cortical connectivity associated with sleep have been widely investigated in functional magnetic resonance (fMRI) studies, but their relationship with consciousness remains unclear (Klimova 2014). Results point to a general impairment of functional connectivity in the thalamocortical system (Spoormaker et al. 2010, 2012); in particular, long-range connectivity was shown to be affected by sleep (Tagliazucchi et al. 2013) and connectivity networks in NREM sleep showed increased local clustering when compared to wakefulness (Boly et al. 2012). Studies investigating early NREM sleep with combined

transcranial magnetic stimulation and EEG approaches (Massimini et al. 2005*a*) report a break-down of large-scale connectivity in the sleeping brain.

EEG is often preferred in studies of brain connectivity from a practical point of view (Sitt et al. 2014), as it can be applied relatively easily at the bed-side and at low cost, and EEG-based systems can be used in routine clinical work in the home or ward, as well as in intensive care units or operating theatres where assessment of level of consciousness may be carried out. Functional connectivity in the EEG during sleep has been investigated in previous studies mainly using correlation or coherence analyse, with inconclusive results (Achermann & Borbély 1998, Corsi-Cabrera et al. 2003) that show sometimes contrasting findings. The limitation of these studies is that, as demonstrated in the previous sections, Coh is a symmetric measure not able to convey directional information and which is strongly affected by volume conduction artifacts. A series of more recent EEG sleep studies have investigated the functional interactions between EEG channels by mean of the Synchronization Likelihood index (Ferri et al. 2005, 2008); the SL quantifies dynamical interdependencies in time, without an analysis in the frequency domain where important information of the EEG signal is coded. Except for one early application of DTF for a topographic analysis of EEG activity during overnight sleep (that will be considered in more detail in the Discussion, section 5.4), to the best of our knowledge there is no study that applies advanced functional connectivity techniques such as DC or PDC to the analysis of sleep EEG.

The aim of the current study is therefore to describe changes in strength and direction of functional connectivity associated with NREM sleep using Directed Coherence and to propose and test indexes of brain connectivity based on DC that could distinguish between states of consciousness. The performances of the proposed approaches will be compared to more established spectral measures and assessed in discriminating between NREM sleep and wakefulness. Since the overarching aim of any proposed measure is to assess the level of consciousness in individual subjects, performance is assessed against each subject's own time-line through the sleep stages. The focus on individual variability is an important and distinctive feature of this work, since the majority of previous studies investigating correlates of consciousness (where consciousness is diminished or reduced) have focussed on the analysis of the average values across the cohort. This work is also original in investigating directional connectivity (using Directed Coherence) on EEG, where previously the strength of connection, rather than the direction of information flow, was the focus. In view of applying the proposed methodology to the monitoring of anaesthesia we are looking for an index that is computationally convenient and suitable for online monitoring of individuals. In this sense the spontaneous EEG is preferable to fMRI and TMS approaches as it represents a low cost and easily implementable method with good temporal resolution.

5.2 Methods

5.2.1 Subjects, Experimental Protocol and Preprocessing

Sixteen healthy subjects participated in the sleep study. However only ten subjects (three females and seven males, aged between 22 and 30 years) underwent N1, N2 and N3 stages of NREM sleep and so were included in the analysis. The experiment was approved by the local ethics committee (University of Southampton) and following informed consent, and conformed to requirements of the Declaration of Helsinki. In order to exploit the circadian sleep drive, the experiment was performed in the afternoon after lunch. The subjects were asked to refrain from drinking coffee or tea on the day of the experiment. The subjects, lying with eyes closed on a reclining chair, were invited to sleep. After they spontaneously woke up again, they were asked to rest with eyes closed (REST W) and then to perform mental arithmetic with eyes closed (ACTIVE W). EEG was collected using a 32 channel system with active electrodes (Biosemi BV, Amsterdam) placed according to the international 10-20 system. Additional electrodes were used to record the electrooculogram (EOG) and the chin electromyogram (EMG) (see figure 5.1). The plan was to use auditory stimulation in order to investigate the changes of MLR features: however in pilot work subjects woke up when stimulated, demonstrating the challenges of using this procedure, which was then removed from the protocol. It was therefore decided to focus the analysis on brain connectivity for the sleep study. Sleep stages were scored by visual inspection of contiguous epochs of 30 s according to the standard criteria (Iber et al. 2007), summarized in table 5.1. The signals traces in the time domain were inspected and classified with the help of an EEG sleep atlas (Bonnet et al. 1992). Data were then downsampled to 250 Hz and digitally referenced with respect to the average of T7 and T8 channels (linked mastoids), as in previous EEG functional connectivity studies during wakefulness and sleep (Kus et al. 2005, De Gennaro et al. 2004). This reference choice is recommended for functional connectivity estimation (Kaminski & Blinowska 2014) and it was shown to reduce long distance volume conduction effects in our simulations and previous literature results (Srinivasan et al. 1998, Nunez et al. 1997). Moreover this choice preserves the symmetry of data, in contrast to a single non-midline electrode reference. The EEG time series were band pass filtered (1-45 Hz) and additionally notch filtered at the mains frequency using zero phase filters[1]. Only continuous and artifact-free epochs were selected and included in the following analysis.

[1]The filter order n_filt used was automatically selected by the function firfilt.m implemented in EEGlab and equal to

$$n_filt = \frac{3.3 * Srate}{0.25 * passband} \tag{5.1}$$

, where $Srate$ is the EEG sampling rate (250Hz in this case).

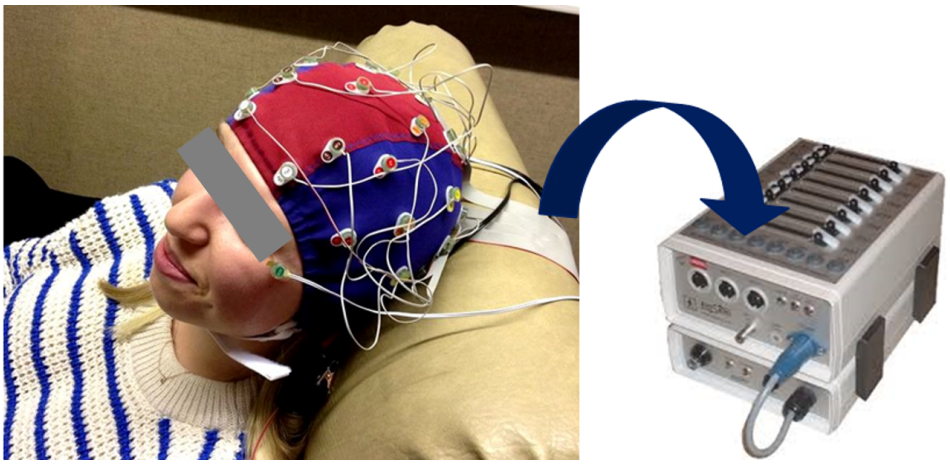

FIGURE 5.1: Electrodes set up for polysomnographic recordings.

Stage	EEG	EMG	EOG
Wakefulness	More than the 50% of the epoch has 50% α rhythm over the occipital region	High muscle tone	Present if the subject has open eyes. Irregular conjugate rapid eye movement
Sleep N1	Reduced α rhythm (with respect to wakefulness). Mixed frequency activity. Vertex sharp waves over the central region	Lower amplitude (with respect to wakefulness)	Possible Slow eye movements
Sleep N2	K complex (negative sharp wave immediately followed by a positive component) in frontal and central electrodes. Sleep spindles (train of waves f = 8-16 Hz and duration $<$ 2s) in central electrodes	Variable amplitude (lower than N1, as low as REM)	Not present
Sleep N3	Slow wave activity (SW) (f=0.5-2 Hz, amplitude $>$ $75\mu V$, frontal electrodes) in more than the 20% of the epochs. Spindle activity can persist.	Variable amplitude (lower than N1, as low as REM)	Not present

TABLE 5.1: Summary of sleep scoring rules from Iber et al. (2007)

5.2.2 Brain connectivity estimation and significance assessment

A subset of 12 electrodes (Fp1, Fp2, F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, O2) evenly distributed on the midline scalp was considered for connectivity estimation, in order to reduce the computational cost typical of multivariate connectivity estimation. Traces of continuous data of 60 s (i.e. 15000 samples in each channel) that were not interrupted by artifacts or high-level noise (according to visual inspection) were considered for connectivity analysis in order to include a sufficient number of data points for the estimation of the free parameters (see section 3.3.1). Therefore only segments with two consecutive 30 s epochs of the same sleep stage were included in the study while isolated epochs were excluded from the following analysis. For each segment a scMVAR model was estimated using the LS estimator implemented in ARfit (Matlab,(Schneider & Neumaier 2001)) and the DC and PDC computed as a function of the scMVAR coefficients in the frequency domain as extensively described in section 3.2. In particular the squared PDC and DC were used due to their higher accuracy and stability with respect to corresponding not squared measures (Astolfi et al. 2006) and their useful interpretation in terms of spectrum (or inverse spectrum) transfer (section 3.2).
When making inferences about EEG connectivity it is crucial to assess the significance of the estimator producing confidence intervals of statistical thresholds. In this study the significance of the causal links was assessed using a surrogate statistics based on a shuffling procedure. The shuffling was performed in the frequency domain and generated a set of surrogate data in which any temporal correlation between channels is removed. To this end the Fourier Transform of the time series was performed and their phases randomly and independently shuffled between frequencies, while keeping the magnitude of the Fourier coefficients unchanged. In this way the surrogate data have the same power spectrum as the original time series but the temporal order, and therefore the causality between signals, is removed. An MVAR model was fitted on the surrogate data and the PDC/DC values estimated. In order to obtain a reliable null distribution the shuffling procedure was repeated 1000 times, as in previous studies (Astolfi et al. 2006). In this way an empirical null PDC/DC distribution was obtained for each frequency and pair of channels. The significance of causal links was assessed comparing the estimated connectivity with the null distribution setting the statistical significance level at 0.01. Correction for multiple comparisons was performed using the false discovery rate (FDR) approximation for dependent measurements (Benjamini & Yekutieli 2001). Only links that were thus found to be statistically significant were included in the subsequent analysis and in the calculation of EEG indexes of connectivity.

A widespread practice in connectivity analysis is to threshold connectivity matrices to

remove weak or spurious connections and retain only a small percentage of the strongest links (Sporns 2013). In this study the connectivity matrices were thresholded to retain either 10% or 30% of the strongest connections, as in Chennu et al. (2014). The connectivity matrices were then averaged in the four frequency bands of interest delta (δ) [1-4] Hz, theta (θ) [4-7] Hz, alpha (α) [8-13] Hz and beta (β) [13- 25] Hz. In order to specifically investigate changes in long-range connectivity, the PDC and DC links were subdivided with respect to the 3D Euclidian interelectrode distances, computed using default channel coordinates. Distance thresholds were set for differentiating between three groups of channel pairs in roughly the same proportions with respect to the total number of possible links: 35% of short-range links (interchannel distances below 10 cm), 32% of medium range links (between 10 and 14 cm) and 33% of long-range links (above 14 cm).

5.2.3 Computation of EEG indexes

For each 60 s epoch of different sleep stages and for each subject, a series of EEG measures were jointly extracted. We organized indexes into two classes: Spectral measures (as commonly used in many previous studies) and connectivity based measures. The power spectral density (PSD) for each epoch and electrode was estimated using the Welch method (Hanning window 7.5 s long, 50% overlap); the power in each frequency band was calculated as the integral of the PSD within each frequency band. The spectral analysis was focused on δ, θ and α bands since previous studies of neural correlates of consciousness report major changes in these bands (Koch et al. 2016, Chennu et al. 2014). To allow for differences in power between EEG channels, we estimated the normalized power in these three frequency bands by the total power (1-45 Hz) in each time epoch and for each electrode. Normalized power is thought to be a more reliable estimator because it encompasses the individual variances in the absolute EEG power caused, for instance, by variations in electrodes impedances (Sitt et al. 2014). We assessed connectivity through indexes quantifying the strength of the connectivity networks and indexes estimating the direction of information flow. The rationale for the former is to be found in the large number of studies showing that long-range connectivity is significantly affected by the level of consciousness. The rationale for the latter comes from published results showing a prevalence of frontal EEG activation in sleep as opposed to a strong posterior activation in wakefulness (Brown et al. 2010), with indications that the direction of long-range connectivity may represent a prominent feature of sleep as compared to wakefulness (De Gennaro et al. 2004). To this end we assessed the number of significant connections from centro-posterior (O1, O2, P3, P4, C3, C4, Cz) to anterior (Fp1, Fp2, F3, F4, Fz) electrodes and vice versa. We thus defined an

index that quantifies the dominant direction of information flow on the front-posterior axis ($Dir_{P\to A}$) as the normalized differences of the number of links in the two opposite directions over the rostro-caudal axis:

$$Dir_{P\to A} = \frac{\sum_i \sum_j num(|DC|^2_{P\to A}) - num(|DC|^2_{A\to P})}{\sum_i \sum_j num(|DC|^2_{P\to A}) + num(|DC|^2_{A\to P})} \tag{5.2}$$

where the sums are taken over all pairwise connections between the posterior and anterior channels ($P \to A$) and anterior to posterior channels ($A \to P$), respectively and num(.)=1 when that connection is significant, and zero otherwise.
In the following, first some preliminary results on 5 subjects where the most promising connectivity features were identified will be presented. Then results at a group (cohort) level, showing differences in connectivity based measures between the different sleep stages and frequency bands, with associated statistical analysis will be shown. Finally, results at an individual subject level are presented, and the most promising measures are correlated (Spearman coefficient) with the manual scoring of sleep stages (hypnogram) in each individual, using the indexes from consecutive 1-minute segments. The presentation of results from individual subjects, in addition to the statistical analysis across the cohort, provides insight into the potential of the method in monitoring individual patients.

5.3 Results

According to standard criteria (Iber et al. 2007), hypnograms were generated for each subject, using consecutive one-minute artefact-free signal segments. The hypnograms represent the sleep stages visually identified from off-line scoring, and their assessment was carried out blinded to the connectivity analysis that followed. The following percentages in time (mean $\pm$ standard deviation) spent in the sleep and wakefulness stages were obtained across subjects: sleep N3 $21 \pm 9\%$, sleep N2 $27 \pm 4\%$, sleep N1 $16 \pm 7\%$, REST W $19 \pm 6\%$, and ACTIVE W $16 \pm 4\%$. Since volunteers were allowed to spontaneously wake up from the post-prandial sleep, sleep duration was highly variable across subjects (16.2 ± 5.9 minutes of sleep were analysed per subject).
As for the previous chapters, statistical tools were applied to check the validity of MVAR model estimation. A MVAR model was estimated for each of 60 s segments the experimental stages were divided into and the validation tests applied for each independent MVAR estimation; it is therefore unpractical to show all the results and general findings will be commented on. Residuals of the MVAR estimation were found not white both with the Ljung-Box Portmanteau (LJB) test and the Autocorrelation Function (ACF)

test in the majority of stages for all the subjects. The average probability to reject the whiteness hypothesis was around 0.2 for all the conditions and subjects (ACF test). This suggests that the full lagged correlation structure of the data was not always accurately estimated. However an examination of the variance of residuals revealed that the prediction error was extremely small (in the order of $10^{-4}\mu V^2$) if compared with the power of EEG data, which is in the order of $100 - 150\mu V^2$. For each stage the residuals covariance matrix was inspected and its non diagonality considered as an indication of the significance of zero-lag correlations among time series. Typical results are showed in figures 5.2 and 5.3 for ACTIVE W and SLEEP N2 respectively. In general, the residual covariance matrices were not perfectly diagonal, suggesting that some instantaneous effects existed among signals; however in the majority of cases only the covariances of residuals from adjacent leads were significantly different from zero, suggesting that the zero-lag effects were mainly introduced by volume conduction in adjacent electrodes but were negligible for distant scalp locations (long-range connectivity).

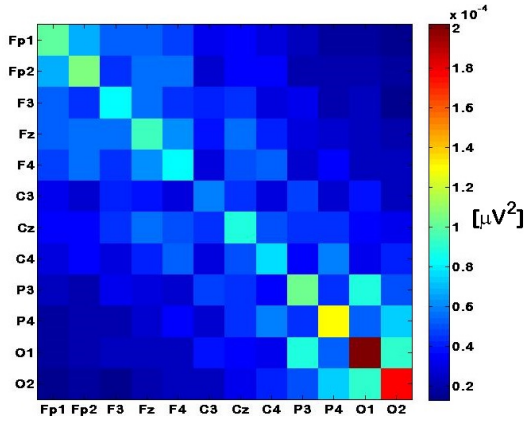

FIGURE 5.2: Example of the residuals covariance matrix for a 1- minute epoch of ACTIVE W in one subject. The matrix is not perfectly diagonal, indicating that some instantaneous effects existed among signals, in particular for adjacent electrodes. This suggest that the zero-lag effects are mainly introduced by volume conduction in contiguous electrodes.

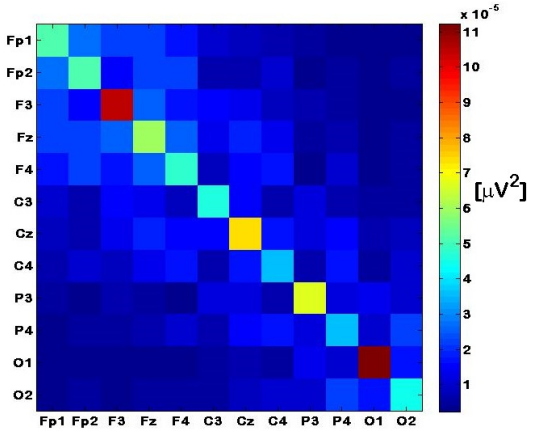

FIGURE 5.3: Example of the residuals covariance matrix for a 1- minute epoch of SLEEP N2 in one subject. Compared to wakefulness (figure 5.2), the residuals matrix in sleep appears less affected by zero-lag correlations since it is closer to be diagonal.

As a first step of the analysis, preliminary results from few subjects were analysed in order to identify the most discriminative connectivity measures.

Before analysing results at group level individual DC and PDC pattern were inspected. Figure 5.4 shows connectivity scalp plots estimated with DC for a representative subject (1). The scalp plot were obtained using the software eConnectome (Bin et al. 2011), a MATLAB toolbox for functional brain connectivity imaging. The DC connectivity networks in wakefulness (REST W and ACTIVE W) were found to be characterized by a

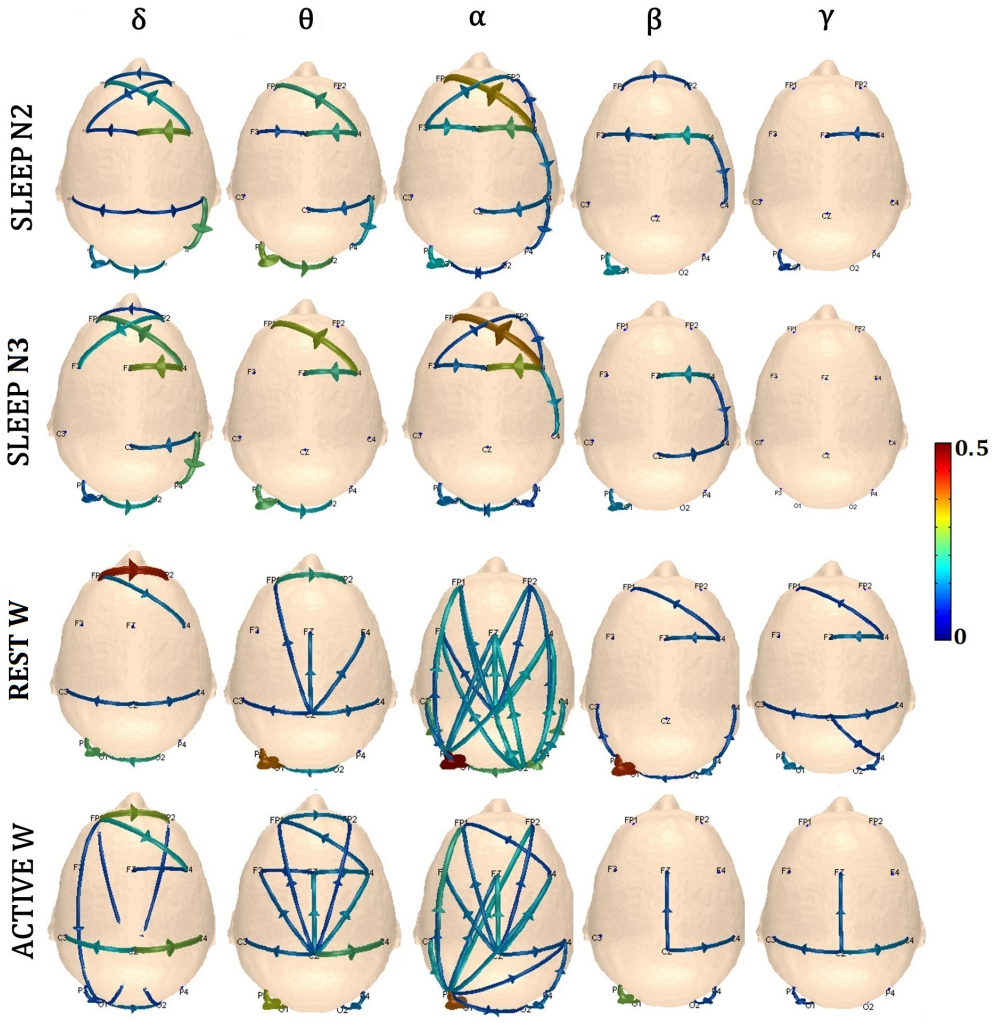

FIGURE 5.4: Examples of scalp maps of DC from a representative subject (subject 1). DC patterns averaged in the five frequency bands of interest δ (1-4 Hz), θ (4-8 Hz), α (8-13 Hz), β (13-30 Hz) and γ (30-45 Hz) for the four experimental stages SLEEP N2, SLEEP N3, REST W and ACTIVE W are showed. The strength of the connections is coded by the colour and size of arrows. Only significant links above the threshold of 0.1 are shown.

denser network of statistically significant connections with respect to sleep, in particular in the θ and α bands. These networks are characterized by the presence of long-range connections linking occipital, central and frontal areas in both hemispheres. The number of long-range connections considerably drops in SWS, in particular in the α range, and the connectivity patterns are localized, linking adjacent electrodes mainly in the frontal and occipital regions. In the δ band the number of links elicited during sleep is comparable to wakefulness. The topographies of the connectivity networks elicited in the two sleep stages are very similar. Sparse and weak connections are elicited in the β and γ bands in all the stages. A dense posterior to frontal spread of α rhythms can be observed in both in REST W and ACTIVE W, in analogy with results shown in section 4.1.4 and previous published results (Kamiński et al. 1997, Faes et al. 2013).

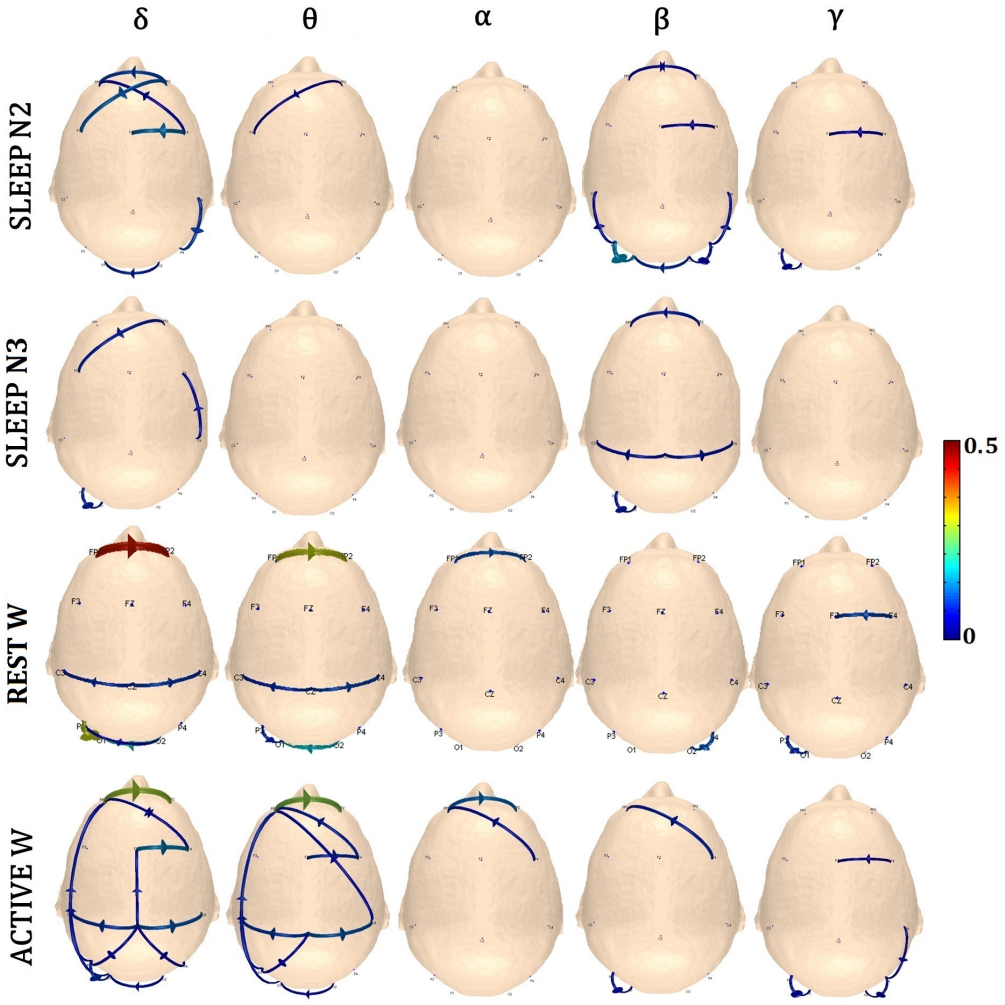

FIGURE 5.5: Examples of scalp maps of PDC from a representative subject (subject 1). PDC patterns averaged in the five frequency bands of interest δ (1-4 Hz), θ (4-8 Hz), α (8-13 Hz), β (13-30 Hz) and γ (30-45 Hz) for the four experimental stages SLEEP N2, SLEEP N3, REST W and ACTIVE W are showed. The strength of the connections is coded by the colour and size of arrows. Only significant links above the threshold of 0.05 are shown.

Figure 5.5 is a matrix plot of connectivity networks in sleep and wakefulness and estimated with the PDC (subject 1). Similarly to what observed for the DC, the connectivity networks elicited in sleep present sparse and localized connections while in ACTIVE W more long-range links can be observed in the delta and theta bands. However the number and strength of significant PDC links is significantly lower if compared with DC connections because the PDC estimator takes into account only direct pathways between structures. As a results the majority of scalp plots show fragmented or absent networks also in other subjects (results not shown).

Figures 5.6 and 5.7 show the number of significant causal links in two bands of interest (θ and α) estimated with, respectively, DC and PDC and averaged across all the available epochs of each experimental stage for five subjects. A trend of increasing number

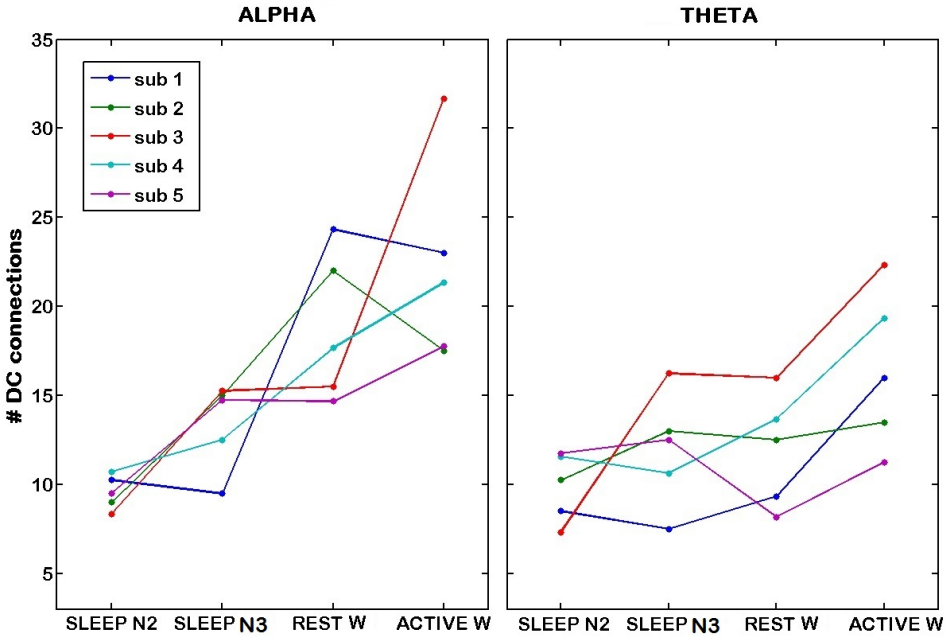

FIGURE 5.6: Mean number of significant DC links (averaged across all the epochs of each stage) as a function of the experimental state in the α (left) and θ (right) bands. The different curves represent results from 5 subjects of the study and show a trend of increasing DC links strength in the transition between sleep and wakefulness in the α band.

of connections with the level of arousal can be observed for the DC averaged in the α band (figure 5.6, left plot). It is interesting how the number of DC links exhibits a small scatter among subjects in sleep and a larger variability in wakefulness. This may be the results of EEG patterns in sleep being more stereotyped while in wakefulness brain activity presents a larger inter-subject variability. The number of PDC connections show similar trends, dramatically increasing in the transition sleep-wakefulness in the θ band (figure 5.7, right plot); however for few subjects the number of significant PDC causal links is close to zero in all the experimental stages, making the identification of a trend and the interpretation of results difficult (i.e. the number of PDC links in the alpha band in subjects 2, 3 and 5 shows little change, Figure 5.7). For this reason, and in order to reduce the number of parameters considered, we only included connectivity estimated with DC in the following analysis.

We calculated the average strength of magnitude squared DC links in the different experimental stages and then averaged this across subjects. In this case we differentiated short and long-range links on the basis of the interelectrode distance. Figure 5.8 shows DC average strength (across subjects and electrodes) as a function of the distance threshold and the sleep stages. The strength of connectivity links exhibit different trends across sleep stages depending on the distance range considered. Long-range connections are

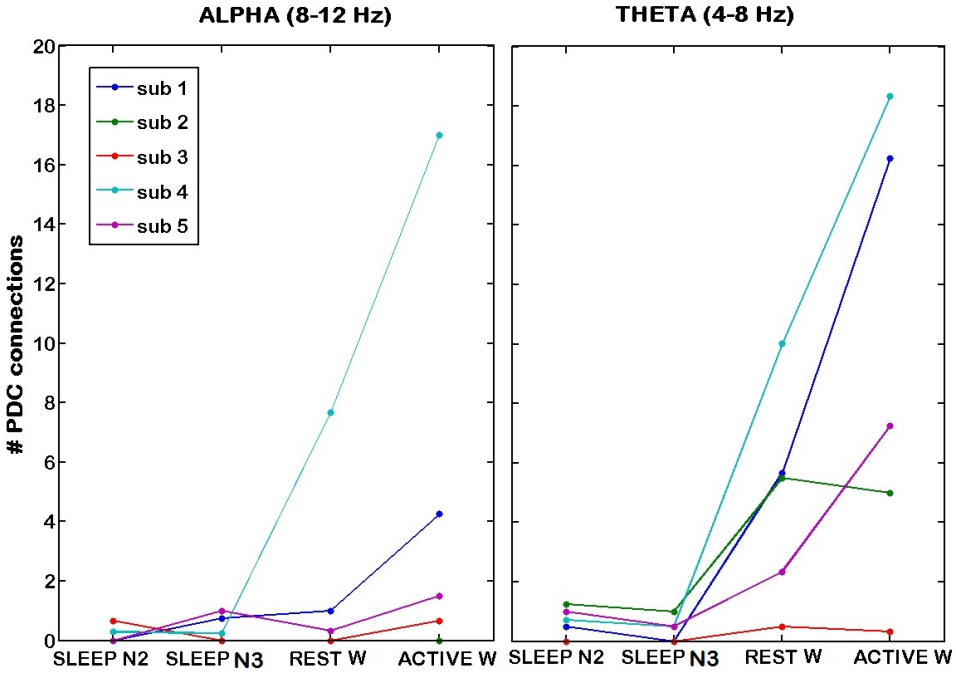

FIGURE 5.7: Mean number of significant PDC links (averaged across all the epochs of each stage) as a function of the experimental state in the α (left) and θ (right) bands. In order to reduce the great number of features available, a preliminary analysis on 5 subjects was carried out to identify the most promising features. For some subjects the mean PDC curves are close to zero in all the experimental stages. Due to the predominance of fragmented or absent networks estimated with PDC, we only included connectivity estimated with DC in the following analysis.

generally disrupted in NREM sleep where connectivity networks gain a more localized character (there is a prevalence of short-range links in sleep N2 and N3, in particular in the δ band) and long-range connectivity in the α band showed the best performance in distinguishing sleep from wakefulness. This is in accordance with previous works (Chennu et al. 2014). In order to further reduce the number of parameters investigated and for the sake of the clarity of displayed results, we therefore only included indexes relative to the α band as connectivity markers in the later results.

Figure 5.9 shows the group topographic characteristics of a number of features derived from the EEG (band power, strength of long range functional connections and the direction of functional connections) as a function of experimental stages, along with mean and standard error plots for those features. In order to reduce dimensionality and quantify the discriminative power of the different measures, we summarized the topographic information by averaging across electrodes and investigated whether the global indexes were able to discriminate NREM sleep from wakefulness (two level analysis) and also specific experimental stages (multilevel analysis). For the two level analysis a Mann-Whitney test explored whether the different markers in the two wakefulness stage (average of ACTIVE W and REST W) significantly differed from NREM sleep (average of N2 and

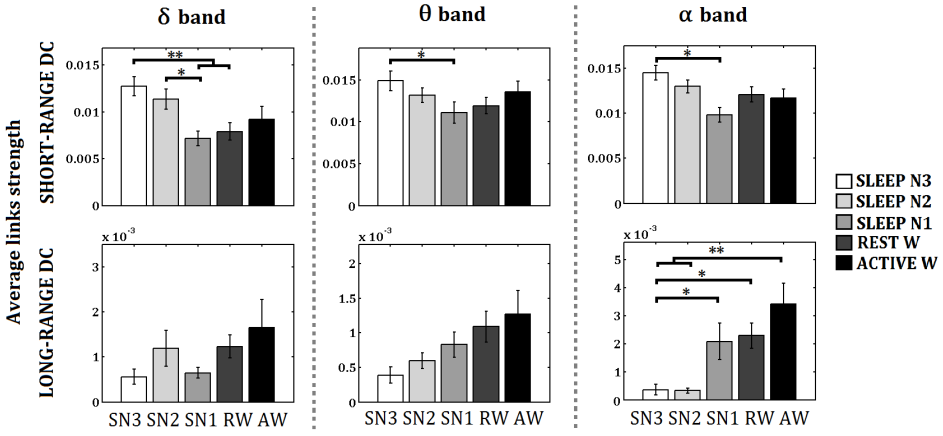

FIGURE 5.8: Short and long-range connectivity (as measured by DC) in δ, θ and α bands. Top bar row: short-range connections. Bottom row: long-range connections. Bars indicate average strength of 10% strongest DC (magnitude squared) links across subjects (N=10). The error bars represent the within group standard error. The asterisks specify that the two means designated by the brackets significantly differ (Friedman test with post-hoc analysis, $*p < 0.05$, $**p < 0.01$).Short-range connectivity is dominant in NREM sleep (N2+N3) while the strength of long-range links is reduced as compared to wakefulness, in particular in the α band.

N3). To test whether the measures could differentiate the specific sleep or wakefulness stages, a Friedman test was conducted, with post-hoc testing using Tukey's honestly significant difference (HSD) test. In figure 5.9 changes in normalized power can be seen that are consistent with those commonly reported in the literature: low frequency (δ and θ) power (rows 1 and 2 of figure 5.9) is dominant in NREM sleep and gradually decreases from SLEEP N3 to ACTIVE W. The two wakefulness states are characterized by a dominant occipital α rhythm (row 3 of figure 5.9). The power spectrum in all three bands significantly distinguished wakefulness from NREM sleep, as found when averaging the result of N2 and N3, and comparing these with the average of both stages of wakefulness (rest and active) ($p < 10^{-4}$, Mann-Whitney test). However, the normalized δ and θ power more efficiently discriminate NREM stages N2 and N3 from sleep N1 and wakefulness. The two-level analysis (sleep N2 and N3 vs. awake) follows the approach used by Massimini et al. (2010). As described in chapter 2, theoretical models of consciousness and experimental results obtained in sleep, anaesthesia and disorders of consciousness predict that the long-range information sharing is essential to maintain consciousness. In agreement with these findings we observed that the average strength of long-range connections in the α band gradually increases in the progression from deep sleep (N3) to ACTIVE W (rows 4 and 5 of figure 5.9). The difference between sleep and wakefulness is more marked if only a small percentage (10%) of strongest connections is included in the analysis (row 5 of figure 5.9). This index is able to significantly discriminate SLEEP N3 from SLEEP N1 and the two wakefulness stages ($p < 0.01$), and SLEEP N2 from active wakefulness ($p < 0.01$).

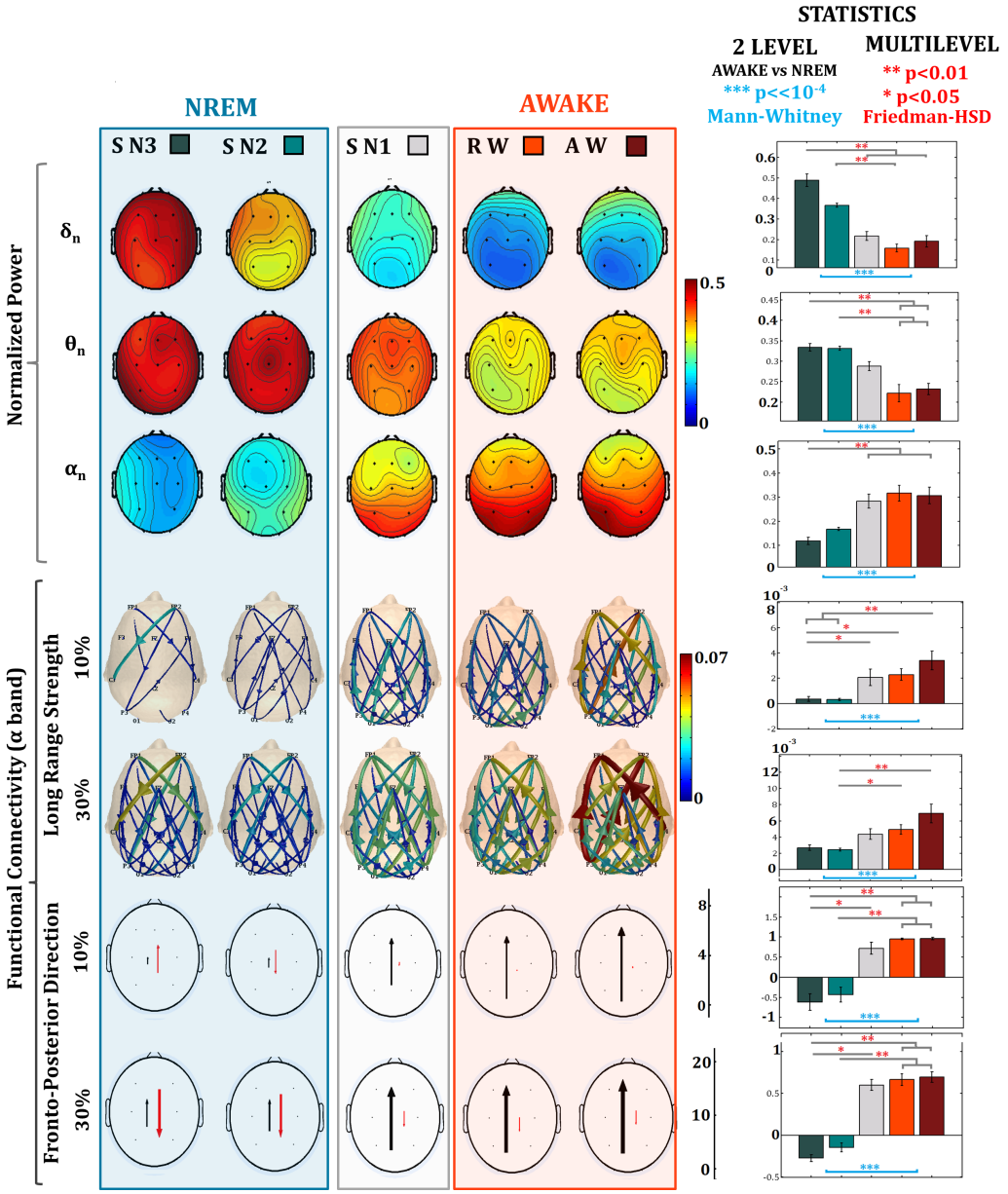

FIGURE 5.9: Scalp topography of the different EEG measures (rows), averaged across all 10 subjects, with associated statistics. Rows 1 to 3 show the normalized power distributions (δ_n, θ_n and α_n) across sleep stages. Rows 4 and 5 show the Grand Average of the strength of long-range connections in the α band plotted for the 10% and 30% strongest connections respectively. Rows 6 and 7 indicate the average number of postero-anterior (black) and antero-posterior (red) connections in the α band coded by the length and thickness of the arrows for 10% and 30% strongest connections respectively. Columns 1 to 5 indicate the experimental stage. The last column on the right indicates whether the indexes averaged across electrodes significantly discriminate wakefulness (REST W and ACTIVE W) from NREM sleep (N2 and N3) as assessed with a two level Mann-Whitney test (p value indicated in blue) and shows results of a Tukey's HSD test on indexes averages across electrodes to assess significant differences across all the stages. The asterisks specify that the two means designated by the brackets significantly differ ($*p < 0.05$, $**p < 0.01$, $***p < 10^{-4}$) (results showed for 10% strongest connections fourth row, last column- are repeated from figure 1, to aid comparison). Power distributions are consistent with results reported in literature. The most significant findings regarding the reorganization of connectivity patters are a increase in strength and number of connections in wakefulness (compared to sleep) and an inversion of the main direction of links on the fronto-posterior axis. Abbreviations: AW-ACTIVE W, RW- REST W, SN1, SN2 and SN3 NREM sleep stages.

We also assessed changes in the direction of the information flow over the rostro-caudal axis. The most notable trends were observed in the α band (which also showed that greatest changes in connectivity strength as seen in figure 5.8) and plotted in the two bottom rows (6 and 7) of figure 5.9. The number of posterior to anterior links dramatically increases from NREM sleep to wakefulness and reaches its maximum in ACTIVE W. In contrast, the number of connection in the antero-posterior direction gradually shrinks in the progression from NREM to wakefulness. Thus the $Dir_{P\to A}$ index, that summarizes the dominant direction of information flow, is particularly efficient in discriminating between stages, as revealed by the multilevel analysis. This provides evidence of a significant inversion of information flow in the α band from frontal to posterior vs posterior to frontal in the progression from sleep to wakefulness.

5.3.1 Individual analysis

A clinically useful index of consciousness needs to distinguish between sleep stages at an individual and not only at the group level. In order to investigate whether the changes observed were both consistent at the individual level and able to correlate with the experimental stage across the individual sleep-wake cycle, we show in figure 5.10 the different indexes for each subject and epoch of the experimental time-line. This epoch by epoch analysis allowed comparison with the individual hypnograms.

Among the connectivity measures, we have plotted the indexes that in the group analysis showed best discriminatory performances: The average strength of the 10% strongest long-range links and the $Dir_{P\to A}$ computed from the 30% strongest connectivity links. Given that the shift in EEG power toward lower frequencies is a well-known and prominent feature of NREM sleep (sleep δ waves) and that the δ and θ power showed similar discriminative properties at group level, we have only plotted the normalized δ power from the spectral measures (the plots were inverted to facilitate the comparison with the hypnogram and the connectivity derived indexes). The experimental stages were assigned a value as a measure of the level of consciousness, ranging from 0 (SLEEP N3) to 4 (ACTIVE W), and Spearman correlation was computed between each parameter and the individual hypnogram. Table 5.2 shows the resulting correlation values for individual subjects. All the indexes considered exhibit dramatic changes as a function of the experimental stages at the individual level. As shown in figure 5.10, the normalized δ power follows the experimental time-line in the majority of subjects. However in two subjects (2 and 4) the changes in δ power do not track the hypnogram and do not significantly correlate with the level of consciousness (table 5.2). Similar results are obtained for the θ power, with a significant negative correlation with the experimental time-line only in 7 of the 10 subjects. The average strength of long-range connections is severely

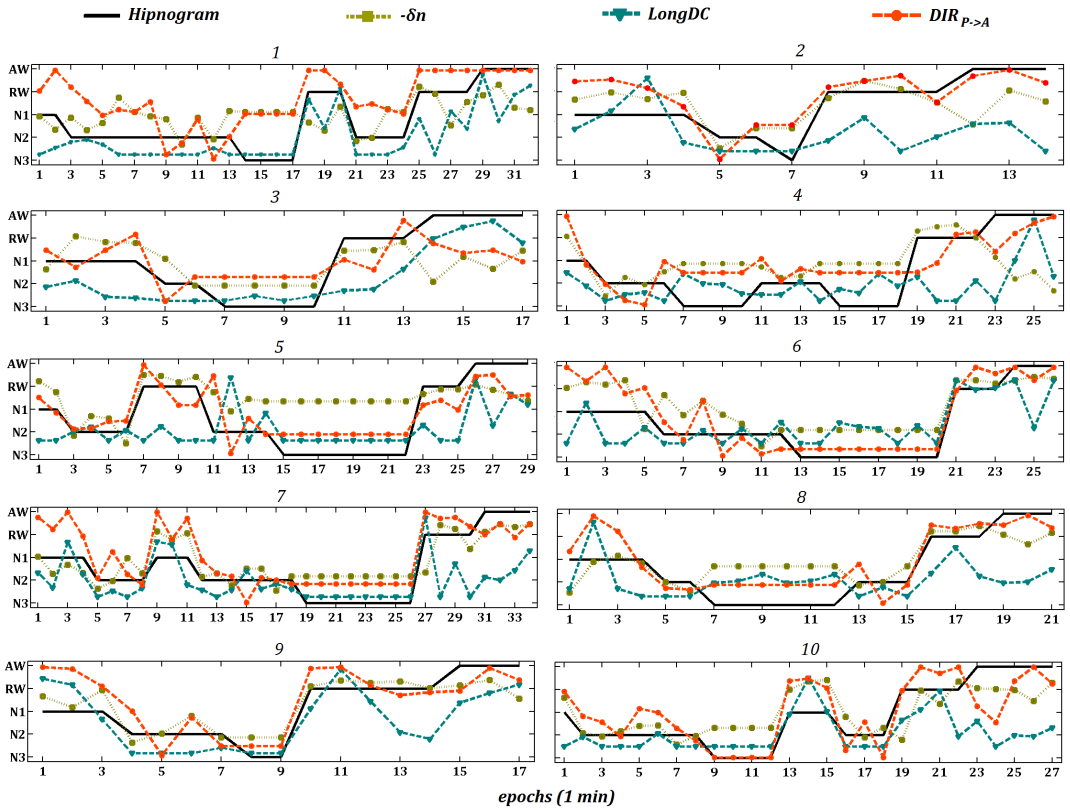

FIGURE 5.10: Individual trends over the experimental timeline. Each epoch is 60 seconds in duration. For each of the 10 subjects the amplitude (magnitude squared) of significant long-range DC links (blue dashed line), the $Dir_{P\rightarrow A}$ index (orange dashed line) in the α band and the normalized power in the δ band (green dotted line) are plotted and can be compared to the manually scored hypnogram (solid black line). The individual plots suggest that the changes observed at group level are broadly consistent at single subjects' level and that the connectivity features have a good degree of correlation with the hypnogram (see also table 5.2). Abbreviations: AW-ACTIVE W, RW- REST W, N1, N2 and N3 NREM sleep stages. For ease of visualization, all plots were rescaled, and δ power was inverted.

reduced in NREM sleep at an individual level. It shows performance similar to the normalized power indexes, highly correlating with the level of consciousness in the majority of subjects, but failing to do so in three of them. Of all the parameters considered, the $Dir_{P\rightarrow A}$ showed the best performances in 'tracking' the individual hypnogram, with a high and significant correlation in each of the subjects and the highest mean correlation value. It also is able to significantly discriminate between NREM sleep and wakefulness at an individual level in all the subjects (table 5.2, last column), as assessed by a Wilcoxon test across the epochs of each stage.

Index	**Subject**										**Mean**	**% Subjs**
	1	2	3	4	5	6	7	8	9	10		2 level
Power δ_n	*-0.36*	-0.44	**-0.52**	0	**-0.67**	**-0.89**	**-0.79**	**-0.64**	**-0.83**	**-0.63**	-0.58±0.26	100%
Power θ_n	**-0.54**	-0.15	**-0.51**	-0.38	**-0.50**	**-0.72**	**-0.76**	**-0.87**	*-0.55*	**-0.80**	-0.58±0.2	90%
Power α_n	0.34	0.2	0.13	0.02	0.1	*0.41*	**0.45**	0.22	*0.58*	**0.51**	0.29±0.18	80%
DC_{long} 10%	**0.77**	0.27	**0.84**	0.08	**0.49**	*0.48*	**0.69**	0.16	**0.72**	**0.63**	0.51±0.25	70%
DC_{long} 30%	**0.83**	0.21	**0.75**	**0.51**	-0.01	**0.63**	*0.38*	-0.23	**0.75**	*0.46*	0.43±0.34	70%
$Dir_{P\to A}$ 10%	**0.56**	**0.73**	-0.34	**0.87**	**0.86**	**0.67**	**0.78**	**0.77**	**0.85**	**0.89**	0.66±0.37	90%
$Dir_{P\to A}$ 30%	**0.82**	**0.74**	**0.68**	**0.70**	**0.85**	**0.84**	**0.82**	**0.79**	**0.68**	**0.80**	0.77±0.06	100%

TABLE 5.2: Spearman correlation coefficient computed between the hypnogram and each EEG index for individual subjects. Values in bold correspond to a significance level of $p < 0.01$, values in italic to a significance level $p < 0.05$; other values are not significant. Results relative to the $Dir_{P\to A}$, long-range DC links and δ power were highlighted in bold, to facilitate comparison with figure 5.10. The last column shows the percentage of subjects where the specific EEG index was able to significantly discriminate wakefulness (rest and active) from NREM sleep (N2 and N3) at individual level, as assessed by a Wilcoxon signed rank test ($p < 0.05$). Among the different features considered, the $Dir_{P\to A}$ showed the best performance, significantly correlating with the individual hypnogram for all the subjects.

5.4 Discussion

The objective of the current work was to propose and test indexes of brain connectivity that could distinguish between states of consciousness. These indexes should also be computationally relatively simple for clinical application such as in sleep studies or depth of anaesthesia monitoring. As discussed in Chapter 2, theoretical and experimental findings suggest that the ability of the brain to integrate information is critical to maintain consciousness (Tononi & Massimini 2008) and that unconsciousness may be associated to a disruption of large scale connectivity. Given this conceptual model, we focussed on connectivity measures, that include directional information. In assessing the performance of these measures, we compared them to the more established power spectral indexes taken from the EEG. To this end we collected polysomnographic recordings from a sample of 10 healthy subjects undergoing post-prandial sleep and extracted the EEG indexes across the sleep-wake cycle at individual and group level. Those indexes were chosen exploiting previous theoretical and experimental findings on neural correlates of consciousness and our preliminary results on a subgroup of 5 subjects. These included normalized power and DC based indexes. We found that the proposed amplitude of long-range connections across the scalp and especially the $Dir_{P\to A}$ index (that quantifies the dominant direction of information flow in the rostro-caudal axis) showed a monotonic change with level of consciousness. In the current sample, the $Dir_{P\to A}$

index showed the best performance in tracking the individual experimental time line: It consistently correlated well with the hypnogram, and significantly discriminated NREM sleep from wakefulness in each of the subjects in the sample. Its performance was found to be superior (Mann-Whitney test, p=0.041) to that of the power in the δ frequency band, which has been widely used in the past, and still represents the gold standard to stage deep sleep (N3).

5.4.1 Methodological considerations

In this study an advanced method (DC) for the estimation of functional connectivity that is able to infer directed causal information was chosen. The direction of connectivity networks has not previously been well explored in finding indexes of level of consciousness, even though it has been suggested that it could provide important insights into neural correlates of consciousness (Sitt et al. 2014). As mentioned in chapter 3, functional connectivity estimators based on an MVAR model have been shown to be robust to noise (Blinowska 2011) and to perform well even in the case of some non-linear interactions (Winterhalder et al. 2005), and have also been widely applied for estimating functional connectivity from multichannel EEG in different experimental conditions other than sleep (Astolfi et al. 2008, Blinowska 2011).

Despite DC having become well established, caution is required in the interpretation of its results as necessarily indicating causal links connecting underlying cortical sources. We have seen in chapter 4 how volume conduction effects lead to spreading of electrical activity to a number of electrodes, which could be confused with functional (neurological) connectivity between these brain regions. Our results, as well as those of Faes and colleagues (Faes & Nollo 2010, 2011, Faes et al. 2013), have suggested that while DC and DTF do not eliminate volume conduction effects, they do reduce them, when compared to conventional coherence. In particular our results (chapter 4) showed that volume conduction induced DC connectivity is close to zero for interelectrode distances larger than 10 cm. In this study, volume conduction effects are less of an issue since the analysis focused on long-range connectivity (>14 cm), whereas short-range connections (which are likely to be dominated by the spread due to electrical conduction in tissue and bone) were disregarded. Non-significant connectivity was also removed from the study using surrogate data analysis.

In the analysis of brain connectivity, the many relatively weak links can obscure important connections. Thresholds are thus usually recommended and applied (Rubinov & Sporns 2010) to only select connections deemed to be important. The choice of the threshold is somewhat arbitrary (Sporns 2013), but statistical significance (i.e. a threshold set at the critical value) should always be satisfied. In the current work either the

10% or the 30% strongest connections were analysed, following some preliminary investigations. The choice of threshold reflects a compromise between including too many connections that may only be weakly related to LOC and discarding connections that might hold useful information. Further work in optimizing the threshold for specific tasks such as assessing depth of anaesthesia or sleep stage should be carried out on a larger sample of recordings.

5.4.2 The relationship of our findings to previous studies

To the best of our knowledge there is only one study investigating EEG networks using directional measures of connectivity, but this used the closely related approach of DTF in sleep (Kamiński et al. 1997). DC, which includes information on signal power flowing between different regions of the brain, as well as the transfer function of the linear 'filters' linking these regions (as used in DTF), would seem to be more appropriate than DTF for quantifying the functional connections between brain regions: DC performs better than DTF in the case of time-series with different variances (Baccalà & Sameshima 2007) and has a straightforward interpretation in terms of power transfer (Faes & Nollo 2011). The current study thus goes beyond previous work in describing DC patterns during the change from wakefulness to NREM sleep, extending the analysis to different frequency bands and refining the methods with rigorous statistical significance assessment of the estimated links. Furthermore, we differentiate connectivity links with respect to their interelectrode distance and we provide an assessment of performance in individual subjects, as well as that at the cohort-level. Our findings are in line with results from the study of Kamiński et al. (1997) that showed more complex and denser connections in wakefulness than in NREM sleep and a prevalence of posterior sources during wakefulness (figure 5.11). Despite the strong topological similarities, in the study of Kaminski, connectivity networks exhibited a notably larger number of connections, possibly as a result of not performing a significance test (or using a different choice of threshold criterion) for including the estimated DTF in the final analysis, and also because connectivity was integrated over a larger frequency range (0-30 Hz) than in the current study.

The underlying conjecture of our as well as a number of other studies (Kamiński et al. 1997, Tononi 2008) etc. is that changes in level of consciousness are critically associated with a dynamic reorganisation of large-scale connectivity patterns. In line with previous results in disorders of consciousness (Chennu et al. 2014) the connectivity networks in the α band showed the best discriminative performance between sleep stages. Long-range connectivity in the α band was shown to be impaired in NREM sleep, when networks are active but characterized mainly by short-range links. Our results thus support previous

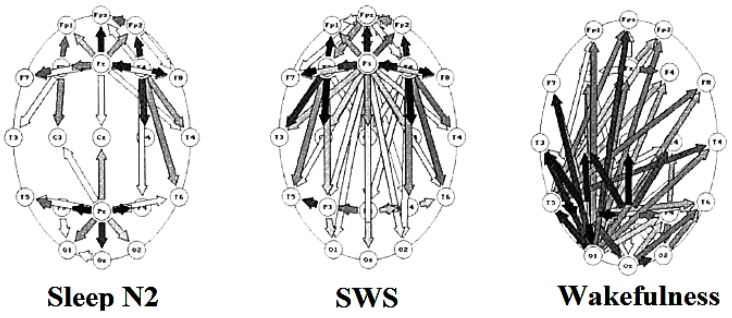

FIGURE 5.11: Grand Average (N=8 subjects) of EEG connectivity estimated by means of DTF integrated in the 0-30 Hz band. Only DTF links greater than 0.06 are shown. The connectivity patters estimated with DTF from Kaminski and colleagues show strong topological similarities with our results. Adapted from Kamiński et al. (1997).

studies using different methods and protocols suggesting that the disruption of wakefulness in NREM sleep may be a result of impaired information sharing among cortical areas. Massimini and colleagues (Massimini et al. 2005*a*, 2010) used a pertubational approach involving Transcranical Magnetic Stimulation (TMS) and EEG recording to investigate how TMS triggered neural activity spread from the stimulation site. During wakefulness the TMS elicited complex patterns of scalp waves spreading to distant cortical areas. During NREM sleep TMS evoked a stereotypical and local response, thus indicating a *breakdown of long-range effective connectivity*. The findings of Massimini et al. (2010) and our data suggesting that long range connections reduce during sleep also fit well into a wide range of evidence from a growing literature investigating fMRI brain connectivity in altered states of consciousness such as NREM sleep (Spoormaker et al. 2010, 2012) , general anaesthesia (Boly et al. 2011, Schrouff et al. 2011) and vegetative states (King et al. 2013, Boly & Seth 2012). It appears that those states share, among other major features, a suppression of functional connectivity.

In SLEEP N1 all the EEG indexes showed values intermediate between deeper sleep (N2 and N3) and wakefulness. Often spectral and networks features elicited in SLEEP N1 were more similar to wakefulness than to NREM sleep. SLEEP N1 represents the transition between wakefulness and sleep and it is considered unstable sleep (Klimova 2014). When awakened from SLEEP N1 subjects often report dream-like experiences or claim they were awake (Nir et al. 2013). Experimental results suggest preserved long-range connectivity in this stage (Massimini et al. 2005*a*). For this reason we have considered only sleep N2 and N3 trials to characterize stable NREM sleep in the two level analysis, as used by Massimini et al. (2010).

The findings of recent seminal works in Network Physiology (Bartsch et al. 2015, Liu

et al. 2015, Bashan et al. 2012) broadly align with our study. Network physiology is a new research field that aims to characterize how the various physiological systems dynamically integrate their functions in different physiological (and pathological) states (Ivanov et al. 2016). A series of studies investigating changes in networks of interactions between (Bartsch et al. 2015) and within (Liu et al. 2015) different physiological systems (muscular, cardiac, respiratory and central nervous systems) during the transition from wakefulness to sleep and across sleep stages have provided important progress in this emergent field, contributing to the realization of an atlas of global network physiology in sleep and wakefulness. In these studies network connectivity was assessed using an estimator based on the stability of the delay between signals (time delay stability -TDS), which quantifies in fine temporal detail the undirected strength of coupling

Despite the differences in approach, EEG networks estimated from six channels and in different frequency bands in Liu et al. (2015) and Bartsch et al. (2015), showed important similarities with those shown in our study for the α band: they observed a significant decrease in the strength and number of links (in particular of long-range fronto-occipital connections) in deep sleep (as compared to wakefulness and light sleep) and a remarkable symmetry between the two hemispheres, characterizing all physiological states. Another important analogy with our results regards EEG networks elicited in sleep N1, whose features are more similar to wakefulness than to deeper sleep. As Liu and colleagues have observed this is an interesting result, given that sleep N1 is commonly classified as belonging to the same macro state (NREM sleep) of sleep N2 and N3. In their analysis of brain networks Liu and colleagues have also reported that while local connections (frontal/frontal, central/central and occipital/occipital) are reduced but preserved in deep sleep, fronto-occipital and occipital-frontal networks show practically no connections in deep sleep: these findings agree with the significant impairment of long-range connectivity we observed in sleep N2 and N3.

Our study provides important new contributions beyond this work (and previous studies) in the characterization of brain networks by assessing the direction of links and the consequent ability to identify a switch in the direction of information flow with sleep onset, that constitutes the most characteristic change observed in DC patterns. Group analysis reveals a significant inversion of the direction of posterio-frontal networks with state. The marked posterior to anterior spread of rhythm in wakefulness is reversed in NREM sleep (N2 and N3) that is characterized by a dominance of frontal sources of activity. An inversion of information flow from frontal-posterior in sleep to posterior-frontal in wakefulness has also been found in a previous analysis of sleep onset (De Gennaro et al. 2004).

Another important original contribution of this work is the investigation, and presentation, of patterns for individual subjects, in addition to aggregated cohort results. The

majority of studies investigating EEG markers of LOC have only performed group analysis or found the disagreement between behavioural and EEG-based measures too high for a reliable individual assessment (Sitt et al. 2014). The aim of this study is to identify an EEG index able to 'track' the consciousness level of the subjects, with a view to assess depth of sleep or anaesthesia. The performance of the proposed measures in individuals provides an indication of the potential of the proposed approach. Among all the markers considered, the $Dir_{P\to A}$ index was found to be the most reliable in tracking the level of consciousness during individual sleep experiments, strongly correlating with the hypnogram in all the subjects. It should be noted that the staging of the EEG time series was performed taking in account the proportion of δ waves, as recommended in standard sleep staging criteria, therefore high correlation between measured δ power and sleep stages is only to be expected. Given that EEG spectral features are used to characterize sleep (Corsi-Cabrera et al. 2003) and to monitor hypnotic level in anaesthesia (Myles et al. 2004), it is interesting that the $Dir_{P\to A}$ index, which reflects very different features of the EEG signals, showed better performance than the δ power in correlating with the sleep stage at an individual level. These results suggest that the inversion of information flow represents a promising indicator of the descent into deep sleep. We speculate that this may also be seen in other states of altered consciousness, such as anaesthesia. In order to arrive at an even more powerful index of LOC, it may be beneficial to combine power-spectral measures with those obtained from DC and this approach is explored in the next chapter.

5.4.3 Limitations

The current work was carried out on a relative small sample (10 subjects) recorded during a nap, not all-night sleep. Thus only between 13 and 31 one-minute segments were available from each subject in this cohort (see figure 5.10). The rationale for this sample size and for the choice of the nap recordings is to be found in the exploratory character of this study. While the results showed the power of the $Dir_{P\to A}$ index, this could be specific to this small sample. The current work should thus be considered as an investigation in which the hypothesis of superior performance of $Dir_{P\to A}$ was generated, but cannot be robustly tested, on the same small sample. Further tests on an independent sample and whole night sleep recordings collected in a sleep lab should be carried out. Such a study might also include a wider range of indexes as well as their combination in order to independently statistically test relative performance measures.

Chapter 6

EEG directed connectivity as a monitor of anaesthetic depth

6.1 Introduction

6.1.1 Background and challenges of the anaesthetic study

As a last stage of the project, EEG connectivity measures were tested in an anaesthetic study involving ten patients scheduled for surgery. Before that, considerable effort was spent in the design and ethics submission for a study involving healthy volunteers. The rationale for the volunteer study was to obtain high quality EEG recording to test connectivity measures. We thought that collecting EEG in a controlled environment from volunteers would have provided better quality data if compared with experiments performed in a clinical operating theatre, contaminated by electrical environmental 'noise'. Moreover our previous experience suggested that carrying out the study on patients would present a series of difficulties and constraints: One of them is that we needed a period of time in the anaesthetic room, before surgery, to collect the EEG data during a slow induction of anaesthesia. This was very difficult to organise in an operating theatre schedule. We needed to know, before we consented a patient for the study, that the previous patient on the list was going to have surgery of long enough duration to allow us to have the necessary time to collect data. Otherwise the risk was to cut the data collection before the end of the designed protocol. These motivations were not judged by the local ethics board (Research Governance Office, University of Southampton) to be strong enough to justify the risk of exposing healthy volunteers to the small (but serious) risk of anaesthesia. As a result our application for a volunteer study was rejected by the ethics committee.

We proceeded therefore to a patient study for which an ethics approval for electrophysiological recording from patient (with and without auditory stimulation) was already in place. Patients scheduled for elective surgery were recruited by one of the research team, Dr. David Smith. We found recruitment of patients for the clinical anaesthetic study quite challenging. The majority of patients suitable for our study were admitted on the day of surgery, having been seen previously in a pre-assessment clinic. The logistics of the recruitment and consent process in this situation have been quite frustrating. We have therefore extended the duration of the study to try to ameliorate the recruitment difficulties.

6.1.2 The Anaesthetic Study on Patients

In this work we assessed EEG directional connectivity between brain regions in a sample of 10 patients using DC during a slow anaesthetic induction in which the concentration of propofol in the brain (Effect-site concentration, ESC) was increased gradually to a peak level of 4 $\mu g \cdot ml^{-1}$.
Given that theoretical considerations and experimental findings suggest that brain connectivity is critically related to LOC and considered our results in sleep (chapter 5) our hypothesis is that measures of LOC based on connectivity can be expected to be more effective than measures based on local neuronal activity. A few recent studies have investigated EEG directed connectivity during anaesthesia using phase synchronization (Lee et al. 2009) , Information theory entropy (Ku et al. 2011) and Granger Causality (time domain) (Nicolaou & Georgiou 2014) approaches. However it is not clear how directed connectivity measures correlate with other DoA indexes. The objectives of this work are therefore:

1. to identify EEG directed connectivity features that robustly reflect anaesthetic-induced LOC at group as well as at an individual level
2. to test the performances of DC features (in comparison with other DoA indexes) in discriminating wakefulness from anaesthesia at different depths
3. to propose an index based on EEG connectivity that could be incorporated into future DoA monitors. The latter would add connectivity to current measures based on activity in individual EEG channels.

We compared the performances of the DC with more established DoA indicators (BIS and MLR) in discriminating 'awake' from 'anaesthetized' states. We investigated the group and individual trends of the different indexes and assessed their discriminative

performance using a linear and non-linear classifier to test the use of directed connectivity as an indicator of adequacy of anaesthesia.
This study goes beyond previous work in three important points:

1. We used a slow target-controlled induction of propofol anaesthesia that allowed investigation of graded changes in directed connectivity for increasing, stable anaesthetic depths ranging from light to deep (surgical) sedation. Because of its rapid action, propofol is ideal for ultraslow step-wise induction of general anaesthesia (Mhuircheartaigh et al. 2013*b*). This is in contrast with previous EEG connectivity studies that are carried out in settings where general anaesthesia induction is performed at once (Lee et al. 2009, Barrett et al. 2012, Untergehrer et al. 2014, Nicolaou et al. 2012) (typically with a bolus of intravenous drugs), causing the crucial transition to LOC to occur within 30-60 s. Moreover, we have focused only on the hypnotic effect of one agent (propofol), avoiding confounding factors (i.e. the administration of muscle relaxants).

2. We described, together with group results, single subject trends and individual variability in connectivity features. On the other hand the majority of studies investigating neural correlate of consciousness have focussed on average results across the cohort. The rationale for the individual analysis is to be found in the potential practical applications: we aim to assess adequacy of anaesthesia using DC features and in this sense the measure should be able to assess anaesthetic effect in individual subjects, rather than on 'average'.

3. To the best of our knowledge, a comparison between directed connectivity and other DoA indexes has not been performed. We assessed DC performances in discriminating 'wakefulness' from 'anaesthesia' in comparison with BIS and MLR jointly extracted from 60 s EEG epochs in stable propofol ESC.

6.2 Methods

6.2.1 Subjects and Experimental Protocol

The study was approved by the Southampton and Southwest Hampshire Research Ethics Committee (ref 002/98) and included ten patients: Three females, seven males, aged between 44 and 79 years (mean 63.8 $\pm$ 8.1). The rationale for this sample size is based on effect size considerations of a similar study carried out by our group (Bell et al. 2006). They were selected from the cardiac surgical operating schedule the day before their planned operation, based on their ability to tolerate an additional 90 min

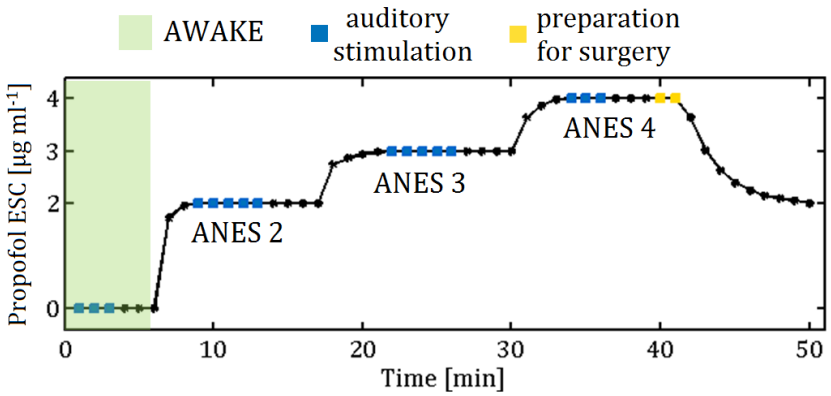

FIGURE 6.1: Experimental protocol schematic: Slow induction of general anaesthesia with computer-controlled infusion of propofol to achieve brain ESC of 2, 3, 4 $\mu g \cdot ml^{-1}$. For each stable anaesthetic level EEG was collected with and without auditory stimulation (blue squares). At a peak level of 4 $\mu g \cdot ml^{-1}$ the target ESC concentration was set to 2 $\mu g \cdot ml^{-1}$ and the patient prepared for surgery (yellow squares).

period of anaesthesia before their surgery commenced and they all provided written informed consent. A standard pure tone audiometry test was then performed using an AS608 audiometer (Interacoustics, Assens, Denmark) before the anaesthetic induction. We used a computer-controlled infusion to achieve central nervous system effect-site propofol concentrations (ESC) of 2, 3, 4 $\mu g \cdot ml^{-1}$ as in Mhuircheartaigh et al. (2013*b*) and Purdon et al. (2013) in order to cover a range of anaesthetic depths from light to deep anaesthesia (figure 6.1). During the experiment EEG was collected at different stable propofol ESC: the recording protocol included an equilibration period and then the maintenance of stable propofol ESC for 10 minutes: in the first 5 min subjects were presented with auditory stimulation and in the last 5 min stimulation-free EEG was recorded. The target propofol effect-site concentration was increased to 3 $\mu g \cdot ml^{-1}$ and then to 4 $\mu g \cdot ml^{-1}$, with the EEG recording protocol repeated at each target level, again allowing 5-10 minutes for stability to be achieved (stability was considered 'achieved' once the patient ESC reached the target level, as verified by the experimenters). EEG was collected using a 32 channel system with active electrodes (Biosemi BV, Amsterdam) placed according to the international 10-20 system. Auditory stimuli were delivered binaurally by means of a computer controlled interface (CEDmicro1401), a headphone amplifier (Creek Audio Ltd, OBH 21) and through ER-2 insert headphones (Etymotic Research). Chirps sweeping from 0.1 Hz to 10 kHz in 10.4 ms were delivered at 60 dB HL with a stimulation rate of 143 Hz using maximum length sequences (MLS). MLS are pseudo-random binary sequences that enable evoked potentials to be acquired at stimulation rates beyond conventional values (Bell et al. 2001). Results in literature have shown that the combination of chirp stimuli at high MLS rates produces the highest improvement of SNR, corresponding to a tenfold reduction in test time if compared with conventional recording (Bell et al. 2002).

The EEG recording protocol was repeated at each ESC target. At the end of the induction, at a propofol ESC of 4 $\mu g \cdot ml^{-1}$, the target propofol ESC was reduced to $2\mu g \cdot ml^{-1}$ and the patients were prepared for surgery. EEG was continuously monitored until the ESC reached $2\mu g \cdot ml^{-1}$.

6.2.2 MLR estimation and analysis

The MLR is usually recorded placing the active electrode either on the midline forehead (Fz) or on the vertex (Cz), with the ground electrode on the low forehead. The position of the reference electrode often represents an issue because of the myogenic interference of the postauricular muscle (PAM). Tooley and colleagues (Tooley et al. 2004) showed that the vertex-inion electrode site gave the best results against the PAM artifact both in awake and anaesthetized subjects and reference electrode close to the nape of the neck are also recommended in Bell et al. (2004). In this study data were referenced with respect to the occipital electrode Oz and the channels Fz and Cz were selected for MLR estimation. Data were then down-sampled to 1 kHz and band pass filtered (15-250 Hz), as recommended by Bell and colleagues (Bell et al. 2004). In order to remove the 50 Hz mains interference and its harmonics a bank of notch filters at 50, 150 and 250 Hz was also applied. All the filters were applied forwards and backwards in order not to introduce phase distortion. Data were then segmented in 105 ms epochs aligned with the stimulus onset and an amplitude threshold criterion (20 μV) was applied in order to remove artifactual trials. The number of epochs included in the estimation of MLR was then set to 2500 for all the experimental stages and subjects. In order to estimate the MLR evoked by a MLS sequence from the overlapped response a deconvolution algorithm (Thornton et al. 1998) was additionally applied to the matrix of trials and then the coherent average performed (see Appendix C for a description of the MLS properties and deconvolution algorithm).

The SNR of the MLR was assessed using the F-value at a single point (F_{sp}) (Erberling & Don 1984). The F_{sp} statistic is the ratio of the variance of the reconstructed MLR and the variance of the background noise (estimated as the variance of a single point at a fixed latency across trials). In this study we selected a latency of 30 ms from the stimulus onset for the single point and a time window of [30-80] ms to compute the variance of the signal, according to the assumption that the brainstem response (within the 20 ms from the end of the 10 ms chirp stimuli) is not affected by the level of arousal of the subject.

To determine if real MLR responses were present in recordings that were significantly different from background noise, a bootstrap procedure was carried out for all the experimental stages and for each of the subjects. Each epoch of the original set of EEG

data was de-trended and rotated starting for a random sample point (different for each epoch). 1000 epochs for each propofol ESC were considered for the bootstrap procedure. A 'null MLR' estimate was obtained first applying the deconvolution algorithm and then the coherent average of the randomly rotated epochs. A F_{sp} value for each of the bootstrap iterations was estimated as well. The same procedure was performed for 500 times to obtain a null distribution for the MLR and the F_{sp} value. The significance of the estimated MLR was assessed comparing the F_{sp} value with the null F_{sp} distribution obtaining, using the Davison and Hinkley (1997) formula, an empirical p-value. Before applying it to the study data, we have validated the bootstrap procedure assessing its false positive rate on Gaussian noise.
Literature results report a significant reduction of MLR amplitude and an increase of the N_b latency during anaesthesia as compared to responsiveness (Tooley et al. 1996, Thornton & Sharpe 1998, Loveman et al. 2001, Tooley et al. 2004, Bell et al. 2006). We therefore assessed changes in MLR amplitude (power) and N_b latency. As a first step the MLR was computed using all the available non artifactual recordings for each experimental stage, i.e using 2500 sweeps, corresponding to 4.3 min of recordings for each propofol ESC. This improves the quality of MLR estimation but on the other hand, due to the relatively long averaging period, may smear MLR changes in time and reduces the time resolution. In order to overcome these limitations and to allow comparison with other EEG indexes, a second analysis was performed to obtain a time resolution of 1 min, therefore considering 600 epochs for the computation of the MLR.

6.2.3 Estimation of BIS index

All the methodologies for BIS estimation were designed to reproduce as far as possible the proprietary algorithm for the computation of BIS and the published computation of BIS sub-parameters (Rampil 1998, Miller et al. 2004). The latest version of the BIS monitor has four electrodes placed on the forehead corresponding to the bipolar montage F3-F7 and Fz-F7 in the 10-20 international system (Johansen 2006), we therefore used electrodes F3 and Fz (with F7 as reference) to compute BIS. The signal was downsampled at 256 Hz, high pass filtered (0.5 Hz) and notch filtered with zero phase filters. The raw signals were divided in 2 s long segments and epochs with signals whose amplitude exceeded 100 μV were rejected. The BIS is a combination EEG parameters in time (burst suppression ratio, BSR) and frequency domain (beta ratio β_R and SynchFastSlow $SynFS$). Due to the random nature of the EEG signals all the sub-parameters are computed for the 2 s epochs and then smoothed by averaging in 60 s segments (the BIS index value is therefore updated each minute).

The BSR was computed as the fraction of epoch length in which the EEG was suppressed ($\|EEG\| < 5\mu V$) for more than 0.5 s (Rampil 1998). The epochs were then preprocessed for frequency domain transformation by multiplication by a Blackman window. The spectral power was computed using direct FFT method and then averaging the complex product of FFT across 30 2 s long epochs; it was then integrated in empirical frequency bands to estimate the β_R , as indicated in the equation 6.1

$$\beta_R = log_{10}\frac{(Power)_{30-47Hz}}{(Power)_{11-22Hz}} \tag{6.1}$$

The bispectrum was computed using direct FFT method using the Matlab function bispecd.m (1 Hz resolution, 75% of overlap) as in Miller et al. (2004). The $SynFS$ parameter was then computed according to the equation:

$$SynFS = log_{10}\frac{(BispectralPower)_{0.5-47Hz}}{(Power)_{40-47Hz}} \tag{6.2}$$

Because of the symmetry properties of the bispectrum (Miller et al. 2004), the latter was computed only for the non-redundant subset of frequencies represented by the triangle with vertices $(0,0), (\frac{fs}{4}, \frac{fs}{4}), (\frac{fs}{2}, 0)$.
As for the BSR, the β_R and the $SynFS$ were computed for each 2 s epoch and then averaged in 1-min long segments. The BIS proprietary algorithm combines the subparameters with weights extracted from a multivariate model based on a database of EEG recordings matched to corresponding hypnotic drug levels. The weights are assigned using a non-linear function and they vary depending on the anaesthetic stage: the β_R is dominant in light sedation, the $SynFS$ during EEG activation and surgical level of hypnosis while the BSR detects deep anaesthesia (Rampil 1998). It has been shown that with increasing anaesthetic levels the BSR increases while the $SynFS$ and β_R decrease (Morimoto et al. 2004). In this analysis we assessed the changes of the single subparameters and also combined them to obtain a BIS equivalent index ($eqBIS$) using the simple formula:

$$eqBIS = -BSR + \beta_R + SynFS \tag{6.3}$$

Since the algorithm that combines the different sub-parameters to obtain the BIS value is proprietary, in this study the analysis was focused on the trends and variability of the different BIS sub-parameters and their sum (eqBIS).

6.2.4 Estimation of Connectivity

Directional connectivity was assessed estimating DC from the MVAR model of multichannel EEG time-series (60 s long), following the procedure described in the previous chapters (chapter 3). As for previous works (and for the sleep study), a subset of $M = 12$ electrodes that are fairly evenly distributed across the scalp (Fp1, Fp2, F3, Fz, F4, C3, Cz, C4, P3, P4, O1, O2) was selected and connectivity was estimated for epochs of 60 s, in order to reduce the computational cost typical of multivariate connectivity estimation and to ensure a number of samples sufficient to accurately fit the MVAR model. The statistical significance of each DC link was tested with surrogate analysis (phase shuffling procedure, $p < 0.01$) and corrected for multiple comparison using the false discovery rate (FDR) approximation for dependent measurements (Benjamini & Yekutieli 2001). As for the sleep study, in addition to the statistical threshold, we applied thresholds to retain the strongest connections. Since the choice of the threshold is somewhat arbitrary, it is useful to investigate results over a range of plausible thresholds (Bullmore & Sporns 2009). In this study we retained respectively the 10%, 30% and 50% of strongest connections. In order to specifically investigate changes in long-range connectivity, DC connections were also grouped on the basis of the 3D Euclidian interelectrode distances into 1) short-range links connecting adjacent electrodes (interchannel distances below 10 cm) and 2) long-range links.

6.2.4.1 Computation of EEG Connectivity Features

Building on our previous work on sleep (chapter 5) and previous literature results (Chennu et al. 2014), we focussed the analysis on the α band and long-range links only, as these parameters were shown to be most sensitive to changes in the level of consciousness. We therefore assessed connectivity using two measures: The dominant direction of information flow and the strength of long-range connectivity networks. The rationale for the former comes from the literature (Purdon et al. 2013, De Gennaro et al. 2004, Barrett et al. 2012, Lee et al. 2009) and our previous work on sleep, indicating that the direction of fronto-parietal connectivity represents a promising indicator of the level of consciousness. The rationale for the latter is to be found in the large number of studies showing that long-range connectivity is significantly modulated by the level of consciousness. To this end we assessed the number of significant connections from centro-parietal (P3, P4, C3, C4, Cz) to frontal (Fp1, Fp2, F3, F4, Fz) electrodes and vice versa. As for the sleep study, we defined an index that quantifies the dominant direction of information flow on the front-posterior axis ($Dir_{P \to A}$, see equation 5.2) as the normalized differences of the strength of links in the two opposite directions over the

rostro-caudal axis. In order to assess the statistical significance at group level, a Friedman test (followed by a Tukey's honestly significant difference HSD test) was performed across subjects considering the experimental stages (AWAKE , ANES $2\mu g \cdot ml^{-1}$, ANES $3\mu g \cdot ml^{-1}$, ANES $4\mu g \cdot ml^{-1}$) as independent variables.

Together with a cohort analysis we have compared network feature trends in individual subjects with their ESC time-line. In order to summarize the individual DC features in a unique parameter, for each subject and 60 s epoch e we summed the strength of long-range links and the $Dir_{P\to A}$ to obtain what we have called the DC_{index} (equation 6.4). While the $Dir_{P\to A}$ is a normalized index varying from -1 to 1 for all the subjects, the strength of long-range DC can vary from one subject to another as a result of individual networks differences. For this reason we normalized the strength of DC links in each epoch by the average strength of long-range connectivity across all epochs in each subject as follows

$$DC_{index}(e) = \frac{\overline{|DC(e)|^2}}{\frac{1}{L}\sum_{e=1}^{L}\overline{|DC(e)|^2}} + DIR_{P\to A}(e), e = 1, ..., L \tag{6.4}$$

where $\overline{|DC(e)|^2}$ is the average strength (across electrode pairs) of connectivity links for the epoch e and L is the number of the 60 s segments considered for each individual.

6.2.5 Wakefulness vs Anaesthesia Classification and Assessment of Performances

The aim of the study was to investigate the performances of the different EEG indexes in discriminating wakefulness from the different anaesthetic levels (ESC). To this end we used a binary classification procedure based on support vector machine (SVM) (Brereton & Lloyd 2010). A linear SVM classifier estimates the optimal linear combination of features (in this case MLR, BIS and DC sub-parameters) that separates the samples in distinct classes in the hyperspace of features (in other words, it finds the hyperplane that separates the classes and has the maximum distance from the features of the two groups). The optimal weights depend on the nature of the data and are usually found by training the SVM model on a number of datasets (also called 'folds') and then testing the model on the remaining data ('multifolded cross-validation'). In this study we used a *leave-one-out* approach to cross-validate the SVM models: the model was trained on the entire set of observations except those relative to the specific subject, and then it was tested on the selected subject. The data in input to the classifier for the cross-validation were the different sub-parameters (EEG features) arranged on a matrix X $[N \times f]$ (N, number of observations; f, number of features) and a vector Y of N elements corresponding to the classes. The SVM was estimated by using a Sequential Minimal Optimization (SMO) routine (Fan et al. 2005). The classifier's outputs are

the probabilities of each observation (series of features) of belonging to a specific class. The posterior probability is estimated from the output scores of the classifier using a transform function (Platt 1999).

We used a linear SVM binary classifier to distinguish wakefulness (AWAKE) from light anaesthesia (ANES $2\mu g \cdot ml^{-1}$) and deep anaesthesia (ANES $3\mu g \cdot ml^{-1}$, ANES $4\mu g \cdot ml^{-1}$), therefore using three separate binary classification procedures. Another aim of the analysis was to test if a combination of the various EEG features (MLS, BIS and DC sub-parameters) would improve the classification performances or if the performances of the different measures combined were comparable with those of one 'optimal' index. To this end, after having tested the single indexes, we have applied the paradigm to all the EEG indexes combined. The cross-validation/classification procedure was implemented using the matlab functions fitcsvm.m and predict.m.

We have also more generally assessed the performances of the different indexes in distinguish 'Wakefulness' vs 'Anaesthesia' (including in the binary classification all the anaesthetic stages), as it is a crucial feature for a monitor of the adequacy of anaesthesia. To this end we compared the performances of the EEG indexes using the SVM linear model with a non-linear classifier based on a multilayer neural networks (NN), to investigate if a more complex classifier would improve performances.

Artificial NN are mathematical models inspired to the stimuli processing in the brain (Krogh 2008). They are composed by simple units that are defined by connection weights w, a threshold t and an activation (or transfer) function g. The total input to the unit neuron is a weighted sum of all the inputs $x_i, i = 1, ...N$ it receives from other neurons or sources: the output can be expressed as

$$g(\sum_{i=1}^{N} w_i * x_i - t). \tag{6.5}$$

If the activation function is a step function then the output will be 1 if the weighed sum of inputs is above the threshold t and 0 otherwise. A neuron whose activation function is the step function is called a threshold unit, or perceptron. The points in the input space that satisfy the condition $\sum_{i=1}^{N} w_i * x_i > t$ define an hyperplane.

Classification problems can be solved by a threshold unit NN if the classes are separable by hyperplanes: i.e. if they are linearly separable. However many classifications problems are non-linear: to separate classes in these problems more hyperplanes are required therefore more neurons are introduced in the networks. This results in the basic multilayer NN structure that includes input, hidden and output layers of neurons and is able to solve more complex pattern recognition and decision making problems. The weights and thresholds of the neuron units are adjusted to the particular model by means of a learning process. Commonly NN are trained by supervised learning where a

series of training inputs together with the correct outputs are iteratively provided to the network: for each training set an error vector can therefore be defined and the weights (and thresholds) adjusted in such a way that the error is reduced. The learning rule that updates the weights in order to decrease the cost function (mean squared error) is called gradient descent algorithm (Krogh 2008) and it is usually applied to the error of each layer starting from the NN output and back-propagating the error to the input layer (back-propagation).

Once the network is trained it can be applied to unseen data to test for its classification and generalization performances. The generalization is the ability of the trained NN to extend the performances to a new set of data: if the model is overfitted the error for the training data may be very small but when the NN is applied to new data it may have scarce classification performances. For this reason a common practice during training is to test the NN on a new set (validation set): if the accuracy over the training data increases but the performances on the validation set worsen then the training is stopped. This procedure is called 'early stopping' and represents one way to minimize model overfitting. In this work the NN was implemented using the Matlab NN toolbox. We used the 85% of the available data for trainining and the 15% for validation and we tested the NN with a 'leave-one-out' procedure.

The design of a NN include the choice of the activation function of the unit neurons and of the particular learning algorithm and also the number of neurons in the hidden layer. We used the default Matlab standard functions for patterns recognition (sigmoid function for the hidden layer, softmax for the output, conjugate gradient descent algorithm) which provide optimal performances. The number of neurons in the hidden layer is normally an undefined parameter and has to be tested. We repeated the classification procedure for 2, 5, 10, 15, 20 and 25 neurons in the hidden layer, selecting the NN with best performances as in Nicolaou & Georgiou (2013).

To reliably and completely characterize the performances of a DoA monitor, they should be described in terms of Sensitivity, Specificity and global Accuracy (Pandit & Cook 2014). Specificity refers to number of true positives (in this case the number of 'wakefulness' observations correctly classified as AWAKE) while Sensitivity was assessed as the fraction of observations in 'anaesthesia' correctly classified. The Accuracy summarizes the two properties and represent the fraction of correctly classified examples.

6.3 Results

EEG signals at different anaesthetic levels presented the characteristic changes associated with increasing depth of anaesthesia (Murphy et al. 2011, Rampil 1998, Marchant et al. 2014). Typical example of EEG traces from one subject are shown in figure 6.2.

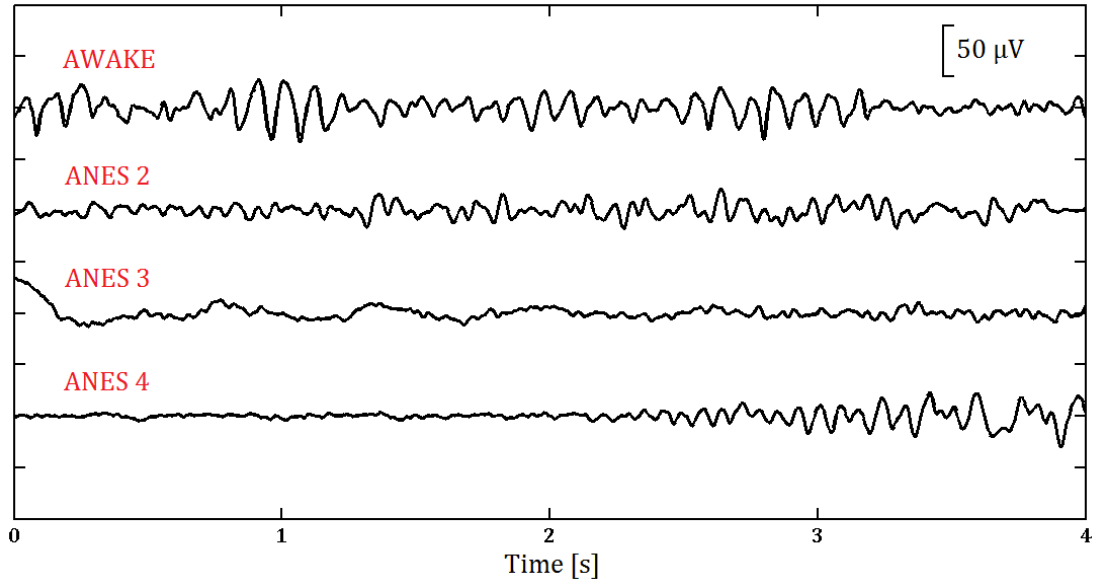

FIGURE 6.2: Representative examples of 4 s long EEG recordings in wakefulness (AWAKE) and for increasing ESC levels (ANES 2, ANES 3, ANES 4) in one subject (subject 1). Note the α (8-13 Hz) oscillation in wakefulness (first trace) and the increased β (13-30 Hz) activity in ANES 2 (second trace). In ANES 3 the high amplitude slow waves are dominant (third trace) while ANES 4 is characterized by an alternation of burst (last second of the recording) and suppression (first three seconds) patterns.

For ESC of 2 $\mu g \cdot ml^{-1}$ the EEG time series showed increased activation in frequencies between 13 and 30 Hz (β) typical of light anaesthesia. For ESC 3 $\mu g \cdot ml^{-1}$ or 4 $\mu g \cdot ml^{-1}$ the EEG was characterized by slow wave (δ) activity. In most of the subjects a burst suppression pattern, typical of deep anaesthesia, was observed only for ESC 4 $\mu g \cdot ml^{-1}$. We can therefore argue that the slow induction of anaesthesia included a series of levels ranging from light to deep anaesthesia.

In the following, we will first present findings relative to the single EEG features group and individual trends. Results comparing the different EEG indexes with associated statistical analysis will be then showed. The discrimination performances of the EEG indexes will be eventually compared in the last section of the results, with a particular focus on the global 'wakefulness vs anaesthesia' classification.

6.3.1 Changes in connectivity topography associated with increasing anaesthetic levels

Based on literature results (Chennu et al. 2014), previous work carried out on sleep by our group (Lioi et al. 2017) and for the sake of the clarity, we will only show results for the most discriminative connectivity features, focusing on long-range connectivity networks in the α band.

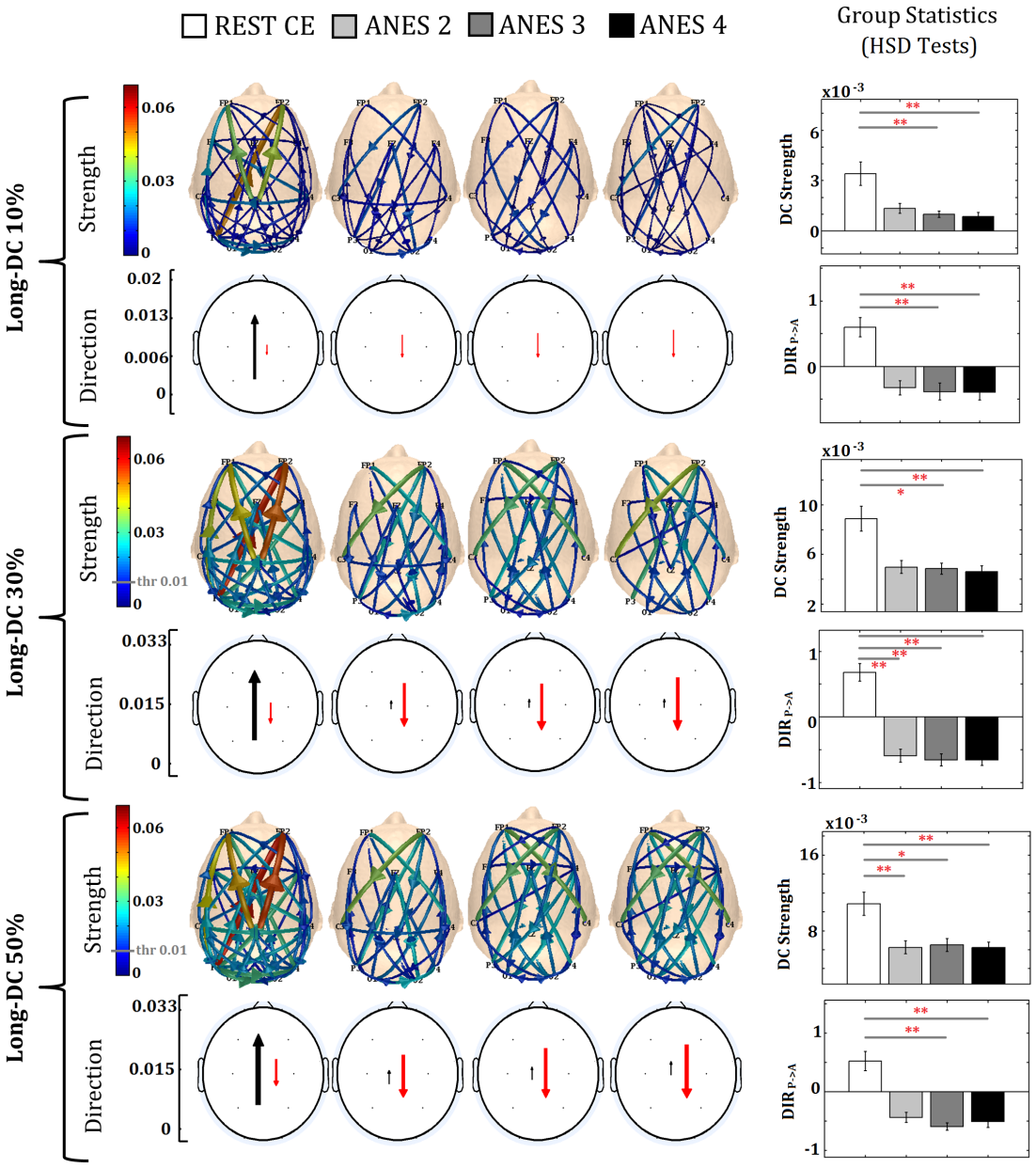

FIGURE 6.3: Scalp topography of connectivity networks, averaged across all 10 subjects, with associated statistics, plotted for the 10% (first two rows) 30% (rows 3 and 4) and 50% (last two rows) strongest connections. The first row of each subplot represents the Grand Average of long-range connections (average across subjects of long range $|DC|^2$ matrices) , with the color and size of arrows coding for the average strength of the specific link. The second row indicates the average strength of postero-anterior (black) and antero-posterior (red) connections in the α band coded by the length and thickness of the arrows. The bar plots on the right hand side show the mean and standard error (across subjects) of the respective features (strenght and $Dir_{P\rightarrow A}$). The asterisks specify that the two means designated by the brackets significantly differ (*, $p < 0.05$; **, $p < 0.01$), as revealed by Friedman and Tukey's HSD test. Connectivity scalp plots were obtained using eConnectome imaging software (Bin et al. 2011).This figure shows how, for all the percentages of strongest connections included in the analysis, two significant trends are observed. Firstly, an abrupt drop in the strength of long-range connectivity occurs at the onset of anaesthesia. Secondly the main direction of connectivity links switches from posterio-frontal in wakefulness to fronto-posterior in anaesthesia.

Figure 6.3 represents scalp plot of DC networks in wakefulness and the in three anaesthetic levels. Theoretical models of consciousness and experimental results obtained in sleep, anaesthesia and disorders of consciousness predict that large-scale information sharing is essential to maintain consciousness and that anaesthetic induced LOC is associated with a breakdown of long-range connectivity. In agreement with these findings we observed that the strength of long-range links in the α band is notably reduced at all the anaesthetic levels (figure 6.3, last three column of each subplot) as compared to wakefulness (figure 6.3, first column of each subplot) and significantly distinguishes AWAKE from deep anaesthesia (ANES 3 $\mu g \cdot ml^{-1}$ and ANES 4 $\mu g \cdot ml^{-1}$) at group level, as indicated by the group statistics bar plots in the right hand side of figure 6.3. A threshold effect can be observed: in light anaesthesia (ANES 2 $\mu g \cdot ml^{-1}$) the strength of long-range links decreases if compared with wakefulness, and then shows a plateau remaining relatively constant across the increasing propofol ESC.
We also assessed changes in the direction of information flow over the rostro-caudal axis. The average strength of parieto-frontal and fronto-parietal links across experimental stages are shown in figure 6.3 (second row of each subplot). The strength of posterior to anterior links dramatically decreases during anaesthesia as compared to wakefulness, while the contribution of fronto-posterior connections is dominant in anaesthesia, following an opposite trend. As a consequence the $Dir_{P\rightarrow A}$ index, that summarizes the dominant direction of functional links, shows a switch associated with the onset of anaesthesia. This indicates a significant inversion of information flow form posterio-frontal in wakefulness to fronto-posterior in anaesthesia (bar graphs, last column in figure 6.3). Also for the $Dir_{P\rightarrow A}$ a step effect is observable, rather than a gradual change with anaesthesia deepening, that significantly distinguishes AWAKE from the different anaesthetic stages. The connectivity networks estimated considering the 10%, 30% or 50% of strongest connections exhibit very similar trends, indicating that the choice of the threshold does not critically affect the results. For the sake of brevity, in the following analysis we will only show results for the DC networks obtained retaining the 30% strongest links.

DC_{index} individual trends A reliable and useful index of adequacy of anaesthesia should be able to distinguish between anaesthetic stages for an individual and not only at the group level; we therefore looked at single subject trends and individual variability. To this end, we have assessed DC_{index} values across the whole experiment, including the anaesthetic induction phase and the transitions between stable anaesthetic levels. This allowed comparison with the individual propofol ESC time-line as indicated by the pharmacokinetic model. During the transition periods we assumed that the propofol ESC followed an exponential time course between measured concentrations, in line with the

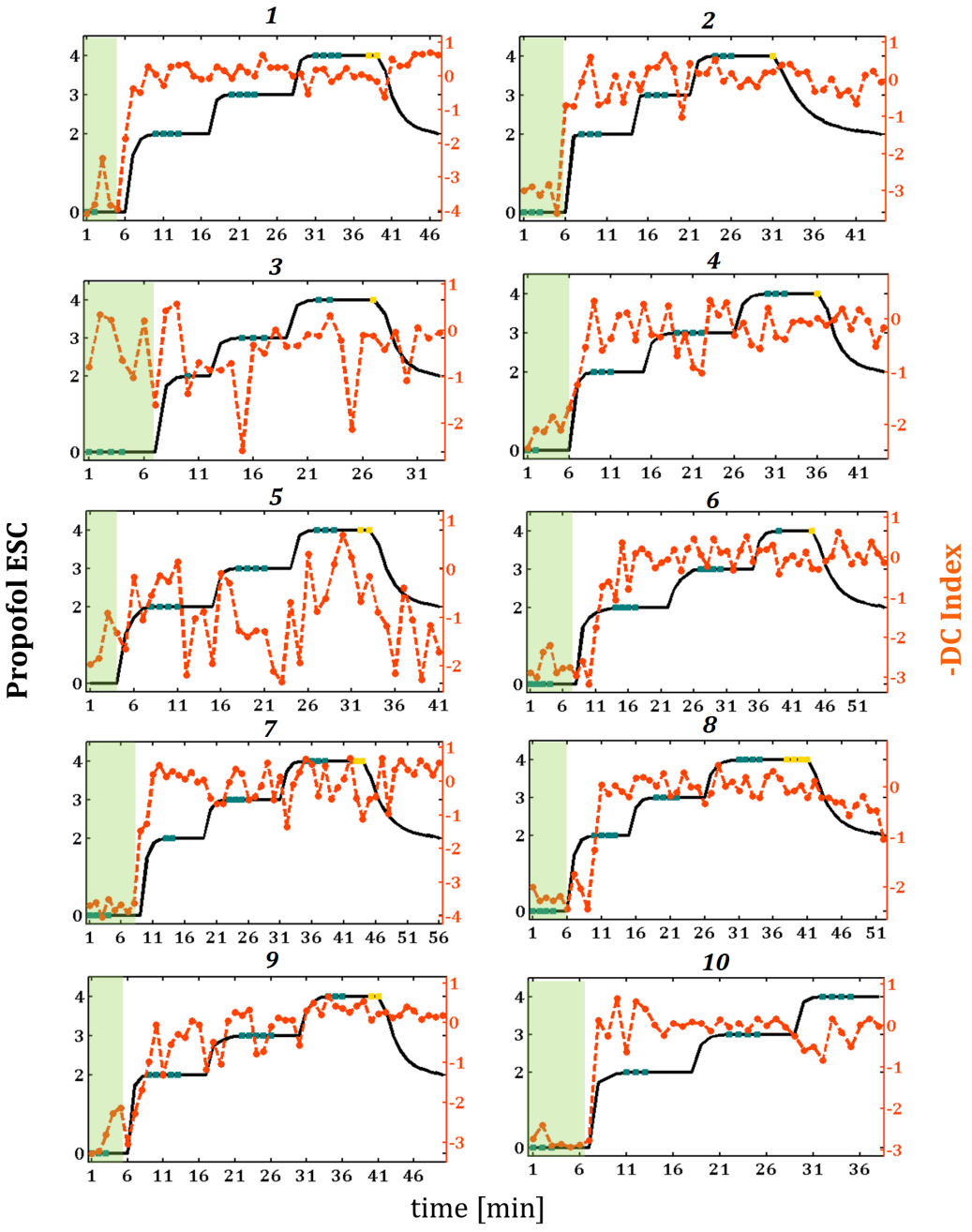

FIGURE 6.4: DC_{index} individual trends (dashed orange line) compared with ESC time-line (black solid line) for all the length of the recording and for all the subjects. DC_{index} is inverted to facilitate comparison with the propofol ESC. The epochs in wakefulness (ESC=0) are highlighted in green and each time point refer to a a 60 s epoch. The epochs where auditory stimulation was delivered are indicated on the ESC time-line by blue markers while the epochs in deep anaesthesia in which muscle relaxant was administered are indicated by yellow markers. In the last subject (10) the recovery period was was heavily contaminated by artefacts and therefore this period was excluded from the analysis.

typical pharmacokinetics of anaesthetic drugs (Roberts & Freshwater-Turner 2007); during the recovery to light anaesthesia the times when the ESC reached intermediate values (3.5, 3 and 2.5 $\mu g \cdot ml^{-1}$) were recorded and the ESC values followed an exponential wash-out curve (figure 6.4). In order to reduce computational effort, the DC links estimated for the transition and recovery periods were not tested for statistical significance and only the threshold to retain the 30% of strongest connections was applied. The shuffling procedure takes the 95% of the total connectivity computation time. If we exclude the shuffling procedure, the time required to estimate connectivity for a 1 min recording with 12 electrodes is roughly 20 s, when using Matlab and a typical Windows-based PC. For some subjects (and for the grand average of DC links) we compared results with and without significance assessment and found very similar values, indicating that the 30% strongest links are likely to be significant (see figure 6.5). In figure 6.4 individual DC_{index} trends for all the subjects are presented. With the exception of subjects 3 and 5, a steep transition at the onset of anaesthesia is noticeable. In the majority of subjects an abrupt change occurs soon after the start of propofol administration, sometimes (i.e. subjects 6 and 8) with 3-4 minutes of delay. The step change in DC values shows then a plateau (with some oscillations) with increasing propofol ESC that does not appear to be significantly affected by the injection of muscle relaxants (indicated with a yellow square in figure 6.4), the following intubation, preparation for surgery nor recovery to ESC 2 $\mu g \cdot ml^{-1}$.

6.3.2 MLR changes during a slow induction of anaesthesia

With the exception of one patient that had normal hearing, the majority of them presented mild hearing loss (HL) at low frequencies and 7 patients were affected by moderate to severe HL at high frequencies. However, a clear response was evoked during wakefulness in all the subjects (with the exception of subject 5 when we could not assess the MLR in wakefulness because of technical issues during the experiment). The relative F_{sp} values are in fact significantly different from the empirical null distribution obtained with the bootstrap procedure ($p < 0.01$). On the other hand MLR evoked in anaesthesia is not always significant in all subjects. Figure 6.6 shows MLR results for subject 1: MLR waveforms for different stable anaesthetic levels are showed (left plot) together with the relative F_{sp} value (as compared with the bootstrap distribution). Due to the strong similarities between trends obtained from the two channels considered for MLR estimation (Fz and Cz), only results for Fz are showed. Results in figure 6.6 show that the MLR morphology changes abruptly from wakefulness to anaesthesia: the Pa-N_b complex amplitude dramatically decreases and in general the MLR features disappear in all the anaesthetic levels. As a result, the responses evoked in anaesthesia are not

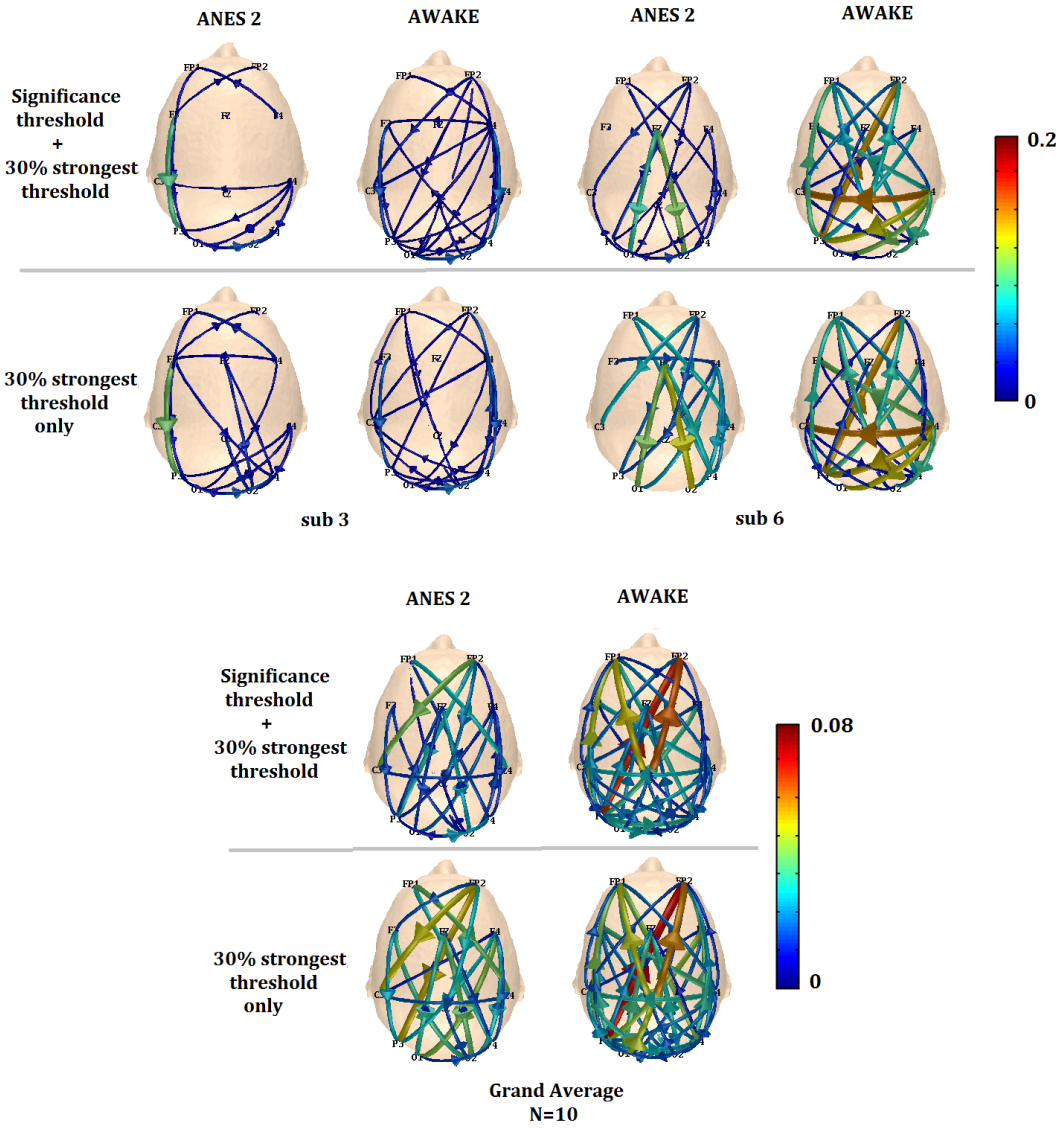

FIGURE 6.5: Comparison of DC connectivity networks obtained applying the significance and the strongest connections (30%) thresholds (first row) or only the 30% strongest connections threshold (second row) in two representative subjects (top graphs) and for the Grand Average across subjects (N=10) (bottom graphs). This figure indicates that both at single subject level and at cohort level the connectivity networks obtained with and without applying the significance threshold and selecting the 30% strongest connections are very similar. This suggests that the 30% strongest links are likely to be significant and that in future applications the computational demanding procedure to assess significance of links may be omitted.

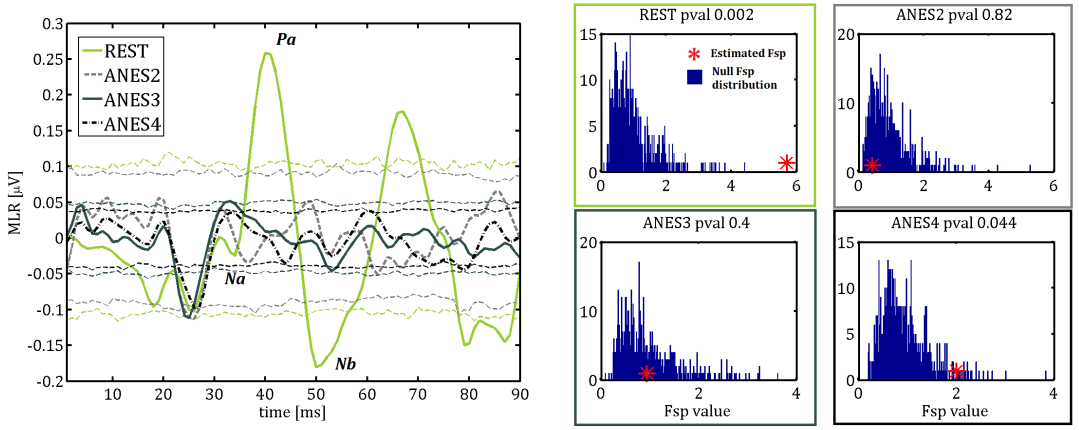

FIGURE 6.6: Representative example of MLR for increasing ESC levels in subject 1. Left plot: MLR waveforms estimated from 2500 sweeps (4 minutes of recording) in AWAKE (green solid line), ANES 2 $\mu g \cdot ml^{-1}$ (light gray dashed line), ANES 3 $\mu g \cdot ml^{-1}$ (gray solid line), ANES 4 $\mu g \cdot ml^{-1}$ (black dashed-dotted line). The respective 95% critical values are showed in the same colours. Right Plot: Estimated F_{sp} (red star) with relative null distribution (blue histogram) for each experimental stage. The empirical p-value obtained comparing the null distribution with the estimated F_{sp} is indicated for each stage in the subplot title.

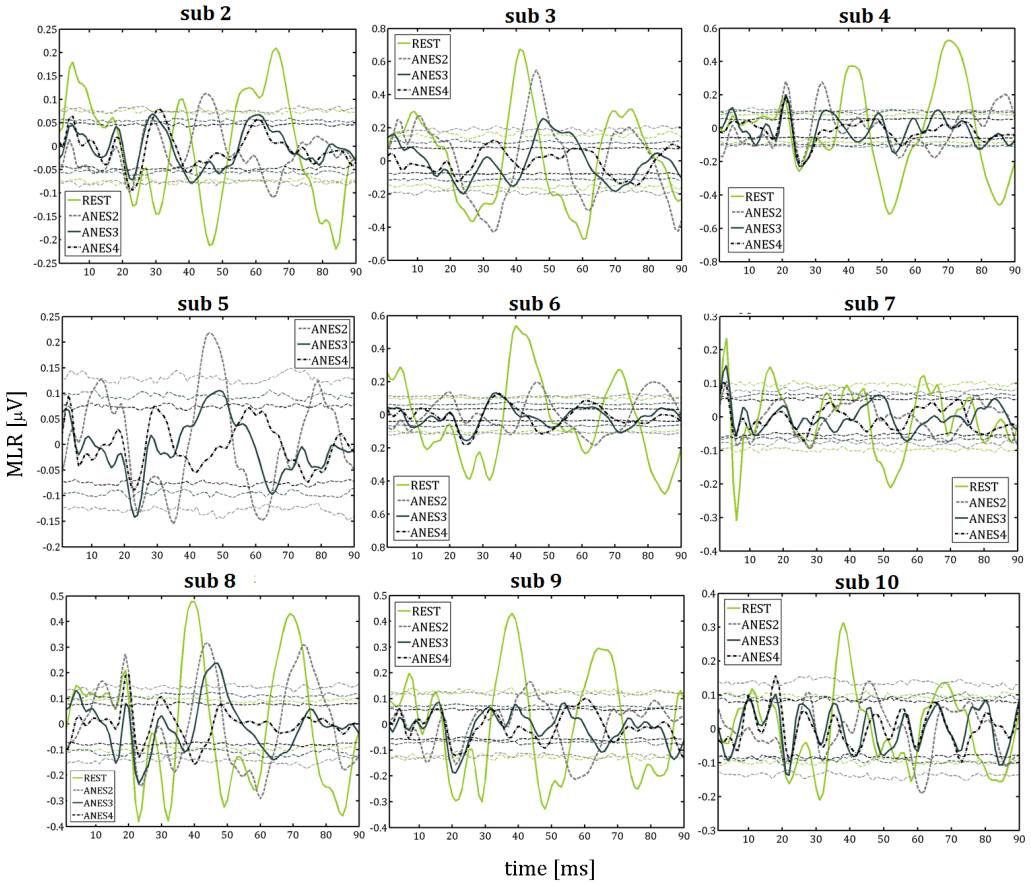

FIGURE 6.7: MLR for increasing propofol ESC in subjects 2 to 10. MLR waveforms were estimated from 2500 sweeps (4 minutes of recording) in AWAKE (green solid line), ANES 2 $\mu g \cdot ml^{-1}$ (light gray dashed line), ANES 3 $\mu g \cdot ml^{-1}$ (gray solid line), ANES 4 $\mu g \cdot ml^{-1}$ (black dashed-dotted line). The respective 95% critical values are showed in the same colours.

Subject	REST CE	ANES 2 $\mu g \cdot ml^{-1}$	ANES 3 $\mu g \cdot ml^{-1}$	ANES 4 $\mu g \cdot ml^{-1}$
1	40	**NA**	**NA**	**NA**
2	36	55	69	67
3	50	52	60	66
4	42	45	44	46
5	**NA**	53	56	**NA**
6	51	54	62	69
7	42	**NA**	**NA**	**NA**
8	39	50	54	56
9	38	47	55	59
10	48	51	**NA**	**NA**
Mean ± Std	43.8 ± 5.4	50.8 ± 3.4	57.1 ± 7.7	60.5 ± 8.6

TABLE 6.1: N_b latency (ms) measured by visual inspection of the estimated MLR for all the subjects of the sample and experimental stages. The **NA** symbol indicates the cases where it was not possible to objectively assess the N_b latency.

significantly different from the background noise (e.g. the estimated F_{sp} values are not significantly different -$p < 0.01$- from the null distribution). Figure 6.7 shows MLR trends for all the other subjects of the sample. In some of them the MLR exhibit an abrupt change during anaesthesia as observed in subject 1. However a more gradual effect was observed in other subjects (e.g. 3, 5, 8, 9 in figure 6.7) where the MLR amplitude more gradually decreases with the deepening of anaesthesia and a shift in the peaks latency is clearly identifiable.

In table 6.1 we show changes in N_b latency for increasing anaesthetic depths for all the subject of the study. In the majority of subjects, we observed a gradual shift in N_b latency with increasing levels of anaesthesia, as previously reported in literature (Thornton et al. 1992, Loveman et al. 2001, Tooley et al. 2004). In average, a N_b latency shift of around 7 ms was observed between AWAKE and ANES 2 $\mu g \cdot ml^{-1}$ and between ANES 2 $\mu g \cdot ml^{-1}$ and ANES 3 $\mu g \cdot ml^{-1}$(table 6.1). A smaller shift of about 3 ms characterized the transition between the deeper anaesthetic levels. However, in a considerable number of subjects the MLR evoked in anaesthesia was within the null distribution 95% critical values or very small, thus making the identification of the N_b peak quite subjective. As a result we could assess N_b latency only in 8 (out of 10) subjects for ANES 2 $\mu g \cdot ml^{-1}$, in 7 for ANES 3 $\mu g \cdot ml^{-1}$ and only in 6 subjects in ANES 4 $\mu g \cdot ml^{-1}$ [1]. There is therefore an indication of N_b latency shift with increasing propofol ESC but it can't be consider it as a robust indicator of the anaesthetic level.

[1] In one subject (5) we could not assess the MLR in wakefulness because of technical issues during the experiment

6.3.3 Comparison of EEG indexes changes associated with increasing ESC levels

Changes in BIS features with anaesthesia have been previously described in literature (Rampil 1998, Miller et al. 2004, Morimoto et al. 2004). In this study we aim at comparing the performances of these different EEG indexes in discriminating the ESC levels during a slow anaesthetic induction. In this section results at cohort level for the different EEG sub-parameters will be shown.

Figure 6.8 shows the cohort averages (and standard error) of connectivity features (first two rows), MLR variance and latency (third row) and BIS sub-parameters (last row), in wakefulness and in the three stable anaesthetic stages.

The MLR power (MLR variance calculated between 30 and 80 ms after the stimulus onset, as in Bell et al. (2006)) and N_b latency are able to significantly distinguish wakefulness from deep anaesthesia but not from ANES 2 $\mu g \cdot ml^{-1}$. On the other hand both the DC_{index} and the $eqBIS$ exhibit significant changes between wakefulness and all the three anaesthetic levels. The two indexes show a different trend with the deepening of anaesthesia: the DC_{index} (and its single features) is characterized by a step change, abruptly decreasing in light anaesthesia (ANES 2 $\mu g \cdot ml^{-1}$) and showing a near plateau for increasing anaesthetic depths. The $eqBIS$ index exhibits a more gradual change as a function of the anaesthetic level, more closely 'tracking' the propofol ESC. The BIS sub-parameters (represented in the last subplot of figure 6.8) shows properties that are in agreement with those reported in literature: the BSR efficiently detects deep anaesthesia, the β_R is sensitive to light anaesthesia and the $SynFS$ more generally differentiate 'activated EEG' in wakefulness from hypnosis (Rampil 1998).

These trends appear to be consistent at individual level. Figure 6.9 shows trends of the three indexes ($eqBIS$, MLR power and DC_{index} in the last subplot) in wakefulness and stable ESC for all the subjects of the sample. The MLR power values show great variability across subjects and in some of them the values in wakefulness and ANES 2 $\mu g \cdot ml^{-1}$ overlap. As also observed at group level, individual DC_{index} values undergo a dramatic drop in anaesthesia and then remain relatively constant with increasing propofol ESC in all the subjects, with the exception of one outlier (subject 3, in purple in figure 6.9, bottom plot). Individual $eqBIS$ values exhibit a more gradual trend with the deepening of anaesthesia in the majority of subjects and a relatively good consistence across segments of the same stable propofol ESC.

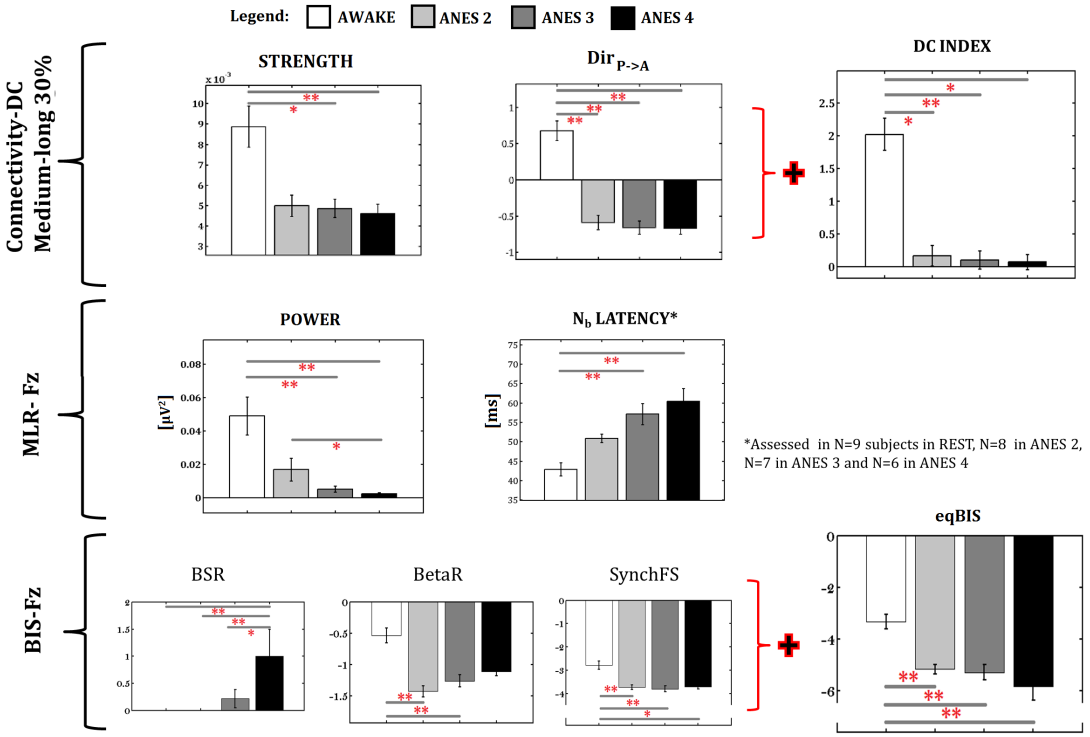

FIGURE 6.8: EEG indexes changes associated with increasing anaesthetic depth. Each bar plot represents the mean and standard error (N=10 subjects) of the specific EEG feature, with relative statistics (Friedman test). The asterisks specify that the two means designated by the brackets significantly differ (*, $p < 0.05$; **, $p < 0.01$), as revealed by Tukey's HSD test. The EEG features are distinguished in three classes, indicated by brackets on the left. The plots in the first bracket represent connectivity networks results (30% strongest connections, α band): the first row shows bar plots of the long-range DC average strength (with associated standard error across subjects) while the second row shows the mean and standard error of the $Dir_{P\rightarrow A}$. The two subparameters were summed to obtain a single index (DC_{index}) whose results (averaged across electrodes) and statistics are showed in the last plot on the first row. The second subplot (third row) show results for the MLR variance and N_b latency. The last bracket include results relative to the BIS sub-parameters (BSR, β_R and $SynFS$) and the global index $eqBIS$. Results obtained from the two channels (F3 and Fz) considered for BIS estimation were very similar, therefore only results for Fz are showed. Both the DC_{index} and the $eqBIS$ exhibit significant changes between wakefulness and all the three anaesthetic levels. The two indexes are however characterized by different trends: the DC_{index} is characterized by a step change while the $eqBIS$ index exhibits a more gradual change as a function of the anaesthetic level.

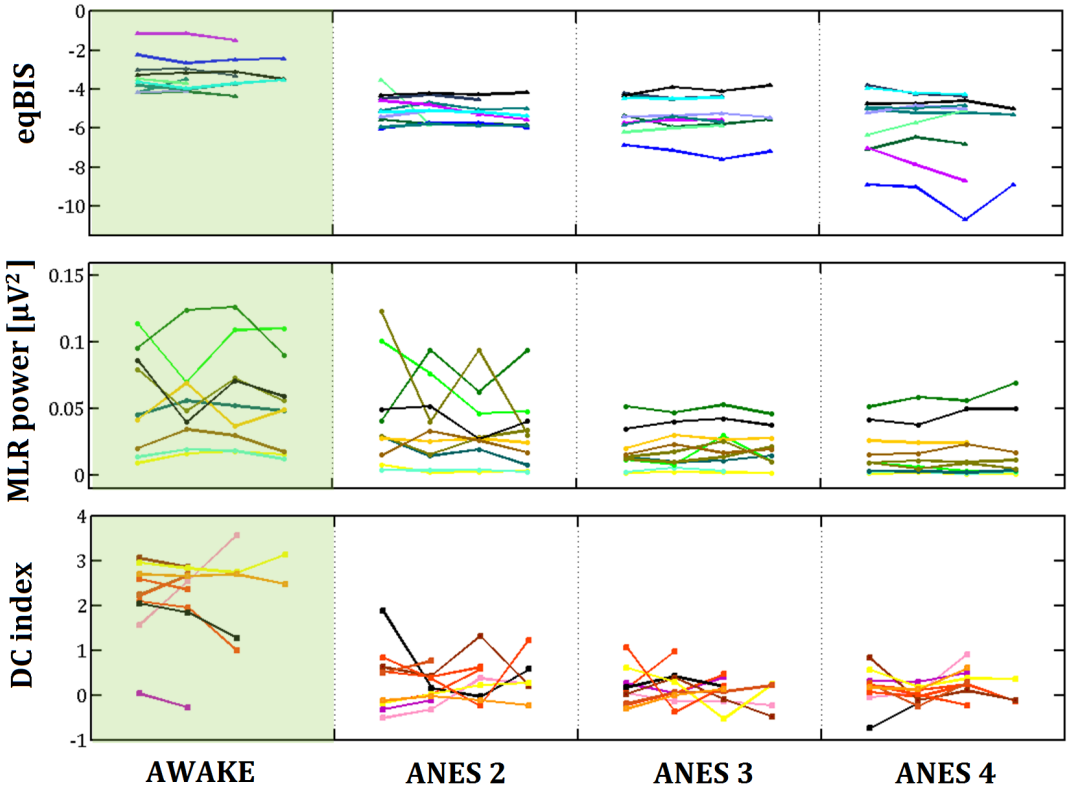

FIGURE 6.9: MLR power, eqBIS and DC_{index} individual trends (in a different color for each subject, N=10) in wakefulness and in stable anaesthetic levels. Each time point represents results from a 60 s epoch.The MLR power values show great variability across subjects with overlapping values in wakefulness and anaesthesia. As also observed at group level, individual DC_{index} values undergo a dramatic drop in anaesthesia and then remain relatively constant with increasing propofol ESC while individual $eqBIS$ values exhibit a more gradual trend with the deepening of anaesthesia.

6.3.4 'Wakefulness vs Anaesthesia' classification performances

We tested the performances of the different EEG indexes in distinguish wakefulness from the different propofol ESC using a SVM binary classifier trained with a 'leave-one-out' cross-validation procedure. In the analysis of the classification properties of MLR the N_b latency was excluded as we could not objectively assess it in all the subjects. Table 6.2 indicates the classification accuracy obtained using the different EEG features (MLR, BIS, DC). In figure 6.10 we show the respective confusion matrices showing the specificity and sensitivity for the different binary classifications.

The SVM classifier trained on MLR power shows the poorest performances, with a percentage of correctly classified epochs lower than 80% for all the binary classifications (table 6.2). In particular it highly misclassifies wakefulness as anaesthesia, as indicated by the confusion matrices in figure 6.10 (first row). The percentage of correctly classified epochs in the binary classification AWAKE vs ANES 2 $\mu g \cdot ml^{-1}$ is only 54%. Among the

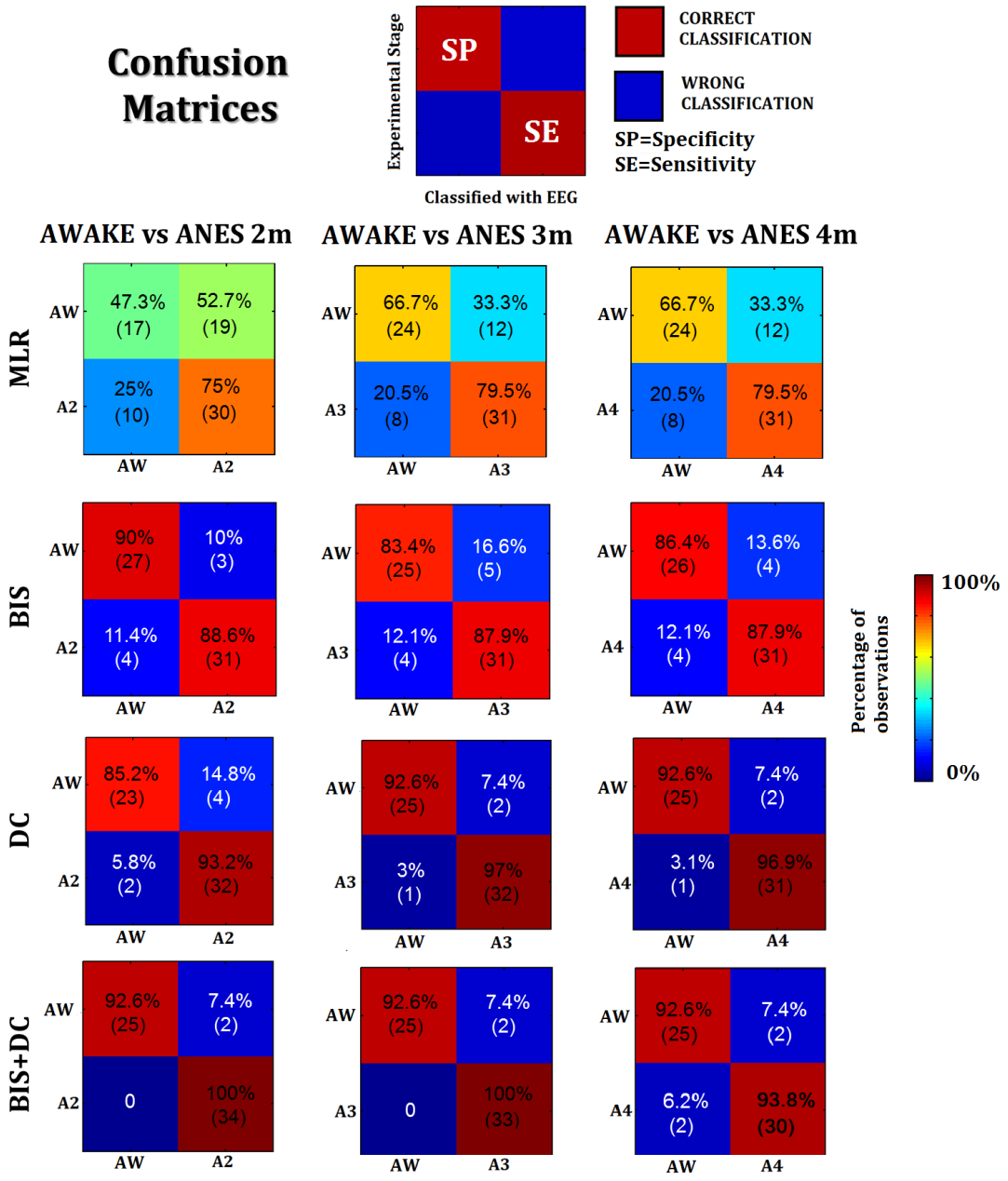

FIGURE 6.10: Comparison of SVM classification based on EEG indexes and experimental stages assessed by propofol ESC. The confusion matrices show on the y-axis the experimental stage and on the x-axis the prediction of the SVM classifier. The percentage of observations classified in each stage (and the associated number of observations in brackets) are reported in each cell. Elements on the main diagonal represent the percentage of correct classifications of 'Awake' observations (Specificity, SP, first diagonal element) and 'Anaesthesia' (Sensitivity, SE, second diagonal element), while on the opposite diagonal are indicated misclassifications. Different columns show results for the three binary classifications, while the rows refer to the type of EEG predictor used in the classification. Results for each of the different EEG indexes (MLR power, BIS and DC, first three rows) and for the combination of DC and BIS subparameters (last row) are showed.

	MLR	BIS	DC	BIS+DC	BIS+DC+MLR
AWAKE vs ANES 2	54.1 (61)	89.2 (65)	90.2 (61)	96.7 (61)	91.8 (61)
AWAKE vs ANES 3	63.3 (60)	85.7 (63)	95 (60)	96.7 (60)	93.3 (60)
AWAKE vs ANES 4	62.7 (59)	87.3 (63)	94.9 (59)	93.2 (59)	87.9 (58)
AWAKE vs ANES (all)	77.9 (154)	90 (131)	94.9 (126)	96.8 (126)	94.3 (124)

TABLE 6.2: Classification accuracy results. The percentage of correctly classified observations (Accuracy) and corresponding number of observations tested (in brackets) are indicated for each of the four different binary classifications (rows) and for different features (columns). The features considered were MLR power, BIS and DC sub-parameters separately (first three columns) or the combination of BIS and DC parameters (fourth column) and MLR, BIS and DC features (last column). Of all the parameters considered, the MLR features showed the poorest classification performances; on the other hand DC features (alone or in combination with BIS) gave the highest classification accuracy, with more than 90% of correctly classified epochs for all the binary classifications.

different single EEG features, the DC_{index} shows best classification performances with a percentage of correctly classified observations above 90% for all the classifications and very low mis-classification percentages (figure 6.10, third row of matrix plots). DC features exhibit significantly higher classification accuracies than MLR (Kruskal Wallis test across subjects, p= 0.001). When the BIS and DC parameters are combined to train the SVM model, the performances slightly improve and the percentage of correct classifications increases above 93% for all the cases (see table 6.2). Moreover the misclassifications of anaesthetic stages as wakefulness are strongly reduced (to 0% in AWAKE vs ANES 2 $\mu g \cdot ml^{-1}$ and AWAKE vs ANES 3 $\mu g \cdot ml^{-1}$) and the number of AWAKE observations correctly classified (specificity) increases as well, as showed in figure 6.10. If the MLR variance is included in the predictors, together with BIS and DC features, the SVM classifier performances are not further improved (see table 6.2, last column) and for most of the classifications the classifier is outperformed by the SVM model trained only on DC features.

We have performed the 'wakefulness' vs 'anaesthesia' classification with the linear SVM model and using a non-linear classifier based on NN, using in both cases a 'leave-one-out' training procedure. A comparison of the average performances of the linear (SVM) and non-linear (NN) models is summarized in table 6.3 (for the sake of brevity we have not reported results for MLR). The SVM and NN classifiers show remarkably similar performances, indicating that the BIS and DC features are linearly separable and the use of a more complex non-linear classifier does not add anything to the discrimination performances. The similarities between the performances of the SVM and NN classifier are considerable also across subjects, we have therefore reported results for single subjects only for one classifier (NN) in table 6.4. The DC performs significantly better than the

	Accuracy			Specificity			Sensitivity		
	# TOT	SVM	NN	# AWAKE	SVM	NN	# ANES	SVM	NN
BIS	131	0.901	0.90	30	0.767	0.716	101	0.95	0.941
DC	126	0. 945	0.964	27	0.815	0.866	99	0.99	0.988
BIS+DC	126	0.966	0.964	27	0.889	0.833	99	0.99	1

TABLE 6.3: Average classification ('Awake' vs 'Anaesthetized') performances (expressed as the fraction of correctly classified epochs) for the linear SVM classifier and the non-linear NN classifier. For each performance descriptor (accuracy, specificity and sensitivity) is indicated the number of epochs it was tested on (total number of epochs for the accuracy, number of observations in wakefulness and anaesthesia for specificity and sensitivity, respectively). Results indicate that DC features exhibit higher classification performances than the BIS and that results obtained with a linear SVM classifier are very similar to those relative to the non-linear NN classifier.

	Accuracy			Specificity			Sensitivity		
Subject	BIS	DC	BIS-DC	BIS	DC	BIS-DC	BIS	DC	BIS-DC
1	0.785	0.928	0.928	0	0.666	0.666	1	1	1
2	0.916	1	1	0.5	1	1	1	1	1
3	0.900	0.800	0.800	1	0	0	0.875	1	1
4	0.916	1	0.916	1	1	0.666	0.888	1	1
5	1	0.916	1	1	1	1	1	0.888	1
6	1	1	1	1	1	1	1	1	1
7	0.812	1	1	1	1	1	0.75	1	1
8	0.818	1	1	0	1	1	1	1	1
9	0.923	1	1	0.666	1	1	1	1	1
10	0.928	1	1	1	1	1	0.9	1	1
Mean	**0.909**	**0.964**	**0.964**	**0.716**	**0.866**	**0.833**	**0.941**	**0.988**	**1**

TABLE 6.4: 'Wakefulness' vs 'anaesthesia' classification performances (in terms of accuracy, specificity and sensitivity) for all the subjects using the non-linear NN classifier trained on BIS subparameters, DC sub-parameters and a combination of them. This table shows classification performance at individual level and indicate that the cohort results showed in previous tables and figures are quite consistent across subjects. This is particularly relevant in view of a possible application of the DC_{index} as a clinical monitor.

BIS (Mann-Whitney test across subjects, p=0.039). The ability of DC (and combined DC+BIS) to correctly classify anaesthesia is particularly high: the sensitivity is equal to 1 in 9 subject out of 10 if only the DC features are used (second to last column in table 6.4) and in all the subjects if both the BIS and DC sub-parameters are included in the classification (last column in table 6.4).

Another consideration regards the optimal linear weights estimated with the validation of the linear SVM model. In order to reduce the dependence of the estimated linear weights from the different subjects dataset we standardized all the EEG features before the cross-validation procedure. We then obtained an array of optimal weights and a linear bias

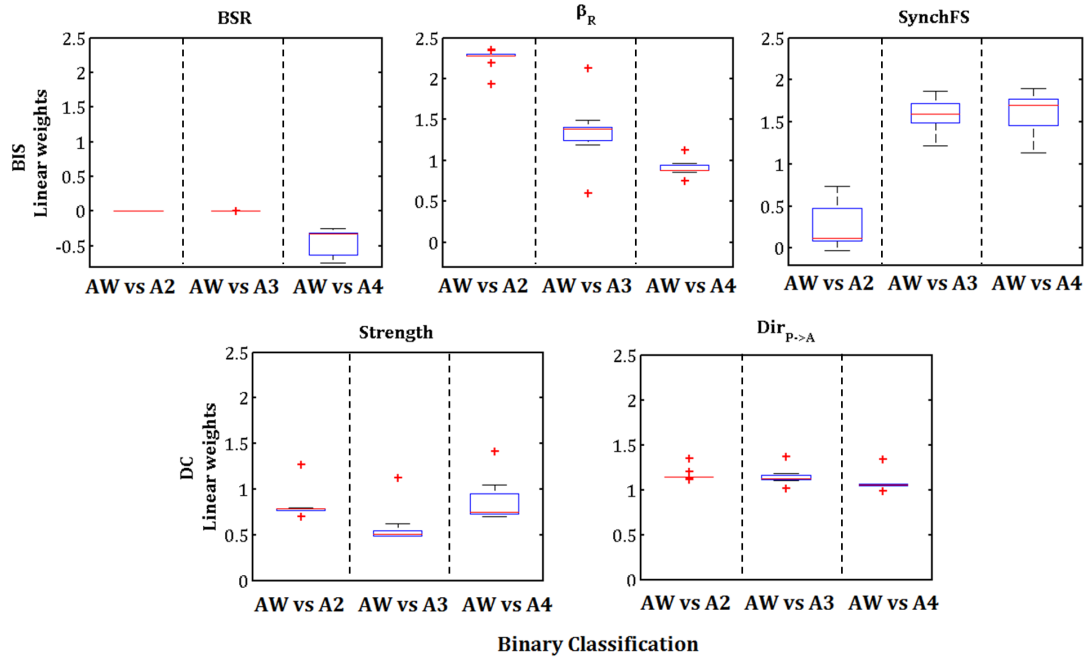

FIGURE 6.11: Optimal linear coefficients distribution (across subjects) of the SVM model trained on BIS sub-parameters for the three binary classifications (AWAKE vs ANES 2, AWAKE vs ANES 3 and AWAKE vs ANES 4). The box plots indicate the first and third quartiles (blue) and the median (in red) and the whiskers represent the 99.3% of the data distribution (outliers are indicated by a red cross).

for each EEG index and for all the subjects and binary classifications. The box plots in figure 6.11 show the distribution (across subjects) of the optimal weights as a function of the binary classification for the BIS and DC indexes. The weights of the different BIS sub-parameters notably vary as a function of the anaesthetic depth classified against wakefulness (figure 6.11, first row). In particular the BSR optimal weight is different from zero only in the classification of AWAKE against ANES 4 $\mu g \cdot ml^{-1}$, while the contribution of the β_R to the global index is twice larger (in average) in light anaesthesia (ANES 2 $\mu g \cdot ml^{-1}$) than in deep anaesthesia (ANES 4 $\mu g \cdot ml^{-1}$). On the other hand, the optimal weights estimated for the normalized strength of long-range DC and the $Dir_{P \to A}$ exhibit more uniform values (the median is close to 1 for both sub-parameters) across the different binary classifications. This confirm that the BIS features have different discriminative power for different stages of anaesthesia and indicate that the weights used to compute the BIS from its subparameters need to be 'tuned' depending on the anaesthetic level, which is quite challenging (and controversial) to realize in real-time monitoring. On the other hand the DC features have broadly the same 'weight' across anaesthetic levels: This may represent an advantage in the implementation of a more stable index for anaesthesia monitoring.

6.4 Discussion

The objectives of this study were to characterize changes in EEG directed connectivity during a slow induction of anaesthesia and to test the ability of connectivity features to discriminate wakefulness from anaesthesia at different depths in comparison with other DoA indexes. The overarching aim was to propose and test a novel index that could be integrated as a new signal processing tool in future DoA monitors. To this end we collected EEG data (with and without auditory stimulation) from a sample of 10 patients undergoing a slow induction of anaesthesia and extracted different EEG indexes in stable propofol ESC. In addition to well established indicators of depth of anaesthesia (BIS and MLR), we proposed a novel index based on functional connectivity features. The rationale behind this index is found in theoretical models predicting that anaesthetic induced LOC is associated with a disruption of brain circuits with a consequent impairment of 'cognitive binding' (Tononi 2008, Mashour 2006, Alkire et al. 2008) and experimental findings showing that brain connectivity is significantly affected in states where consciousness is diminished or suppressed (Ferrarelli et al. 2010, Gómez et al. 2013, Boly et al. 2011). In order to quantify the discrimination performances of the different indexes we have implemented a linear SVM classification analysis by which we also tested if a combination of indexes could result in improved performances. To test for results robustness, we have also used a non-linear classification approach based on artificial NN. This may be able to utilise non-linear dependencies that the SVM cannot. Our results show that connectivity networks undergo dramatic changes with anaesthesia and long-range fronto-posterior connectivity is particularly affected. The most characteristic changes observed in DC during anaesthesia are a significant reduction in the strength of long-range DC links and a marked inversion of direction of information flow, from posterio-frontal in wakefulness to antero-posterior in anaesthesia. These findings appear to be consistent at individual level for the majority of subjects, where a brisk change in DC features is observed at the onset of anaesthesia.

We found that all the different EEG indexes (MLR, BIS and DC) exhibit significant changes in general anaesthesia as compared to wakefulness. The $eqBIS$ and DC index show different trends but they are both particularly efficient in discriminating wakefulness from anaesthesia. Of all the EEG index considered, however, the DC features have the highest accuracy, specificity and sensitivity in discriminating 'awake' vs 'anaesthesia'.

6.4.1 Relation of our results to previous findings

Directed connectivity changes during anaesthesia In order not to interfere with the natural fading of consciousness during the slow anaesthetic induction, we did not assess the responsiveness of patients. Moreover previous works alert on the accuracy of verbal assessment when the protocol involves auditory stimulation, because the insert headphones may interfere with normal hearing (Tooley et al. 1996). Other studies using a similar ESC regimen have assessed the responsiveness of subjects with randomized auditory tasks and have identified an ESC around 2 $\mu g \cdot ml^{-1}$ as the threshold for loss of behavioral responsiveness (Mhuircheartaigh et al. 2013*a*, Purdon et al. 2009, 2013, Forrest et al. 1994). Similarly, in Forrest et al. (1994) a blood propofol concentration (at equilibrium[2]) of 2.3 $\mu g \cdot ml^{-1}$ was required for suppression of consciousness (50% probability of no response to verbal command and eyelash reflex). We therefore expect that in our protocol LOC occurred somewhere between start of propofol infusion and achieving an ESC of 2 $\mu g \cdot ml^{-1}$.

The decrease of the long-range connectivity strength observed in anaesthesia is in line with a wide scope of experiments in the literature that report an impairment of large-scale information flow across brain areas in anaesthetic-induced LOC. A general impairment of brain network integration (with fronto-parietal connectivity particularly affected) has been reported in fMRI and TMS studies of propofol anaesthesia (Schrouff et al. 2011, Boveroux et al. 2010, Massimini et al. 2010). Also activity in the fronto-parietal associative network is systematically altered in other states of diminished consciousness (such as vegetative states, coma or NREM sleep (Massimini et al. 2005*a*, Spoormaker et al. 2010, King et al. 2013, Sitt et al. 2014)). Together with these findings, our results support the important role of the fronto-parietal association cortices in the maintenance of consciousness (Boly et al. 2008) and the hypothesis that the breakdown of information flow may affect signaling between the sensory posterior areas and the associative frontal cortices that is essential for a conscious experience (Boveroux et al. 2010).

The inversion of information flow from parieto-frontal in wakefulness to fronto-parietal during anaesthesia deserves a special comment. The observed switch in the direction of connectivity from wakefulness is consistent with results reported in studies investigating EEG directional connectivity during anaesthesia. Nicolaou and colleagues (Nicolaou et al. 2012, Nicolaou & Georgiou 2013) used a bivariate Granger Causality approach to assess directed connectivity in anaesthesia and they found a significant increase in fronto-posterior causality in anaesthetic induced LOC. They have not observed, on the other hand, a significant decrease in the opposite direction. DC is grounded in Granger

[2] At equilibrium, the effect site concentration is expected to be the same as the blood concentration since the whole system (concentration in blood, fat, muscle, brain, etc.) is in equilibrium.

theory, therefore the two estimators are conceptually similar, however they are characterized by substantial differences: Granger Causality is a pairwise measure in the time domain, while DC is a multivariate measure of causal interactions in the frequency domain and it therefore captures the causality structure of the whole electrodes set. A dominance of fronto-posterior coupling in the cingulate cortex was also observed in a Granger Causality study of propofol anaesthesia (Barrett et al. 2012). Even though fronto-parietal coupling has been suggested as a central mechanism for consciousness, contrasting results about its direction have been reported in the literature. Some studies have shown an impairment of fronto-posterior connectivity in LOC (Boly et al. 2012, Lee et al. 2009) and in some cases two different methodologies applied to the same dataset gave divergent results regarding the dominant direction of coupling (Ku et al. 2011). These conflicting findings are likely to be the results of the use of different estimators of directed connectivity and they are a warning that it may be difficult to interpret results from different brain activity models.

Comparison with the NREM sleep study Results from our previous study on NREM sleep report DC changes associated with deep sleep (N2 and N3) very similar to those observed in anaesthesia. The reduction of long range connectivity observed in NREM sleep as compared to wakefulness has roughly the same proportion in anaesthesia (compare figures 5.9 and 6.3). A switch in the direction of information flow was also observed in NREM sleep (from posterior-frontal in wakefulness to fronto-posterior in sleep N2 and N3); however the inversion observed in anaesthesia appear more marked than in NREM sleep (compare figures 5.9 and 6.3). Similarly, Massimini and colleagues (Massimini et al. 2005*b*, Ferrarelli et al. 2010) observed a remarkable resemblance between EEG patterns elicited by TMS in NREM sleep and midazolam-induced LOC, as compared to wakefulness. A high-density EEG sleep study (Massimini et al. 2004) showed that deep sleep is characterized by a travelling wave that originates in frontal regions and propagates in the anteroposterior direction. Given the similarities of sleep and anaesthesia neurobiology (Mashour 2010) it has been suggested (Nicolaou et al. 2012) that a similar behaviour characterizes the anaesthetic slow wave and that the observed increase in fronto-posterior coupling may be a result of these mechanisms. These results also suggest that DC networks exhibit consistent changes with the level of consciousness of the subject and that these changes are scarcely influenced by the mechanisms (physiological or drug-induced) by which LOC was induced.

MLR changes during anaesthesia In agreement with previous studies (Bell et al. 2006, Loveman et al. 2001, Thornton et al. 1992, Tooley et al. 1996, 2004), we found a significant change in MLR amplitude and a shift in N_b peak latency with anaesthesia.

Bell and colleagues (Bell et al. 2006) investigated MLR changes associated with changes in responsiveness to command and reported a switch in MLR power with the onset of anaesthetic induced unresponsiveness but they did not report any latency shift. On the other hand, in the study of Loveman and colleagues (Loveman et al. 2001) observed a shift in average N_b latency from 48.4 ms during responsiveness to 54.6 ms in periods of propofol (in combination with fentanyl) induced unresponsiveness. In an earlier study using volatile agents (Thornton et al. 1989) a threshold N_b latency of 44.5 ms distinguished responsiveness from unresponsiveness epochs in four out of seven patients. Interestingly enough, this threshold would separate wakefulness from anaesthesia in 7 of the 10 subjects in our study (table 6.1). In two studies Tooley and colleagues investigated relationships between MLR features, subject's responsiveness and blood concentration of propofol alone (Tooley et al. 1996) or in combination with alfentanil (Tooley et al. 2004): N_b latency was identified in both studies as the best indicator of unconsciousness with an optimal threshold of 53 ms for propofol alone and 46 ms in combination with alfentanil. In contrast with the majority of works but in line with the present study, Tooley et al. (1996) also report cases where MLR was too suppressed to allow reliable estimation of N_b. Taken together, these results report a large variation of the N_b cut-off threshold and suggest that changes in N_b latency strongly depend on the combination of drug used. We found a significant change both in MLR amplitude and latency during anaesthesia. However we couldn't reliably assess the N_b latency in all the subjects, especially in deep anaesthesia, because the evoked MLR was not significantly different from noise and/or the N_b peak not objectively identifiable. Moreover, together with an abrupt change of the MLR amplitude, we also observed in few subjects a more graded decrease in amplitude. This may be a result of the gradual induction of anaesthesia used in this study but it emphasizes that there is a critical individual variability in the effect of anaesthesia on the MLR. As also previous results have shown, we found a large range of individual N_b latencies and MLR amplitudes that complicates the identification of a general cut-off point. This is also reflected in the scarce 'awake' vs 'anesthetized' classification performances of the MLR. These findings suggest that the relation between anaesthetic concentration and AEP is complex and not fully understood and alert on the limitations of the clinical application of MLR as an anaesthetic monitor.

Comparison of EEG indexes performances Both BIS and connectivity features efficiently distinguish wakefulness form light and deep anaesthesia. The BIS subparameters exhibit values and performances that are expected, if we consider previous literature (Morimoto et al. 2004, Miller et al. 2004, Rampil 1998). As also previous studies have showed, BIS presents a gradual change with anaesthesia deepening. It has

Accuracy	Prediction Probability	Index
0.98		Granger Causality
	0.92	Narcotrend Monitor
0.86		Recurrence quantification analysis
	0.86	Approximate Entropy
	0.86	Spectral Edge Frequency
	0.78	Median Frequency
	0.82	BIS Monitor
	0.87	Permutation Entropy
	0.87	Order Recurrence Rate
	0.87	Phase coupling of order patterns
0.69		Time-Encoded Signal Processing and Recognition
	0.89	State Entropy
	0.88	Response Entropy
0.93		Complexity based on Lempel-Ziv
0.76		Spectral Entropy
0.64		Median Frequency

TABLE 6.5: Quantitative comparison of classification performances with other methods reported in the literature. Adapted from Nicolaou et al. (2012)
. Even if a rigorous comparison is not possible due to the different methodologies used to assess classification accuracy, this table gives an indication about the performance of various depth of anaestesia indexes proposed in literature as compared to the results obtained using the DC_{index}.

been suggested that BIS actually measures the anaesthetic induced suppression of EEG activity and thus may be an indicator of the clearance of drugs instead of the state of arousal (Gajraj et al. 1999), which is the results of a complex balance of anaesthetic (hypnosis) level, analgesic level and surgical stimulation (Schneider et al. 2002, Chan & Gin 2000). BIS values show a high variability during stable physiological conditions and in some studies have been demonstrated to be unable to detect the transition from unconsciousness to the recovery of awareness (Gajraj et al. 1999). The ability of BIS in measuring the hypnotic level was also questioned in a study on 20 patients monitored with the isolated forearm technique (IFT) (Schneider et al. 2002) where the BIS index could not reliably distinguish between responsive and unresponsive patients.

The DC index shows a substantial different trend from BIS and MLR and other indexes that have been proposed as DoA indicators. The majority of commercially available indexes usually exhibit monotonic graded changes with increasing anaesthetic doses, as reflected by the use of an index ranging from 0 (cortical silence) to 100 (subject awake and oriented). Nicolaou and colleagues Nicolaou et al. (2012) have investigated EEG directed connectivity assessed with Granger causality and found a similar categorical

changes of Granger Causality associated with LOC. They have suggested that such an abrupt change may more closely reflect the point at which consciousness is lost.

Should we consider consciousness as a 'switch' or expect a gradual change from wakefulness towards LOC? There is supporting evidence in literature that anaesthetic (and slow-wave sleep) induced LOC is a switch-like phenomenon because it is related to the switch of thalamic cells from a tonic-firing pattern, characteristic of vigilance, to a burst-firing behaviour (Llinás & Steriade 2006). This switch in the thalamic neurons occurs coincident with a change in EEG patterns from low voltage high frequency 'Activated EEG' to slow waves activity and it is caused by an hyperpolarization of thalamic cells that block the transmission of sensory input to the cortex (Alkire et al. 2000). Whether this thalamic switch is a direct effect of anaesthetics or it is mediated by cortical activity is an open question (Alkire et al. 2008): it is hypothesized that primary effects of anaesthesia occur at the cortical level (with higher order cortical areas more sensitive than the lower order ones) and then secondary effects occur in subcortical areas (Velly et al. 2007), however the exact sequence of events remains unclear. We observed at individual level that the DC_{index} is scarcely influenced by the administration of muscle relaxants, auditory stimulation or the intubation of the patient: this supports the speculation that it reflects the general physiological mechanism of hypnosis, rather than, for instance, the level of anti-nociception (Marchant et al. 2014, Nicolaou et al. 2012).

Irrespective of its physiological interpretation, the DC_{index} step change identifies a clear boundary between the DC features in 'awake' and anaesthesia', leading to very high classification performances. DC performs better than BIS in distinguish wakefulness from light ananesthesia, deep anaesthesia and anaesthesia in general. The DC index is particularly efficient in discriminating wakefulness from anaesthesia if compared with the qualitative performances of other commercially available devices regardless from the type of classifier (linear and non-linear) used (see, table 6.5 from Nicolaou et al. (2012) for a comparison of different DoA indexes; note that, due to the use of different methods to assess performances- Prediction Probability vs Accuracy-, a rigorous comparison between different methods is not possible). Despite the inter-subject variability in connectivity networks that one would expect, the DC_{index} performances are robust across subjects (see table 6.4).

A monitor of anaesthetic adequacy should perform equivalently well in detecting all the episodes of awareness (specificity) and in identifying when a patient is adequately anaesthetized (sensitivity). The classifier based on the DC features show both very high sensitivity and specificity. In particular its ability to correctly detect adequate anaesthesia its ideal (100%) in all the subjects of the sample but one. When the DC subparameters are combined with BIS features the classification performances slightly (but not significantly) increase. The ability to detect 'awareness' is a fundamental feature of a monitor of the adequacy of anaesthesia. EEG-derived DoA monitors are now

recommended in the clinical practice with the specific purpose of reducing the risk of intraoperative awareness (Association of Anaesthetists of Great Britain and Ireland 2016). However the ability of these devices to discriminate episodes of unexpected awareness is limited (Pandit & Cook 2014, Shepherd et al. 2012) and need to be improved (Marchant et al. 2014). In view of possible application to clinical setting, the DC_{index} 'on-off' response may be useful during anaesthetic onset, but potentially less useful during offset. However, if the DC_{index} switch becomes 'on' before the patient is conscious, then it could help to ensure the adequacy of anaesthesia during the maintenance phase and could give warnings that the patient will soon awaken during the recovery phase. Further studies are required to determine when the switchover occurs in relation to conscious responsiveness (during induction and recovery from general anaesthesia) and to test the use of the DC_{index} (in combination to other DoA indexes) in monitoring of anaesthetic adequacy. Our exploratory work indicates that connectivity features may be useful if applied in the clinical practice alone or in combination with established DoA monitors to improve their performances. In this sense an important advantage of our methodology (with respect to other brain imaging technique) is its clinical applicability, due to its relatively cheap technology, the possibility to be applied at bedside and the relatively short computational time.

6.4.2 Limitations

The binary classification wakefulness vs anaesthesia performed with the SVM and NN models deserves an additional comment. Due to the design of the experimental protocol, the number of epochs available in wakefulness and anaesthesia was highly unbalanced (27 and 99 respectively). This may bias the classifier performances towards a higher sensitivity to the detriment of the specificity. In our study in fact the sensitivity obtained with the DC_{index} is ideal in 9 subjects out of 10, while in some subjects the specificity is zero, because the number of observations in wakefulness is very low if compared with the anaestesia examples (i.e. in subject 3, that shows a DC_{index} specificity equal to zero, the observations in wakefulness are only two, table 6.4). Nonetheless, the average values of specificity are quite high (around 87% in average and 100% in 7 out of 10 subjects), thus indicating that despite the relatively scarce number of observations in wakefulness, DC shows promising performances that may improve if a higher number of wakefulness instances is considered.

In this work we have compared the performances of a novel index based on connectivity features with established DoA monitors as the BIS. In interpreting our findings some limitations in the computation of the BIS index must be taken in account. Using the commercially available BIS monitor requires the application of a large strip of electrodes

on the front of the patient. In this exploratory study we decided not to compromise the EEG recording from frontal electrodes for connectivity estimation and therefore to estimate the BIS values a posteriori from frontal EEG recordings. The exact algorithm that combines with predetermined weights the different BIS sub-parameters is proprietary (Rampil 1998). In addition the optimal weights are changing depending on the level of anaesthesia thus making the reproduction of BIS values more complicated. The aim of this work was, however, to assess changes and trends in the BIS sub-parameters and the performances of an optimal combination of these features in discriminating wakefulness from anaesthesia. Future work will include an anaesthetic study where EEG recording and BIS® monitoring will be performed simultaneously in order to more reliably compare their performances.

The methodology used to assess brain connectivity suffers from some limitations. DC quantifies the normalized amount of spectrum transferred from one signal to another through all the possible paths in the multivariate set and therefore it is not able to distinguish between direct and indirect linear causality. This may represent an issue in the typical case where two channels are influenced by a third one, if the aim of the analysis is to specifically assess the direct connection between the former two. DC however presents, differently from estimators of direct causality developed in the same framework (i.e. Partial Directed Coherence) the advantage of offering a straightforward interpretation in terms of frequency bands (Faes & Nollo 2011): This is particular relevant in the analysis of EEG time series that are characterized by specific brain rhythms associated with behavioural states (Klimesch 1999). Moreover, in order to infer the precise structural causality of the dataset, all the sources of influences must be considered, together with the effect of the volume conduction of cortical sources across tissues of the brain to the scalp. This is obviously not practicable in a clinical contest. However if, as in the present study, the interest is on the broad changes observed on scalp connectivity rather than the investigation of specific effective connections, DC represent a valid (Kaminski & Blinowska 2014), relatively simple methodology, that performs well also in the case of non-linear interactions and non-stationary signals (Winterhalder et al. 2005).

An important consideration regards the number of electrodes considered, a parameter that affects the computational times and the clinical applicability of this methodology. In future, considerable effort will be spent in refining the methodology through the identification of a reduced number of electrodes and the simplification of the signal processing pipeline. In this work we tested the significance of DC links by means of a shuffling procedure that significantly increases the computation time (it takes approximately the 95% of the total connectivity estimation time, that for 1 min epoch and 12 electrodes is roughly of 6 min, using a Windows based pc). Our findings (an example is

given in figure 6.5) however indicate that the network features obtained applying only the 30% threshold on strongest links are very similar to those obtained using also the significance threshold. This suggests that in the computation of the DC index the computationally costly shuffling procedure may not be required, thus making the real-time implementation more efficient.

A critical limitation of this study is that we did not assess the behavioural responsiveness of the patients during anaesthetic induction: the rationale for this choice was not to interfere with the natural fading of consciousness by asking the subject to respond to a verbal command and to reduce to the minimum the movement artifacts during EEG collection. We considered as a gold standard measure of the anaesthetic depth the ESC that quantifies, through pharmacokinetic models, the concentration of the anaesthetics in the brain. The choice of a reference measure presents the intrinsic problem of the absence of a gold standard measure of anaesthetic depth (Boly & Seth 2012). As a consequence indirect parameters must be used. Critics moved to the choice of defining the level of consciousness on the basis of the ability to respond to a command are that it relies also on the subject willingness to respond (Ferrarelli et al. 2010) and that unresponsiveness and consciousness are not causally linked (consciousness occurs in case of non-responsiveness: for example in REM sleep subjects are conscious of their dreams but are unresponsive to sensory stimuli and incapable of moving because of brainstem induced paralysis (Sanders et al. 2012)). Generally DoA monitors performances are investigated with respect to the estimated or measured anaesthetics' concentration, however more parameters (or anaesthetic induced 'effects') should be explored to thoroughly describe the properties of an index. This need a detailed understanding of anaesthetic effects on consciousness (subjective experience), connectedness (awareness of external stimuli) and responsiveness (goal-directed behaviour). These concepts are not clearly distinguished in literature (Marchant et al. 2014) and further work should be carried out in future to test how DC (and other EEG indexes) perform in relation to these different endpoints.

Finally, the sample size of the current anaesthetic study is relatively small and measurement were performed in the electrically noisy environment of a clinical department (although this does represent a real world situation): A larger study is therefore needed to confirm these initial findings.

Chapter 7

Conclusions and future works

In this thesis a novel approach to monitor anaesthetic depth based on EEG connectivity measures has been developed and tested. The work aimed at proposing and testing such methods, but also addresses with some more basic signal analysis issues that might confound the results in the application to multichannel EEG signals. In the following we will summarize the main contributions first. A related discussion and suggestions for future works will follow.

7.1 Summary of Original Contributions

- **Limitations in the proposed eMVAR approaches to account for instantaneous connectivity in the EEG.** Classical MVAR models do not include zero-lags coefficients, therefore they disregard the presence of instantaneous connectivity that may affect the estimation of lagged causality. With EEG signals, instantaneous connectivity is expected due to strong volume conduction effects and this has the potential to confound the analysis of causal connectivity. The eMVAR framework has been introduced to deal with this issue and has been tested in previous works on simple connectivity models where it was shown to effectively eliminate spurious causality introduced by zero-lag effects. In this work we tested the eMVAR approach on simulations that more realistically model EEG signals and on EEG recordings and showed that in this case the eMVAR algorithm fails to reliably estimate instantaneous effects. Our findings suggest that the characterization of instantaneous causality between EEG time-series remains a challenging issue and that caution is required when interpreting results from the estimation of EEG connectivity.

- **Characterization of instantaneous volume conduction effects on scalp DC and PDC by means of a realistic head model.** The effects of volume conduction on the estimation of scalp connectivity are not clear and some authors claim that PDC and DC are not affected by the zero-phase (instantaneous) mixing of cortical sources. By means of a series of simulation studies using a realistic head model we demonstrated that DC and PDC are affected by volume conduction when source signals are coloured. We also characterized volume-conducted, non-physiological connectivity as a function of interelectrode distance and reference choice: our preliminary results show that DC and PDC considerably reduce the effects of volume conduction if compared with Coh and PCoh, in particular for interelectrode distances larger than 10 cm.

- **Correlation of directed connectivity performance with the individual level of consciousness during NREM sleep.** We have assessed the performances of different EEG indexes (PDC and DC directed connectivity and normalized spectra in different frequency bands) in relation to sleep stages in a NREM sleep study. Our results show that among all of the EEG measures tested, a proposed index of the direction of information flow on the rostro-caudal axis that is based on DC performed well at a group level and gave the highest correlation with individuals sleep stage and hence level of consciousness.

- **EEG directed connectivity changes during a slow induction of propofol anaesthesia -group and individual trends**. We assessed changes in multivariate EEG connectivity estimated with DC during a target-controlled slow induction of propofol anaesthesia with a view to proposing a connectivity-based measure of depth of anaesthesia. We observed an inversion of directed connectivity from posterio-frontal in wakefulness to fronto-posterior in anaesthesia. We have identified a step change of connectivity features with the onset of anaesthesia (in contrast with a more gradual trend with increasing propofol ESC observed in BIS and MLR) that is broadly consistent at individual level and is relevant in terms of a physiological interpretation of anaesthetic-induced LOC.

- **Proposal of a novel and promising index of anaesthetic depth based on EEG directed connectivity features and assessment of its performances in comparison with BIS and MLR**. We proposed a novel DoA index based on DC features (DC_{index}). We assessed DC_{index} performance in discriminating wakefulness from anaesthesia in a clinical setting and compared it with MLR and BIS features extracted at stable anaesthetic brain concentrations (ESC). We showed that directed connectivity features have the best performances in discriminating wakefulness from anaesthesia, as compared with MLR and BIS, with an average

accuracy of 96% and results robust across subjects. Our results indicate the potential for directed connectivity to be integrated into future DoA monitors (possibly in combination with other EEG features) to improve the detection of intraoperative awareness.

7.2 Discussion

7.2.1 Signal Analysis Methods

PDC and DC are well established and relatively straightforward methods that quantify the direction and strength of linear interactions between signals. They are computationally simple and represent a convenient solution that does not require an a-priori functional model of interactions and allows a straightforward interpretation in terms of EEG rhythm content. The relationship between PDC/DC scalp connectivity and effective connectivity among underlying cortical and subcortical sources remains however unclear, mainly as a result of the confounding effects of volume conduction. The literature suggests that these measures are relatively immune to the inevitable volume conduction effects (Kaminski & Blinowska 2014, Schlögl & Supp 2006), but the published works are based on not very realistic simulations, confirming the potential limitations of these connectivity measures. Furthermore, solutions proposed in the literature using the eMVAR model are based on unrealistic assumptions (Shimizu et al. 2006, Hyvarinen et al. 2010, Faes & Nollo 2010): we showed that as a result the application of this more sophisticated approach (eMVAR) to remove instantaneous causality among EEG derivations is not reliable and the estimation of instantaneous causality in this case remains a challenge. At the same time an attempt to define the conditions under which the volume conductions effects on PDC/DC could be considered negligible was carried out: we showed that MVAR based directed estimators importantly reduce the spurious connectivity generated by instantaneous source mixing if compared to Coh and PCoh and that PDC/DC are minimally affected by volume conduction for interelectrode distances larger than 10 cm. These results, even if preliminary, provided useful guidelines to interpret the physiological significance of DC/PDC links for the experimental studies that have followed and oriented the choice to focus on the analysis of long-range connectivity (i.e. interelectrode distance larger than 10 cm) in the analysis of experimental data that followed.

Limitations As discussed in the final sections of chapters 4, the simulations studies carried out in this thesis suffer from some limitations and leave unresolved questions. For instance, in the characterization of volume conduction effects on DC and PDC scalp

connectivity, we used only a source dipole distribution and white or coloured uncorrelated sources signals. Further work is required to test how volume-conducted scalp connectivity varies as a function of different cortical source distributions (different LFM matrices) and if realistic cortical dipole signals are used.

A key limitation of the methodologies used in this work is that they quantify only linear dependencies between EEG time-series. In biological systems, in particular in very complex ones like brain networks, we cannot assume linearity or homogeneity. Both linear and non-linear processes take place in the brain (Nunez 1981). Phenomena like volume conduction of cortical source generators are accurately modelled under the assumption of linearity (Nunez et al. 1991), but whether a linear model is an accurate approximation of neuronal connectivity is debated (Friston 1994).

As mentioned in chapter 3, several functional connectivity estimators that quantifies non-linear coupling have been proposed in literature (i.e. Mutual Information and Transfer Entropy (Schreiber 2000), Generalized Synchronization (Rulkov et al. 1995), etc.). Estimators based on information theory quantify statistical dependencies between time series and requires the computation of probability density which can be computationally demanding and requiring very large dataset (Winterhalder et al. 2005). Similarly, General Synchronization entails higher computational effort compared to linear estimators. Whether non-linear measures give an important contribution to the study of synchronization in EEG signals has been investigated in a study on recorded EEG (Quiroga et al. 2002): despite their sensitivity to different properties of the signals all the measures gave similar results in estimating EEG synchronization, except for the Mutual Information which was not robust due to the limited data length available. There is no consensus about the best method to characterise neuronal couplings among EEG channels: each estimator has a different sensitivity in detecting connectivity that depends on the spectral characteristics of the interaction and the nature of the coupling (David et al. 2004). Linear modelling may outperform alternatives when the assumption of linearity is (approximately) valid, but is likely to underperform when the assumption is grossly violated. Linear models represent a straightforward, computationally convenient method, with a good temporal resolution. These properties are important in view of an application for anaesthetic monitoring where quick updates on the current DoA of the patients are needed. Moreover MVAR based estimator have been shown to be robust to noise (Blinowska 2011) and to perform well even in the case of some nonlinear interactions (Winterhalder et al. 2005) and have also been widely applied in different experimental conditions other than sleep or anaesthesia (Astolfi, Bakardjian, Cincotti, Mattia, Marciani, De Vico Fallani, Colosimo, Salinari, Miwakeichi, Yamaguchi, Martinez, Cichocki, Tocci & Babiloni 2007, Ginter et al. 2001, Kuś et al. 2008, Brzezicka et al. 2011) showing good agreement with physiological considerations. If, as in the present study, the interest is on the broad changes observed on scalp connectivity and

to find an efficient measure of LOC rather than elucidating physiological connections in detail, the linear solution was deemed to be a good starting point from the authors and it was shown to be a valid and promising methodology when applied to the analysis of sleep and anaesthesia. Even if in future work non-linear methods are to be explored, the linear models will be a useful baseline to which non-linear approaches can be compared. Only if their performance is very clearly better, will the added complexity and computational effort be justified.

7.2.2 Experimental Findings

As a first stage of the experimental work a sleep study was designed in order to explore changes in EEG connectivity during the descent into NREM sleep, the macro-stage most likely associated with LOC. This normative study of EEG changes during sleep could be conducted in a laboratory environment, whereas a similar study using anaesthetic-induced LOC is experimentally and ethically much more challenging. Our attempt to obtain ethics approval for the volunteer study was rejected, based on perceived balance of risk and scientific need. In the sleep study the connectivity estimators and features that were most sensitive to changes in the individual level of consciousness were identified among a set of connectivity (Coh, PDC, DC) and spectral (normalized power in EEG bands) measures: DC showed more robust results if compared with PDC, that exhibited fragmented networks in some subjects. Among the different network features, the strength of long-range links and the dominant direction of connectivity in the posterio-anterior axis were most sensitive to the effect of NREM sleep. The rationale behind the investigation of EEG connectivity is in the physiological understanding indicating that the level of consciousness is closely associated with brain connectivity (Tononi & Massimini 2008, Prichep et al. 2004, Alkire et al. 2008, Massimini et al. 2005*a*, 2010, Mashour 2004). Results from the sleep and the anaesthetic study confirmed the research hypothesis showing that brain connectivity is significantly modulated by the individual's level of consciousness (sleep stage or anaesthetic depth). The changes observed in DC patterns during a slow administration of propofol in the anaesthetic study that followed show remarkable similarities with those observed in deep sleep: a comparable reduction in the strength of long-range links and a slightly more marked inversion of the direction of connections was observed during anaesthesia. These findings are in line with the widely acknowledged hypothesis that brain networks that generates sleep are modulated by anaesthetics (Mashour 2010).

As discussed in the last section of chapter 6, the results of the anaesthetic study are also in line with a considerable number of other studies that used neuroimaging techniques sensitive to different aspects of brain activity (fMRI, PET, TMS) and reported a

general impairment of cortico-cortical connectivity; fronto-parietal networks appear to be specifically affected both in anaesthesia (Schrouff et al. 2011, Boveroux et al. 2010, Massimini et al. 2010) and other states of altered consciousness (Massimini et al. 2005*a*, Spoormaker et al. 2010, King et al. 2013, Sitt et al. 2014). These findings support the various models of anaesthetic actions proposed in the literature (Tononi & Massimini 2008, John & Prichep 2005, Alkire et al. 2000) suggesting that the breakdown of cortical information sharing is a central mechanism of anaesthetic induced LOC.

The changes in EEG directed connectivity observed at cohort level appear to be consistent at individual level for the majority of subjects, where a brisk change in DC features is observed at the onset of anaesthesia. The DC_{index} is then scarcely affected by the increasing dose of anaesthetic, the administration of muscle relaxants, auditory stimulation or the intubation of the patient. Interestingly, the changes observed in EEG connectivity patterns in NREM sleep and anaesthesia remain consistent also when the underlying EEG rhythms vary considerably: For instance in deep anaesthesia (ESC 4 $\mu g \cdot ml^{-1}$) EEG time-series are characterized by burst-suppression while in sleep N2 spindle activity is dominant but the connectivity features in these two stages are similar. Taken together, these findings support the speculation that connectivity changes reflect the general physiology of LOC, independently of the mechanism by which it is induced. While conventionally the EEG patterns have received more attention, because they are more easily identified visually, connectivity may be more informative regarding LOC. Methods based on individual channels of EEG (spectral features, AEPs) may be sensitive to other aspect of sleep or anaesthesia and may disregard a key element of LOC: the information sharing between different areas of the brain.

As a final step of the anaesthetic study, the performance of the DC_{index} in predicting anaesthetic-induced LOC were compared with those of two established DoA indexes: BIS and MLR. In combination with a SVM (or NN) classifier, directed connectivity features showed the best performances in discriminating wakefulness from anaesthesia at different depths. This is a result of the step change in DC strength and direction that occurs at the onset of anaesthesia, in contrast to a more gradual trend observed in MLR and BIS for increasing anaesthetic depth. In view of a possible application to the clinical setting, the DC_{index} categorical 'on-off' response may represent a useful feature during anaesthetic onset that supplements the more gradual trends of other DoA indexes, provided that it occurs in correspondence to the physiological transition from wakefulness to LOC. Even if our results suggest that the step transition occurs for a ESC level that has been associated with behavioural LOC in other studies, further work is required to confirm this.

Another advantage of connectivity based index is that it can reflect changes in brain patterns during anaesthesia. This may be better justified from clinicians than a single index that does not have a straightforward interpretation in terms of changes in brain state. However, there may not be a unique optimal index that effectively monitor the different components of GA (Marchant et al. 2014). GA is defined as the presence of hypnosis (unconsciousness), amnesia and immobility (in response to surgical stimulation) that are the results of anaesthetic actions at different sites. An ideal DoA monitor should specifically target all these behavioural end-points and it is difficult to evaluate which of them is the most relevant. In this view a solution may be to combine a variety of parameters that optimally correlate with the different anaesthetic goals and could give specific indication to clinicians. AEP and EMG entropy (Response Entropy-RE) may provide information about the level of anti-nociception (response to a surgical stimulation) (Marchant et al. 2014); on the other hand, given that primary effects of hypnosis are thought to occur at cortical level and that LOC is thought to be associated to a disruption of cortical networks, connectivity-based indexes may provide information about hypnosis, functioning as a marker of the 'consciousness-switch'.

Limitations The results of this thesis were obtained on a relative small sample (10 subjects for the NREM sleep study and 10 patients for the study on propofol anaesthesia), therefore they could also be specific to the modest sample considered and not generalize on a larger independent dataset or when data quality is worse. The rationale behind a relatively small cohort lies in the exploratory character of the study: given the novelty of the approach proposed to monitor anaesthesia, the experimental studies were designed to explore and generate hypothesis that could then possibly be robustly tested on a much larger cohort. Moreover the analysis of a limited number of subjects also allowed the investigation and presentation of individual trends, which were deemed important in view of a clinical application of the methodologies proposed.

Another critical point of the present and related work is the intrinsic problem of the absence of a gold standard of anaesthetic depth or level. The lack of a reference measure to compare connectivity (and other DoA indexes) with, makes it difficult to rigorously assess the validity of the proposed index. As discussed earlier in this paragraph and in chapter 6, since general anaesthesia is has different endpoints (hypnosis,amnesia, analgesia), then no single component of anaesthesia can be used to define overall depth. In order to assess the validity of the proposed index several test and studies should be performed. DoA monitors performances are usually (Bruhn et al. 2006) investigated with respect to the estimated or measured anaesthetics' concentration (as we did in the present work) in *correlational studies*. The ability of DoA monitors to prevent

awareness (or response to surgical stimulation or to verbal command) is usually assessed as well. In addition, large randomized clinical trials comparing DoA-based anaesthetic administration with routine care are usually carried out to investigate the ability of the proposed index to reduce drug administration, recovery times and comorbidities and assess the clinical utility of the DoA monitor. From these considerations it clearly emerges that the development and testing of the EEG connectivity index proposed here is at a very early stage. Further tests on independent and larger samples should be carried out to statistically test the performance of the proposed index and to address questions that this thesis leaves open:

1. *How do DC patterns correlate with the different anaesthetic endpoints?* We investigated how DC relates to changes in ESC (estimated concentrations of propofol in the brain), but not to changes in other anaesthetic effects such as patient behavioural responsiveness, memory formation or response to noxious stimulation.

2. *What are the DC features changes associated with recovery of consciousness? Does the DC perform equally in detecting anaesthesia induction and emerging from anaesthesia?* We investigated loss of consciousness in a slow induction from wakefulness to anaesthesia however we did not explore changes in DC during the opposite transition (recovery of consciousness). Results in the literature indicate that changes in connectivity during anesthesia induction and recovery do not mirror each other (Hudetz 2012) emphasizing the need to investigate the properties of DC connectivity during emergence from anaesthesia in future work.

3. *How does the DC_{index} perform in comparison to the commercial monitors in distinguish responsive from unresponsive patients?* We have compared the performances with MLR and BIS sub-parameters extracted from EEG recordings using our own implementations of algorithms, but not with the proprietary monitor values (aepEX ® and BIS®) and other commercial DoA indexes whose full algorithms are not publically available.

4. *Is the DC index clinically applicable for routine monitoring of anaesthesia?* Results indicate that in future the signal processing pipeline may be simplified (for instance removing the surrogate data generation procedure), however we have not systematically addressed this issue. A clinical applicable signal processing tool should also include a robust and automated artefacts rejection procedure, while in this work artefactual epochs were identified by visual inspection.

7.3 Future works

Considering the unresolved questions listed in the previous section, future research should develop in two directions: firstly a series of experimental studies should explore the properties of the DC_{index} on a larger sample (undergoing different anaesthetic protocols) and with respect to different anaesthetic endpoints. In parallel with the experimental work, different ways to simplify the signal processing procedure should be explored in order to test the applicability of the methodology in a clinical environment for routine monitoring. Some specific indications for future work are listed in the following.

7.3.1 Signal Analysis Methods

- Simplifying the signal processing procedure to estimate connectivity for a possible application in a clinical environment for routine monitoring by means of a

  (a) Reduction of the number of electrodes used to estimate DC features to reach an optimal compromise between monitoring efficacy and clinical applicability (including reduced computational times). To some extent, by focussing on long-range connections on the fronto-posterior axis we have already identified a limited number of features. The efficacy of those feature estimated from a reduced number of electrodes (i.e. 2 frontal, 2 parietal) in monitoring anaesthesia should be tested in the future.

  (b) Investigation of the effects of removing the computationally heavy surrogate data generation procedure to assess the significance of links and substituting it with an empirical threshold. Our preliminary results suggest that the 30% of strongest connections are in most case significant. Further work should assess if setting a threshold based on the percentage of strongest connections give robust results.

  (c) Design of an automated artefacts rejection procedure. The commercial DoA monitors have a sophisticated artifact rejection procedure that guarantees the quality of EEG collected in the noisy environment of the surgical theatre. These artifacts detection algorithms are proprietary and their description is typically vague (Bruhn et al. 2006). They usually have two main blocks: the first identifies specific artifacts as cardiac activity (EKG), pacemaker spikes, muscle activity and eye blinks. The second removes noisy epochs whose variance exceeds the average variance of EEG epochs previously processed. A similar signal processing tool that automatically rejects artifactual epochs before the estimation of connectivity should be developed in future work.

7.3.2 Experimental Work

- In order to assess the changes in DC features also during emergence from anaesthesia and performance of DC in correlating with the individual responsiveness an experimental protocol that includes induction of anaesthesia (LOC) and full recovery of consciousness (ROC) should be designed. The two opposite transitions LOC and ROC should be continuously monitored with EEG and other commercial monitors. The responsiveness of the subject could be assessed on a fine time scale presenting (i.e. every 10-20 s) a list of pre-recorded words (Purdon et al. 2013). The task could be for example to identify the stimulus type (i.e. name of persons or of cities) by pressing an appropriate button. In this way a curve of the individual responsiveness (Purdon et al. 2013) during the experiment would be available with a fine time resolution. Such a protocol would allow the investigation of changes in DC features in relation to probability of behavioural response (assessed across subjects) and individual responsiveness, and a comparison of DC performance with commercial monitors such as BIS® and aepEX ®.
- Some DoA monitors are 'blind' to some anaesthetics (i.e. their are not sensitive to their action) (Barr et al. 1999, Pandit & Cook 2014) or exhibit different trends depending on the drug used to induce GA (Olejarczyk et al. 2017). In order to investigate if the changes in DC features observed during propofol anaesthesia are agent specific, an assessment of EEG directed connectivity changes during anaesthesia induced with different agents (i.e. comparison intravenous and volatile anaesthetics) should be carried out.

In summary, as a continuation of this thesis, future work should test measures of brain connectivity with respect to another components of anaesthesia such as behavioural responsiveness. It should be assessed if connectivity features are able to robustly discriminate responsive from unresponsive patients and if they can be effectively implemented for routine anaesthesia monitoring. If proved to be efficient, this methodology may respond to the need of improving the commercially available DoA monitors in detecting unexpected awareness and have an impact on the development of more efficient DoA monitors and therefore on patient care and hospital costs.

Appendix A

Partial Coherence computation as a function of the inverse spectral matrix

In this section we will provide the rationale behind the use of the inverse spectral matrix to compute PCoh.

The pairwise Partial Coherence $PCoh$ between two signals $x_i(n)$ and $x_j(n)$ is defined as follows:

$$PCoh_{i,j}(f) = \frac{S_{ij|(X/ij)}(f)}{\sqrt{S_{ii|(X/ij)}(f)S_{jj|(X/ij)}(f)}} \tag{A.1}$$

where $S_{ij|(X/ij)}(f)$ is the partial cross-spectral density function of $x_i(n)$ and $x_j(n)$ and is defined as:

$$S_{ij|(X/ij)}(f) = S_{ij}(f) - \boldsymbol{S}_{i(X/ij)}(f)\boldsymbol{S}^{-1}_{(X/ij)(X/ij)}(f)\boldsymbol{S}_{(X/ij)j}(f) \tag{A.2}$$

and can be interpreted as the cross-spectrum between $x_i(n)$ and $x_j(n)$ once the linear effects of all the other time series of the process X have been removed, so to obtain a measure of the *direct* linear relations between the two signals neglecting the confounding effects of other components.

The PCoh can be calculated recursively in terms of lower order PCoh that in turn may be eventually computed as a function of the power spectra using equation A.2 (Bendat & Piersol 2000). An alternative approach, that is more computationally efficient for large PCoh orders, is to compute PCoh in one-step as a function of the inverse of the spectral density matrix $\boldsymbol{S}(f)$. The validity of this procedure has been demonstrated by

(Dahlhaus 2000) and will be outlined in this section.

Given the spectral density matrix $\boldsymbol{S}(f)$ and its inverse $\boldsymbol{G}(f) = \boldsymbol{S}^{-1}(f)$ we define a matrix $\boldsymbol{D}(f)$ to be

$$\boldsymbol{D}(f) = -\boldsymbol{H}(f)\boldsymbol{G}(f)\boldsymbol{H}(f) \tag{A.3}$$

where $\boldsymbol{H}(f)$ is a diagonal matrix whose elements are the square root of the inverse spectral matrix diagonal elements $g_{i,i}(f)^{-\frac{1}{2}}$, then the $PCoh$ between $x_i(n)$ and $x_j(n)$ is obtained as the negative value of the the ith and jth element of the rescaled inverse of the spectral matrix as follows:

$$PCoh_{i,j}(f) = -D_{i,j}(f) \tag{A.4}$$

$$\frac{S_{ij|(X/ij)}(f)}{\sqrt{S_{ii|(X/ij)}(f)S_{jj|(X/ij)}(f)}} = -\frac{g_{i,j}(f)}{\sqrt{g_{i,i}(f)g_{j,j}(f)}} \tag{A.5}$$

To demonstrate this theorem we consider $i = 1$ and $j = 2$ and the spectral density matrix in the following form, without loss of generality:

$$\boldsymbol{S}(f) = \begin{bmatrix} \boldsymbol{S}_{X1,X2}(f) & \boldsymbol{S}_{X1,Y}(f) \\ \boldsymbol{S}_{X2,Y}(f) & \boldsymbol{S}_{Y,Y}(f) \end{bmatrix} \tag{A.6}$$

with

$$\boldsymbol{S}_{X1,X2}(f) = \begin{bmatrix} \boldsymbol{S}_{1,1}(f) & \boldsymbol{S}_{1,2}(f) \\ \boldsymbol{S}_{2,1}(f) & \boldsymbol{S}_{2,2}(f) \end{bmatrix} \tag{A.7}$$

Direct verification gives that the inverse of the spectral matrix is:

$$\boldsymbol{G}(f) = \boldsymbol{S}^{-1}(f) = \begin{bmatrix} \boldsymbol{E}^{-1}(f) & -\boldsymbol{E}^{-1}(f)\boldsymbol{F}(f) \\ -\boldsymbol{G}(f)\boldsymbol{E}^{-1}(f) & \boldsymbol{S}_{Y,Y}^{-1}(f) + \boldsymbol{G}(f)\boldsymbol{E}^{-1}(f)\boldsymbol{F}(f) \end{bmatrix} \tag{A.8}$$

where

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{S}_{X1,X2}(f) - \boldsymbol{S}_{X1,Y}(f)\boldsymbol{S}_{Y,Y}^{-1}(f)\boldsymbol{S}_{X2,Y}(f) \\ \boldsymbol{F} &= \boldsymbol{S}_{X1,X2}(f)\boldsymbol{S}_{Y,Y}^{-1}(f) \\ \boldsymbol{F} &= \boldsymbol{S}_{Y,Y}^{-1}(f)\boldsymbol{S}_{X2,X1}(f) \end{aligned} \tag{A.9}$$

$\boldsymbol{E}$ is then a 2×2 matrix

$$\boldsymbol{E} = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} \tag{A.10}$$

whose rescaled inverse is

$$\boldsymbol{E}^{-1}_{rescaled} = \begin{bmatrix} 1 & \frac{-e_{12}}{(e_{11}e_{22})^{\frac{1}{2}}} \\ \frac{-e_{21}}{(e_{11}e_{22})^{\frac{1}{2}}} & 1 \end{bmatrix} \tag{A.11}$$

which, being the expression of $\boldsymbol{E}$ as given in equation A.9 proves the equality in equations A.4 and A.5.

Appendix B

The inverse EEG problem

The activity of a small region of the brain produces potentials that are spread out over the scalp, overlapping potentials generated by other sources. Hence the causal dependence between scalp sensors result both from correlation among cortical sources and volume conduction through brain, cerebrospinal fluid, skull and scalp (Sanei & Chambers 2007). This mixing effect can be reduced through the reconstruction of the putative electric sources in the brain, therefore through the solution of the so-called inverse problem.
The problem of finding the current sources inside the head is strongly ill-posed, i.e. for a given set of scalp measurements there are infinite configurations of intracranial sources compatible with it; only by introducing a priori constraints from assumptions about the source statistical distribution and volume conductor we can solve the inverse problem in a unique way (Michel et al. 2004) and the nature of these a priori assumptions strongly determines the quality of estimated data (i.e. if they actually give neurophysiologic information about signals generated in the brain).

Several approaches have been proposed. Methods assuming a certain number of cortical sources as generating the surface measurements are called *overdetermined* models since, in order to warrant a unique solution for the inverse problem, the number of fixed sources is less or equal to the number of scalp recordings. This approach is mainly used for AEP and epileptic foci's identification, where the number and position of sources can be set following physiological considerations. Considered that the exact number of dipoles cannot be determined a priori, a different approach has been developed. Distributed source models estimate the electric activity of the cortex in each point of a 3D grid of solution points. The number of points is much larger than the sensor point, which makes the inverse problem undetermined. In this thesis we have used this last

approach (chapter 4) using the algorithm implemented in the eConnectome toolbox. In the following models and methods employed will be described.

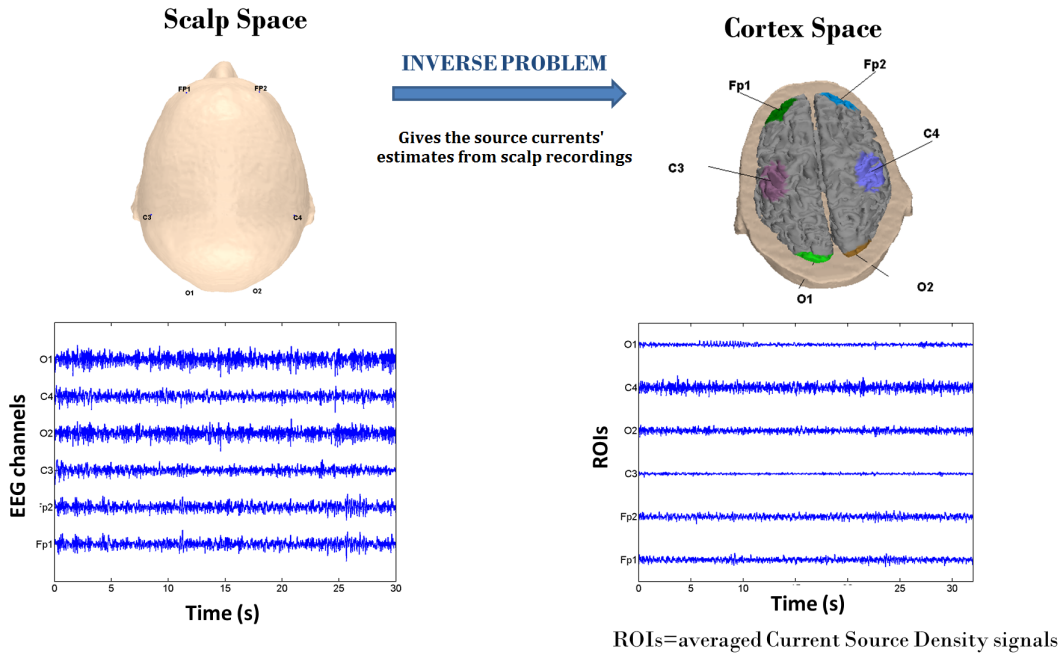

FIGURE B.1: Schematic representation of the inverse problem solution to infer the current source time series from the spatiotemporal profile of volume conducted scalp potentials. The electrodes considered for the causality analysis and the respective cortical ROI locations are shown, together with an example of the EEG traces and the respective ROIs waveforms. ROIs refer to the averaged current source densities underneath the corresponding electrodes.

B.1 The Inverse Model

To solve the problem of finding cortical sources from a scalp voltage distribution, we used an underdetermined, distributed source model: It does not need an a priori assumption on the number of sources and is based on the reconstruction of the brain electric activity in each point of a 3-D grid of solution points (much more numerous than the measured ones). The underdetermined nature of the problem necessitates some assumptions in order to identify the most likely solution. All the different models proposed in the literature differ in the choice of these assumptions, which can be purely mathematical or based on physiological and anatomical knowledge. In this work we used the cortical current density source model CCD (Dale & Sereno 1993) to solve the inverse problem: this method involves a linear approach with the integration of multiple constraints. Based on the assumption that for the range of frequencies typical of EEG analysis the electric and magnetic field can be well described by the quasi-static Maxwell equations,

a linear relationship between the electric recordings and the component of the cortical dipoles was considered.

$$x_i = \sum_{j=1}^{N} a_{ij} s_j + n_j \tag{B.1}$$

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{n} \tag{B.2}$$

where x_i is the potential at the i^{th} electrode, s_j is the strength of the j_{th} dipole component, N is the number of volume element in which the brain is divided and $\boldsymbol{M}$ represent the Lead Field Matrix (LFM). The lead field matrix is a non-linear function of the electrode locations, source locations and the permeability and conductivity of the head model (see chapter 4, section 4.2). The i_{th} row of $\boldsymbol{M}$ represent the how the potential of each electrode varies with the strength of each dipole, while the j_{th} column of $\boldsymbol{M}$ is the gain vector for the j_{th} dipole (i.e. how the potential at each electrode varies with the strength of the j_{th} dipole). $\boldsymbol{n}$ is a zero-mean random vector representing the additive noise at the sensors.

B.1.1 Minimum norm (MN) algorithm

The inverse problem can be solved if a priori information exists about the statistical distribution of the dipole moments and the sensor noise with algorithms that minimize the difference between the estimated and correct solution.

$$Err_w = |\boldsymbol{W}\boldsymbol{x} - \boldsymbol{s}|^2 \tag{B.3}$$

where $\boldsymbol{W}$ is the linear operator (demixing matrix) that maps the recorded potentials $\boldsymbol{x}$ in the estimated source vector $\hat{\boldsymbol{s}}$. Developing equation B.3 we obtain

$$Err_w = |\boldsymbol{W}(\boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}) - \boldsymbol{s}|^2 = |(\boldsymbol{W}\boldsymbol{A} - \boldsymbol{I})\boldsymbol{s} + \boldsymbol{W}\boldsymbol{s}|^2; \tag{B.4}$$

If the error Err_w is minimised by tacking the gradient and setting it to zero the optimal inverse linear operator is

$$\boldsymbol{W} = \boldsymbol{R}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{R}\boldsymbol{A}^T + \boldsymbol{C})^{-1} \tag{B.5}$$

where $\boldsymbol{R}$ and $\boldsymbol{C}$ are respectively the covariance matrix of the dipole strength and the sensor noise vector. If the dipole components are assumed to be independent and with same variance σ^2_{source} (i.e. if $R = \sigma^2_{source}\boldsymbol{I}$) and if the same assumption is made for the noise ($C = \sigma^2_{sensor}\boldsymbol{I}$, i.e. the noise is assumed to be spatially uniform across channel sites), this solution is equivalent to the minimum-norm solution (MN). This solution requires no a priori information but the only assumption that the current 3D distribution has the smallest L2-norm; therefore it favours weak and localised activation patterns (superficial sources).

To compensate the MN trend towards superficial sources different weighting strategies have been proposed. The approach implemented in eConnectome toolbox (and used for PDC source estimation in Astolfi et al. (2006), Astolfi, Cincotti, Mattia, Marciani, Baccala, de Vico Fallani, Salinari, Ursino, Zavaglia, Ding, Edgar, Miller, He & Babiloni (2007), Toppi et al. (2012) is the lead field weighting minimum norm (WMN) (Fuchs et al. 1999), based on the norm of the columns of the lead field matrix. Equations B.5 is still valid but with the variation that the metric of the sensor space is equal to the identity matrix $\boldsymbol{C} = \boldsymbol{I}$ and the covariance of the source space is given by the following metric:

$$\begin{aligned} R_{ij} &= |\boldsymbol{A}_i|^{-2}, i = j \\ R_{ij} &= 0, otherwise \end{aligned} \tag{B.6}$$

being $|\boldsymbol{A}_i|^2$ the norm of the i_{th} column of the lead field matrix as in Babiloni et al. (2005). In this way the variance of the sources is taken to be proportional to the inverse of the norm of the gain column and as a consequence the variance of deeper sources will be larger with respect to the superficial ones (deep-weighting).

B.1.2 Regularization

Ill-posed problems suffer to be highly sensitive to high-frequency perturbations and require the application of more sophisticated methods in order to compute a meaningful solution. This is the goal of the regularization methods (Hansen & Zaglia 1993). The dominating approach to regularize ill-posed problems and obtain a useful and stable solution is requiring the L2-norm of the solution to be small. The most common method

is the Tikhonov regularization whose idea is to obtain an optimal solution not only minimising the estimation error as in equation B.3 but a weighted combination of the residual norm and the solution norm

$$Err_s = |(\boldsymbol{As} - \boldsymbol{x})|^2 + \lambda^2 |\boldsymbol{s}|^2 \tag{B.7}$$

where λ is a regularization parameter introduced to account for the noise in the data and provides stability to the solution, such that small variations in the data do not lead to large variations in the source configuration. Therefore the optimal MN or WMN solution derived using Tikhonov regularization in the Regularization Toolbox (Hansen & Zaglia 1993) is the following

$$\boldsymbol{W} = \boldsymbol{RA}^T(\boldsymbol{ARA}^T + \lambda \boldsymbol{C})^{-1} \tag{B.8}$$

The optimal value for the regularization parameterλ is obtained following the L-curve approach (for a detailed description of this method see Hansen & Zaglia (1993)).

B.1.3 Head Model

As previously observed, the lead field matrix is a function of the electrical (i.e. conductivity, permeability) and geometrical (shape) properties of the volume conductor and the relative positions of sensors and sources; to obtain an accurate estimate of the lead field matrix is necessary

1. EEG-MRI co-registration: first is necessary to match the sensor positions to the scalp surface, in the same reference system of the cortical and subcortical sources. This is performed defining a wire-frame representation the cortical surface (and eventually subcortical structures) with non-invasive imaging techniques, typically MRI. eConnectome use one single template MRI (MNI brain from the Montreal Neurological Institute), and assume a standard 10-20 electrode coordinate system.

2. Head model: an accurate head model is necessary for a likely solution of the inverse model. The simplest one is a spherical model, with uniform conductivity properties. eConnectome uses a realistic head model with the help of the boundary element method (BEM[1], as described in Hämäläinen & Sarvas (1989)). In this work

[1]This method assumes that conductivity is isotropic and homogeneous in a tissue volume (brain, skull, skin) but considers 3 conductivity discontinuities at the boundaries of the volumes (surface brain-skull, skull-scalp, scalp-air)

the head is modelled as a multilayer structure, with 3 surfaces (scalp, skull, brain) separating volumes with different isotropic conductivities. The surfaces S1, S2, and S3 should take the form of the scalp-air, the skull-scalp, and the skull-brain interfaces, respectively. This information is acquired from high quality MRI scans of the MNI brain. The cortical surface obtained from the segmentation of the MRI images was triangulated and a high resolution cortical surface (downsampled from 41136 to 7850 triangles) formed the source space (the MRI scans are segmented into white and grey matter and sources are constrained to reside only on the grey matter volume). The skull and scalp surface were also segmented and reconstructed from the MNI brain and a scalp surface consisting of 2054 triangles formed the sensor space (Collins et al. 1994). Thus an high resolution LFM (2054x7850) relating the scalp triangles to the sources is pre-computed and used for the solution of the inverse problem.

The ambiguity of the inverse solution can be reduced introducing anatomical information, i.e. considering in the forward solution only those dipole locations consistent with anatomical and physiological information. As suggested in Dale & Sereno (1993), the EEG is mainly produced by currents in the apical dendrites of the cortical pyramidal cells, that have a columnar orientation, therefore the resulting dipolar moment is roughly oriented perpendicularly to the cortical surface. Hence the dipoles were constrained to the segmented grey matter, with their orientations perpendicular to the local cortical surface (triangle). In particular an orthogonal equivalent current dipole was placed in each node of the triangulated surface, with direction parallel to the vector sum of the normal to the surrounding triangles. The solution of the inverse problem is a vector of continuous time course for cortical sources strength, one for each dipole position, therefore the estimated source data matrix will have around 20000 time series. For further computation (i.e. connectivity or coherence estimation) is useful to reduce the computational cost, selecting regions of interest (ROI) source signals, whose waveform is computed by averaging the estimated cortical sources in the ROI.

Appendix C

Maximum Length Sequences (MLS) properties and deconvolution algorithm

In this section the properties of the MLS will be briefly described and the deconvolution algorithm for the extraction of evoked responses will be illustrated. Details of MLS generation and deconvolution have been published in Davis (1996) and the first audiological application (on ABR) of MLS was given by Eysholdt & Schreiner (1982). A MLS is a quasi-random binary sequence assuming values equal to +1 or -1. If n is the order of the sequence then its length is $L = 2^n - 1$ and the number of stimuli is $N_s = 2^{n-1}$. MLS have a number of attractive properties. The most important is that with the exception of a DC error, the autocorrelation is a perfect impulse:

$$\sum_{i=0}^{L-1} mls(i) = -1; \tag{C.1}$$

$$\sum_{i=-L}^{L} mls(i)mls(i+k) = \begin{cases} L & \text{if } k = 0 \\ -1 & \text{otherwise} \end{cases}; \tag{C.2}$$

As a result, a linear system with impulsive response h, when stimulated with an MLS sequence, produces a response y given by the convolution sum:

$$y(k) = \sum_{i=0}^{L-1} h(i)mls(k-i); \tag{C.3}$$

In the case of AEPs evoked using MLS, y represents the overlapped response, h is the 'true' response of the system (hence the AEP itself) and $mls(i)$ is one element of the stimulation sequence. If a MLS is used the original response h can be recovered from the overlapped averaged waveform by a deconvolution procedure.

Since the MLS signal is periodic of period L the expression in equation C.2 is a circular convolution between the periodic mls and the periodic system response h. More precisely the auditory system is stimulated by a MLS_S, where the subscript S stands for stimulation sequence. It differs from the original MLS (recovery sequence) from having 0 instead of -1 (where 0 represent the absence of stimulus) (see equation C.4 and figure C.1).

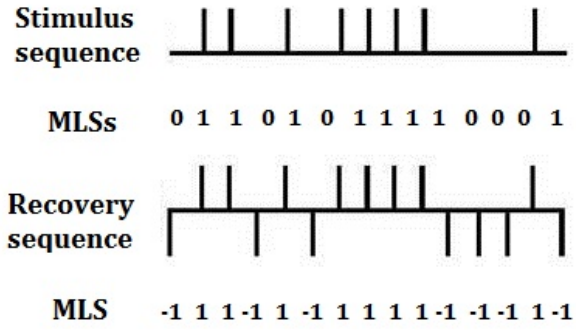

FIGURE C.1: An example of a MLS stimulation sequence and its respective recovery sequence.

$$mls_s(i) = \frac{mls(i)+1}{2}; \tag{C.4}$$

Hence the overlapped response is:

$$y(k) = \sum_{i=0}^{L-1} h(i) mls_s(k-i) = h(i) \otimes mls_s(i); \tag{C.5}$$

The deconvolution algorithm consist the circular convolution of the overlapped response with the temporal reverse of the original MLS:

$$y(k) \otimes mls(-k) = h(k) \otimes mls_s(k) \otimes mls(-k); \tag{C.6}$$

In fact, due to the properties of the MLS (see equation C.1), the following equivalence is true

$$mls_s(k) \otimes mls(-k) = \frac{L+1}{2}\delta(k); \tag{C.7}$$

therefore

$$y(k) \otimes mls(-k) = \frac{L+1}{2}h(k); \tag{C.8}$$

The AEP response is therefore equivalent to

$$h(k) = \frac{2}{L+1} sum_{i=0}^{L-1} y(i) mls(i-k); \tag{C.9}$$

or, in matrix form

$$\begin{bmatrix} h(0) \\ h(1) \\ \cdots \\ h(L-1) \end{bmatrix} = \begin{bmatrix} mls(0) & mls(1) & \cdots & mls(L-1) \\ mls(-1) & mls(0) & \cdots & mls(L-2) \\ \cdots & \cdots & \cdots & \cdots \\ mls(1) & mls(2) & \cdots & mls(0) \end{bmatrix} \times \begin{bmatrix} y(0) \\ y(1) \\ \cdots \\ y(L-1) \end{bmatrix} \tag{C.10}$$

The auditory evoked response is obtained multiplying the overlapped response by the matrix obtained left shifting the element of the original MLS sequence and dividing by the number of the stimuli occurred.

IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **38** (2017) 1802–1820 https://doi.org/10.1088/1361-6579/aa81b5

Directional connectivity in the EEG is able to discriminate wakefulness from NREM sleep

G Lioi[1], S L Bell[1], D C Smith[1,2] and D M Simpson[1]

[1] Institute for Sound and Vibration Research, University of Southampton, Southampton, United Kingdom

[2] Department of Anesthesia, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom

E-mail: gl4g13@soton.ac.uk and lioi.giulia@gmail.com

Received 31 January 2017, revised 27 June 2017
Accepted for publication 24 July 2017
Published 22 August 2017

Abstract

A reliable measure of consciousness is of great interest for various clinical applications including sleep studies and the assessment of depth of anaesthesia. A number of measures of consciousness based on the EEG have been proposed in the literature and tested in studies of dreamless sleep, general anaesthesia and disorders of consciousness. However, reliability has remained a persistent challenge. Despite considerable theoretical and experimental effort, the neural mechanisms underlying consciousness remain unclear, but connectivity between brain regions is thought to be disrupted, impairing information flow. *Objective*: The objective of the current work was to assess directional connectivity between brain regions using directed coherence and propose and assess an index that robustly reflects changes associated with non-REM sleep. *Approach*: We tested the performance on polysomnographic recordings from ten healthy subjects and compared directed coherence (and derived features) with more established measures calculated from EEG spectra. We compared the performance of the different indexes to discriminate the level of consciousness at group and individual level. *Main results*: At a group level all EEG measures could significantly discriminate NREM sleep from waking, but there was considerable individual variation. Across all individuals, normalized power, the strength of long-range connections and the direction of functional links strongly correlate with NREM sleep stages over the experimental timeline. At an individual level, of the EEG measures considered, the direction of functional links constitutes the most reliable index of the level of consciousness, highly correlating with the individual experimental time-line of sleep in all subjects. *Significance*: Directed coherence provides a promising new means of assessing level of consciousness, firmly based on current physiological understanding of consciousness.

Keywords: EEG, directed coherence, sleep, consciousness, functional connectivity

(Some figures may appear in colour only in the online journal)

1. Introduction

Neural correlates of consciousness have attracted considerable interest in recent years (De Graaf *et al* 2012). This has motivated a series of studies that contrast brain activity in wakefulness, where consciousness is typically present, with conditions where it is diminished or suppressed, such as in dreamless sleep (Massimini *et al* 2005, Spoormaker *et al* 2010, Siclari *et al* 2016) or anaesthesia (Ferrarelli *et al* 2010, Gómez *et al* 2013). In parallel, a series of theoretical advances have predicted that consciousness is critically related to functional connectivity that enables widespread information sharing among distant brain areas (Tononi 2008). A series of putative markers of consciousness have been proposed in the literature that include event related potentials evoked by auditory or visual stimuli (Sergent *et al* 2005, Sitt *et al* 2014), spectral patterns (Mhuircheartaigh *et al* 2013) gamma synchrony and measures of information sharing across distant brain areas (Tononi and Massimini 2008). A recent critical review (Koch *et al* 2016) has highlighted the importance of EEG as a fundamental clinical tool to discriminate conscious from unconscious subjects, however, it has suggested that only few of these measures are promising neurophysiologic correlates of consciousness. The low-frequency high amplitude EEG (usually referred to as slow-waves) that characterizes the loss of consciousness in physiological, drug-induced or pathological conditions, as opposite to high-frequency 'Activated-EEG' in wakefulness, remains one of the oldest and most reliable markers of awareness (Koch *et al* 2016). Other promising approaches appear to be measures of brain connectivity. These are thought to indicate the ability of the brain to integrate information (Tononi 2008).

Consciousness naturally fades during deep non-rapid eye movement (NREM) sleep, in particular in the early night (Tononi and Massimini 2008), when reports after awakening refer to little or absent conscious experience (Stickgold *et al* 2001); thus the onset of NREM sleep may represent an opportunity to relate changes in brain activity to changes in consciousness. During REM sleep, in contrast, reports of dream-like experiences are common. A nap generally includes only NREM sleep (N1, N2 and N3) with sleep N2 representing the largest proportion and the time spent in N3 increasing with the nap-length for naps longer than 10 min (Brooks and Lack 2006). Naps, as also used by Massimini *et al* (2005), therefore represent a convenient condition to investigate loss of consciousness at the onset of NREM sleep, with the advantage of only requiring a simple experimental setup that does not entail the practical challenges of overnight sleep recordings.

Changes in cortical connectivity associated with sleep have been widely investigated in functional magnetic resonance (fMRI) studies, but their relationship with consciousness remains unclear (Klimova 2014). Results point to a general impairment of functional connectivity in the thalamocortical system (Spoormaker *et al* 2010, 2012); in particular, long-range connectivity was shown to be affected by sleep (Tagliazucchi *et al* 2013) and connectivity networks in NREM sleep show increased local clustering when compared to wakefulness (Boly *et al* 2012). Recent studies investigating early NREM sleep with combined transcranial magnetic stimulation (TMS) and EEG approaches (Massimini *et al* 2005) show a break-down of large-scale connectivity in the sleeping brain.

EEG is often preferred in studies of brain connectivity from a practical point of view (Sitt *et al* 2014), as it can be applied relatively easily at the bed-side and at low cost, and EEG-based

systems can be used in routine clinical work in the home or ward, as well as in intensive care units or operating theatres where assessment of level of consciousness (LOC) may be carried out. The strength of frequency-dependent relationships between EEG channels in sleep have mainly been investigated in previous studies with conventional coherence (COH) estimates (Achermann and Borbély 1998, Corsi-Cabrera *et al* 2003) and Synchronicity (Ferri *et al* 2005, 2007). COH and Synchronicity are symmetrical measures unable to convey directional information, therefore losing some functional significance. A more advanced measure of functional connectivity, the directed coherence (DC) (Baccalá *et al* 1998) has been proposed to overcome this limitation, by using a model-based approach involving causal linear filters quantifying the interactions between channels. The DC estimator is consistent with the framework of Granger Causality and provides information about the strength, direction and spectral content of linear dependencies. It therefore has potential to give additional information about the direction of functional links in the brain compared to standard COH. It has been demonstrated that for Gaussian and quasi-Gaussian distributions (Hlaváčková-Schindler 2011), GC is equivalent to transfer entropy, but has the advantages of being simpler to understand and interpret and easier to apply, providing a straightforward decomposition in frequency (Barnett *et al* 2009). This property is of particular relevance for EEG applications, where specific brain rhythms are dominant in behavioural states such as sleep stages, or when performing cognitive tasks (Klimesch 1999).

The aim of the current study is therefore to describe changes in strength and direction of functional connectivity associated with NREM sleep using DC and to propose and test indexes of brain connectivity based on DC that could distinguish between states of consciousness. Performance of the proposed approaches will be compared to more established spectral measures and assessed in discriminating between NREM sleep and wakefulness. Since the overarching aim of any proposed measure is to assess the LOC in individual subjects, performance is assessed against each subject's own time-line through the sleep stages. The focus on individual variability is an important and distinctive feature of this work, since the majority of previous studies investigating correlates of consciousness (where consciousness is diminished or reduced) have focussed on the analysis of the average values across the cohort. This work is also original in investigating directional connectivity (using DC) on EEG, where previously the strength of connection, rather than the direction of information flow, was the focus. Our interest in individual variability is motivated by a potential clinical application: we hope to assess depth of anaesthesia in future work and so we are looking for an index that is computationally convenient and suitable for online monitoring of individuals and can show changes in individual subjects, not just statistical differences between groups. In this sense the spontaneous EEG is also preferable to fMRI and TMS approaches as it represents a low cost and easily implementable method with good temporal resolution.

2. Methods

2.1. Subjects, protocol and preprocessing

Sixteen healthy subjects participated in the sleep study. However only ten subjects (three females and seven males, aged between 22 and 30 years) underwent N1, N2 and N3 stages of NREM sleep and so were included in the analysis. The experiment was approved by the local ethics committee and following informed consent, and conformed to requirements of the Declaration of Helsinki. In order to exploit the circadian sleep drive, the experiment was performed in the afternoon after lunch. The subjects were asked to refrain from drinking coffee or tea on the day of the experiment. The subjects, lying with eyes closed on a reclining chair, were

invited to sleep. After they spontaneously woke up again, they were asked to rest with eyes closed (REST W) and then to perform mental arithmetic with eyes closed (ACTIVE W). EEG was collected using a 32 channel system with active electrodes (Biosemi BV, Amsterdam) placed according to the international 10–20 system. Additional electrodes were used to record the electrooculogram (EOG) and the chin electromyogram (EMG). Sleep stages were scored by visual inspection of contiguous epochs of 30 s according to the standard criteria (Iber *et al* 2007). Data were then downsampled to 250 Hz and digitally referenced with respect to the average of T7 and T8 channels (linked mastoid), as recommended for functional connectivity estimation (Kamiński and Blinowska 2014). The EEG time series were band pass filtered (1–45 Hz) and additionally notch filtered at the mains frequency using zero phase filters. Only continuous and artifact-free epochs were selected and included in the following analysis.

2.2. Multivariate connectivity estimation

The DC is obtained from multivariate autoregressive (MVAR) model parameters. A MVAR process describes each multi-channel EEG time series $\boldsymbol{x}$ as a sum of p previous samples from the set of M-signals (here the EEG channels), weighted by model coefficients, plus a noise component, as given in the following equation (1):

$$\boldsymbol{x}(\boldsymbol{n}) = \sum_{l=1}^{p} \boldsymbol{A}(l)\boldsymbol{x}(n-l) + \varepsilon(n), n = 1, \ldots .N \tag{1}$$

where $\boldsymbol{x}(n)$ is the M dimensional vector of the EEG channels time-series at time lag n, N is the number of samples in the signals, $\boldsymbol{A}(l)$ is the $M \times M$ coefficient matrix (weights) describing the linear interactions between channels at lag l, p is the model order, and $\varepsilon(n)$ is the vector of white innovations, with the non-singular residual covariance matrix:

$$\Sigma_{\varepsilon} = \left\{\sigma_{ij}^2\right\}. \tag{2}$$

σ_{ij}^2 is the cross-covariance between innovations signals $\varepsilon(n)$ for channels i and j.

The weights relate the present sample of one signal to the past of another (and itself) and capture the directed influence between signals that can be interpreted in the sense of the Granger Causality (Granger 1969). It should further be pointed out that equation (1) explicitly excludes instantaneous connections (with zero time-lag). By transformation into the frequency domain, the MVAR process is modelled as a filter with transfer matrix $\boldsymbol{H}(f)$ and white noise $\boldsymbol{E}(f)$ as an input:

$$\boldsymbol{X}(f) = \boldsymbol{A}(f)\boldsymbol{X}(f) + \boldsymbol{E}(f) = \boldsymbol{H}(f)\boldsymbol{E}(f). \tag{3}$$

In equation (3), $\boldsymbol{A}(f)$ is the Fourier Transform of the matrix of parameters and $\boldsymbol{H}(f) = [\boldsymbol{I} - \boldsymbol{A}(f)]^{-1}$ (where $\boldsymbol{I}$ is the identity matrix), which conveys information about the linear dependencies between signals and their spectral features. The DC from signal j to signal i of the M-variate dataset is defined as follows (Baccalá *et al* 1998):

$$DC_{i,j}(f) = \frac{\sigma_{jj} H_{i,j}(f)}{\sqrt{\sum_{m=1}^{M} \sigma_{mm}^2 |H_{i,m}(f)|^2}} \tag{4}$$

and, because of the normalization, it quantifies the linear coupling from x_j to x_i as compared to all the other contributions the signal x_i receives from other structures of the M-variate dataset. In particular it has been shown (Faes *et al* 2013) that the squared modulus of $DC_{i,j}(f)$ measures the normalized portion of the autospectrum of x_i at frequency f due to the signal

x_j (or transferred from x_j via the transfer function $H_{i,j}(f)$ to x_i). In other words, $|DC_{i,j}(f)|^2$ is a measure of the portion of the autospectrum of x_i at frequency f due to the signal x_j. Due to its relatively straightforward interpretation in term of spectral content, in this work we will use the squared modulus of DC to quantify functional connectivity. This differs from the related formulation of the Directed Transfer Function (DTF, Kaminiski and Blinowska (1991)); as DC includes the variance of the residuals σ_*, it brings the advantage of robustness against different signal scaling (Baccalà and Sameshima 2007). When all residuals variances are equivalent, the DC reduces to the DTF. One may view this as DTF reflecting the existence of (directional) connections, while DC also quantifies how these connections are used and it is interpretable in terms of signal power content.

When estimating an MVAR model of order p from a dataset, it is important that segments of EEG data of adequate length are collected to ensure that the number of samples is sufficient to accurately fit the model. Given a M-variate dataset, a minimum of M^2p data points is required for the model fitting, since there are M^2p parameters to estimate; however in practice a much higher number is recommended (typically 10 times the minimum number) for an accurate estimate (Schlogl and Supp 2006). In order to follow this recommendation and to reduce computational cost, which is always of concern in multivariate connectivity estimation, a reduced number of electrodes was considered for connectivity analysis, as did Toppi *et al* (2012) and Marinazzo *et al* (2014). A subset of $M = 12$ electrodes that are fairly evenly distributed across the scalp (Fp1, Fp2, F3, Fz, F4, C3, Cz, C4, P3, P4, O1, O2) was selected and connectivity was estimated for epochs of 60 s (i.e. 15000 samples in each channel) that were not interrupted by artefacts or high-level noise (according to visual inspection). Therefore only segments with two consecutive 30 s epochs of the same sleep stage were included in the study while isolated epochs were excluded from the following analysis.

When making inferences about EEG connectivity, only statistically significant estimates should be considered. In this study the significance of DC links was assessed using surrogate data based on a phase shuffling of the EEG signals, with 1000 repetitions that generated a set of surrogate data in which any temporal correlation between channels was removed, but autocorrelation (and thus the spectrum) of each signal was maintained. DC was then estimated from the surrogate dataset in order to obtain an empirical null distribution for each pair of signals at all frequencies. The significance of causal links was assessed comparing the estimated connectivity with the null distribution, setting the significance level at $p < 0.01$. Correction for multiple comparisons was performed using the false discovery rate (FDR) approximation for dependent measurements (Benjamini and Yekutieli 2001). Only links that were thus found to be statistically significant were included in the subsequent analysis and in the calculation of EEG indexes of connectivity.

A widespread practice in functional connectivity analysis is to threshold connectivity matrices to remove weak or spurious connections and retaining only a small percentage of the strongest connections (Sporns 2013). In this study the connectivity matrices were thresholded to retain either 10% or 30% of the strongest connections (as in Chennu *et al* (2014)) and then averaged in the four physiologically relevant frequency bands delta (δ) (1–4) Hz, theta (θ) (4–7) Hz, alpha (α) (8–13) Hz and beta (β) (13–25) Hz.

In order to specifically investigate changes in long-range connectivity, the DC links were subdivided with respect to the 3D Euclidian interelectrode distances, computed using default channels coordinates. Distance thresholds were set for differentiating between three groups of channel pairs in roughly the same proportions with respect to the total number of possible links: 35% of short-range links (interchannel distances below 10 cm), 32% of medium range links (between 10 and 14 cm) and 33% of long-range links (above 14 cm).

2.3. Computation of EEG indexes

For each 60 s epoch of different sleep stages and for each subject, a series of EEG measures were then extracted. We organized indexes into two classes: spectral measures (as commonly used in many previous studies) and connectivity based measures. The power spectral density (PSD) for each epoch and electrode was estimated using the Welch method (Hanning window 7.5 s long, 50% overlap); the power in each frequency band was calculated as the integral of the PSD within each frequency band. The spectral analysis was focused on δ, θ and α bands since previous studies on neural correlates of consciousness had reported major changes in these bands (Chennu *et al* 2014, Koch *et al* 2016). To allow for differences in power between EEG channels, we estimated the normalized power in these three frequency bands by the total power (1–45 Hz) in each time epoch and for each electrode. The normalized power is thought to be a more reliable estimator because it encompasses the individual variances in the absolute EEG power caused, for instance, by variations in electrodes impedances (Sitt *et al* 2014).

We assessed connectivity through indexes quantifying the strength of the connectivity networks and indexes estimating the direction of information flow. The rationale for the former is to be found in the large number of studies showing that long-range connectivity is significantly affected by the LOC. The rationale for the latter comes from published results showing a prevalence of frontal EEG activation in sleep as opposed to a strong posterior activation in wakefulness (Brown *et al* 2010), with indications that the direction of long-range connectivity may represent a prominent feature of sleep as compared to wakefulness (De Gennaro *et al* 2004). To this end we assessed the number of significant connections from centro-posterior (O1, O2, P3, P4, C3, C4, Cz) to anterior (Fp1, Fp2, F3, F4, Fz) electrodes and vice versa. We thus defined an index that quantifies the dominant direction of information flow on the front-posterior axis ($\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$) as the normalized differences of the number of links in the two opposite directions over the rostro-caudal axis:

$$\mathrm{Dir}_{\mathrm{P}->\mathrm{A}} = \frac{\sum_i \sum_j \mathrm{num}(\mathrm{DC}_{ij\mathrm{P}->\mathrm{A}}) - \sum_i \sum_j \mathrm{num}(\mathrm{DC}_{ij\mathrm{A}->\mathrm{P}})}{\sum_i \sum_j \mathrm{num}(\mathrm{DC}_{ij\mathrm{P}->\mathrm{A}}) + \sum_i \sum_j \mathrm{num}(\mathrm{DC}_{ij\mathrm{A}->\mathrm{P}})} \tag{5}$$

where the sums are taken over all pairwise connections between the posterior and anterior channels ($P \rightarrow A$) and anterior to posterior channels ($A \rightarrow P$), respectively and $num(\,\cdot\,) = 1$ when that connection is significant, and zero otherwise.

In the following, we will first present results at a group (cohort) level, showing differences in connectivity based measures between the different sleep stages and frequency bands, with associated statistical analysis (Friedman tests). The average topographic distribution of spectral parameters and connectivity measures is then assessed and the ability of these and derived indexes to distinguish between sleep stages at a group-level is tested. We then also present results at an individual subject level, and the most promising measures are correlated with the manual scoring of sleep stages (hypnogram) in each individual, by Spearman correlation using the indexes from consecutive 1 min segments. The presentation of results from individual subjects, in addition to the statistical analysis across the cohort, provides insight into the potential of the method in monitoring individual patients.

3. Results

3.1. Group analysis

According to standard criteria (Iber *et al* 2007), hypnograms were generated for each subject, using consecutive 1 min artefact free signal segments. The hypnograms represent the sleep

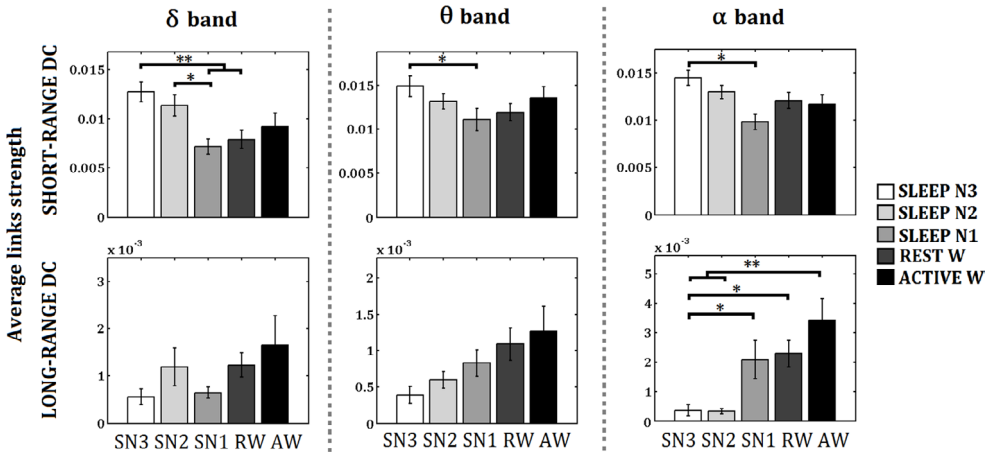

Figure 1. Short and long-range connectivity (as measured by DC) in δ, θ and α bands. Top row: short-range connections. Bottom row: long-range connections. Bars indicate average strength of 10% strongest DC (magnitude squared) links across subjects ($N = 10$). The error bars represent the within group standard error. The asterisks specify that the two means designated by the brackets significantly differ (Friedman test with post-hoc analysis, $*$ $p < 0.05$, $**$ $p < 0.01$). Short-range connectivity is dominant in NREM sleep (N2 + N3) while the strength of long-range links is reduced as compared to wakefulness, in particular in the α band.

stages visually identified from off-line scoring, and their assessment was carried out blinded to the connectivity analysis that follows. The following percentages of time spent in the sleep and wakefulness stages were obtained across subjects: sleep N3 21 $\pm$ 9%, sleep N2 27 $\pm$ 4%, sleep N1 16 $\pm$ 7%, REST W 19 $\pm$ 6%, and ACTIVE W 16 $\pm$ 4%. Since volunteers were allowed to spontaneously wake up from the post-prandial sleep, sleep duration was highly variable across subjects (16.2 $\pm$ 5.9 min of sleep were analysed per subject).

We calculated the average strength of DC links in the different experimental stages and then averaged this across subjects. Figure 1 shows DC average strength (across subjects and electrodes) as a function of the distance threshold and the sleep stages. The strength of connectivity links exhibit different trends across sleep stages depending on the distance range considered. Long-range connections are generally disrupted in NREM sleep where connectivity networks gain a more localized character (there is a prevalence of short-range links in sleep N2 and N3, in particular in the δ band) and long-range connectivity in the α band showed the best performance in distinguishing sleep from wakefulness. This is in accordance with previous works (Chennu *et al* 2014, Lioi *et al* 2016). In order to reduce the number of parameters investigated and for the sake of the clarity, we therefore only included indexes relative to the α band as markers of connectivity in the later results.

Figure 2 shows the group topographic characteristics of a number of features derived from the EEG (band power, strength of long range functional connections and the direction of functional connections) as a function of sleep stages, along with mean and standard error plots for those features. In order to reduce dimensionality and quantify the discriminative power of the different measures, we summarized the topographic information by averaging across electrodes and investigated whether the global indexes were able to discriminate NREM sleep from wakefulness (two level analysis) and also specific sleep stages (multilevel analysis). For the two level analysis a Mann–Whitney test explored whether the different markers in the two wakefulness stage (average of ACTIVE W and REST W) significantly differed from NREM

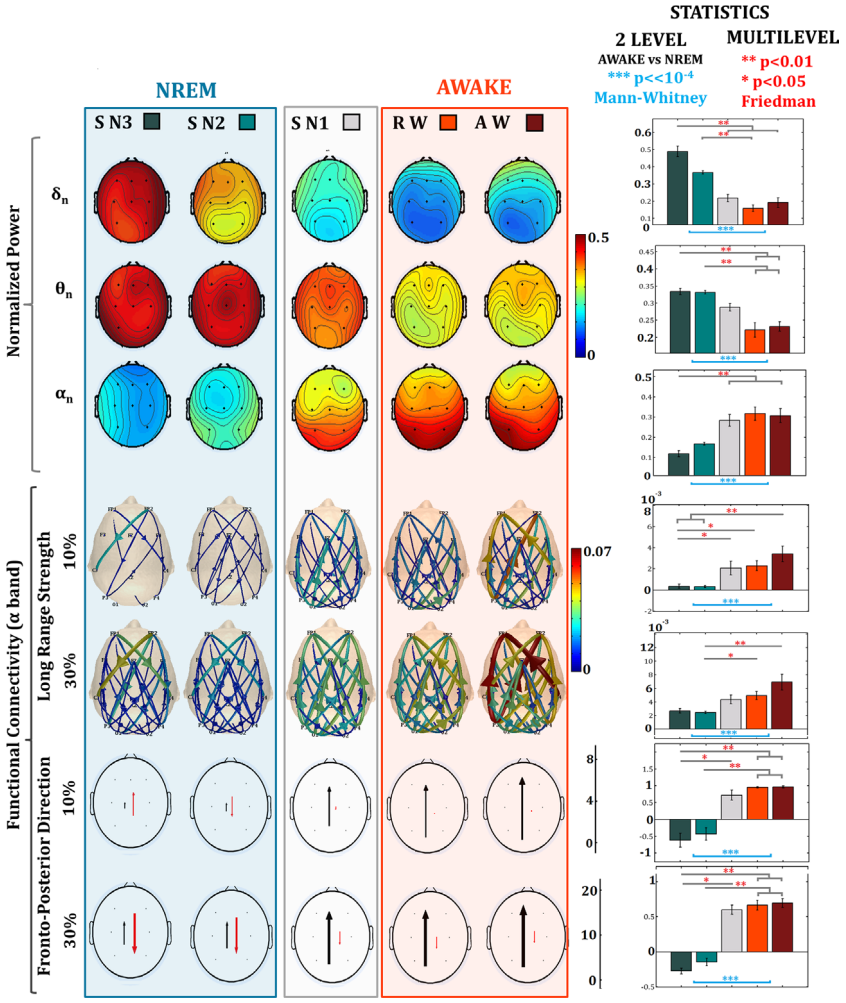

Figure 2. Scalp topography of the different EEG measures (rows), averaged across all 10 subjects, with associated statistics. In the left hand box, rows 1–3 show the normalized power distributions (δ_n, θ_n and α_n) across sleep stages. Rows 4 and 5 show long range connections with the grand average of the strength of long-range connection in the α band plotted for the 10% and 30% strongest connections respectively. Rows 6 and 7 indicate the average number of postero–anterior (black) and antero–posterior (red) connections in the α band coded by the length and thickness of the arrows in the bottom rows for 10% and 30% strongest connections respectively. Columns 1 to 5 indicate the experimental stage (from SN3 to AW respectively). On the right hand side, the last column indicates whether the indexes averaged across electrodes significantly discriminate wakefulness (REST W and ACTIVE W) from NREM sleep (N2 and N3) as assessed with a two level Mann–Whitney test (the p value is indicated in light blue) and shows results of a multifactor Friedman test on index averages across electrodes to assess significant differences across all the stages. The asterisks specify that the two means designated by the brackets significantly differ ($*$, $p < 0.05$; $**$, $p < 0.01$, $***$ $p < 10^{-4}$), as revealed by Tukey's HSD test (results showed for 10% strongest connections–fourth row, last column–are repeated from figure 1, to aid comparison) The connectivity scalp plots were obtained using the eConnectome imaging software (Bin *et al* 2011). Abbreviations: AW-ACTIVE W, RW-REST W, SN1, SN2 and SN3-NREM sleep stages.

sleep (average of N2 and N3). To test whether the measures could differentiate the specific sleep or wakefulness stages, a Friedman test was conducted, with post-hoc analysis using Tukey's honestly significant difference (HSD) test.

In figure 2 changes in normalized power can be seen that are consistent with those commonly reported in the literature: low frequency (δ and θ) power (rows 1 and 2 of figure 2) is dominant in NREM sleep and gradually decreases from SLEEP N3 to ACTIVE W. The two wakefulness states are characterized by a dominant occipital α rhythm (row 3 of figure 2). The power spectrum in all three bands significantly distinguished wakefulness from NREM sleep, as found when averaging the result of N2 and N3, and comparing these with the average of both stages of wakefulness (rest and active) ($p < 10^{-4}$, Mann–Whitney test). However, the normalized δ and θ power more efficiently discriminate NREM stages N2 and N3 from sleep N1 and wakefulness. The two-level analysis (sleep N2 and N3 versus awake) follows the approach used by Massimini *et al* (2010). Theoretical models of consciousness and experimental results obtained in sleep, anaesthesia and disorders of consciousness predict that the long-range information sharing is essential to maintain consciousness. In agreement with these findings we observed that the average strength of long-range connections in the α band gradually increases in the progression from deep sleep (N3) to ACTIVE W (rows 4 and 5 of figure 2). The difference between sleep and wakefulness is more marked if only a small percentage (10%) of strongest connections is included in the analysis (row 5 of figure 2). This index is able to significantly discriminate SLEEP N3 from SLEEP N1 and the two wakefulness stages ($p < 0.01$), and SLEEP N2 from active wakefulness ($p < 0.01$).

We also assessed changes in the direction of the information flow over the rostro-caudal axis. The most notable trends were observed in the α band (which also showed the greatest changes in connectivity strength as seen in figure 1) and plotted in the two bottom rows (6 and 7) of figure 2. The number of posterior to anterior links dramatically increases from NREM sleep to wakefulness and reaches its maximum in ACTIVE W. In contrast, the number of connection in the antero-posterior direction gradually shrinks in the progression from NREM to wakefulness. Thus the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ index, that summarizes the dominant direction of information flow, is particularly efficient in discriminating between stages, as revealed by the multilevel analysis. This provides evidence of a significant inversion of information flow in the α band from frontal to posterior versus posterior to frontal in the progression from sleep to wakefulness.

3.2. Individual analysis

A clinically useful index of consciousness needs to distinguish between sleep stages at an individual and not only at the group level. In order to investigate whether the changes observed were both consistent at the individual level and able to correlate with the experimental stage across the individual sleep-wake cycle, we show (figure 3) the different indexes for each subject and epoch of the experimental time-line. This epoch by epoch analysis allowed comparison with the individual hypnograms.

Among the connectivity measures, we have plotted the indexes that in the group analysis showed best discriminatory performances: the average strength of the 10% strongest long-range links and the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ computed from the 30% strongest connectivity links. Given that the shift in EEG power toward lower frequencies is a well-known and prominent feature of NREM sleep (sleep δ waves) and that the δ and θ power showed similar discriminative properties at group level, we have only plotted the normalized δ power from the spectral measures (the plots were inverted to facilitate the comparison with the hypnogram and the connectivity derived indexes). The experimental stages were assigned a value as a measure of LOC,

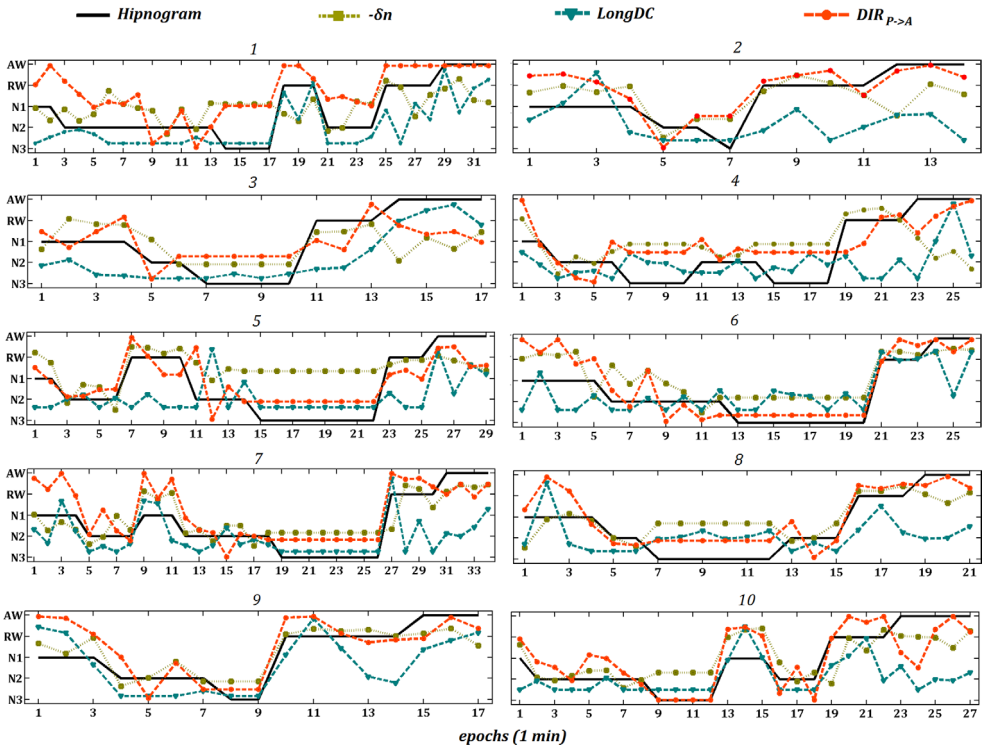

Figure 3. Individual trends over the experimental timeline. Each epoch is 60 s in duration. For each of the 10 subjects the amplitude (magnitude squared) of significant long-range DC links (blue dashed line, triangle marker), the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ index (orange dashed line, circle marker) in the α band and the power in the δ band (green dotted line, square marker) are plotted and can be compared to the manually scored hypnogram (solid black line). Note that plots are rescaled/inverted to facilitate comparison with the hypnogram. Abbreviations: AW-ACTIVE W, RW-REST W, N1, N2 and N3-NREM sleep stages. For ease of visualization, all plots were rescaled, and delta power was inverted.

ranging from 0 (SLEEP N3) to 4 (ACTIVE W), and Spearman correlation was computed between each parameter and the individual hypnogram. Table 1 shows the resulting correlation values for individual subjects.

All the indexes considered exhibit dramatic changes as a function of the sleep stages at the individual level. As shown in figure 3, the normalized δ power follows the experimental timeline in the majority of subjects. However in two subjects (2 and 4) the changes in δ power do not track the hypnogram and do not significantly correlate with the LOC (table 1). Similar results are obtained for the θ power, with a significant negative correlation with the experimental time-line only in 7 of the 10 subjects. The average strength of long-range connections is severely reduced in NREM sleep at an individual level. It shows performances similar to the normalized power indexes, highly correlating with the LOC in the majority of subjects, but failing to do so in three of them. Of all the parameters considered, the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ showed the best performances in 'tracking' the individual hypnogram, with a high and significant correlation in each of the subjects and the highest mean correlation value. It also is able to significantly discriminate between NREM sleep and wakefulness at an individual level in all the subjects (table 1, last column), as assessed by a Wilcoxon test across the epochs of each stage.

Table 1. Spearman correlation coefficient computed between the hypnogram and each EEG index for individual subjects. Values in bold correspond to a significance level of $p < 0.01$, values in bold-italic to a significance level $p < 0.05$; other values are not significant. The last column shows the percentage of subjects where the specific EEG index was able to significantly discriminate wakefulness (rest and active) from NREM sleep (N2 and N3) at individual level, as assessed by a Wilcoxon signed rank test ($p < 0.05$).

	Subject											% Subjs
	01	02	03	04	05	06	07	08	09	10	Mean ± Std	$p < 0.05$ 2 level
Norm power δ	***−0.36***	−0.44	***−0.52***	0	**−0.67**	**−0.89**	**−0.79**	**−0.64**	**−0.83**	**−0.63**	−0.58 ± 0.26	100%
Norm power θ	**−0.54**	*−0.15*	**−0.51**	*−0.38*	*−0.50*	**−0.72**	**−0.76**	**−0.87**	***−0.55***	**−0.80**	−0.58 ± 0.2	90%
Norm power α	0.34	0.2	0.13	0.02	0.1	***0.41***	**0.45**	0.22	***0.58***	**0.51**	0.29 ± 0.18	80%
Long DC α (10%)	**0.77**	0.27	**0.84**	0.08	**0.49**	***0.48***	**0.69**	0.16	**0.72**	**0.63**	0.51 ± 0.25	70%
Long DC α (30%)	**0.83**	*0.21*	**0.75**	**0.51**	*−0.01*	**0.63**	***0.38***	*−0.23*	**0.75**	***0.46***	0.43 ± 0.34	70%
$Dir_{P\rightarrow A}$ (10%)	**0.56**	**0.73**	−0.34	**0.87**	**0.86**	**0.67**	**0.78**	**0.77**	**0.85**	**0.89**	0.66 ± 0.37	90%
$Dir_{P\rightarrow A}$ (30%)	**0.82**	**0.74**	**0.68**	**0.70**	**0.85**	**0.84**	**0.82**	**0.79**	**0.68**	**0.80**	0.77 ± 0.06	100%

4. Discussion

The objective of the current work was to propose and test indexes of brain connectivity that could distinguish between states of consciousness. These indexes should also be computationally relatively simple for clinical application such as in sleep studies or depth of anaesthesia monitoring. Given the conceptual model that the loss of consciousness is associated with a loss in the brain's ability to integrate information (Tononi 2008), we focussed on connectivity measures that include directional information. In assessing the performance of these measures, we compared them to the more established power spectral indexes taken from the EEG.

To this end we collected polysomnographic recordings from a sample of 10 healthy subject undergoing post-prandial sleep and extracted the EEG indexes across the sleep-wake cycle at individual and group level. Those indexes were chosen exploiting previous theoretical and experimental findings on neural correlates of consciousness and included normalized power and connectivity based indexes. We found that the proposed amplitude of long-range connections across the scalp and especially the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ index (that quantifies the dominant direction of information flow in the rostro-caudal axis) showed a monotonic change with the LOC. In the current sample, the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ index showed the best performances in tracking the individual experimental time line, and consistently correlated well with the hypnogram, and significantly discriminated NREM sleep from wakefulness in each of the subjects in the sample. Its performance was found to be superior (Mann–Whitney test, $p = 0.041$) to that of the power in the δ frequency band, which has been widely used in the past.

4.1. Methodological considerations

In this study an advanced method (DC) for the estimation of functional connectivity that is able to infer directed causal information was chosen. The direction of connectivity networks has not previously been well explored in finding indexes of LOC, even though it has been suggested that it could provide important insights into neural correlates of consciousness (Sitt *et al* 2014). Functional connectivity estimators based on an MVAR model have been shown to be robust to noise (Blinowska 2011) and to perform well even in the case of some nonlinear interactions (Winterhalder *et al* 2005), and have also been widely applied for estimating functional connectivity from multichannel EEG in different experimental conditions other than sleep (Astolfi *et al* 2008, Blinowska 2011). In this context the term 'causality' has been used to refer to Granger causality, i.e. indicating that one signal predicts another in the MVAR model of simultaneously observed signals. We disregarded more complex measures of functional connectivity such as those based on information theory (e.g. transfer entropy) in order to achieve the necessary computational simplicity and temporal resolution required for on-line monitoring (Barnett *et al* 2009).

Despite DC having become well established, caution is required in the interpretation of its results as necessarily indicating causal links connecting underlying cortical sources. Volume conduction effects lead to spreading of electrical activity to a number of electrodes, which could be confused with functional (neurological) connectivity between these brain regions. One distinguishing feature of volume conduction effects is that it is virtually instantaneous, without the delay typical of neuronal activity. The supposition that DC is not affected by volume conduction because it is sensitive only to phase differences between channels (Kamiński and Blinowska 2014), is still debated (Faes and Nollo 2011). Our own studies (Lioi *et al* 2016), as well as those of Faes and Nollo (2010, 2011), Faes *et al* (2013), have suggested that while DC and DTF do not eliminate volume conduction effects, they do reduce them,

when compared to COH. That being said, DC networks estimated during motor (Ginter *et al* 2001), attention and memory tasks (Kuś *et al* 2008, Brzezicka *et al* 2011) show a remarkable agreement with evidence obtained from anatomical considerations and other neuroimaging techniques (Kamiński and Blinowska 2014). In this study, volume conduction effects are less of an issue since the analysis was focused on long-range connectivity, whereas short-range connections (which are likely to be more strongly dominated by the spread due to electrical conduction in tissue and bone) were disregarded. Non-significant connectivity was also removed from the study using surrogate data analysis. Caution must also be exercised in interpreting DC as reflecting direct connections between cortical regions because the method cannot remove the confounding effect of sources (typically deep in the brain) that spread activity to pairs of electrodes on the scalp. From the signals recorded on the scalp it may be impossible to determine whether there are direct neuronal pathways between the corresponding two cortical regions or if both are driven by another (unmeasured) source.

In this exploratory study of connectivity measures, it was important to test the statistical significance of each DC connectivity estimate, which was achieved with surrogate data analysis. This precaution has not always been taken in previous work. However the results (not shown) indicated that the 30% strongest DC connections calculated from 1 min segments were almost always statistically significant (i.e. DC was larger than obtained with the surrogate data under the null hypothesis of no connections). This suggests that in future work the computationally costly surrogate data analysis may not be required, when using the proposed indexes. This would make real-time implementation computationally feasible. The surrogate data generation takes the 95% of the total computation time in estimating connectivity. To give an indication of computational cost, if we exclude the shuffling procedure, the time required to estimate connectivity for a 1 min recording with 12 electrodes and 1 min epochs, is roughly 20 s, when using Matlab® and a typical Windows-based PC.

In the analysis of brain connectivity, the many relatively weak links can obscure important connections. Thresholds are thus usually recommended and applied (Rubinov and Sporns 2010) to only select connections deemed to be important. The choice of the threshold is somewhat arbitrary (Sporns 2013), but statistical significance (i.e. a threshold set at the critical value) should always be satisfied. In the current work either the 10% or the 30% strongest connections were analysed, following some preliminary investigations. The choice of threshold reflects a compromise between including too many connections that may only be weakly related to the LOC and discarding connections that might hold useful information. Further work in optimizing the threshold for specific tasks such as assessing depth of anaesthesia or sleep stage should be carried out on a larger sample of recordings.

4.2. The relationship of our findings to previous studies

To the best of our knowledge there is only one previous study investigating EEG networks using directional measures of connectivity, but this used the closely related approach of DTF (see Methods section) in sleep (Kamiński *et al* 1997). DC, which includes information on signal power flowing between different regions of the brain, as well as the transfer function of the linear 'filters' linking these regions (as used in DTF), would seem to be more appropriate than DTF for quantifying the functional connections between brain regions (Baccalà and Sameshima 2007, Faes and Nollo 2011). The current paper thus goes beyond previous work in describing DC patterns during the change from wakefulness to NREM sleep, extending the analysis to different frequency bands and refining the methods with rigorous statistical significance assessment of the estimated links. Furthermore, we differentiate connectivity links

with respect to their inter-electrode distance and we provide an assessment of performance in individual subjects, as well as that at the cohort-level. Our findings are in line with results from the study of Kamiński *et al* (1997) that showed more complex and denser connections in wakefulness than in NREM sleep and a prevalence of posterior sources during wakefulness. Despite the strong topological similarities, in the study of Kamiński, connectivity networks exhibited a notably larger number of connections, possibly as a result of not performing a significance test (or using a different choice of threshold criterion) for including the estimated DTF in the final analysis, and also because connectivity was integrated over a larger frequency range (0–30 Hz) than in the current study.

The underlying conjecture of our as well as a number of other studies (Kamiński *et al* 1997, John and Prichep 2005, Mashour 2006, Tononi 2008) is that changes in LOC are critically associated with a dynamic reorganisation of large-scale connectivity patterns. In line with previous results in disorders of consciousness (Chennu *et al* 2014) the connectivity networks in the α band showed the best discriminative performance between sleep stages. Long-range connectivity in the α band was shown to be impaired in NREM sleep, when networks are active but characterized mainly by short-range links. Our results thus support previous studies using different methods and protocols suggesting that the disruption of wakefulness in NREM sleep may be a result of impaired information sharing among cortical areas. Massimini and colleagues (Massimini *et al* 2005, 2010) used a pertubational approach involving transcranical magnetic stimulation (TMS) and EEG recording to investigate how TMS triggered neural activity spread from the stimulation site. During wakefulness the TMS elicited complex patterns of scalp waves spreading to distant cortical areas. During NREM sleep, TMS evoked a stereotypical and local response, thus indicating a 'breakdown of long-range effective connectivity'. The findings of Massimini *et al* (2010) and our data suggesting that long range connections reduce during sleep also fit well into a wide range of evidence from a growing literature investigating fMRI brain connectivity in altered states of consciousness such as NREM sleep (Spoormaker *et al* 2010, 2012), general anaesthesia (Boly *et al* 2011, Schrouff *et al* 2011) and vegetative states (Boly and Seth 2012, King *et al* 2013). It appears that those states share, among other major features, a suppression of functional connectivity.

In SLEEP N1 all the EEG indexes showed values intermediate between deeper sleep (N2 and N3) and wakefulness. Often spectral and networks features elicited in SLEEP N1 were more similar to wakefulness than to NREM sleep. SLEEP N1 represents the transition between wakefulness and sleep and it is considered 'unstable sleep' (Klimova 2014). When awakened from SLEEP N1 subjects often report dream-like experiences or claim they were awake (Nir *et al* 2013). Experimental results suggest preserved long-range connectivity in this stage (Massimini *et al* 2005). For this reason we have considered only sleep N2 and N3 trials to characterize stable NREM sleep in the two level analysis, as used in Massimini *et al* (2010).

The findings of recent seminal works in Network Physiology (Bashan *et al* 2012, Bartsch *et al* 2015 and Liu *et al* 2015) broadly align with our study. Network physiology is a new research field that aims to characterize how the various physiological systems dynamically integrate their functions in different physiological (and pathological) states (Bashan *et al* 2012, Ivanov *et al* 2016). A series of studies investigating changes in networks of interactions between (Bartsch *et al* 2015) and within (Liu *et al* 2015) different physiological systems (muscular, cardiac, respiratory and central nervous systems) during the transition from wakefulness to sleep and across sleep stages have provided important progress in this emergent field, contributing to the realization of an atlas of global network physiology in sleep and wakefulness. In these studies network connectivity was assessed using an estimator based on the stability of the time delay between signals (time delay stability -TDS), which quantifies in fine temporal detail the undirected strength of coupling. Despite the differences in approach, EEG networks

estimated from six channels and in different frequency bands in Liu *et al* (2015) and Bartsch *et al* (2015), showed important similarities with those shown in our study for the α band: they observed a significant decrease in the strength and number of links (in particular of long-range fronto-occipital connections) in deep sleep (as compared to wakefulness and light sleep) and a remarkable symmetry between the two hemispheres, characterizing all physiological states. Another notable analogy with our results regards EEG networks elicited in sleep N1, whose features are more similar to wakefulness than to deeper sleep. As Liu and colleagues have observed, this is an interesting result, given that sleep N1 is commonly classified as belonging to the same macro state (NREM sleep) as sleep N2 and N3. In their analysis of brain networks Liu and colleagues have also reported that while local connections (frontal/frontal, central/central and occipital/occipital) are reduced but preserved in deep sleep, fronto-occipital and occipital-frontal networks show practically no connection in deep sleep: these findings agree clearly with the significant impairment of long-range connectivity we observed in sleep N2 and N3.

Our study provides important new contributions beyond this work (and previous studies) in the characterization of brain networks by assessing the direction of links and the consequent ability to identify a switch in the direction of information flow with sleep onset, that constitutes the most characteristic change in DC patterns. Group analysis reveals a significant inversion of the direction of posterio-frontal networks with state. The marked posterior to anterior spread of α rhythm in wakefulness is reversed in NREM sleep (N2 and N3) that is characterized by a dominance of frontal sources of activity. An inversion of information flow from frontal-posterior in sleep to posterior-frontal in wakefulness has also been found in a previous analysis of sleep and general anaesthesia onset (De Gennaro *et al* 2004, Nicolaou *et al* 2012).

The importance of quantifying, together with the strength, the direction of links has been highlighted in a recent work where the TDS estimator originally proposed in the framework of network physiology has been extended to the concept of delay-correlation landscape (DCL) (Lin *et al* 2016). While the TDS is computed considering the delay corresponding to the maximum (absolute) correlation between signals, the DCL estimator retains information about the delay dependence of the cross-correlation, which carries important information about the directionality of physiologic interactions. The DCL approach has been specifically designed to quantify interactions between the outputs of different physiological systems and it was shown to efficiently grasp the directed correlation between EEG power in different bands and the cardiac signal (Lin *et al* 2016). If, however, the aim of the analysis is to investigate interactions between signals of the same subsystem (brain-brain in this case), DC may be more efficient in capturing the direction and strength of links as it summarizes the different delays, dependences using a simple metric (direction of links) that can be synthetically represented with a directed arrow. On the other hand, in the case of DCL the directional information is coded in a more complex and rich 'landscape' of different delays for positive and negative correlations. While conveying important information about the type of dependence (negative or positive) at different time lags, the DCL estimator is probably more difficult to interpret and present in a succinct manner. In addition, in the case of EEG signals, correlation (including DCL, as well as the coherence in the frequency domain) is likely to be more heavily affected by confounding volume conduction effects than the DC (Faes *et al* 2013).

In view of an application in the framework of Network Physiology, DC presents important features that are supplementary to the proposed measures based on time-delays. In the case of EEG signals, the two estimators measure different aspects of the interaction between signals. While TDS quantifies connectivity (as stable time delay) between EEG spectral power (brain rhythms) in different frequency bands, the DC quantifies linear connectivity (Granger causality) between channels of the original EEG time series and then transforms it into the frequency

domain. The application of DC may be therefore promising for the analysis of sub-system connectivity and provide new insights related to the network physiology atlas in physiological and pathological states. DC is based on linear modelling, and may thus outperform alternatives when this assumption is (approximately) valid, but is likely to underperform when the assumption is grossly violated. Further work is required to test the application of the MVAR approach to model interaction between signals on different time scales and from different physiological systems.

Another important original contribution of this work is the investigation, and presentation, of patterns for individual subjects, in addition to aggregated cohort results. The majority of studies investigating EEG markers of LOC have only performed group analysis or found the disagreement between behavioural and EEG-based measures too high for a reliable individual assessment (Sitt *et al* 2014). Our long-term aim is to identify an EEG index able to 'track' the consciousness level of the subjects, with a view to assess depth of sleep or anaesthesia. The performance of the proposed measures in individuals provides an indication of the potential of the proposed approach.

Among all the markers considered in the current paper, the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ index was found to be the most reliable in tracking the LOC during individual sleep experiments, strongly correlating with the hypnogram in all the subjects. It should be noted that the staging of the EEG time series was performed taking in account the proportion of δ waves, as recommended in standard sleep staging criteria, therefore high correlation between measured δ power and sleep stages is only to be expected. Given that EEG spectral features were used to characterize sleep (Corsi-Cabrera *et al* 2003) and to monitor hypnotic level in anaesthesia (Myles *et al* 2004), it is interesting that the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ index, which reflects very different features of the EEG signals, showed better performance than the δ power in correlating with the sleep stage at an individual level. These results suggest that the inversion of information flow represents a promising indicator of the descent into deep sleep. We speculate that this may also be seen in other states of altered consciousness, such as anaesthesia. In order to arrive at an even more powerful index of LOC, it may be beneficial to combine power-spectral measures with those obtained from DC and this approach is currently being pursued.

4.3. Limitations

The current work was carried out on a relative small sample (10 subjects) recorded during a nap, not all-night sleep. Thus only between 13 and 31 1 min non-artifactual segments were available from each subject in this cohort (see figure 3). While the results showed the power of the $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ index, this could be specific to this small sample. The current work should thus be considered as exploratory, in which the hypothesis of superior performance of $\mathrm{Dir}_{\mathrm{P}\rightarrow\mathrm{A}}$ was generated, but cannot be robustly tested, on the same small sample. Further tests on an independent sample, that might include a wider range of indexes as well as their combination should be carried out to independently statistically test relative performance measures.

Conclusions

We have assessed the performances of different EEG indexes to predict LOC during a NREM sleep study. Our results show that the EEG signal includes many features able to discriminate NREM sleep from wakefulness at a group level, but correlation with individual hypnograms varied across subjects. In agreement with theoretical consideration, EEG indexes relying on directional connectivity assessment have proven to be particularly promising. Among all of the EEG measures tested, a proposed index of the direction of information flow on the

rostro-caudal axis that is based on DC performed well at a group level and gave the highest correlation with individuals' LOC.

Acknowledgments

This work was supported by the National Institute of Academic Anaesthesia (UK) and the Kerkut Trust (University of Southampton, UK).

ORCID iDs

G Lioi https://orcid.org/0000-0001-8231-0886

References

Achermann P and Borbély A 1998 Coherence analysis of the human sleep electroencephalogram *Neuroscience* **85** 1195–208

Astolfi L *et al* 2008 Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators *IEEE Trans. Bio-Med. Eng.* **55** 902–13

Baccalà L A and Sameshima K 2007 Generalized partial directed coherence *Proc. 2007 15th Int. Conf. on Digital Signal Processing* pp 162–6

Baccalá L A, Sameshima K, Ballester G, Do Valle A C and Timo-Iaria C 1998 Studying the interaction between brain structures via directed coherence and granger causality *Appl. Signal Process.* **5** 40–8

Barnett L, Barrett A B and Seth A K 2009 Granger causality and transfer entropy are equivalent for Gaussian variables *Phys. Rev. Lett.* **103** 1–10

Bartsch R P, Liu K K L, Bashan A and Ivanov P C 2015 Network physiology: how organ systems dynamically interact *PLoS One* **10** e0142143

Bashan A, Bartsch R P, Kantelhardt J W, Havlin S and Ivanov P C 2012 Network physiology reveals relations between network topology and physiological function *Nat. Commun.* **3** 702–9

Benjamini Y and Yekutieli D 2001 The control of the false discovery rate in multiple testing under depencency *Ann. Stat.* **29** 1165–88

Bin H, Astolfi L and Babiloni F 2011 eConnectome: a MATLAB toolbox for mapping and imaging brain functional connectivity *J. Neurosci. Methods* **195** 261–9

Blinowska K J 2011 Review of the methods of determination of directed connectivity from multichannel data *Med. Biol. Eng. & Comput.* **49** 521–9

Boly M, Garrido M I, Gosseries O, Bruno M-A, Boveroux P, Schnakers C, Massimini M, Litvak V, Laureys S and Friston K 2011 Preserved feedforward but impaired top-down processes in the vegetative state *Science* **332** 858–62

Boly M, Perlbarg V, Marrelec G, Schabus M, Laureys S, Doyon J, Pelegrini-Issac M, Maquet P and Benali H 2012 Hierarchical clustering of brain activity during human nonrapid eye movement sleep *Proc. Natl. Acad. Sci.* **109** 5856–61

Boly M and Seth A K 2012 Modes and models in disorders of consciousness science *Arch. Ital. Biol.* **150** 172–84

Brooks A and Lack L 2006 A brief afternoon nap following nocturnal sleep restriction: which nap duration is most recuperative? *Sleep* **29** 831–40

Brown E N, Lydic R and Shiff N D 2010 General anesthesia, sleep and coma *New Engl. J. Med.* **363** 2638–50

Brzezicka A *et al* 2011 Information transfer during a transitive reasoning task *Brain Topography* **24** 1–8

Chennu S *et al* 2014 Spectral signatures of reorganised brain networks in disorders of consciousness *PLoS Comput. Biol.* **10** e1003887

Corsi-Cabrera M, Miró E, Del-Río-Portilla Y, Pérez-Garci E, Villanueva Y and Guevara M A 2003 Rapid eye movement sleep dreaming is characterized by uncoupled EEG activity between frontal and perceptual cortical regions *Brain Cognit.* **51** 337–45

De Gennaro L, Vecchio F, Ferrara M, Curcio G, Rossini P M and Babiloni C 2004 Changes in fronto-posterior functional coupling at sleep onset in humans *J. Sleep Res.* **13** 209–17

De Graaf T A, Hsieh P J and Sack A T 2012 The 'correlates' in neural correlates of consciousness *Neurosci. Biobehav. Rev.* **36** 191–7
Faes L and Nollo G 2010 Extended causal modeling to assess Partial Directed Coherence in multiple time series with significant instantaneous interactions *Biol. Cybern.* **103** 387–400
Faes L and Nollo G 2011 Multivariate frequency domain analysis of causal interactions in physiological time series *Biomedical Engineering, Trends in Electronics, Communications and Software* ed A Laskovski (Rijeka: InTech)
Faes L, Erla S, Porta A and Nollo G 2013 A framework for assessing frequency domain causality in physiological time series with instantaneous effects *Phil. Trans. R. Soc.* A **371** 20110618
Ferrarelli F, Massimini M, Sarasso S, Casali A, Riedner B A, Angelini G, Tononi G and Pearce R A 2010 Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness *Proc. Natl. Acad. Sci. USA* **107** 2681–6
Ferri R, Rundo F, Bruni O, Terzano M G and Stam C J 2005 Dynamics of the EEG slow-wave synchronization during sleep *Clin. Neurophysiol.* **116** 2783–95
Ferri R, Rundo F, Bruni O, Terzano M G and Stam C J 2007 Small-world network organization of functional connectivity of EEG slow-wave activity during sleep *Clin. Neurophysiol.* **118** 449–56
Ginter J *et al* 2001 Phase and amplitude analysis in time-frequency space-application to voluntary finger movement *J. Neurosci. Methods* **110** 113–24
Gómez F *et al* 2013 Changes in effective connectivity by propofol sedation *PLoS One* **8** e71370
Granger C W J 1969 Investigating causal relations by econometric models and cross-spectral methods *Econometrica* **37** 424–38
Hlaváčková-Schindler K 2011 Equivalence of Granger causality and transfer entropy: a generalization *Appl. Math. Sci.* **5** 3637–48
Iber C, Ancoli-Israel S, Chesson A L and Quan S F 2007 *The AASM Manual for the Scoring of Sleep and Associated Events* (Westchester, IL: American Academy of Sleep Medicine)
Ivanov P C, Liu K K L and Bartsch R P 2016 Focus on the emerging new fields of network physiology and network medicine *New J. Phys.* **18** 100201
John E R and Prichep L S 2005 The anesthetic cascade. A theory of how anesthesia suppresses consciousness *Anesthesiology* **102** 447–71
Kamiński M J and Blinowska K J 1991 A new method of the description of the information flow sources *Biol. Cybern.* **65** 203–10
Kamiński M and Blinowska K J 2014 Directed Transfer Function is not influenced by volume conduction-inexpedient pre-processing should be avoided *Front. Comput. Neurosci.* **8** 61–3
Kamiński M, Blinowska K and Szclenberger W 1997 Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness *Electroencephalogr. Clin. Neurophysiol.* **102** 216–27
King J-R, Sitt J D, Faugeras F, Rohaut B, El Karoui I, Cohen L, Naccache L and Dehaene S 2013 Information sharing in the brain indexes consciousness in noncommunicative patients *Curr. Biol.* **23** 1914–9
Klimesch W 1999 EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis *Brain Res. Brain Res. Rev.* **29** 169–95
Klimova M 2014 What is lost during dreamless sleep: the relationship between neural connectivity patterns and consciousness *J. Eur. Psychol. Stud.* **5** 56–65
Koch C, Massimini M, Boly M and Tononi G 2016 Neural correlates of consciousness: progress and problems *Nat. Rev. Neurosci.* **17** 307–21
Kuś R *et al* 2008 Transmission of information during continuous attention test *Acta Neurobiologiae Experimentalis* **68** 103–12
Lin A, Liu K K L, Bartsch R P and Ivanov P C 2016 Delay-correlation landscape reveals characteristic time delays of brain rhythms and heart interactions *Phil. Trans. R. Soc.* A **374** 20150182
Lioi G, Bell S L and Simpson D M 2016 *XIV Mediterranean Conf. on Medical and Biological Engineering and Computing 2016: MEDICON 2016 (Paphos, Cyprus, 31 March–2 April 2016)* ed E Kyriacou *et al* (Berlin: Springer) vol 57, pp 3–8
Liu K K L, Bartsch R P, Lin A, Mantegna R N and Ivanov P C 2015 Plasticity of brain wave network interactions and evolution across physiologic states *Front. Neural Circuits* **9** 1–15
Marinazzo D, Gosseries O, Boly M, Ledoux D, Rosanova M, Massimini M, Noirhomme Q and Laureys S 2014 Directed information transfer in scalp electroencephalographic recordings: insights on disorders of consciousness *Clin. EEG Neurosci.* **45** 33–9
Mashour G A 2006 Integrating the science of consciousness and anesthesia *Anesth. Analg.* **103** 975–82

Massimini M, Ferrarelli F, Huber R, Esser S K, Singh H and Tononi G 2005 Breakdown of cortical effective connectivity during sleep *Science* **309** 2228–32
Massimini M, Ferrarelli F, Murphy M J, Huber R, Riedner B A, Casarotto S and Tononi G 2010 Cortical reactivity and effective connectivity during REM sleep in humans *Cognit. Neurosci.* **1** 176–83
Mhuircheartaigh N R, Warnaby C, Rogers R, Jbabdi S and Tracey I 2013 Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans *Sci. Transl. Med.* **5** 208ra148
Myles P S, Leslie K, McNeil J, Forbes A and Chan M T V 2004 Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial *Lancet* **363** 1757–63
Nicolaou N, Hourris S, Alexandrou P and Georgiou J 2012 EEG-based automatic classification of 'awake' versus 'anesthetized' state in general anesthesia using Granger causality *PLoS One* **7** e33869
Nir Y, Massimini M and Boly M 2013 *Neuroimaging of Consciousness* (Berlin: Springer) 133–81
Rubinov M and Sporns O 2010 Complex network measures of brain connectivity: uses and interpretations *NeuroImage* **52** 1059–69
Schlogl A and Supp G 2006 Analyzing event-related EEG data with multivariate autoregressive parameters *Prog. Brain Res.* **159** 135–47
Schrouff J *et al* 2011 Brain functional integration decreases during propofol-induced loss of consciousness *NeuroImage* **57** 198–205
Sergent C, Baillet S and Dehaene S 2005 Timing of the brain events underlying access to consciousness during the attentional blink *Nat. Neurosci.* **8** 1391–400
Siclari F, LaRocque J J, Bernardi G, Postle B R and Tononi G 2016 The neural correlates of consciousness in sleep: a no-task, within-state paradigm *bioRxiv* 012443
Sitt J D, King J R, El Karoui I, Rohaut B, Faugeras F, Gramfort A, Cohen L, Sigman M, Dehaene S and Naccache L 2014 Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state *Brain* **137** 2258–70
Spoormaker V I, Gleiser P M and Czisch M 2012 Frontoparietal connectivity and hierarchical structure of the brain's functional network during sleep *Front. Neurol.* **3** 1–10
Spoormaker V I, Schröter M S, Gleiser P M, Andrade K C, Dresler M, Wehrle R, Sämann P G and Czisch M 2010 Development of a large-scale functional brain network during human non-rapid eye movement sleep *J. Neurosci.* **30** 11379–87
Sporns O 2013 Structure and function of complex brain networks *Dialogues Clin. Neurosci.* **15** 247–62
Stickgold R, Malia A, Fosse R, Propper R and Hobson J A 2001 Brain-mind states: I. Longitudinal field study of sleep/wake factors influencing mentation report length *Sleep* **24** 171–9
Tagliazucchi E, von Wegner F, Morzelewski A, Brodbeck V, Jahnke K and Laufs H 2013 Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep *Proc. Natl Acad. Sci. USA* **110** 15419–24
Tononi G 2008 Consciousness as integrated information *Biol. Bull.* **215** 216–42
Tononi G and Massimini M 2008 Why does consciousness fade in early sleep? *Ann. New York Acad. Sci.* **1129** 330–4
Toppi J, Petti M, De Vico Fallani F, Vecchiato G, Maglione A G, Cincotti F, Salinari S, Mattia D, Babiloni F and Astolfi L 2012 Describing relevant indices from the resting state electrophysiological networks *Conf. Proc. IEEE Engineering Medical Biology Society* vol 2012 pp 2547–50
Winterhalder M *et al* 2005 Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems *Signal Processing* **85** 2137–60

Changes in EEG directional connectivity during slow induction of anaesthesia

Giulia Lioi[1] Steven L Bell[1] David C Smith[2] David M Simpson[1]

1 Institute of Sound and Vibration Research
University of Southampton
Highfield
Southampton
SO15 1BJ

2 Department of Anaesthesia
Southampton General Hospital
Southampton
SO16 6YD

Corresponding Author: G Lioi
gl4g13@soton.ac.uk

Running title: EEG connectivity during anaesthesia

Abstract

Background: The analysis of brain connectivity during general anaesthesia has the potential to shed light on the mechanisms of anaesthetic-induced loss of consciousness (LOC) and the neural correlates of consciousness and thus act as a marker of level of consciousness (LOC). A variety of neuroimaging techniques and anaesthetic protocols have associated a general suppression of long-range connectivity with anaesthetic-induced LOC and emphasized the important role of the cortical fronto-parietal network in the maintenance of consciousness.

Method: We used a slow induction of propofol anaesthesia in 10 patients to assess graded changes in EEG directional connectivity for increasing propofol effect-site concentrations (ESC). Connectivity was estimated from multichannel EEG recordings using Directed Coherence (DC), a multivariate directed spectral estimator based on Granger Causality. We investigated changes at cohort and individual level to identify brain network features that exhibit robust changes with anaesthetic-induced LOC.

Results: We found a significant reduction in the strength of long-range DC links and a switch in the direction of information flow from markedly postero-frontal in wakefulness to fronto-posterior during anaesthesia. These changes occur at the onset of light anaesthesia (at a propofol ESC of 2 mcg ml^{-1}) and remain relatively constant as infusion rate increases, consistent with a step change in DC with anaesthesia, rather than a gradual change with increasing anaesthetic dose.

Conclusion: These results give a novel insight into the reorganization of EEG directed connectivity during anaesthesia and are particularly relevant to physiological interpretation of anaesthetic-induced LOC and with potential for exploitation in future monitors of LOC.

Keywords

anaesthesia, depth; anaesthetics i.v., propofol; monitoring, electroencephalography

The neurobiology mechanisms by which anaesthetics induce loss of consciousness remain unclear[1 2]. The sensitivity of EEG time and spectral features to anaesthetic-induced changes has prompted the development of a series of EEG-derived depth of anaesthesia (DoA) monitors[3 4]. Although these indexes correlate well with the delivered anaesthetic concentration, they provide only limited insight into the cerebral mechanisms underlying anaesthetic action and may reflect the clearance of anaesthetic drugs rather than the state of arousal of the brain[5 6].

The potential of brain connectivity to quantify the global organized behaviour of neural circuits[7 8] and provide insight into the neural mechanisms underlying LOC[9] has prompted a series of investigations of anaesthetic modulation of brain connectivity. Disruption of long-range connectivity[10 11] and changes in fronto-parietal coupling[12-14] have been reported as crucial mechanisms, although some studies report contrasting results[15 16]. These findings are supported by suggestions that consciousness arises from large-scale information sharing among brain areas, and that anaesthesia induces unconsciousness by disrupting cortical integration and information processing[17 18], in particular by functional uncoupling of parieto-frontal cortical activity[2]. The majority of these studies used functional magnetic resonance imaging (fMRI), positron emission tomography (PET) or combined transcranial magnetic stimulation (TMS) and EEG approaches. However recording the EEG from scalp electrodes has practical advantages[19], as it can be applied relatively easily at the bedside and at low cost, and is already recognized as an important tool to discriminate conscious from unconscious subjects[20]. Among the different methodologies proposed to assess EEG connectivity, Directed Coherence (DC) is a well-established method that provides information about the strength, direction and spectral content of multivariate linear dependencies between EEG time series[21]. The objectives of this work are therefore to:

1) identify novel DC features that reliably reflect anaesthetic-induced LOC

2) test the robustness of these features at group and individual level

3) propose indexes of depth of anaesthesia that could be incorporated into future depth of anaesthesia monitors.

Methods

The study was approved by the Southampton and Southwest Hampshire Research Ethics Committee (ref 002/98) and all patients provided written informed consent. Ten patients (three females, seven males, aged between 44 and 79 years) participated in the study. They were selected from the cardiac surgical operating schedule the day before their planned operation, based on their ability to tolerate an additional 90 min of anaesthesia before their surgery commenced. Patients did not receive premedication.

Routine monitoring with 12-lead ECG and pulse oximetry were started on arrival in the anaesthetic room. Under local anaesthesia, a 14G cannula was inserted into a forearm vein, and a 20G cannula into a radial artery for direct arterial blood pressure measurement. EEG was collected throughout the experimental period using a 32-channel system with active electrodes (Biosemi BV, Amsterdam) built into a headcap (similar to a swimming cap) according to the International 10-20 system. Once the EEG monitoring set-up was completed, the patients were asked to lie quietly with their eyes closed for 10 min, representing the AWAKE state. A target-controlled infusion of propofol 1% (B Braun, Melsungen, Germany) was then started by syringe driver (Alaris PK, Carefusion, Sheffield, UK) to achieve an initial effect-site concentration (ESC) of 2 μg ml^{-1} using the Marsh pharmacokinetic model. All patients breathed spontaneously via a Hudson mask, supplemented with oxygen at 4-6 l min. We allowed an equilibration period (of around 5-10 min) to reach a stable ESC that was then maintained for 10 min. This procedure was repeated for propofol ESC of 3 μg ml^{-1} and subsequently 4 μg ml^{-1}. During the first 5 min of each stable ESC period, auditory stimulation was presented and evoked responses (AER) measured, however this paper refers only to EEG connectivity measurements and disregards the analysis of AER (see Supplementary Materials for more detail on the experimental protocol). After 10 min of recording at a propofol ESC of 4 μg ml^{-1}, fentanyl 1 μg ml^{-1} and pancuronium 0.2 mg kg^{-1} were given, after which the patient's trachea was intubated and mechanical ventilation commenced. Five minutes later, the target propofol ESC was reduced to 2 μg ml^{-1}; EEG recording continued until the 2 μg ml^{-1} target was achieved. The patients were then prepared for surgery.

Connectivity estimation

EEG recordings were down-sampled to 250 Hz and digitally referenced with respect to the average of the T7 and T8 channels (linked mastoid), as recommended for functional connectivity estimation[22]. The EEG time series were band pass filtered (1-45 Hz) and zero-phase notch filtered at the 50 Hz mains frequency. Only continuous and artifact-free epochs were included in the following analysis.

Directional connectivity was estimated from the multivariate (MVAR) model of multi-channel EEG time-series as previously reported[23]. A MVAR process describes each multi-channel EEG time series as the sum of a defined number of previous samples from the set of available signals (here the individual EEG channels), weighted by model coefficients, plus a noise component. The elements of the dataset relate the present of one signal to the past of another and capture the causal or directed influence between signals. Due to its straightforward interpretation in terms of spectral content[24], we used the squared modulus of DC to estimate functional connectivity. Connectivity scalp plots were obtained using eConnectome imaging software[25].

A subset of 12 electrodes that are fairly evenly distributed across the scalp (Fp1, Fp2, F3, Fz, F4, C3, Cz, C4, P3, P4, O1, O2) was selected and connectivity was estimated for epochs of 60 s, in order to reduce the computational cost typical of multivariate connectivity estimation and to ensure that the number of MVAR model parameters to be estimated was appropriate for the number of EEG samples available in each epoch[26]. The statistical significance of each DC link was tested with surrogate analysis[27] (see Supplementary Materials for more details) and all subsequent analysis was only performed on the significant connections. We also corrected for multiple comparisons using the false discovery rate (FDR) approximation for dependent measurements[28]. In addition to the statistical threshold, we applied higher thresholds to retain only the strongest connections. This procedure is widely recommended to discard weak (but significant) connections that may obscure the topology of strong links[29]. Since the choice of the threshold is somewhat arbitrary, results over a range of plausible thresholds[30] were investigated. In this study we retained respectively the 10%, 30% and 50% of strongest connections. The DC matrices were then averaged in the four standard EEG

frequency bands. In order to specifically investigate changes in long-range connectivity, DC connections were also grouped on the basis of the 3D Euclidian inter-electrode distances into short-range links connecting adjacent electrodes (inter-channel distance below 10 cm) and long-range links.

Building on previous work[31] [23], we focused the analysis on the dominant direction of information flow in the *α* band [8-13 Hz] [32] [33] [15] [16] and long-range links[10] [11], as these parameters have been found to be most sensitive to changes in the level of consciousness. We assessed the number of statistically significant connections from centro-parietal (P3, P4, C3, C4, Cz) to frontal (Fp1, Fp2, F3, F4, Fz) electrodes and vice versa. In order to quantify the dominant direction of information flow on the front-posterior axis we defined an index ($Dir_{P\rightarrow A}$), given by the normalized differences of the number of links in the two opposite directions over the rostro-caudal axis. In order to assess the changes with LOC at group level, a Friedmann test (followed by a Tukey's HSD test) was performed across subjects considering the experimental stages (AWAKE, ANES 2 µg ml^{-1}, ANES 3 µg ml^{-1}, ANES 4 µg ml^{-1}) as independent variables and the strength and dominant direction of DC links as dependent variables. In order to summarize the individual DC features in a unique parameter for each subject and 60 s epoch, we summed the normalized strength of long-range links and the $Dir_{P\rightarrow A}$ to obtain what we have called the *DCindex* (see Supplementary Materials for more details on the computation of $Dir_{P\rightarrow A}$ and DCindex). Differences in this index between the experimental stages were again assessed using the Friedman test.

As well as a cohort analysis we compared network feature trends in individual subjects with their ESC time-line, in order to assess if all (or most) individuals showed the same response as the average across the cohort. Such analysis at the individual level will indicate the potential to apply this in monitoring individual patients.

Results

EEG signals at different propofol ESC presented the characteristic changes associated with increasing depth of anaesthesia[3 34 35] (see Supplementary Materials): for ESC of 2 μg ml^{-1} the EEG time series showed increased activation in the beta band typical of light anaesthesia. For ESC 3 μg ml^{-1} or 4 μg ml^{-1} the EEG was characterized by slow wave (δ) activity. In most of the subjects a burst suppression pattern, typical of deep anaesthesia, was observed only for ESC 4 μg ml^{-1}. We can therefore argue that the slow induction of anaesthesia achieved a series of levels ranging from light to deep anaesthesia.

Figure 1 shows scalp plots of DC networks in wakefulness and at the three anaesthetic levels for the three thresholds of strongest connections. Visual analysis suggests that the strength of long-range links decreases at ESC 2 μg ml^{-1} as compared to wakefulness and then remains relatively constant with increasing propofol ESC, significantly (Friedman test, $p<0.01$) distinguishing AWAKE from deep anaesthesia at group level. The strength of postero-anterior links decreases during anaesthesia as compared to wakefulness, while the contribution of fronto-posterior connections becomes dominant. As a consequence the $Dir_{P\rightarrow A}$ index, that summarizes the dominant direction of functional links, shows a switch associated with the onset of anaesthesia. Also for the $Dir_{P\rightarrow A}$, a step effect is observable that significantly distinguishes AWAKE from the different propofol ESCs, rather than a gradual change with increasing anaesthetic depth. The connectivity networks estimated considering the 10%, 30% or 50% of strongest connections exhibit very similar trends, indicating that the choice of threshold does not critically affect the results. For the sake of brevity, in the following analysis we will only show results for the DC networks obtained retaining the 30% strongest links.

In order to investigate whether the changes observed for the different EEG features were both consistent at the individual level and able to correlate with the subject's propofol ESC time-line we present results for the DC parameters in each patient (figure 2). The *DCindex* exhibits some variability across subjects, especially at ESC 2 μg ml^{-1}. As observed at group level, individual *DCindex* values undergo a marked decrease at the onset of anaesthesia and then remain relatively constant with increasing propofol dose rate in all but one patient.

Finally, we have assessed *DCindex* values for all individuals across the whole experiment to allow comparison with the individual propofol ESC time-line (figure 3). With the exception of subjects 3 and 5, a steep transition is noticeable at the onset of anaesthesia. In the majority of subjects this change occurs soon after the start of propofol administration, sometimes with a 3-4 minutes delay (i.e. subjects 6 and 8). The *DCindex* then plateaus (with some oscillations) with increasing propofol ESC and is not significantly affected by auditory stimulation, the injection of muscle relaxants, intubation, or recovery to ESC 2 μg ml^{-1}.

Discussion

In this study we have assessed changes in the strength and direction of EEG connectivity during a slow step-wise induction of propofol anaesthesia. The most characteristic changes observed in DC during anaesthesia were a significant reduction in the strength of long-range DC links and a marked inversion of direction of information flow, from postero-frontal in wakefulness to antero-posterior in anaesthesia. These findings are consistent at individual level for the majority of subjects, where a brisk change in DC features is observed at the onset of anaesthesia.

In order to not interfere with the natural fading of consciousness during the slow anaesthetic induction, we did not assess the responsiveness of patients. Other studies using a similar ESC regimen have assessed the responsiveness of subjects with randomized auditory tasks and have identified an ESC around 2 μg ml^{-1} as the threshold for loss of behavioral responsiveness[32 36]. We therefore expect that in our protocol LOC occurred somewhere between the start of propofol infusion and achieving an ESC of 2 μg ml^{-1}. The decrease of the long-range connectivity strength observed in anaesthesia is in line with reports of an impairment of large-scale information flow and a general impairment of brain network integration (with fronto-parietal connectivity particularly affected) in fMRI and TMS studies of propofol anaesthesia[37-39]. Activity in the fronto-parietal associative network is also systematically altered in other states of diminished consciousness such as vegetative states, coma or NREM sleep[19 39-41]. Together with these findings, our results support the important role of the fronto-parietal association cortices in the maintenance of consciousness[42] and the hypothesis that the breakdown of information flow may affect signaling between the sensory posterior areas and the associative frontal cortices that is essential for a conscious experience[38]. DC estimates obtained from multichannel EEG and indexes derived from this were shown to be effective in measuring the changes in connectivity.

This study goes beyond previous work in two important points. First, we used a slow target-controlled induction of anaesthesia that allowed investigation of graded changes in directed connectivity for stepwise increasing propofol ESCs ranging from light to deep sedation. This is in contrast with previous connectivity EEG studies that are typically carried out with a bolus of

intravenous drugs[12 14-16], causing the crucial transition to LOC to occur within 30-60 s. Also we focus only on the hypnotic effect of one agent (propofol), avoiding confounding factors such as the administration of muscle relaxants or the use of different anaesthetic protocols and agents. Second, we describe both group results and single subject variability in connectivity features, while the majority of previous studies have focused on average results across the cohort. The rationale for the individual analysis is to be found in the potential practical applications: we aim to assess adequacy of anaesthesia using DC features and in this sense the measure should be able to assess anaesthetic effect in individual subjects, rather than on 'average'.

The inversion of information flow from parieto-frontal in wakefulness to fronto-parietal during anaesthesia deserves a special comment. The observed switch in the direction of connectivity from wakefulness is consistent with results reported in studies investigating EEG directional connectivity during anaesthesia. Nicolaou and colleagues[14] used a bivariate (i.e. based on only two EEG time series) Granger Causality (GC) approach to assess directed connectivity in anaesthesia and found a significant increase in fronto-posterior causality in anaesthetic induced LOC, but did not observe a significant decrease in the opposite direction. DC is grounded in Granger theory, therefore the two estimators are conceptually similar, however they are characterized by substantial differences. GC is a pairwise measure in the time domain, while DC is a multivariate measure of causal interactions in the frequency domain and it therefore captures the causality structure of the whole electrode set. A dominance of fronto-posterior coupling in the cingulate cortex was also observed in a GC study of propofol anaesthesia[16]. Furthermore, a high-density EEG sleep study[43] showed that deep sleep is characterized by a traveling wave that originates in frontal regions and propagates in the anteroposterior direction. Given the similarities of sleep and anaesthesia neurobiology[1] it has been suggested[14] that a similar behavior characterizes the anaesthetic slow wave and that the observed increase in fronto- posterior coupling may be a result of these mechanisms. In support of this hypothesis, results from our previous study on NREM sleep[23] show DC changes associated with deep sleep (N2 and N3) very similar to those observed during anaesthesia. Even though fronto-parietal coupling has been suggested as a central

mechanism for consciousness, contrasting results regarding its direction have been reported. Some studies have reported an impairment of fronto-posterior connectivity in LOC[15 44] and in some cases two different methodologies applied to the same dataset gave divergent results regarding the dominant direction of coupling[13]. These conflicting findings are likely to be the results of the use of different estimators of connectivity, and they are a warning that it may be difficult to interpret or compare results from different brain activity models that quantify different effects.

We assessed DC features at different anaesthetic levels ranging from light to deep anaesthesia. The changes in directed connectivity networks showed a step change with the onset of anaesthesia and then a near plateau with anaesthesia deepening. These changes were consistent for most individuals. The DC features show a substantially different trend from other EEG-derived indexes of DoA. The commercially available indexes (i.e. BIS, Approximate Entropy, AEPindex) usually exhibit graded changes with increasing anaesthetic doses, reflected by an index usually ranging from 0 to 100. They highly correlate with the anaesthetic drug concentration in the body and are thought to reflect the gradual clearance of anaesthetic drugs, rather than the level of consciousness of the patient that results from a more complex balance of hypnosis, analgesia, and external stimulation[5]. On the other hand the 'switch' transition of the *DCindex* that occurs at anaesthetic onset may more closely reflect the physiological mechanisms of anaesthetic-induced LOC[14]. Moreover, we have observed at individual level that the *DCindex* is scarcely influenced by the administration of muscle relaxants, auditory stimulation, or the intubation of the patient: this supports the speculation that it reflects the general physiological mechanism of hypnosis, rather than, for instance, the level of anti-nociception[14 35]. In view of possible application to the clinical setting, such an 'on-off' response may be useful during anaesthetic onset, but potentially less useful during offset unless the *DCindex* switch becomes 'on' before the patient is responsive. Further studies are required to determine when the switchover occurs in relation to conscious responsiveness and to test the use of the *DCindex* (in combination with other DoA indexes) in monitoring of anaesthetic adequacy.

Some consideration should be given to the interpretation of DC, which has become a well-

established method to assess directed connectivity. It has been applied to the study of different behavioural tasks[45] [46] showing remarkable agreement with anatomical and neuroimaging evidence[22]. However caution is required in interpreting the causality estimated by DC as indicating effective axonal links between underlying cortical sources. In the case of EEG signals, volume conduction effects may lead to highly connected adjacent electrodes as a result of unmeasured deep cortical sources that instantaneously drive many superficial electrodes. Although it has been proposed that DC is not affected by volume conduction because it is sensitive only to phase differences between channels[22], this is still debated[24]. In this study we have limited the spurious effects of volume conduction by disregarding short-range (<10 cm) connections, which are likely to be dominated by the spread due to electrical conduction, and included in the analysis only long-range connectivity. Furthermore, additional work (not shown) has demonstrated that DC is much less sensitive to volume conduction effects than some of the alternative measures of EEG connectivity. To overcome the problem of hidden sources and resolve the whole structural causality of the dataset, one must measure and include all the relevant sources. However, this is not feasible in practice and would not be compatible with a possible online clinical monitoring application.

The current study should be considered as exploratory, since results were obtained from a relatively small number of patients and tested only on one anaesthetic (propofol). Further work is required to investigate DC on a large sample of patients undergoing different anaesthetic protocols to test the reliability of directed connectivity as an indicator of LOC associated with anaesthesia. Future work will also compare DC measures with behavioral responsiveness during anaesthesia to more reliably assess changes associated with anaesthetic-induced LOC.

Figures

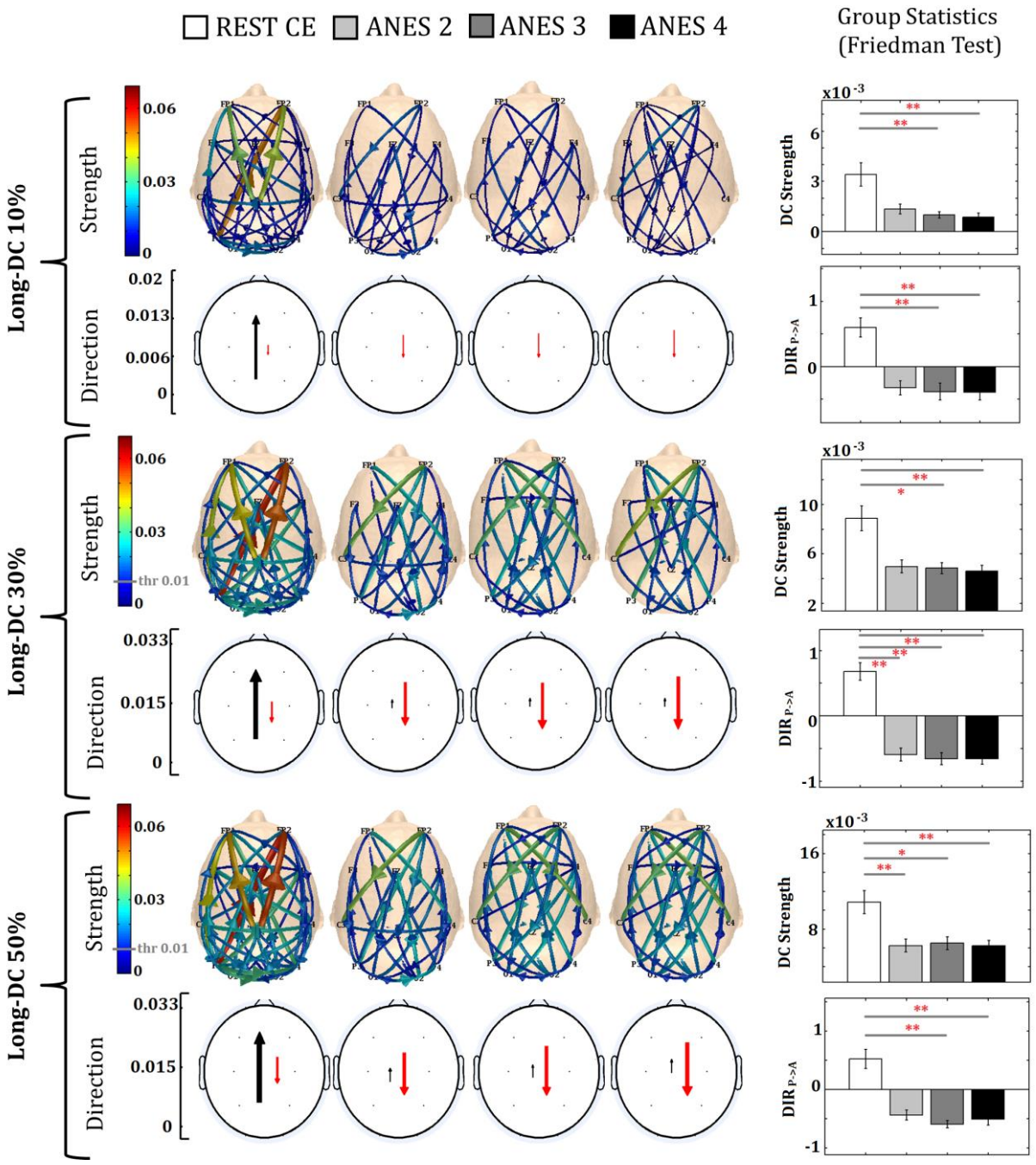

Figure 1

Scalp topography of connectivity networks, averaged across all 10 subjects, with associated statistics, plotted for the 10% (first two rows) 30% (rows 3 and 4) and 50% (last two rows) strongest connections (all connections shown are also statistically significant). The first row of

each subplot represents the Grand Average across subjects of long-range connections, with the color and size of arrows coding for the average strength of the specific link. The second row indicates the average strength of postero-anterior (black) and antero-posterior (red) connections in the α band coded by the length and thickness of the arrows. The bar plots on the right show the mean and standard error (across subjects) of the respective features (long-range links strength and $Dir_{P \rightarrow A}$). * : $p < 0.05$; ** : $p < 0.01$ (Friedman test).

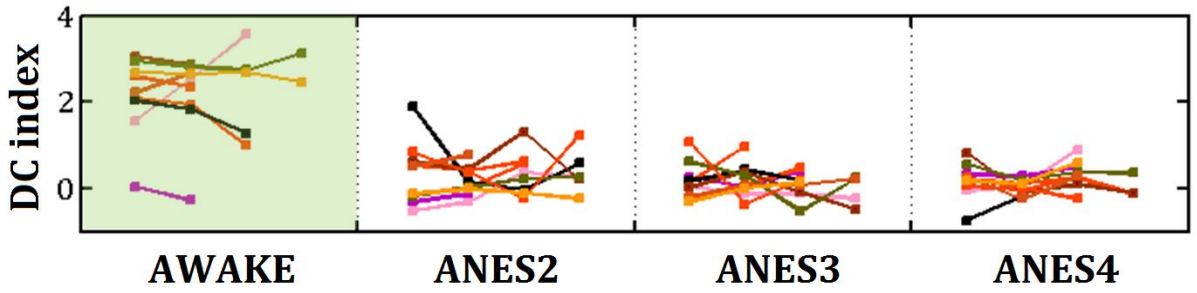

Figure 2

DCindex individual trends (in a different color for each subject, N=10) in wakefulness and at stable anaesthetic levels for the 30% strongest connections. At each level of anaesthesia, there were at most four epochs of 1 min available for analysis. Each time point represents results from a 60 s epoch.

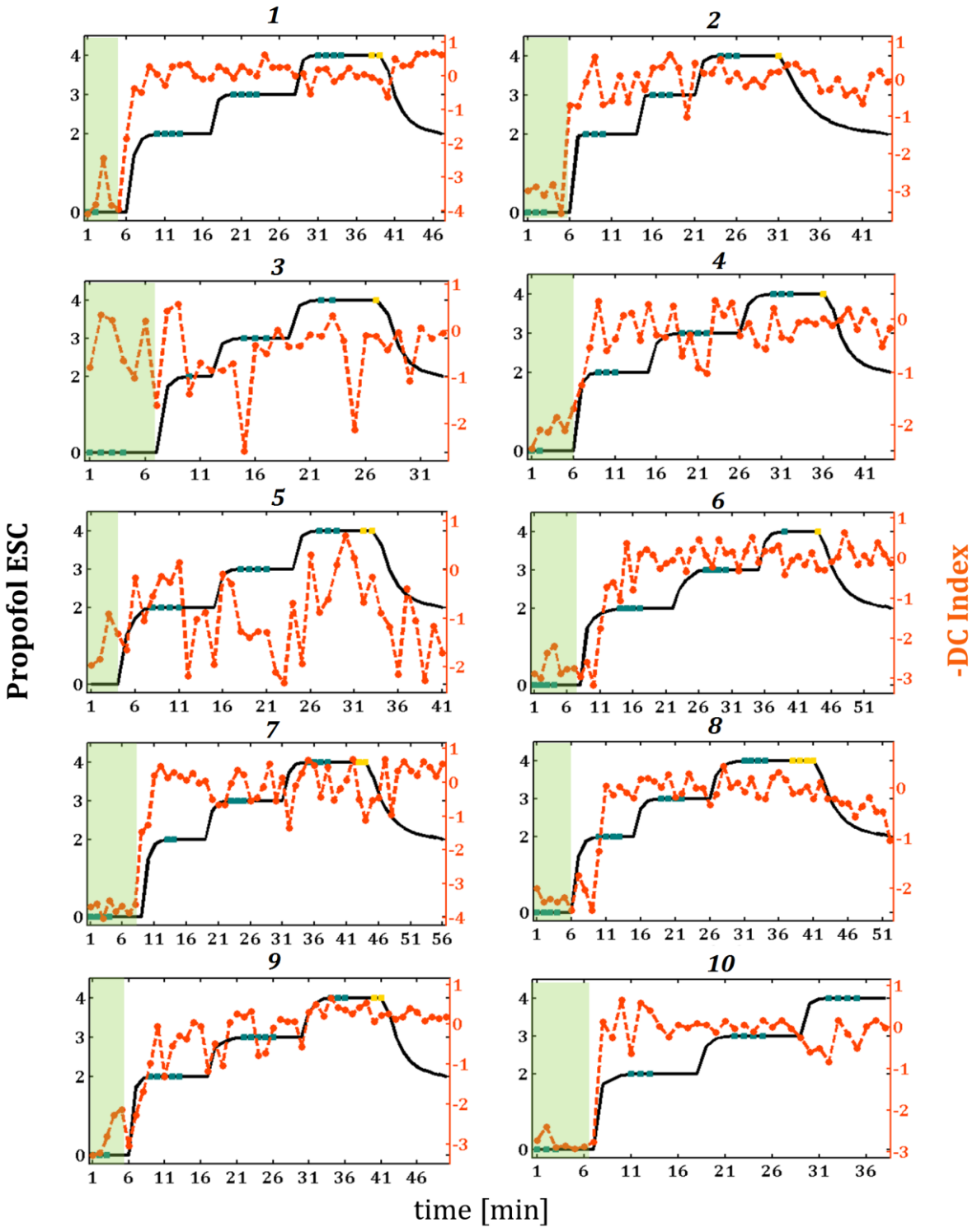

Figure 3

DCindex individual trends (dashed orange line) compared with ESC time-line (black solid line) for the whole length of the recording and for all the subjects. The *DCindex* is inverted to facilitate comparison with the propofol ESC. The epochs in wakefulness (ESC=0) are

highlighted in green and each time point refers to a 60 s epoch. The epochs where auditory stimulation was delivered are indicated on the ESC time-line by blue markers while the epochs in deep anaesthesia in which muscle relaxant was administered are indicated by yellow markers. In subject 10 the recovery period was heavily contaminated by artifacts and was excluded from the analysis.

Declaration of interests

DCS has received equipment loans and / or honoraria from GE Healthcare, Aspect Medical, and Medical Device Management. He was specialist committee member for NICE DG6, and national moderator for the NAP5 project.

Funding

This work was funded by the National Institute of Academic Anaesthesia (AAGBI/Anaesthesia Foundation Small Research Grant) and the Gerald Kerkut Trust (University of Southampton, UK).

Authors' contributions

GL: performed the EEG data collection and analysis, wrote the first draft of the manuscript.

SLB: obtained partial funding, supervised data analysis/research, revised the manuscript.

DCS: obtained ethical approval and partial funding, designed the anaesthetic protocol, revised the manuscript.

DMS: supervised data analysis/research, revised the manuscript.

References

1. Mashour GA. *Consciousness, Awareness, and Anesthesia*. New York: Cambridge University Press, 2010: 1-15; 35-43

2. John ER, Prichep LS. The anesthetic cascade. A theory of how anesthesia suppresses consciousness. *Anesthesiology* 2005; **102**: 447–71

3. Rampil I. A primer for EEG signal processing in anesthesia. *Anesthesiology* 1998; **89**: 980–1002

4. Kreuer S, et al. The Narcotrend – a new EEG monitor designed to measure the depth of anaesthesia. A comparison with bispectral index monitoring during propofol-remifentanil anaesthesia. *Anaesthesiology* 2001; **50**: 921–5

5. Schneider G, et al. Bispectral Index (BIS) may not predict awareness reaction to intubation in surgical patients. *J Neurosurg Anesth* 2002; **14**: 7–11

6. Gajraj RJ, et al. Comparison of bispectral EEG analysis and auditory evoked potentials for monitoring depth of anaesthesia during propofol anaesthesia. *Br J Anaesth* 1999; **82**: 672-8

7. Hudetz AG. General anesthesia and human brain connectivity. *Brain Connect* 2012; **2**: 291–302

8. David O, Cosmelli D, Friston KJ. Evaluation of different measures of functional connectivity using a neural mass model. *Neuroimage* 2004; **21**: 659–73

9. Nallasamy N, Tsao DY. Functional connectivity in the brain: effects of anesthesia. *Neuroscientist* 2011; **17**: 94–106

10. Ferrarelli F, et al. Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc Natl Acad Sci USA* 2010; **107**: 2681-6

11. Gómez F, et al. Changes in effective connectivity by propofol sedation. *PLoS One*

2013; **8**: e71370

12. Untergehrer G, et al. Fronto-parietal connectivity is a non-static phenomenon with characteristic changes during unconsciousness. *PLoS One* 2014; **9**: e87498

13. Ku S-W, et al. Preferential inhibition of frontal-to-parietal feedback connectivity is a neurophysiologic correlate of general anesthesia in surgical patients. *PLoS One* 2011; **6**: e25155

14. Nicolaou N, et al. EEG-based automatic classification of 'awake' versus 'anesthetized' state in general anesthesia using Granger causality. *PLoS One* 2012; **7**: e33869

15. Lee U, et al. The directionality and functional organization of frontoparietal connectivity during consciousness and anesthesia in humans. *Conscious Cogn* 2009; **18**: 1069–78

16. Barrett AB, et al. Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. *PLoS One* 2012; **7**: e29072

17. Tononi G. Consciousness as integrated information. *Biol Bull* 2008; **215**: 216–42

18. Alkire MT, Hudetz AG, Tononi G. Consciousness and anesthesia. *Science* 2008; 322.5903: 876–80

19. Sitt JD, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 2014; **137**: 2258–70

20. Koch C, et al. Neural correlates of consciousness: progress and problems. *Nat Rev Neurosci* 2016; **17**: 307–21

21. Baccalá LA, et al. Studying the interaction between brain structures via Directed Coherence and Granger Causality. *Appl Signal Process* 1998; **5**: 40–8

22. Kaminski M, Blinowska KJ. Directed Transfer Function is not influenced by volume conduction - inexpedient pre-processing should be avoided. *Front Comput Neurosci* 2014; **8**: 1-3

23. Lioi G, Bell SL, Smith DC, Simpson DM. Directional Connectivity in the EEG is able to discriminate wakefulness from NREM sleep. *Physiol Meas* 2017; **38**: 1802–1820

24. Faes L, Nollo G. Multivariate frequency domain analysis of causal interactions in physiological time series. In: Laskowski A (Ed), *Biomedical Engineering, Trends in Electronics, Communications and Software ,* INTECH Open Access Publisher, 2011: 403-428

25. Bin H, Astolfi L, Babiloni F. eConnectome: a MATLAB toolbox for mapping and imaging brain functional connectivity. *J Neurosci Methods* 2011; **195**: 1–9

26. Schlögl A, Supp G. Analyzing event-related EEG data with multivariate autoregressive parameters. *Prog Brain Res* 2006; **159**: 135–47

27. Astolfi L, et al. Estimate of causality between independent cortical spatial patterns during movement volition in spinal cord injured patients. *Brain Topogr* 2007: **19**: 107–23.

28. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 2001; **29**: 1165–88

29. Sporns O. Structure and function of complex brain networks. *Dialogues Clin Neurosci* 2013; **15**: 247–62

30. Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Publ Gr* 2009; **10**: 186–98

31. Chennu S, et al. Spectral signatures of reorganised brain networks in disorders of consciousness. *PLoS Comput Biol* 2014; **10**: e1003887

32. Purdon PL, et al. Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proc Natl Acad Sci USA* 2013; **110**: E1142–51

33. De Gennaro L, et al. Changes in fronto-posterior functional coupling at sleep onset in humans. *J Sleep Res* 2004; **13**: 209–17

34. Murphy M, et al. Propofol anesthesia and sleep: a high-density EEG study. *Sleep* 2011; **34**: 283–91A

35. Marchant N, et al. How electroencephalography serves the anesthesiologist. *Clin EEG Neurosci* 2014; **45**: 22–32

36. Purdon PL, et al. Simultaneous electroencephalography and functional magnetic resonance imaging of general anesthesia. *Ann N Y Acad Sci* 2009; **1157**: 61–70

37. Schrouff J, et al. Brain functional integration decreases during propofol-induced loss of consciousness. *Neuroimage* 2011; **57**: 198–205

38. Boveroux P, et al. Breakdown of within- and between-network resting state during propofol-induced loss of consciousness. *Anesthesiology* 2010; **113**: 1038–53

39. Massimini M, et al. Cortical reactivity and effective connectivity during REM sleep in humans. *Cogn Neurosci* 2010; **1**: 176–83

40. Massimini M, et al. Breakdown of cortical effective connectivity during sleep. *Science* 2005; **309**: 2228–32

41. Spoormaker VI, et al. Development of a large-scale functional brain network during human non-rapid eye movement sleep. *J Neurosci* 2010; **30**: 11379–87

42. Boly M, et al. Intrinsic brain activity in altered states of consciousness. *Ann N Y Acad Sci* 2008; **1129**; 119–29

43. Massimini M, et al. The SSO as a traveling wave. *J Neurosci* 2004; **24**: 6862–70

44. Boly M, et al. Hierarchical clustering of brain activity during human non-rapid eye movement sleep. *Proc Natl Acad Sci* 2012; **109**; 5856–61

45. Kus R, et al. Transmission of information during continuous attention test. *Acta Neurobiol Exp (Wars)* 2008; **68**: 103–12

46. Brzezicka A, et al. Information transfer during a transitive reasoning task. *Brain Topogr*

Bibliography

Achermann, P. & Borbély, A. (1998), 'Coherence analysis of the human sleep electroencephalogram', *Neuroscience* **85**(4), 1195–1208.

Alkire, M. T., Haier, R. J., Barker, S. J., Shah, N. K., Wu, J. C. & J, K. (1995), 'Cerebral Metabolism during Propofol Anaesthesia in Humans Studied with Positron Emission Tomography', *Anesthesiology* **82**, 393–403.

Alkire, M. T., Haier, R. J. & Fallon, J. H. (2000), 'Toward a unified theory of narcosis: brain imaging evidence for a thalamocortical switch as the neurophysiologic basis of anesthetic-induced unconsciousness.', *Conscious. Cogn.* **9**(3), 370–86.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/10993665*

Alkire, M. T., Hudetz, A. G. & Tononi, G. (2008), 'Consciousness and anesthesia.', *Science* **322**(5903), 876–880.

Association of Anaesthetists of Great Britain and Ireland (2016), 'Recommendations for standards of monitoring during anaesthesia and recovery (4th Ediction)', *Anaesthesia* **71**, 85–93.
URL: *http://www.aagbi.org/sites/default/files/standardsofmonitoring07.pdf*

Astolfi, L., Bakardjian, H., Cincotti, F., Mattia, D., Marciani, M. G., De Vico Fallani, F., Colosimo, A., Salinari, S., Miwakeichi, F., Yamaguchi, Y., Martinez, P., Cichocki, A., Tocci, A. & Babiloni, F. (2007), 'Estimate of causality between independent cortical spatial patterns during movement volition in spinal cord injured patients.', *Brain Topogr.* **19**(3), 107–23.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/17577652*

Astolfi, L., Cincotti, F., Mattia, D., De Vico Fallani, F., Tocci, a., Colosimo, a., Salinari, S., Marciani, M. G., Hesse, W., Witte, H., Ursino, M., Zavaglia, M. & Babiloni, F. (2008), 'Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators.', *IEEE Trans. Biomed. Eng.* **55**(3), 902–13.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/18334381*

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccalà, L. a., de Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M. & Babiloni, F. (2006), 'Assessing cortical functional connectivity by partial directed coherence: simulations and application to real data.', *IEEE Trans. Biomed. Eng.* **53**(9), 1802–12.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16941836*

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. a., de Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M., Ding, L., Edgar, J. C., Miller, G. a., He, B. & Babiloni, F. (2007), 'Comparison of different cortical connectivity estimators for high-resolution EEG recordings.', *Hum. Brain Mapp.* **28**(2), 143–57.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16761264*

Babiloni, F., Cincotti, F., Babiloni, C., Carducci, F., Mattia, D., Astolfi, L., Basilisco, a., Rossini, P. M., Ding, L., Ni, Y., Cheng, J., Christine, K., Sweeney, J. & He, B. (2005), 'Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function', *Neuroimage* **24**(2005), 118–131.

Baccalá, L. a. & Sameshima, K. (2001), 'Partial directed coherence: a new concept in neural structure determination.', *Biol. Cybern.* **84**(6), 463–74.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/11417058*

Baccalà, L. a. & Sameshima, K. (2007), Generalized Partial Directed Coherence, *in* 'Proc. 2007 15th Intl. Conf. Digit. Signal Process.', number 3, pp. 1–4.

Bard, J. W. (2001), 'The BIS monitor: a review and technology assessment.', *AANA J.* **69**(6), 477–83.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/11837151*

Barr, G., Jakobsson, J. G., Owall, A. & Anderson, R. E. (1999), 'Nitrous oxide does not alter bispectral index: Study with nitrous oxide as sole agent and as an adjunct to i.v. anaesthesia', *Br. J. Anaesth.* **82**(6), 827–830.

Barrett, A. B., Murphy, M., Bruno, M.-A., Noirhomme, Q., Boly, M., Laureys, S. & Seth, A. K. (2012), 'Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia.', *PLoS One* **7**(1), e29072.

Bartsch, R. P., Liu, K. K. L., Bashan, A. & Ivanov, P. C. (2015), 'Network physiology: How organ systems dynamically interact', *PLoS One* **10**(11), 1–34.

Bashan, A., Bartsch, R. P., Kantelhardt, J. W., Havlin, S. & Ivanov, P. C. (2012), 'Network physiology reveals relations between network topology and physiological function.', *Nat. Commun.* **3**, 702.
URL: *http://dx.doi.org/10.1038/ncomms1705*

Beecher, H. K. (1947), 'Anesthesia ' s Second Power : Probing the Mind levers into', *Science* **105**, 164–167.

Bell, S. L., Allen, R. & Lutman, M. E. (2001), 'The feasibility of maximum length sequences to reduce acquisition time of the middle latency response', *J. Acoust. Soc. Am.* **109**(3), 1073.
URL: *http://link.aip.org/link/JASMAN/v109/i3/p1073/s1&Agg=doi*

Bell, S. L., Allen, R. & Lutman, M. E. (2002), 'Optimizing the acquisition time of the middle latency response using maximum length sequences and chirps', *J. Acoust. Soc. Am.* **112**(5), 2065.
URL: *http://link.aip.org/link/JASMAN/v112/i5/p2065/s1&Agg=doi*

Bell, S. L., Smith, D. C., Allen, R. & Lutman, M. E. (2004), 'Recording the middle latency response of the auditory evoked potential as a measure of depth of anaesthesia. A technical note.', *Br. J. Anaesth.* **92**(3), 442–5.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/14742332*

Bell, S. L., Smith, D. C., Allen, R. & Lutman, M. E. (2006), 'The auditory middle latency response, evoked using maximum length sequences and chirps, as an indicator of adequacy of anesthesia.', *Anesth. Analg.* **102**(2), 495–8.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16428549*

Bendat, J. S. & Piersol, A. G. (2000), *Random data: analysis and measurement procedures*, 3rd ed. edn, New York.

Benjamini, Y. & Yekutieli, D. (2001), 'The control of the false discovery rate in multiple testing under depencency', *Ann. Stat.* **29**(4), 1165–1188.

Billinger, M., Seeber, M., Brunner, C., Billinger, M., Seeber, M. & Mullen, T. R. (2016), 'Volume Conduction Influences Scalp-Based Connectivity Estimates', *Front. Comput. Neurosci.* (November).

Bin, H., Astolfi, L. & Babiloni, F. (2011), 'eConnectome: a MATLAB toolbox for mapping and Imaging brain functional connectivity', *J Neurosci Methods* **195**(2), 261–269.

Blinowska, K. J. (2011), 'Review of the methods of determination of directed connectivity from multichannel data', *Med. Biol. Eng. Comput.* **49**(5), 521–529.

Boly, M., Garrido, M. I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S. & Friston, K. (2011), 'Preserved feedforward but impaired top-down processes in the vegetative state.', *Science* **332**(6031), 858–62.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21566197*

Boly, M., Perlbarg, V., Marrelec, G., Schabus, M., Laureys, S., Doyon, J., Pelegrini-Issac, M., Maquet, P. & Benali, H. (2012), 'Hierarchical clustering of brain activity during human nonrapid eye movement sleep', *Proc. Natl. Acad. Sci.* **109**(15), 5856–5861.

Boly, M., Phillips, C., Tshibandam, ., Vanhaudenhuyse, A., Schabus, M., Dang-vu, T. T., Hustinx, R., Maquet, P. & Laureys, S. (2008), 'Intrinsic Brain Activity in Altered States of Consciousness', *Ann. N. Y. Acad. Sci.* **1129**, 119–129.

Boly, M. & Seth, A. K. (2012), 'Modes and models in disorders of consciousness science.', *Arch. Ital. Biol.* **150**(2-3), 172–84.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/23165877*

Bonhomme, V., Boveroux, P., Vanhaudenhuyse, a., Hans, P., Brichant, J. F., Jaquet, O., Boly, M. & Laureys, S. (2011), 'Linking sleep and general anesthesia mechanisms: this is no walkover.', *Acta Anaesthesiol. Belg.* **62**(3), 161–71.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/22145259*

Bonnet, M., Carley, D., Consultant, M. C., Consultant, P. E., Chairman, C. G., Harper, R., Hayes, B., Hirshkowitz, M., Keenan, S., Consultant, M. P., Roehrs, T., Smith, J., Weber, S., Westbrook, P., Administrative, A. & Bruce, S. (1992), 'ASDA Report EEG Arousals : Scoring Rules and Examples EEG Arousals : Scoring Rules and Examples A Preliminary Report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association', **15**(2), 173–184.

Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A., Noirhomme, Q., Lauwick, S., Luxen, A., Degueldre, C., Alain, P., Schnakers, C., Phillips, C., ois Brichant, J.-F., Bonhomme, V., Maquet, P., Greicius, M. D., Laureys, S. & Boly, M. (2010), 'Breakdown of within- and between-network Resting State during Propofol-induced Loss of Consciousness', *Anesthesiology* **113**(November), 1038–53.

Brereton, R. G. & Lloyd, G. R. (2010), 'Support vector machines for classification and regression.', *Analyst* **135**(2), 230–267.

Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory*, San Francisco.

Brooks, A. & Lack, L. (2006), 'A brief afternoon nap following nocturnal sleep restriction: which nap duration is most recuperative?', *Sleep* **29**(6), 831–840.

Brown, E. N., Lydic, R. & Shiff, N. D. (2010), 'General Anesthesia, Sleep and Coma', *N. Engl. J. Med.* **363**(27), 2638–50.

Bruchas, R. R., Kent, C. D., Wilson, H. D. & Domino, K. B. (2011), 'Anesthesia awareness: Narrative review of psychological sequelae, treatment, and incidence', *J. Clin. Psychol. Med. Settings* **18**(3), 257–267.

Bruhn, J., Myles, P. S., Sneyd, R. & Struys, M. M. R. F. (2006), 'Depth of anaesthesia monitoring: What's available, what's validated and what's next?', *Br. J. Anaesth.* **97**(1), 85–94.

Bruhn, J., Ropcke, H. & Hoeft, A. (2000), 'Approximate Entropy as an Electroencephalographic Measure of Anesthetic Drug Effect during', *Anesthesiology* **92**, 715–726.

Brzezicka, A., Kamiński M, M., Kamiński M, J. & Blinowska, K. (2011), 'Information transfer during a transitive reasoning task', *Brain Topogr.* **24**(1), 1–8.

Bullmore, E. & Sporns, O. (2009), 'Complex brain networks: graph theoretical analysis of structural and functional systems', *Nat. Publ. Gr.* **10**(3), 186–198.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19190637*

Buzsáki, G., Anastassiou, C. a. & Koch, C. (2012), 'The origin of extracellular fields and currents EEG, ECoG, LFP and spikes', *Nat. Rev. Neurosci.* **13**(June), 407–420.

Chamoun, N. G., Sigl, J. C. & Smith, C. P. (1995), 'Cerebral Biopotential Analysis System and Method'.

Chan, M. T. V. & Gin, T. (2000), 'What does the bispectral EEG index monitor?', *Eur. J. Anaesthesiol.* **17**(3), 146–148.

Chennu, S., Finoia, P., Kamau, E., Allanson, J., Williams, G. B., Monti, M. M., Noreika, V., Arnatkeviciute, A., Canales-Johnson, A., Olivares, F., Cabezas-Soto, D., Menon, D. K., Pickard, J. D., Owen, A. M. & Bekinschtein, T. A. (2014), 'Spectral Signatures of Reorganised Brain Networks in Disorders of Consciousness', *PLoS Comput. Biol.* **10**(10).

Collins, D., Neelin, P., Peters, T. M. & Evans, A. C. (1994), 'Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space.', *J Comput Assist Tomogr.* **18**(2), 192–205.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/8126267*

Corsi-Cabrera, M., Miró, E., Del-Río-Portilla, Y., Pérez-Garci, E., Villanueva, Y. & Guevara, M. a. (2003), 'Rapid eye movement sleep dreaming is characterized by uncoupled EEG activity between frontal and perceptual cortical regions', *Brain Cogn.* **51**(3), 337–345.

Dahlhaus, R. (2000), 'Graphical interaction models for multivariate time series', *Metrika* **51**(2), 157–172.
URL: *http://www.springerlink.com/index/httq33cn094935vy.pdf*

Dahlhaus, R. & Eichler, M. (1997), 'Identification of synaptic connections in neural ensembles by graphical models', *J. Neurosci. Methods* **77**, 93–107.

Dale, A. & Sereno, M. (1993), 'Improved Localization of Cortical Activity By Combining EEG and MEG with MRI Cortical Surface Reconstruction', *J. Cogn. Neurosci.* (5), 162–176.

David, O., Cosmelli, D. & Friston, K. J. (2004), 'Evaluation of different measures of functional connectivity using a neural mass model.', *Neuroimage* **21**(2), 659–73.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/14980568*

Davis, W. D. T. (1996), 'Generation and properties of maximum length sequences.', *Control* **10**(10), 354–5.

De Gennaro, L., Vecchio, F., Ferrara, M., Curcio, G., Rossini, P. M. & Babiloni, C. (2004), 'Changes in fronto-posterior functional coupling at sleep onset in humans.', *J. Sleep Res.* **13**(3), 209–17.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/15339256*

Dodge, Y. & Rousson, V. (2001), 'On Asymmetric Properties of the Correlation Coefficient in the Regression Setting', *Am. Stat.* **55**(1), 51–54.

Eichler, M., Dahlhaus, R. & Sandkühler, J. (2003), 'Partial correlation analysis for the identification of synaptic connections', *Biol. Cybern.* **89**(4), 289–302.

Erberling, C. & Don, M. (1984), 'Quality Estimation Of Auditory Brainstem Responses', *Scand. Audiol.* **13**, 1887–197.

Eysholdt, U. & Schreiner, C. (1982), 'Maximum Length Sequences-A fast method for measuring Brain-Stem Evoked Responses', *Audiology* .

Faes, L., Erla, S., Porta, A. & Nollo, G. (2013), 'A framework for assessing frequency domain causality in physiological time series with instantaneous effects', *Phil. Trans. R. Soc. A* **371**.

Faes, L. & Nollo, G. (2010), 'Extended causal modeling to assess Partial Directed Coherence in multiple time series with significant instantaneous interactions.', *Biol. Cybern.* **103**(5), 387–400.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/20938676*

Faes, L. & Nollo, G. (2011), Multivariate Frequency Domain Analysis of Causal Interactions in Physiological Time Series, *in* Mr Anthony Laskovski, ed., 'Biomed. Eng. Trends Electron. Commun. Softw.', 2011 edn, INTECH Open Access Publisher, chapter 21.

Faes, L., Nollo, G. & Antolini, R. (2001), 'Investigating the level of significance of the coherence function in cardiovascular variability analysis'.
URL: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=977697*

Fan, R.-e., Chen, P.-H. & Lin, C.-j. (2005), 'Working Set Selection Using Second Order Information for Training Support Vector Machines', *J. Mach. Learn. Res.* **6**, 1889–1918.

Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B. a., Angelini, G., Tononi, G. & Pearce, R. a. (2010), 'Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness.', *Proc. Natl. Acad. Sci. U. S. A.* **107**(6), 2681–2686.

Ferri, R., Rundo, F., Bruni, O., Terzano, M. G. & Stam, C. J. (2005), 'Dynamics of the EEG slow-wave synchronization during sleep', *Clin. Neurophysiol.* **116**(12), 2783–2795.

Ferri, R., Rundo, F., Bruni, O., Terzano, M. G. & Stam, C. J. (2008), 'The functional connectivity of different EEG bands moves towards small-world network organization during sleep', *Clin. Neurophysiol.* **119**(9), 2026–2036.

Forrest, F. C., Tooley, M. A., Saunders, P. R. & Prys-Roberts, C. (1994), 'Propofol infusion and the suppression of consciousness : the EEG and dose requirements', *Br. J. Anaesth.* (72), 35–41.

Friston, K. J. (1994), 'Functional and effective connectivity in neuroimaging: A synthesis', *Hum. Brain Mapp.* **2**, 56–78.

Friston, K. J. (2011), 'Functional and effective connectivity: a review.', *Brain Connect.* **1**(1), 13–36.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/22432952*

Fuchs, M., Wagner, M., Kohler, T. & Wischmann, H. (1999), 'Linear and Nonlinear Current Density Reconstructions', *J. Clin. Neurophysiol.* **16**(3), 267–295.

Gajraj, R. J., Doi, M., Mantzaridis, H. & Kenny, G. N. (1999), 'Comparison of bispectral EEG analysis and auditory evoked potentials for monitoring depth of anaesthesia during propofol anaesthesia.', *Br. J. Anaesth.* **82**(5), 672–678.

Gath, I. & Inbar, G. F. (1996), *Advances in Processing and Pattern Analysis of Biological Signals*, Plenum Press, New York, USA.

Ginter, J., Blinowska, K. J., Kamiński, M. & Durka, P. J. (2001), 'Phase and amplitude analysis in time-frequency space–application to voluntary finger movement.', *J. Neurosci. Methods* **110**(1-2), 113–24.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/11564531*

Gómez, F., Phillips, C., Soddu, A., Boly, M., Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A., Gosseries, O., Bonhomme, V., Laureys, S. & Noirhomme, Q. (2013), 'Changes in effective connectivity by propofol sedation.', *PLoS One* **8**(8), e71370.

Gross, J., Timmermann, L., Schnitzler, A., Salmelin, R., Kujala, J. & Ha, M. (2001), 'Dynamic imaging of coherent sources : Studying neural interactions in the human brain', *PNAS* **98**(2), 694–699.

Hagihira, S., Takashina, M., Mori, T., Mashimo, T. & Yoshiya, I. (2001), 'Practical issues in bispectral analysis of electroencephalographic signals.', *Anesth. Analg.* **93**(4), 966–970, table of contents.

Hämäläinen, M. S. & Sarvas, J. (1989), 'Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data.', *IEEE Trans. Biomed. Eng.* **36**(2), 165–71.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/2917762*

Hansen, P. C. & Zaglia, M. R. (1993), 'REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems *', **6**(I 994), 1–35.

Holt, M., Tooley, M. A. & Forrest, F. C. (1998), 'Use of parametric modelling and statistical pattern recognition in detection of awareness during general anaesthesia', *IEE Proc-Sci. Meas. Technol.* **145**(6), 307–316.

Hudetz, A. G. (2012), 'General anesthesia and human brain connectivity.', *Brain Connect.* **2**(6), 291–302.

Hyvarinen, A., Zhang, K., Shimizu, S. & Hoyer, P. O. (2010), 'Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity', *J. Mach. Learn. Res.* **11**, 1709–1731.

Iber, C., Ancoli-Israel, S., Chesson, A. L. & Quan, S. F. (2007), The AASM Manual for the Scoring of Sleep and Associated Events, Technical report, Wesrchester, IL, 60154, U.S.A.

Irwin, M. G., Hui, T. W. C., Milne, S. E. & Kenny, G. N. C. (2002), 'Propofol effective concentration 50 and its relationship to bispectral index.', *Anaesthesia* **57**(3), 242–8.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/11879213*

Ivanov, P. C., Liu, K. K. L. & Bartsch, R. P. (2016), 'Focus on the emerging new fields of network physiology and network medicine', *New J. Phys.* **18**(100201).

Johansen, J. W. (2006), 'Update on Bispectral Index monitoring', *Best Pract. Res. Clin. Anaesthesiol.* **20**(1), 81–99.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S1521689605000613*

John, E. R. & Prichep, L. S. (2005), 'The Anesthetic Cascade. A Theory of How Anesthesia Suppresses Consciousness', *Anesthesiology* **102**(2), 447–471.

Kaminski, M. & Blinowska, K. J. (2014), 'Directed Transfer Function is not influenced by volume conduction-inexpedient pre-processing should be avoided.', *Front. Comput. Neurosci.* **8**(June), 61.

Kamiński, M., Blinowska, K. & Szclenberger, W. (1997), 'Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness.', *Electroencephalogr. Clin. Neurophysiol.* **102**(3), 216–27.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/9129577*

Kaul, H., Kaul, H., Bharti, N. & Bharti, N. (2002), 'Monitoring depth of anaesthesia', *Indian J. Anesth.* **46**(4), 323–332.

Kearse, L., Manberg, P., Chamoun, N., DeBros, F. & Zaslavsky, A. (1994), 'Bispectral analysis of the electroencephalogram correlates with patient movement to skin incision during propofol anesthesia.pdf', *Anesthesiology* **81**, 1365–1370.

Kerssens, C., Klein, J. & Bonke, B. (2003), 'Monitoring versus Remembering What Happened', *Anesthesiology* **99**, 570–575.

King, J.-R., Sitt, J. D., Faugeras, F., Rohaut, B., El Karoui, I., Cohen, L., Naccache, L. & Dehaene, S. (2013), 'Information sharing in the brain indexes consciousness in noncommunicative patients.', *Curr. Biol.* **23**(19), 1914–9.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24076243*

Klimesch, W. (1999), 'EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis.', *Brain Res. Brain Res. Rev.* **29**(2-3), 169–95.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/10209231*

Klimova, M. (2014), 'What Is Lost During Dreamless Sleep : The Relationship Between Neural Connectivity Patterns and Consciousness', *J. Eur. Psychol. Students* **5**(3), 56–65.

Koch, C., Massimini, M., Boly, M. & Tononi, G. (2016), 'Neural correlates of consciousness: progress and problems', *Nat. Rev. Neurosci.* **17**(5), 307–321.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27094080*

Krogh, A. (2008), 'What are artificial neural networks?', *Nat Biotechnol* **26**(2), 195–197.
URL: *http://dx.doi.org/10.1038/nbt1386*

Ku, S.-W., Lee, U., Noh, G.-J., Jun, I.-G. & Mashour, G. a. (2011), 'Preferential inhibition of frontal-to-parietal feedback connectivity is a neurophysiologic correlate of general anesthesia in surgical patients.', *PLoS One* **6**(10), e25155.

Kus, R., Blinowska, K. J., Kaminski, M. & Basinska, S. (2005), 'Propagation of EEG activity during continuous attention test', *Bull. Polish Acad. Sci. Tech. Sci.* **53**(3), 217–222.

Kuś, R., Blinowska, K. J., Kamiński, M. & Basińska-Starzycka, A. (2008), 'Transmission of information during Continuous Attention Test', *Acta Neurobiol. Exp. (Wars).* **68**(1), 103–112.

Lau, K., Matta, B., Menon, D. K. & Absalom, A. R. (2006), 'Attitudes of anaesthetists to awareness and depth of anaesthesia monitoring in the UK', *Eur J Anaesthesiol.* **23**(11), 921–30.

Lee, U., Kim, S., Noh, G.-J., Choi, B.-M., Hwang, E. & Mashour, G. a. (2009), 'The directionality and functional organization of frontoparietal connectivity during consciousness and anesthesia in humans.', *Conscious. Cogn.* **18**(4), 1069–1078.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19443244*

Lioi, G., Bell, S. L., Smith, D. C. & Simpson, D. M. (2017), 'Directional connectivity in the EEG is able to discriminate wakefulness from NREM sleep', *Physiol. Meas.* **38**, 1802–1820.
URL: *https://doi.org/10.1088/1361-6579/aa81b5*

Liu, K. K. L., Bartsch, R. P., Lin, A., Mantegna, R. N. & Ivanov, P. C. (2015), 'Plasticity of brain wave network interactions and evolution across physiologic states', *Front. Neural Circuits* **9**(62), 1–15.

Llinás, R. R. & Steriade, M. (2006), 'Bursting of thalamic neurons and states of vigilance.', *J. Neurophysiol.* **95**(6), 3297–308.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16554502*

Loveman, E., Van Hooff, J. C. & Smith, D. C. (2001), 'The auditory evoked response as an awareness monitor during anaesthesia.', *Br. J. Anaesth.* **86**(4), 513–518.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Berlin and New York.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analyisis*, Springer Berlin Heidelberg, Berlin, Germany.

Lydic, R. & Baghdoyan, H. a. (2005), 'Sleep, anesthesiology, and the neurobiology of arousal state control.', *Anesthesiology* **103**(6), 1268–95.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16306742*

Marchant, N., Sanders, R., Sleigh, J., Vanhaudenhuyse, A., Bruno, M.-A., Brichant, J. F., Laureys, S. & Bonhomme, V. (2014), 'How electroencephalography serves the anesthesiologist.', *Clin. EEG Neurosci.* **45**, 22–32.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24415399*

Mashour, G. a. (2004), 'Consciousness Unbound', *Anaesthesiology* pp. 428–433.

Mashour, G. A. (2006), 'Integrating the science of consciousness and anesthesia', *Anesth. Analg.* **103**(4), 975–982.

Mashour, G. A. (2010), *Consciousness, Awareness, and Anesthesia*, Cambridge University Press, New York, USA.

Mashour, G. A. (2011), 'Sleep, Anesthesia, and Consciousness', *Sleep* **34**(3), 247–248.

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H. & Tononi, G. (2005*a*), 'Breakdown of cortical effective connectivity during sleep.', *Science* **309**(5744), 2228–32.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16195466*

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H. & Tononi, G. (2005*b*), 'Breakdown of cortical effective connectivity during sleep.', *Science* **309**(5744), 2228–32.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16195466*

Massimini, M., Ferrarelli, F., Murphy, M. J., Huber, R., Riedner, B. A., Casarotto, S. & Tononi, G. (2010), 'Cortical reactivity and effective connectivity during REM sleep in humans', *Cogn. Neurosci.* **1**(3), 176–183.

Massimini, M., Huber, R., Ferrarelli, F., Hill, S. & Tononi, G. (2004), 'The SSO as a Traveling Wave', *J. Neurosci.* **24**(31), 6862–6870.

McCleane, G. J. & Cooper, R. (1990), 'The nature of pre-operative anxiety', *Anaesthesia* **45**(2), 153–155.

Medkour, T., Walden, a. T. & Burgess, a. (2009), 'Graphical modelling for brain connectivity via partial coherence', *J. Neurosci. Methods* **180**(2), 374–383.

Mhuircheartaigh, N. R., Warnaby, C., Rogers, R., Jbabdi, S. & Tracey, I. (2013*a*), 'Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans.', *Sci. Transl. Med.* **5**(208), 208ra148.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24154602*

Mhuircheartaigh, R. N., Warnaby, C., Rogers, R., Jbabdi, S. & Tracey, I. (2013*b*), 'Slow-Wave Activity Saturation and Thalamocortical Isolation During Propofol Anesthesia in Humans', *Sci. Transl. Med.* **5**(208).

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L. & Grave de Peralta, R. (2004), 'EEG source imaging.', *Clin. Neurophysiol.* **115**(10), 2195–222.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/15351361*

Miller, A., Sleigh, J. W., Barnard, J. & Steyn-Ross, D. A. (2004), 'Does bispectral analysis of the electroencephalogram add anything but complexity?', *Br. J. Anaesth.* **92**(1), 8–13.
URL: *http://bja.oxfordjournals.org/lookup/doi/10.1093/bja/aeh003*

Morimoto, Y., Hagihira, S., Koizumi, Y., Ishida, K., Matsumoto, M. & Sakabe, T. (2004), 'The relationship between bispectral index and electroencephalographic parameters during isoflurane anesthesia.', *Anesth. Analg.* **98**, 1336–1340, table of contents.

Mullen, T. (2010), Source Information Flow Toolbox (SIFT) Theoretical Handbook and User Manual, Technical report, University of California, San Diego.

Murphy, M., Bruno, M.-A., Riedner, B. a., Boveroux, P., Noirhomme, Q., Landsness, E. C., Brichant, J.-F., Phillips, C., Massimini, M., Laureys, S., Tononi, G. & Boly, M. (2011), 'Propofol anesthesia and sleep: a high-density EEG study.', *Sleep* **34**(3), 283–291A.

Myles, P. S., Leslie, K., McNeil, J., Forbes, a. & Chan, M. T. V. (2004), 'Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial.', *Lancet* **363**(9423), 1757–63.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/15172773*

Nallasamy, N. & Tsao, D. Y. (2011), 'Functional connectivity in the brain: effects of anesthesia.', *Neuroscientist* **17**(1), 94–106.

Nicolaou, N. & Georgiou, J. (2013), 'Monitoring Depth of Hypnosis under Propofol General Anaesthesia Granger Causality and Hidden Markov Models', *Proc. Int. Congr. Neurotechnology, Electron. Informatics* pp. 256–261.

Nicolaou, N. & Georgiou, J. (2014), 'Global field synchrony during general anaesthesia.'.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24169819*

Nicolaou, N., Hourris, S., Alexandrou, P. & Georgiou, J. (2012), 'EEG-based automatic classification of 'awake' versus 'anesthetized' state in general anesthesia using granger causality', *PLoS One* **7**(3).

Nir, Y., Massimini, M. & Boly, M. (2013), *Neuroimaging of Consciousness*, Berlin Heidelberg.
URL: *http://link.springer.com/10.1007/978-3-642-37580-4*

Nolte, G., Ziehe, A., Nikulin, V. V., Schlögl, A., Krämer, N., Brismar, T. & Müller, K. R. (2008), 'Robustly estimating the flow direction of information in complex physical systems', *Phys. Rev. Lett.* **100**(June), 1–4.

Nunez, P. L. (1981), *Electric Fields of theBrain: The Neurophyssics of EEG.*, New York, USA.

Nunez, P. L., Pilgreen, K. L., Westdorp, a. F., Law, S. K. & Nelson, a. V. (1991), 'A visual study of surface potentials and Laplacians due to distributed neocortical sources: computer simulations and evoked potentials.', *Brain Topogr.* **4**(2), 151–168.

Nunez, P. L., Silberstein, R. B., Shi, Z., Carpenter, M. R., Srinivasan, R., Tucker, D. M., Doran, S. M., Cadusch, P. J. & Wijesinghe, R. S. (1999), 'EEG coherency II: Experimental comparisons of multiple measures', *Clin. Neurophysiol.* **110**(3), 469–486.

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B. & Cadusch, P. J. (1997), 'EEG coherency I: Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales', *Electroencephalogr. Clin. Neurophysiol.* **103**(5), 499–515.

Olejarczyk, E., Lipping, T. & Marciniak, R. (2017), Correlation of Depth of Anesthesia Indexes with MAC in Volatile Anesthesia, *in* 'EMBEC NBC 2017, IFMBE Proc. 65', pp. 972–975.

Olofsen, E., Sleigh, J. W. & Dahan, a. (2008), 'Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect.', *Br. J. Anaesth.* **101**(6), 810–821.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/18852113*

Pandit, J. J. & Cook, T. M. (2014), 'National Institute for Clinical Excellence guidance on measuring depth of anaesthesia: Limitations of EEG based technology', *Br. J. Anaesth.* **112**(2), 385–386.

Pearl, J. (1993), *Causality. Models, Reasoning and Inference*, Cambridge, United Kingdom.

Platt, J. C. (1999), 'Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods', *Adv. large margin Classif.* **10**(3), 61–174.

Prichep, L. S., Gugino, L. D., John, E. R., Chabot, R. J., Howard, B., Merkin, H., Tom, M. L., Wolter, S., Rausch, L. & Kox, W. J. (2004), 'The Patient State Index as an indicator of the level of hypnosis under general anaesthesia', *Br. J. Anaesth.* **92**(3), 393–399.

Purdon, P. L., Pierce, E. T., Bonmassar, G., Walsh, J., Harrell, P. G., Kwo, J., Deschler, D., Barlow, M., Merhar, R. C., Lamus, C., Mullaly, C. M., Sullivan, M., Maginnis, S., Skoniecki, D., Higgins, H. A. & Brown, E. N. (2009), 'Simultaneous electroencephalography and functional magnetic resonance imaging of general anesthesia', *Ann. N. Y. Acad. Sci.* **1157**, 61–70.

Purdon, P., Pierce, E., Mukamel, E., Prerau, M., Walsh, J., Wong, K., Salazar-Gomez, A., Harrell, P., Sampson, A., Cimenser, A., ShiNung, C., Kopell, N., Tavares-Stoeckel, C., Habeeb, K., Merhar, R. & Brown, E. (2013), 'Electroencephalogram signatures of loss and recovery of consciousness from propofol.', *Proc. Natl. Acad. Sci. U. S. A.* **110**(12), E1142–51.
URL: *http://www.pnas.org/content/110/12/E1142.long*

Quiroga, R. Q., Kraskov, A., Kreuz, T. & Grassberger, P. (2002), 'Performance of different synchronizazion measures in real data: A case study on electroencephalographic signals', *Phys. Rev. E* **65**, 1–14.

Rampil, I. (1998), 'A Primer for EEG Signal Processing in Anesthesia The Genesis of the EEG Introduction-The Rationale for Monitoring', *Anesthesiology* **89**(October), 980–1002.

Roberts, F. & Freshwater-Turner, D. (2007), 'Pharmacokinetics and anaesthesia', *Contin. Educ. Anaesthesia, Crit. Care Pain* **7**(1), 25–29.

Rubinov, M. & Sporns, O. (2010), 'Complex network measures of brain connectivity: Uses and interpretations', *Neuroimage* **52**(3), 1059–1069.
URL: *http://dx.doi.org/10.1016/j.neuroimage.2009.10.003*

Rulkov, N. f., Sushchik, M. M. & Tsimring, L. S. (1995), 'Generalized synchronization of chaos in directionally coupled chaotic systems', *Phys. Rev. E* **51**(981-994).

Sanders, R. D., Tononi, G., Laureys, S. & Sleigh, J. W. (2012), 'Unresponsiveness unconsciousness', *Anesthesiology* **116**(4), 946–59.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/23095543*

Sandin, R. H., Enlund, G., Samuelsson, P. & Lennmarken, C. (2000), 'Awareness during anaesthesia: a prospective case study.', *Lancet* **355**(9205), 707–11.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/10703802*

Sanei, S. & Chambers, J. A. (2007), *EEG Signal Processing*.

Schlögl, A. (2006), 'A comparison of multivariate autoregressive estimators', *Signal Processing* **86**(9), 2426–2429.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0165168405003993*

Schlögl, A. & Supp, G. (2006), 'Analyzing event-related EEG data with multivariate autoregressive parameters.', *Prog Brain Res.* **159**, 135–47.

Schneider, G., Gelb, A. W., Schmeller, B., Tschakert, R. & Kochs, E. (2003), 'Detection of awareness in surgical patients with EEG-based indices - Bispectral index and patient state index', *Br. J. Anaesth.* **91**(3), 329–335.

Schneider, G., Wagner, K., Reeker, W., Hänel, F., Werner, C. & Kochs, E. (2002), 'Bispectral Index (BIS) may not predict awareness reaction to intubation in surgical patients.', *J Neurosurg Anesth.* **14**(1), 7–11.

Schneider, T. & Neumaier, A. (2001), 'Algorithm 808 : AR FIT A Matlab Package for the Estimation of Parameters and Eigenmodes of Multivariate Autoregressive Models', *ACM Trans. Math. Softw.* **27**(March), 58–65.

Schreiber, T. (2000), 'Measuring information transfer', *Phys. Rev. Lett.* **85**(2), 461–4.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/10991308*

Schrouff, J., Perlbarg, V., Boly, M., Marrelec, G., Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A., Laureys, S., Phillips, C., Pélégrini-Issac, M., Maquet, P. & Benali, H. (2011), 'Brain functional integration decreases during propofol-induced loss of consciousness', *Neuroimage* **57**(1), 198–205.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21524704*

Shepherd, J., Jones, J., Bryant, J., Frampton, G., Baxter, L., Cooper, K., Shepherd, J., Jones, J., Bryant, J., Frampton, G., Baxter, L. & Cooper, K. (2012), Depth of anaesthesia monitoring (E-Entropy , Bispectral Index and Narcotrend), Technical Report 4.

Shepherd, J., Jones, J., Frampton, G. K., Bryant, J., Baxter, L. & Cooper, K. (2013), Health Technology Assessment. Clinical effectiveness and cost-effectiveness of depth of anaesthesia monitoring (E-Entropy, Bispectral Index and Narcotrend): a systematic review and economic evaluation, Technical Report 34.

Shimizu, S., Hoyer, P. O., Hyvarinen, A. & Kerminen, A. (2006), 'A Linear Non-Gaussian Acyclic Model for Causal Discovery', *J. Mach. Learn. Res.* **7**, 2003–2030.

Shimizu, S., Hyvärinen, a. & Kano, Y. (2005), 'Discovery of non-gaussian linear causal models using ICA', *Proc 21st Conf. ...* .

Sitt, J. D., King, J. R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A., Cohen, L., Sigman, M., Dehaene, S. & Naccache, L. (2014), 'Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state', *Brain* **137**(8), 2258–2270.

Spoormaker, V. I., Gleiser, P. M. & Czisch, M. (2012), 'Frontoparietal connectivity and hierarchical structure of the brain's functional network during sleep', *Front. Neurol.* **MAY**(May), 1–10.

Spoormaker, V. I., Schröter, M. S., Gleiser, P. M., Andrade, K. C., Dresler, M., Wehrle, R., Sämann, P. G. & Czisch, M. (2010), 'Development of a Large-Scale Functional Brain Network during Human Non-Rapid Eye Movement Sleep', *J. Neurosci.* **30**(34), 11379–11387.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/20739559*

Sporns, O. (2013), 'Structure and function of complex brain networks', *Dialogues Clin Neurosci* **15**, 247–262.

Srinivasan, R., Nunez, P. L. & Silberstein, R. B. (1998), 'Spatial filtering and neocortical dynamics: Estimates of EEG coherence', *IEEE Trans. Biomed. Eng.* **45**(7), 814–826.

Stedman, T. L. (2005), *Stedman's Medical Dictionary for the Health Professions and Nursing*, Philadelphia.

Sugimoto, H., Ishii, N., Iwata, A., Suzumura, N. & Tomita, T. (1978), 'On the stationarity and normality of the electroencephalographic data during sleep stages', *Comput Programs Biomed* (8), 224–234.

Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., Jahnke, K. & Laufs, H. (2013), 'Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep.', *Proc. Natl. Acad. Sci. U. S. A.* **110**(31), 15419–24.

Thatcher, R. W. (2012), 'Coherence , Phase Differences , Phase Shift , and Phase Lock in EEG / ERP Analyses', *Dev. Neuropsychol.* **37**(6), 476–496.

Thornton, A. R. D., Chambers, J. D. & Folkard, T. J. (1998), 'Deconvolution of MLS Response Data'.

Thornton, C., Barrowcliffe, M. P., Konieczko, K. M., Ventham, P., Dore, C. J., Newton, D. E. F. & Jones, J. G. (1989), 'The Auditory Evoked-Response as an Indicator of Awareness', *Br. J. Anaesth.* **63**(1), 113–115.

Thornton, C., Creagh-Barry, P., Jordan, C., Luff, N. P., Doré, C. J., Henley, M. & Newton, D. E. (1992), 'Somatosensory and auditory evoked responses recorded simultaneously: differential effects of nitrous oxide and isoflurane.', *Br. J. Anaesth.*

68(5), 508–14.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/1642941*

Thornton, C. & Sharpe, R. M. (1998), 'Evoked responses in anaesthesia.', *Br. J. Anaesth.* **81**(5), 771–81.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/10193293*

Tononi, G. (2008), 'Consciousness as Integrated Information', *Biol. Bull.* **215**(December), 216–242.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19098144*

Tononi, G. & Massimini, M. (2008), 'Why does consciousness fade in early sleep?', *Ann. N. Y. Acad. Sci.* **1129**, 330–334.

Tooley, M. A., Greenslade, G. L. & Prys-Roberts, C. (1996), 'Concentration-related effects of propofol on the auditory evoked response', *Br. J. Anaesth.* (77), 720–726.

Tooley, M. A., Stapleton, C. L., Greenslade, G. L. & Prys-Roberts, C. (2004), 'Mid-latency auditory evoked response during propofol and alfentanil anaesthesia', *Br. J. Anaesth.* **92**(1), 25–32.

Toppi, J., De Vico Fallani, F., Vecchiato, G., Maglione, A. G., Cincotti, F., Mattia, D., Salinari, S., Babiloni, F. & Astolfi, L. (2012), 'How the statistical validation of functional connectivity patterns can prevent erroneous definition of small-world properties of a brain connectivity network', *Comput. Math. Methods Med.* **2012**.

Untergehrer, G., Jordan, D., Kochs, E. F., Ilg, R. & Schneider, G. (2014), 'Fronto-parietal connectivity is a non-static phenomenon with characteristic changes during unconsciousness', *PLoS One* **9**(1).

Velly, L. J., Rey, M. F., Bruder, N. J., Gouvitsos, F. A., Witjas, T., Regis, J. M., Peragut, J. C. & Gouin, F. M. (2007), 'Differential dynamic of action on cortical and subcortical structures of anesthetic agents during induction of anesthesia.', *Anesthesiology* **107**(2), 202–12.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/17667563*

Weiskopf, R. B. (2000), 'Monitoring Depth of Anesthesia', *Anesthesiology* **93**(3), 876–882.

Winterhalder, M., Schelter, B., Hesse, W., Schwab, K., Leistritz, L., Klan, D., Bauer, R., Timmer, J. & Witte, H. (2005), 'Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems', *Signal Processing* **85**(11), 2137–2160.