



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΟΝΙΚΗΣ

Επίδραση της ανάλυσης ανεξάρτητων
συνιστωσών στην ανάλυση συνδεσιμότητας
του εγκεφάλου σε δεδομένα EEG

Διπλωματική εργασία
Αλέξανδρος Στεργιάδης

Επιβλέπων: Δημήτρης Κουγιουμτζής
Καθηγητής ΤΗΜΜΥ Α.Π.Θ.

23 Μαρτίου 2022

Περίληψη

Το ηλεκτροεγκεφαλογράφημα (EEG) είναι μία μέθοδος μελέτης της εγκεφαλικής δραστηριότητας που περιλαμβάνει τη χρήση ηλεκτροδίων πάνω στο κρανίο για την μέτρηση της ηλεκτρικής δραστηριότητας του εγκεφάλου. Λόγω της ευκολίας στη χρήση και της υψηλής ακρίβειας στο χρόνο έχει αναπτυχθεί σε μία από τις πιο διαδεδομένες τεχνικές μελέτης του εγκεφάλου. Ωστόσο υπάρχουν και μερικά μειονεκτήματα που συνδέονται με την ανάλυση EEG. Ένα από αυτά είναι η μόλυνση των σημάτων από θόρυβο που μπορεί να προέρχεται από εξωτερικές πηγές, όπως ο θόρυβος που εισάγεται από τη χρήση ελαττωματικών ηλεκτροδίων, ή από εσωτερικές πηγές όπως η ηλεκτρική δραστηριότητα που παράγεται από το ανοιγοκλείσιμο των ματιών ή από τους χαρδιακούς παλμούς. Αυτός ο θόρυβος είναι ιδιαίτερα σημαντικός καθώς συχνά τα σήματα θορύβου έχουν αρκετά μεγαλύτερη ένταση από την εγκεφαλική δραστηριότητα και ο θόρυβος συνήθως μολύνει όλα τα ηλεκτρόδια λόγω της αγωγιμότητας του εγκεφάλου.

Για αυτό το λόγο η ανίχνευση και αφαίρεση του θορύβου θεωρείται απαραίτητο στάδιο της προεπεξεργασίας δεδομένων EEG και έχουν αναπτυχθεί αρκετές μέθοδοι με αυτό το σκοπό, οι οποίες συνήθως επικεντρώνονται στην ανίχνευση και αφαίρεση ενός ή μερικών ειδών θορύβου. Μία οικογένεια μεθόδων καθαρισμού είναι αυτή του τυφλού διαχωρισμού πηγών και ιδιαίτερα της ανάλυσης ανεξάρτητων συνιστώσων / μεταβλητών (ICA). Οι αλγόριθμοι ICA μετασχηματίζουν τα σήματα των αισθητήρων σε ένα γραμμικό σύστημα ανεξάρτητων, μη Γκαουσιανών πηγών.

Αν και η ICA είναι πλέον μία από τις πιο δημοφιλείς μεθόδους καθαρισμού θορύβου σε σήματα EEG, η χρήση της δε συστένεται σε κάθε περίπτωση. Αρχικά, υπάρχει αμφιβολία σχετικά με το κατά πόσο οι υποθέσεις του ολγορίθμου καθώς και ο τρόπος με τον οποίο υπολογίζονται οι ανεξάρτητες συνιστώσες είναι συμβατές με τον τρόπο που λειτουργεί ο εγκέφαλος. Επιπλέον, έχει δειχθεί ότι η χρήση της οδηγεί και στην απόρριψη χρήσιμης εγκεφαλικής δραστηριότητας. Τα προβλήματα αυτά εισάγουν ορισμένες δυσκολίες στην ερμηνεία σημάτων EEG μετά από καθαρισμό δεδομένων με χρήση ICA, ειδικά όταν ο σκοπός της ανάλυσης είναι η μελέτη της συνδεσιμότητας του εγκεφάλου.

Στην παρούσα εργασία μελετάται η επίδραση του καθαρισμού δεδομένων EEG με ICA στη δραστική συνδεσιμότητα του εγκεφάλου (effective brain connectivity) και ιδιαίτερα στις αιτιακές σχέσεις μεταξύ χρονοσειρών δραστηριότητας από διαφορετικές εγκεφαλικές περιοχές. Η ανάλυση συνδεσιμότητας γίνεται και στο επίπεδο των ηλεκτροδίων και στο επίπεδο των υποκείμενων εγκεφαλικών πηγών. Ως δραστική συνδεσιμότητα ορίζονται οι σχέσεις επιρροής μεταξύ κόμβων ενός δικτύου που μοντελοποιεί δυναμική αιτιότητας.

Η μελέτη έγινε στο περιβάλλον προγραμματισμού Matlab με χρήση προσο-

μοιωτικών και πραγματικών δεδομένων και περιλαμβάνει την ανάλυση συνδεσιμότητας σε καθαρά δεδομένα EEG, δεδομένα EEG μολυσμένα από θόρυβο και δεδομένα EEG καθαρισμένα με μία από τέσσερεις μεθόδους: ICA (συγκεχριμένα τον αλγόριθμο FastICA), wavelet enhanced ICA (wICA), stripped ICA (sICA) και noise mixing approximation (NMA). Η δομή της εργασίας διαμορφώνεται ως εξής: αρχικά, στο Κεφάλαιο 1 γίνεται μία εισαγωγή, στην οποία παρατίθεται το θεωρητικό υπόβαθρο του EEG, των μεθόδων καθαρισμού και των μετρικών που χρησιμοποιούνται. Στη συνέχεια, στο Κεφάλαιο 2 αναλύεται το υλικό και οι μέθοδοι που χρησιμοποιήθηκαν. Στο Κεφάλαιο 3 περιγράφεται η μεθοδολογία των πειραμάτων και δίνονται τα αποτελέσματα ενός πειράματος ως παράδειγμα. Κατόπιν, στο Κεφάλαιο 4 αναφέρονται όλα τα αποτελέσματα από τα πειράματα σε προσομοιωτικά δεδομένα. Στο κεφάλαιο 5 αναλύεται η εφαρμογή των μεθόδων σε πραγματικά δεδομένα από δύο πειράματα. Τέλος, στο κεφάλαιο 6 δίνεται μία περίληψη των συμπερασμάτων και γίνονται μερικές προτάσεις για μελλοντικές έρευνες.

Abstract

EEG is a method of studying the brain's electrical activity that involves the usage of electrodes placed either directly on the scalp or on an elastic cap that is then placed on the scalp. Thanks to its ease of use and high temporal precision, it has evolved into one of the most commonly used methods of recording brain activity. However, there are some drawbacks associated with EEG analysis. One of them is the almost certain contamination of EEG signals by noise which can originate from external sources, e.g. noise originating from faulty electrodes, or from internal sources, e.g. noise originating from eyeblinks or heartbeats. The amplitude of the signals which can be classified as noise is often comparable with that of neural signals. Moreover, noise typically affects all electrodes due to the effects of volume conduction. For these reasons detecting and cleaning noise is considered a vital part of data pre-processing in EEG analysis and many methods have been developed for this purpose. One group of such methods is Blind Source Separation (BSS) and a method belonging to that group is Independent Component Analysis (ICA). ICA algorithms transform electrode signals into a linear system of independent, non-Gaussian sources or components.

Even though ICA is one of the most widely used methods of denoising EEG signals, its use is not always recommended. Firstly, it is uncertain whether the algorithm's hypotheses and function are compatible with the way the brain works. Moreover, it's been shown that its use leads to the rejection of useful, neural information.

This uncertainty introduces certain challenges in the interpretation of EEG recordings after denoising with ICA, especially when the goal is brain connectivity analysis.

The purpose of the present study is to examine the effects of ICA denoising on effective brain connectivity analysis with an emphasis on causal relationships between distinct brain regions. Connectivity analysis is performed on both the electrode level and underlying brain sources level. Effective brain connectivity is defined as the causal relationships between nodes of a network modelling causal dynamics.

Analysis was done using the Matlab programming language, using both simulated as well as real data. Connectivity analysis was performed on the initial, clean EEG, the noisy EEG and EEGs produced through cleaning with one of four denoising methods : ICA (specifically the FastICA algorithm), wavelet enhanced ICA (wICA), stripped ICA (sICA) and noise mixing approximation (NMA).

The structure of this thesis is formulated as follows: in chapter 1, an introduction is made to the theoretical background of EEG, denoising methods and connectivity metrics. The methods and materials used are presented in chapter 2. In chapter 3, the methodology of the experiments is described and the results of one experiment are presented as an example. In chapter 4, the results of experiments on simulated data are reported. Chapter 5 includes the results of our analysis on real data. Finally, in chapter 6 a summary of the results is provided and ideas for future research are suggested.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον καθηγητή Δημήτρη Κουγιουμτζή για την εμπιστοσύνη που μου έδειξε αναθέτοντας μου το θέμα της παρούσας διπλωματικής καθώς και για την υποστήριξή του καθ' όλη τη διάρκεια της εκπόνισής της.

Ευχαριστώ ακόμη την οικογένειά μου για την υποστήριξη και καθοδήγησή τους κατά τα φοιτητικά μου χρόνια.

Περιεχόμενα

1 Εισαγωγή - Θεωρία	1
1.1 EEG	1
1.2 Μέθοδοι τυφλού διαχωρισμού	4
1.3 Ανάλυση ανεξάρτητων συνιστώσων	5
1.4 FastICA	7
1.5 Εναλλακτικές και προεκτάσεις της ICA	9
1.5.1 Wavelet enhanced ICA (wICA)	9
1.5.2 Stripped ICA (sICA)	10
1.5.3 Noise mixing approximation (NMA)	10
1.6 Forward and inverse problems	12
1.7 Αυτοπαλίνδρομα μοντέλα - Autoregressive (AR) models	13
1.8 Μετρικές αιτιότητας	14
1.8.1 (υπό συνθήκη) Αιτιότητα κατά Granger / (conditional) Granger causality	15
1.8.2 Επεκτάσεις του (Conditional) Granger causality	16
1.9 Θεωρία γράφων	17
2 Τλικό και μέθοδοι	18
2.1 Τοποθέτηση ηλεκτροδίων	18
2.2 Λύση των forward και inverse problems	19
2.3 Προσομοιωτικά δεδομένα πηγών	19
2.4 Θόρυβος	20
2.5 Καθαρισμός EEG	22
3 Μεθοδολογία πειραμάτων	24
3.1 Περιγραφή πειράματος	24
3.2 Περιγραφή μετρήσεων	26
3.3 Παράδειγμα ανάλυσης πειράματος	31

4 Αποτελέσματα και συζήτηση	40
4.1 Αποτελέσματα στο επίπεδο των 7 περιοχών	40
4.2 Αποτελέσματα στο επίπεδο των 68 περιοχών	43
4.3 Αποτελέσματα στο επίπεδο των 64 ηλεκτροδίων	46
5 Πειράματα σε πραγματικά δεδομένα	52
5.1 Πραγματικά δεδομένα με προσομοιωτικό θόρυβο	53
5.2 Πραγματικά δεδομένα	59
6 Συμπεράσματα και μελλοντικές επεκτάσεις	64
6.1 Περίληψη συμπερασμάτων	64
6.2 Μελλοντικές επεκτάσεις	65
7 Βιβλιογραφία	66

Κατάλογος Σχημάτων

1.1	Είδη όμορύβου σε σήματα EEG	3
1.2	Γραφική απεικόνιση της wICA	10
1.3	Stripped ICA σετ ηλεκτροδίων	11
2.1	2D τοποθέτηση ηλεκτροδίων	18
2.2	3D τοποθέτηση ηλεκτροδίων	18
2.3	Προσομοιωτική χρονοσειρά όμορύβου	21
2.4	Τοπογραφία όμορύβου	22
3.1	Καθαρά δεδομένα, 8 ηλεκτρόδια	25
3.2	Θορυβώδη δεδομένα, 8 ηλεκτρόδια	25
3.3	Καθαρισμένα δεδομένα, 8 ηλεκτρόδια	25
3.4	Confusion matrix από binary classification	28
3.5	Δίκτυο CGCI, καθαρά δεδομένα 7 περιοχών	32
3.6	Δίκτυο PDC, καθαρά δεδομένα 7 περιοχών	33
3.7	Δίκτυο rCGCI, καθαρά δεδομένα 68 περιοχών	33
3.8	Δίκτυο rCGCI, καθαρά δεδομένα 64 ηλεκτροδίων	34
3.9	Δίκτυο rCGCI, θορυβώδη δεδομένα 64 ηλεκτροδίων	36
3.10	Misclassifications rCGCI, θορυβώδη δεδομένα 64 ηλεκτροδίων	37
4.1	Κατανομή misclassifications CGCI, 7 περιοχές, 850 παρατηρήσεις	42
4.2	Κατανομή misclassifications PDC, 7 περιοχές, 850 παρατηρήσεις	43
4.3	Κατανομή false negative PDC, 7 περιοχές, 850 παρατηρήσεις . .	44
4.4	Κατανομή false negative RCGCI, 68 περιοχές, 850 παρατηρήσεις	46
4.5	Κατανομή false positive RCGCI, 68 περιοχές, 850 παρατηρήσεις	47
4.6	50 πιο συχνά misclassifications, 64 ηλεκτρόδια, 5000 παρατηρήσεις	49
4.7	Κατανομή απόλυτου localization error, 64 ηλεκτρόδια	51
5.1	Δίκτυο RCGCI καθαρά δεδομένα 64 ηλεκτροδίων, 5000 παρατηρήσεις	55
5.2	Δίκτυο RCGCI θορυβώδη δεδομένα 64 ηλεκτροδίων, 5000 παρατηρήσεις	55

5.3 Δίκτυο RCGCI καθαρά δεδομένα 64 ηλεκτροδίων, 850 παρατηρήσεις	55
5.4 Δίκτυο RCGCI θορυβώδη δεδομένα 64 ηλεκτροδίων, 850 παρατηρήσεις	55
5.5 Δίκτυο RCGCI καθαρά δεδομένα 68 περιοχών, 850 παρατηρήσεις	56
5.6 Δίκτυο RCGCI θορυβώδη δεδομένα 68 περιοχών, 850 παρατηρήσεις	56
5.7 Κατανομή απόλυτου localization error, 64 ηλεκτρόδια	60
5.8 Δίκτυο rCGCI, επίπεδο ηλεκτροδίων: θορυβώδη δεδομένα	61
5.9 Δίκτυο rCGCI, επίπεδο ηλεκτροδίων: καθαρισμός με ICA	61
5.10 Δίκτυο rCGCI, επίπεδο ηλεκτροδίων: καθαρισμός με sICA	61
5.11 Δίκτυο rCGCI, επίπεδο ηλεκτροδίων: καθαρισμός με NMA	61
5.12 Δίκτυο rCGCI, επίπεδο περιοχών: θορυβώδη δεδομένα	62
5.13 Διαφοροποιήσεις δικτύου rCGCI, επίπεδο περιοχών: καθαρισμός με ICA	62
5.14 Διαφοροποιήσεις δικτύου rCGCI, επίπεδο περιοχών: καθαρισμός με sICA	62
5.15 Διαφοροποιήσεις δικτύου rCGCI, επίπεδο περιοχών: καθαρισμός με NMA	62
5.16 Ηλεκτρόδιο FC1, θορυβώδη δεδομένα	63
5.17 Ηλεκτρόδιο FC1, καθαρισμός με sICA	63

Κατάλογος Πινάκων

3.1	Επίπεδα μετρήσεων	26
3.2	Πίνακας γειτνίασης μοντέλων VAR	31
3.3	Confusion matrices rCGCI, θιρυβώδη δεδομένα 68 περιοχών . .	35
3.4	Confusion matrices rCGCI, θιρυβώδη δεδομένα 64 ηλεκτροδίων	36
3.5	Μετρικές ταξινόμησης, 7 περιοχές, 5000 παρατηρήσεις	37
3.6	Μετρικές ταξινόμησης, 68 περιοχές, 5000 παρατηρήσεις	38
3.7	Μετρικές ταξινόμησης, 64 ηλεκτρόδια, 5000 παρατηρήσεις	38
3.8	Μέσο τετραγωνικό σφάλμα 6 ηλεκτροδίων, 5000 παρατηρήσεις .	38
4.1	Μέσος όρος μετρικών ταξινόμησης, 7 περιοχές, 5000 παρατηρήσεις	40
4.2	Μέσος όρος μετρικών ταξινόμησης, 7 περιοχές, 1000 παρατηρήσεις	41
4.3	Μέσος όρος μετρικών ταξινόμησης, 7 περιοχές, 850 παρατηρήσεις	41
4.4	Μέσος όρος μετρικών ταξινόμησης, 68 περιοχές, 5000 παρατη- ρήσεις	44
4.5	Μέσος όρος μετρικών ταξινόμησης, 68 περιοχές, 1000 παρατη- ρήσεις	44
4.6	Μέσος όρος μετρικών ταξινόμησης, 68 περιοχές, 850 παρατηρήσεις	45
4.7	Αθροισμένα confusion matrices rCGCI, 68 περιοχές, 5000 πα- ρατηρήσεις	45
4.8	Αθροισμένα confusion matrices rCGCI, 68 περιοχές, 1000 πα- ρατηρήσεις	46
4.9	Αθροισμένα confusion matrices rCGCI, 68 περιοχές, 850 πα- ρατηρήσεις	46
4.10	Μέσος όρος μετρικών ταξινόμησης, 64 ηλεκτρόδια, 5000 παρα- τηρήσεις	47
4.11	Μέσος όρος μετρικών ταξινόμησης, 64 ηλεκτρόδια, 1000 παρα- τηρήσεις	48
4.12	Μέσος όρος μετρικών ταξινόμησης, 64 ηλεκτρόδια, 850 παρατη- ρήσεις	48
4.13	Μέσο τετραγωνικό σφάλμα, 5000 παρατηρήσεις	49
4.14	Μέσο τετραγωνικό σφάλμα, 1000 παρατηρήσεις	50

4.15	Μέσο τετραγωνικό σφάλμα, 850 παρατηρήσεις	50
4.16	Mean absolute localization error	50
5.1	Confusion matrices rCGCI, 64 ηλεκτρόδια, 5000 παρατηρήσεις .	56
5.2	Confusion matrices rCGCI, 64 ηλεκτρόδια, 1000 παρατηρήσεις .	56
5.3	Confusion matrices rCGCI, 64 ηλεκτρόδια, 850 παρατηρήσεις .	56
5.4	Μετρικές ταξινόμησης, 64 ηλεκτρόδια, 5000 παρατηρήσεις . . .	57
5.5	Μετρικές ταξινόμησης, 64 ηλεκτρόδια, 1000 παρατηρήσεις . . .	57
5.6	Μετρικές ταξινόμησης, 64 ηλεκτρόδια, 850 παρατηρήσεις . . .	57
5.7	Confusion matrices rCGCI, 68 περιοχές, 5000 παρατηρήσεις . .	57
5.8	Confusion matrices rCGCI, 68 περιοχές, 1000 παρατηρήσεις . .	57
5.9	Confusion matrices rCGCI, 68 περιοχές, 850 παρατηρήσεις . .	58
5.10	Μετρικές ταξινόμησης, 68 περιοχές, 5000 παρατηρήσεις . . .	58
5.11	Μετρικές ταξινόμησης, 68 περιοχές, 1000 παρατηρήσεις . . .	58
5.12	Μετρικές ταξινόμησης, 68 περιοχές, 850 παρατηρήσεις . . .	58
5.13	Μέσο τετραγωνικό σφάλμα, 5000 παρατηρήσεις	59
5.14	Μέσο τετραγωνικό σφάλμα, 1000 παρατηρήσεις	59
5.15	Μέσο τετραγωνικό σφάλμα, 850 παρατηρήσεις	59

Κεφάλαιο 1

Εισαγωγή - Θεωρία

Στο συγκεκριμένο κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο που είναι απαραίτητο για την κατανόηση της παρούσας εργασίας. Αρχικά γίνεται μία σύντομη αναφορά στην έννοια και τον τρόπο λειτουργίας του ηλεκτροεγκεφαλογραφήματος (EEG) καθώς και στην έννοια του θορύβου στο πλαίσιο του EEG. Στη συνέχεια εισάγονται αρχικά οι μέθοδοι τυφλού διαχωρισμού και στη συνέχεια η ICA, ο αλγόριθμος FastICA και μερικές εναλλακτικές μέθοδοι καθαρισμού. Έπειτα εξηγούνται οι έννοιες forward problem και inverse problem. Τέλος γίνεται μία αναφορά στα αυτοπαλίνδρομα διανυσματικά μοντέλα, στις μετρικές αιτιότητας και στη θεωρία γράφων.

1.1 EEG

Η ανακάλυψη των προοπτικών του EEG ξεκινάει από το 1877 με την έρευνα του Richard Caton στους εγκεφάλους κουνελιών και μαϊμούδων. Η πρώτη καταγραφή ανθρώπινης εγκεφαλικής δραστηριότητας με EEG θα δημοσιεύονταν το 1924 από τον Hans Berger [2]. Ο Hans ήταν και ο πρώτος που παρατήρησε ότι τα εγκεφαλικά σήματα παρουσιάζουν σημαντική διαφοροποίηση όταν η ψυχολογική κατάσταση του υποκειμένου αλλάζει, για παράδειγμα από χαλάρωση σε εγρήγορση.

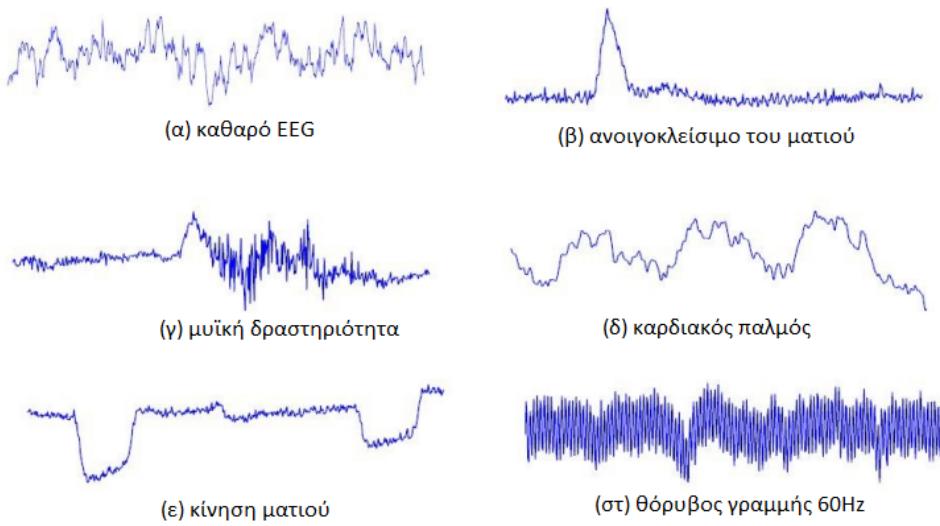
Το EEG μετρά την ηλεκτρική δραστηριότητα που παράγεται από τον εγκέφαλο χρησιμοποιώντας ηλεκτρόδια που τοποθετούνται είτε κατευθείαν στο κρανίο είτε πάνω σε έναν ελαστικό σκούφο που θα τοποθετηθεί στο κρανίο (non invasive EEG). Συχνά εφαρμόζεται και κάποιο αγώγιμο τζελ για τη βελτίωση της επαφής ηλεκτροδίου και κρανίου. Ο αριθμός των ηλεκτροδίων εξαρτάται από το πείραμα και κυμαίνεται από 4 μέχρι και 256 ηλεκτρόδια. Η τοποθέτηση των ηλεκτροδίων γίνεται συνήθως σύμφωνα με το 10-20 international system,

το οποίο είναι ένα παγκοσμίως αναγνωρισμένο στάνταρ επιλογής της θέσης των ηλεκτροδίων στο πλαίσιο των πειραμάτων EEG.

Καθώς τα ηλεκτρόδια μετράνε τάση, η μέτρηση σε κάθε σημείο του χρανίου γίνεται ως προς κάποιο άλλο σημείο, δηλαδή μετράται η διαφορά δυναμικού μεταξύ ενός ηλεκτροδίου και ενός σημείου αναφοράς. Το σημείο αναφοράς μπορεί να είναι διαφορετικό για κάθε ηλεκτρόδιο (*bipolar montage*) ή κοινό για όλα τα ηλεκτρόδια (*referential montage*). Στη δεύτερη περίπτωση, το σημείο αναφοράς μπορεί να είναι κάποιο ηλεκτρόδιο ή ο μέσος όρος όλων των μετρήσεων (*average reference*).

Η εγκεφαλική ηλεκτρική δραστηριότητα που μετράνε τα ηλεκτρόδια παράγεται από πληθυσμούς νευρώνων του εγκεφάλου. Ωστόσο, εκτός από αυτήν, οι μετρήσεις των ηλεκτροδίων επηρεάζονται και από εξωγενείς ως προς τον εγκέφαλο παράγοντες (*artifacts*). Αυτοί μπορεί να προέρχονται από πηγές εσωτερικά του υποκειμένου (*physiological*), για παράδειγμα κινήσεις του ματιού, της καρδιάς, κλπ. ή να είναι εξωτερικοί (*non physiological*), για παράδειγμα ο θόρυβος που εισάγει το όργανο μέτρησης, θόρυβος από το περιβάλλον, κλπ. Μερικά παραδείγματα artifact που εμφανίζονται συχνά σε μετρήσεις EEG είναι τα παρακάτω: [3]

- Ανοιγοκλείσιμο του ματιού (*eyeblink artifact*) : εμφανίζεται πολύ συχνά σε δεδομένα EEG και χαρακτηρίζεται από ένα spike μεγάλου πλάτους το οποίο συνήθως είναι πολλές φορές μεγαλύτερο από τα σήματα ενδιαφέροντος. Αναφέρεται ότι και οι υπόλοιπες κινήσεις των ματιών μπορούν να προκαλέσουν ηλεκτρική δραστηριότητα που θα θεωρούνται θόρυβος στο πλαίσιο του EEG, αλλά οι δύο αυτοί τύποι θορύβου έχουν αρκετά διαφορετική μορφή.
- Σήματα γραμμής (*line noise / power line interference*): Θόρυβος που εκφράζεται από υψηλή ισχύ σε συγκεκριμένες συχνότητες (50 ή 60 Hz) και οφείλεται συνήθως σε ελαττωματικά ηλεκτρόδια η κακή επαφή με το χρανίο.
- Μική δραστηριότητα: προέρχεται από την ηλεκτρική δραστηριότητα που παράγεται κατά τις κινήσεις των μυών, για παράδειγμα στο λαιμό ή στο πρόσωπο.
- Καρδιακός παλμός: προέρχεται από την ηλεκτρική δραστηριότητα που παράγεται κατά τον καρδιακό παλμό.



Σχήμα 1.1: Είδη ψευδών σε σήματα EEG

Η χαρακτηριστική μορφή των παραπάνω φαίνεται στο σχήμα 1.1

Όπως αναφέρθηκε προηγουμένως, η αντιμετώπιση των artifacts είναι ένα σημαντικό κομμάτι του pre-processing για τα δεδομένα EEG και έχουν προταθεί αρκετές μέθοδοι προς αυτό το σκοπό. Ένας από τους πρώτους τρόπους αντιμετώπισης των artifacts που χρησιμοποιήθηκε και συνεχίζει να χρησιμοποιείται ακόμη και σήμερα, είναι η απλή απόρριψη των παρατηρήσεων που επηρεάζονται από το ψευδό, δηλαδή η διαγραφή των περιόδων των μετρήσεων κατά τις οποίες εμφανίζεται artifactual δραστηριότητα. Ωστόσο, με αυτόν τον τρόπο προφανώς χάνεται και η καταγραφή χρήσιμης, εγκεφαλικής δραστηριότητας.

Για αυτό το λόγο, έχουν αναπτυχθεί μέθοδοι που αναγνωρίζουν και αφαιρούν το σήμα artifact ενώ διατηρούν την εγκεφαλική δραστηριότητα. Μία οικογένεια μεθόδων που χρησιμοποιείται για αυτό το σκοπό είναι αυτή των μεθόδων τυφλού διαχωρισμού (BSS). Οι μέθοδοι τυφλού διαχωρισμού επιχειρούν μέσω ενός μετασχηματισμού των μετρήσεων να ανακτήσουν τα σήματα των πηγών από τις οποίες προέρχεται η δραστηριότητα που μετράται. Τα σήματα που υπολογίζονται με αυτές τις μεθόδους, μπορούν να αντιστοιχηθούν είτε σε χρήσιμες, πραγματικές πηγές, είτε σε artifactual πηγές. Στη δεύτερη περίπτωση, τα σήματα που αντιστοιχούν σε ψευδό μπορούν να αφαιρεθούν κατά τον αντίστροφο μετασχηματισμό που θα μας δώσει τα καθαρισμένα δεδομένα των μετρήσεων.

Μία από τις πιο ευρέως χρησιμοποιούμενες μεθόδους BSS είναι η ανάλυση ανεξάρτητων συνιστωσών (independent component analysis - ICA) η οποία μετατρέπει ένα πολυμεταβλητό σήμα σε ένα σύνολο μη Γκαουσιανών και στατιστικά ανεξάρτητων σημάτων (independent components - ICs) μέσω ενός γραφικού μετασχηματισμού.

1.2 Μέθοδοι τυφλού διαχωρισμού

Οι μέθοδοι τυφλού διαχωρισμού είναι μία στατιστική προσέγγιση ανάκτησης των σημάτων πηγών από σήματα μετρήσεων. [2] Η λέξη τυφλού αναφέρεται στο ότι δεν χρειάζεται να έχουμε πληροφορίες σχετικά με τις πηγές. Οι μέθοδοι BSS προουποθέτουν ότι οι μετρήσεις είναι ένας γραμμικός συνδυασμός των πηγών και ότι οι πηγές είναι ασυχέτιστες. Στο πλαίσιο του EEG, το παραπάνω σημαίνει ότι το κάθε ηλεκτρόδιο καταγράφει μία γραμμική μίξη σημάτων από διαφορετικές πηγές του εγκεφάλου. Οι μέθοδοι BSS χρησιμοποιούν τις μετρήσεις όλων των ηλεκτροδίων για να ανακατασκευάσουν τα σήματα των πηγών με τέτοιο τρόπο ώστε οι πηγές αυτές να είναι ασυχέτιστες μεταξύ τους. Ένα μοντέλο BSS συνήθως μπορεί να εκφραστεί από την παρακάτω εξίσωση:

$$\mathbf{X}(t) = \mathbf{A} \times \mathbf{S}(t) \quad (1.1)$$

Όπου $\mathbf{X}(t)$ είναι το $(m \times 1)$ στοιχείο του \mathbf{X} για χρονική t , \mathbf{X} είναι ο $(m \times n)$ πίνακας δεδομένων των μετρήσεων, m και n είναι ο αριθμός των αισθητήρων και ο αριθμός των παρατηρήσεων αντίστοιχα, \mathbf{A} είναι ο (άγνωστος) πίνακας μίξης (mixing matrix) διαστάσεων $(m \times m)$, $\mathbf{S}(t)$ είναι το $(m \times 1)$ στοιχείο του \mathbf{S} για χρονική t και \mathbf{S} είναι ο (άγνωστος) $(m \times n)$ πίνακας των χρονοσειρών της δραστηριότητας των πηγών. Σε αυτό το σημείο γίνεται και μία υπόθεση για τον αριθμό των πηγών. Συγκεκριμένα, οι μέθοδοι BSS υποθέτουν ότι ο αριθμός των πηγών είναι μικρότερος ή ίσος με τον αριθμό των αισθητήρων.

Προκειμένου να υπολογιστούν οι πηγές $\mathbf{S}(t)$ είναι αναγκαίος ο υπολογισμός του un-mixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ έτσι ώστε

$$\hat{\mathbf{S}}(t) = \mathbf{W} \times \mathbf{X}(t) \quad (1.2)$$

Ο un-mixing matrix ουσιαστικά μας δείχνει την επιρροή της κάθε πηγής στον κάθε αισθητήρα.

Αφότου υπολογιστούν οι χρονοσειρές των πηγών, αυτές ταξινομούνται ως θόρυβος ή χρήσιμη πληροφορία. Οι ανεπιθύμητες πηγές (θόρυβος) μπορούν

να μηδενιστούν και στη συνέχεια να γίνει ο αντίστροφος μετασχηματισμός των υπόλοιπων πηγών για να ανακτηθούν τα (καθαρισμένα) δεδομένα. Αυτό γίνεται χρησιμοποιώντας τον ανανεωμένο πίνακα μίζης $\hat{\mathbf{A}}$ στον οποίο έχουμε μηδενίσει τις στήλες που αντιστοιχούν σε πηγές θορύβου.

$$\hat{\mathbf{X}}(t) = \hat{\mathbf{A}} \times \hat{\mathbf{S}}(t) \quad (1.3)$$

Έχουν αναπτυχθεί αρκετοί αλγόριθμοι BSS που χρησιμοποιούν διάφορες προσεγγίσεις για να υπολογίσουν τον un-mixing matrix \mathbf{W} και τις πηγές $\hat{\mathbf{S}}(t)$. Όπως αναφέρθηκε προηγουμένως, μία από τις πιο ευρέως χρησιμοποιούμενες μεθόδους και αυτή με την οποία ασχολείται η παρούσα εργασία είναι αυτή της ICA.

1.3 Ανάλυση ανεξάρτητων συνιστωσών

Η ICA είναι μία μέθοδος BSS που χρησιμοποιεί στατιστική ανώτερης τάξης για να υπολογίσει τις πηγές από τα δεδομένα μέσω ενός γραμμικού μετασχηματισμού. Οι πηγές που υπολογίζονται από την ICA (independent components - ICs) είναι στατιστικά ανεξάρτητες και μη Γκαουσιανές. Έχουν αναπτυχθεί αρκετοί αλγόριθμοι ICA οι οποίοι λειτουργούν είτε μεγιστοποιώντας κάποιο μέτρο non Gaussianity είτε ελαχιστοποιώντας κάποιο μέτρο της εξάρτησης των υπολογισμένων ICs. Στη συνέχεια γίνεται μία αναφορά στις αρχές και υποθέσεις της ICA στο πλαίσιο του EEG.

- Γραμμική αναπαράσταση - Η ICA υποθέτει ότι οι μετρήσεις παράγονται από ένα γραμμικό συνδυασμό ανεξάρτητων πηγών. Στο πλαίσιο του EEG αυτό σημαίνει ότι τα ηλεκτρόδια μετράνε έναν γραμμικό συνδυασμό της ηλεκτρικής δραστηριότητας των διαφόρων πηγών εντός του εγκεφάλου που μεταδίδεται μέχρι το ηλεκτρόδιο χάρις την αγωγιμότητα του κεφαλιού.
- Άμεσο μοντέλο μίζης - Η ICA θεωρεί ότι η μετάδοση του σήματος από την πηγή στο όργανο μέτρησης είναι άμεση. Στην πραγματικότητα, αυτή η υπόθεση είναι λανθασμένη καθώς υπάρχει μία χρονική καθυστέρηση κατά τη μετάδοση σήματος μέσα στο εγκέφαλο ή πάνω στο κρανίο. Ωστόσο αυτή η καθυστέρηση είναι αμελητέα σε σχέση με την περίοδο δειγματοληψίας των οργάνων μέτρησης. Ως εκ τούτου, η υπόθεση ότι τα σήματα μετρώνται από όλα τα ηλεκτρόδια την ίδια στιγμή στην οποία παράγονται γενικά είναι επιτρεπτή στην ανάλυση EEG και αυτός ο τύπος ανάλυσης ονομάζεται instantaneous EEG recording.

- Χωρική στασιμότητα - Οι πηγές που υπολογίζονται από την ICA θεωρούνται στάσιμες στο χώρο. Αυτή η υπόθεση θεωρητικά πληρείται όταν η ICA εφαρμόζεται σε δεδομένα EEG καθώς οι πηγές που υπολογίζονται θεωρούνται ως η δραστηριότητα πληθυσμών νευρώνων σε συγκεκριμένες περιοχές του εγκεφάλου.
- Ανεξάρτησία - Οι πηγές που υπολογίζονται από την ICA είναι στατιστικά ανεξάρτητες και μη Γκαουσιανές. Αυτή η υπόθεση δεν φαίνεται να πληρείται όσον αφορά την εγκεφαλική δραστηριότητα, καθώς για παράδειγμα οι διάφοροι πληθυσμοί νευρώνων εντός του εγκεφάλου συχνά εμφανίζουν συγχρονισμένη δραστηριότητα. Επιπλέον, σύμφωνα με το central limit theorem (CLT), ένα γραμμικό άνθροισμα πολλών ανεξάρτητων τυχαίων μεταβλητών τείνει να είναι Γκαουσιανό. Καθώς ο αριθμός των νευρώνων εντός του εγκεφάλου και σε κάθε περιοχή του είναι πολύ μεγάλος, σύμφωνα με το παραπάνω θεώρημα η δραστηριότητα ενός πληθυσμού νευρώνων θα τείνει να έχει Γκαουσιανή κατανομή. Ωστόσο, οι υποθέσεις ισχύουν για πηγές artifact καθώς οι περισσότερες μη εγκεφαλικές πηγές ηλεκτρικής δραστηριότητας στο EEG έχουν μη Γκαουσιανή κατανομή και είναι ανεξάρτητες από την εγκεφαλική δραστηριότητα (και μεταξύ τους). Επομένως η υπόθεση πληρείται όσον αφορά τις πηγές θορύβου.

Όπως αναφέρθηκε, υπάρχουν διάφοροι αλγόριθμοι ICA που χρησιμοποιούν διαφορετικά μέτρα για να μεγιστοποιήσουν την ανεξάρτησία ή το non-Gaussianity των ICs. Ωστόσο, υπάρχουν δύο αρχικά βήματα που κάνουν σχεδόν όλοι οι αλγόριθμοι ICA για την απλοποίηση των δεδομένων πριν τον υπολογισμό του un-mixing matrix W . Αυτά είναι το κεντράρισμα (centering) και η λεύκανση (whitening).

Το κεντράρισμα είναι η αφαίρεση της μέσης τιμής από ένα παρατηρούμενο σήμα X . Σκοπός του κεντραρίσματος είναι ο μετασχηματισμός του παρατηρούμενου σήματος σε ένα σήμα με μηδενική μέση τιμή.

Η λεύκανση είναι ένας απλός γραμμικός μετασχηματισμός κατά τον οποίο τα κανάλια ενός πολυμεταβλητού παρατηρούμενου σήματος X μετασχηματίζονται σε ασυσχέτιστες μεταβλητές με διακύμανση (variance) ίση με 1. Η λεύκανση πραγματοποιείται πολλαπλασιάζοντας το παρατηρούμενο σήμα X με έναν πίνακα έτσι ώστε ο πίνακας συνδιασποράς του νέου σήματος \tilde{x} να είναι ο μοναδιαίος πίνακας [4], δηλαδή

$$E\{\tilde{x}\tilde{x}^T\} = \mathbf{I} \quad (1.4)$$

Αυτό επιτυγχάνεται συνήθως μέσω διάσπασης ιδιοτιμών (Eigen decomposition) του πίνακα συνδιασποράς του κεντραρισμένου σήματος X , σύμφωνα με την

οποία έχουμε

$$E\{\mathbf{XX}^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^{-1} \quad (1.5)$$

Όπου \mathbf{E} ο ορθογώνιος πίνακας των ιδιοδιανυσμάτων του $E\{\mathbf{XX}^T\}$ και \mathbf{D} ο διαγώνιος πίνακας των ιδιοτιμών $\mathbf{D} = diag(d_1, d_2, \dots, d_n)$.

Από τα παραπάνω προκύπτει ότι το μετασχηματισμένο whitened σήμα δίνεται από τον τύπο

$$\tilde{\mathbf{x}} = \mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{X} \quad (1.6)$$

Μερικοί από τους πιο διαδεδομένους αλγόριθμους ICA είναι οι: AMICA, Infomax, extended-Infomax και FastICA. Στην παρούσα εργασία χρησιμοποιήθηκε ο αλγόριθμος FastICA λόγω της υπολογιστικής του ταχύτητας.

1.4 FastICA

Ο αλγόριθμος FastICA δημοσιεύθηκε από τον A. Hyvärinen το 1999 [5]. Είναι ένας γρήγορος επαναληπτικός αλγόριθμος, βασισμένος στο CLT. Ξεχωρίζει τις πηγές που παράγουν ένα πολυμεταβλητό σήμα μεγιστοποιώντας το non-Gaussianity (ως μέτρο της στατιστικής ανεξαρτησίας). Δύο τρόποι ποσοτικοποίησης του non-Gaussianity είναι η κύρτωση (kurtosis) και η αρνητική εντροπία (negentropy). [6]

Η κύρτωση μίας κατανομής εκφράζει τη διαφορά της κυρτότητας (που εκφράζει το βαθμό συγκέντρωσης των δεδομένων γύρω από τη μέση τιμή) μίας κατανομής σε σχέση με την αντίστοιχη Γκαουσιανή κατανομή. Για ένα σήμα y με μηδενική μέση τιμή υπολογίζεται από τον τύπο

$$kurt(y) = E\{y^4\} - 3E\{y^2\}^2 \quad (1.7)$$

Εάν το σήμα y έχει υποστεί whitening, η διακύμανση είναι ίση με τη μονάδα, δηλαδή $E\{y^2\} = 1$, επομένως η προηγούμενη εξίσωση γίνεται

$$kurt(y) = E\{y^4\} - 3 \quad (1.8)$$

Εάν η κύρτωση είναι ίση με 0, η κατανομή είναι Γκαουσιανή. Εάν η κύρτωση

είναι θετική, η κατανομή είναι υπερ-Γκαουσιανή (super-Gaussian) με εντονότερο spike / υψηλότερη κορυφή από μία Γκαουσιανή κατανομή και τέλος εάν είναι αρνητική, η κατανομή είναι υπο-Γκαουσιανή (sub-Gaussian) με πιο επίπεδη κατανομή από μία Γκαουσιανή. Μερικοί αλγόριθμοι FastICA χρησιμοποιούν το τετράγωνο της κύρτωσης ως μέτρο non-Gaussianity. Ωστόσο, η κύρτωση ως μέτρο είναι ιδιαίτερα ευαίσθητη σε ακραίες τιμές της κατανομής. Για αυτό το λόγο, η κύρτωση δεν θεωρείται ιδιαίτερα αξιόπιστο μέτρο του non-Gaussianity.

Η αρνητική εντροπία είναι μία δεύτερη έκφραση της διαφοράς μίας κατανομής από την Γκαουσιανή κατανομή. Μία σημαντική ιδιότητα της Γκαουσιανής κατανομής, είναι ότι έχει την μεγαλύτερη δυνατή εντροπία σε σχέση με οποιαδήποτε άλλη κατανομή ίσης διασποράς και μέσης τιμής. Ως εκ τούτου, οι μη Γκαουσιανές κατανομές θα έχουν μικρότερη εντροπία από την αντίστοιχη Γκαουσιανή και η διαφορά αυτή εκφράζεται από την ιδιότητα του non-Gaussianity. Η αρνητική εντροπία ενός παρατηρούμενου πολυμεταβλητού σήματος y ορίζεται ως η διαφορά μεταξύ της εντροπίας μίας γκαουσιανής τυχαίας μεταβλητής με αναμενόμενη τιμή και διασπορά ίση με του σήματος y και την εντροπία του σήματος y , δηλαδή

$$J(y) = H(y_{gauss}) - H(y) \quad (1.9)$$

όπου H η (διαφορική) εντροπία μίας μεταβλητής που υπολογίζεται από τον τύπο

$$H(y) = - \int f(y) \log(f(y)) dy \quad (1.10)$$

όπου $f(y)$ η συνάρτηση πυκνότητας πιθανότητας του y .

Η αρνητική εντροπία είναι ίση με 0 για Γκαουσιανές κατανομές και θετική για μη Γκαουσιανές κατανομές.

Ο αλγόριθμος FastICA υπολογίζει επαναληπτικά το διάνυσμα - στήλη W που μεγιστοποιεί το non-Gaussianity της προβολής $w^T X$, μεγιστοποιώντας μία από τις ποσότητες που αναφέρθηκαν προηγουμένως. Εάν ορίσουμε ως $f(u)$ τη συνάρτηση αυτής της ποσότητας, $g(u)$ την πρώτη παράγωγο της και $g'(u)$ τη δεύτερη παράγωγο της, τότε τα βήματα που ακολουθεί ο αλγόριθμος FastICA για τον υπολογισμό μίας στήλης του W είναι τα εξής: [7]

1. Αρχικοποίηση τυχαίου διανύσματος - στήλης w
2. Υπολογισμός της επόμενης επανάληψης του διανύσματος ως $w^+ = E\{\mathbf{X}g(\mathbf{w}^T \mathbf{X})^T\} - E\{g'(\mathbf{w}^T \mathbf{X})\}\mathbf{w}$

3. Αρχικοποίηση του $\mathbf{w} = \mathbf{w}^+ / \|\mathbf{w}^+\|$
4. Έλεγχος σύγκλισης του \mathbf{w} . Εάν δεν έχει συγκλίνει, επιστροφή στο βήμα 2.

1.5 Εναλλακτικές και προεκτάσεις της ICA

1.5.1 Wavelet enhanced ICA (wICA)

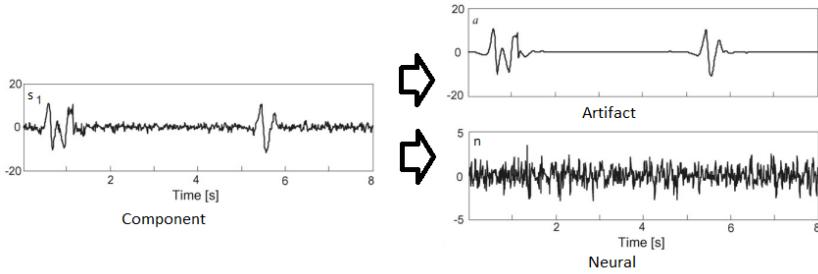
Στο πλαίσιο της παρούσας εργασίας, η απόδοση της ICA στον καθαρισμό σήματος EEG συγκρίνεται με μερικές εναλλακτικές μεθόδους καθαρισμού. Μία από αυτές είναι ο αλγόριθμος wavelet enhanced ICA (wICA).

Στη μελέτη τους το 2006, οι Castellanos και Makarov δείξαν ότι ο καθαρισμός eyeblink artifact από δεδομένα EEG εισάγει σημαντικές διακυμάνσεις και σε περιόδους των μετρήσεων κατά τις οποίες δεν υπάρχει ανοιγοκλείσιμο του ματιού [1]. Για να αντιμετωπίσουν αυτό το πρόβλημα, ανέπτυξαν τον αλγόριθμο wICA. Η wICA είναι μία επέκταση της ICA που βασίζεται στην υπόθεση ότι ο καθαρισμός EEG με ICA απορρίπτει και εγκεφαλική δραστηριότητα ως θόρυβο. Ο συγκεκριμένος αλγόριθμος αναπτύχθηκε συγκεκριμένα για την βελτίωση του καθαρισμού EEG artifact με ICA μέσω της διατήρησης της εγκεφαλικής αυτής δραστηριότητας. Η wICA λειτουργεί “διορθώνοντας” το IC που θεωρείται θόρυβος, δηλαδή χωρίζοντας το σε δύο κομμάτια: ένα κομμάτι χαμηλής έντασης που προέρχεται από εγκεφαλική δραστηριότητα (και η απόρριψη του ως θόρυβος θα ήταν λανθασμένη) και ένα κομμάτι υψηλής έντασης που προέρχεται από artifact. Το παραπάνω απεικονίζεται στο σχήμα 1.2 και εκφράζεται μαθηματικά από τον τύπο

$$s(t) = a(t) + n(t) \quad (1.11)$$

Όπου $s(t)$ το IC που θεωρείται θόρυβος, $a(t)$ το artifactual κομμάτι και $n(t)$ η υπολειπόμενη εγκεφαλική δραστηριότητα του IC. Στη συνέχεια αφαιρείται το artifactual κομμάτι από το IC, και ο αντίστροφος μετασχηματισμός από τα ICs στις χρονοσειρές των ηλεκτροδίων γίνεται χωρίς την απόρριψη κάποιου IC.

Η wICA λειτουργεί υπολογίζοντας ένα wavelet μετασχηματισμό των ICs που προκύπτουν από την ICA. Στη συνέχεια, οι συντελεστές των wavelets συγκρίνονται με ένα κατώφλι και εάν είναι μεγαλύτεροι από αυτό, θεωρούνται artifactual και μηδενίζονται. Ο αντίστροφος wavelet μετασχηματισμός με τους νέους συντελεστές θα επιστρέψει μόνο το εγκεφαλικό κομμάτι της δραστηριότητας $n(t)$.



Σχήμα 1.2: Γραφική απεικόνιση της wICA

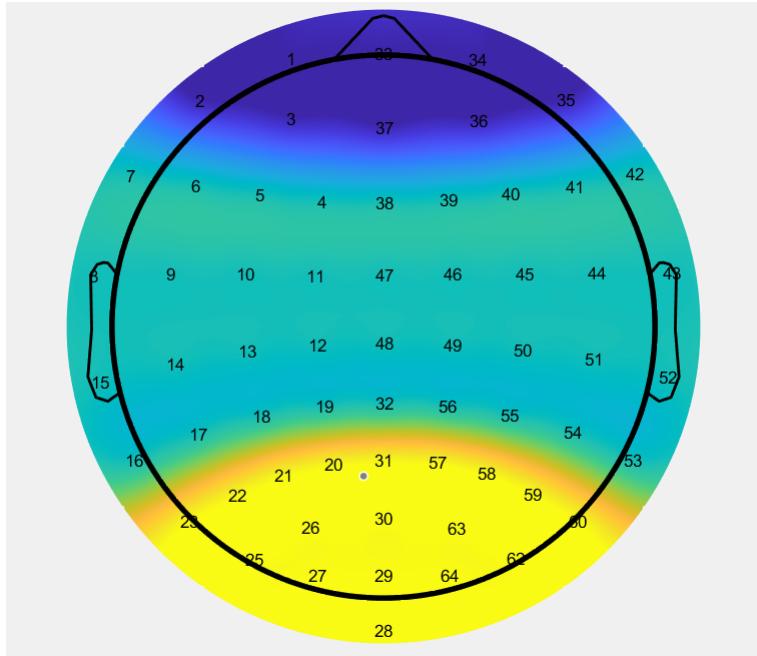
1.5.2 Stripped ICA (sICA)

Η τρίτη μέθοδος καθαρισμού που δοκιμάστηκε είναι ο διαχωρισμός των ηλεκτροδίων σε ομάδες ανάλογα με την απόστασή τους από την πηγή του θορύβου. Η εξής μέθοδος βασίζεται στην παρατήρηση ότι σύμφωνα με τις παραδοχές που γίνονται σχετικά με την αγωγιμότητα του εγκεφάλου, ο θόρυβος που εισάγεται σε δύο ηλεκτρόδια όταν το ένα είναι κοντά στην πηγή του θορύβου ενώ το άλλο μακριά, είναι ίδιας μορφής αλλά πολύ διαφορετικής κλίμακας. Για παράδειγμα, εάν ως πηγή του θορύβου θεωρείται το ανοιχοκλείσιμο του ματιού που προέρχεται από το μπροστινό μέρος του κρανίου, τότε τα σετ ηλεκτροδίων θα μπορούσαν να οριστούν όπως φαίνεται στο σχήμα 1.3

Χωρίζοντας τα ηλεκτρόδια σε διαφορετικά σετ που θα επεξεργαστούν με διαφορετικές εκτελέσεις του αλγορίθμου ICA, εισάγεται η πιθανότητα να μην εμφανιστεί κάποιο IC με την χαρακτηριστική μορφή του θορύβου και μας δίνεται η δυνατότητα να αφήσουμε κάποια ηλεκτρόδια ανεπηρέαστα από τον καθαρισμό (με αυτόματο τρόπο), το οποίο στο πλαίσιο της ανάλυσης συνδεσιμότητας μπορεί να είναι θεμιτό. Επιπλέον, μειώνεται η πολυπλοκότητα του προβλήματος καθώς ο αριθμός συντελεστών που πρέπει να υπολογιστούν σε κάθε εκτέλεση της ICA είναι μικρότερος.

1.5.3 Noise mixing approximation (NMA)

Η τελευταία μέθοδος καθαρισμού έγκειται στον ανεξάρτητο υπολογισμό της στήλης mixing matrix που αντιστοιχεί σε artifactual IC. Για την εφαρμογή αυτής της μεθόδου χρειάζονται δύο στοιχεία



Σχήμα 1.3: Stripped ICA σε ηλεκτροδίων

- Μία χρονοσειρά του θορύβου σε κάποιο ηλεκτρόδιο αναφοράς. Το ηλεκτρόδιο αναφοράς πρέπει να είναι ένα ηλεκτρόδιο για το οποίο θεωρούμε ότι αφαιρέθηκε ένα ικανοποιητικό ποσοστό του θορύβου. Για το συγκεκριμένο στοιχείο μπορεί να χρησιμοποιηθεί η χρονοσειρά του θορύβου που αφαιρέθηκε με εφαρμογή stripped ICA, ICA ή wICA σε κάποιο από τα ηλεκτρόδια που βρίσκονται κοντά στην πηγή του θορύβου.
- Ένα (κανονικοποιημένο) διάνυσμα - στήλη n_b στοιχείων, όπου n_b ο αριθμός των ηλεκτροδίων, το οποίο εκφράζει τη (σχετική) δύναμη της προβολής της πηγής θορύβου σε κάθε ηλεκτρόδιο. Το συγκεκριμένο στοιχείο στην ουσία είναι ισοδύναμο με τη στήλη του mixing matrix που αντιστοιχεί στο artifactual IC κατά την εφαρμογή ICA. Ο υπολογισμός αυτού του διανύσματος μπορεί να γίνει με πολλούς τρόπους. Ένας αρκετά απλός τρόπος, είναι ο υπολογισμός της μέσης διακύμανσης ανά ηλεκτρόδιο που εισάγεται όταν ενεργοποιείται η πηγή του θορύβου. Συγκεκριμένα για το θόρυβο από ανοιγοκλείσιμο του ματιού, αυτό μπορεί να γίνει απομονώνοντας τις παρατηρήσεις των μετρήσεων που καταγράφονται όταν εμφανίζονται τα χαρακτηριστικά spikes, υπολογίζοντας κάποιο μέτρο διακύμανσης για κάθε ηλεκτρόδιο, το οποίο κατά προτίμηση θα μετράει τη διακύμανση που εισάγεται σε σχέση με περιόδους των παρατηρήσεων στις οποίες δεν υπάρχουν spikes, και τέλος κανονικοποιώντας τις

υπολογισμένες τιμές. Το μέτρο διακύμανσης που χρησιμοποιήθηκε στη συγκεκριμένη εργασία είναι το peak to peak distance, το οποίο υπολογίστηκε για κάθε ηλεκτρόδιο ως η διαφορά του μέσου όρου εφτά τιμών γύρω από την κορυφή ενός spike και του μέσου όρου εφτά τιμών γύρω από τη βάση του spike. Το παραπάνω υπολογίστηκε σε κάθε ηλεκτρόδιο και για κάθε spike, ύστερα υπολογίστηκε ο μέσος όρος ανά ηλεκτρόδιο και τέλος έγινε κανονικοποίηση διαιρώντας με τη μέγιστη τιμή.

Σημειώνεται ότι η συγκεκριμένη μέθοδος υπολογισμού του διανύσματος προβολής του θορύβου επηρεάζεται δυστυχώς και από οποιαδήποτε άλλη δραστηριότητα πέρα του θορύβου καταγράφεται από τα ηλεκτρόδια. Συγκεκριμένα, είναι ευαίσθητη σε μη στασιμότητες που εισάγονται κατά την ίδια περίοδο με το θόρυβο.

Έχοντας τα παραπάνω δύο στοιχεία, το επόμενο βήμα είναι ο πολλαπλασιασμός τους για τον υπολογισμό του θορύβου σε κάθε ηλεκτρόδιο. Ο πίνακας που θα προκύψει αφαιρείται από τις (θορυβώδεις) μετρήσεις EEG και έτσι υπολογίζεται το καθαρισμένο EEG.

1.6 Forward and inverse problems

Οι μετρήσεις του EEG προσφέρουν πολύ καλή χρονική ακρίβεια, διότι τα σήματα που παράγονται εντός του εγκεφάλου διαδίδονται με πολύ μεγάλη ταχύτητα σε σχέση με την συχνότητα δειγματοληψίας των οργάνων. Ωστόσο υστερούν σε σχέση με άλλες μεθόδους όσον αφορά τη χωρική ακρίβεια, καθώς οι μετρήσεις του (non invasive) EEG γίνονται στην επιφάνεια του κρανίου. Από αυτές τις μετρήσεις είναι σχετικά εύκολο να βρεθεί η σχετική κατεύθυνση και ο προσανατολισμός των πηγών της δραστηριότητας που μετράται, ωστόσο δεν προσφέρουν αρκετή πληροφορία για τον ακριβή υπολογισμό της θέσης. Παρ' όλα αυτά, έχουν αναπτυχθεί αρκετοί αλγόριθμοι για την ανακατασκευή και τοπικοποίηση των υποκείμενων εγκεφαλικών πηγών (source localization and reconstruction) από δεδομένα EEG.

Αυτοί οι αλγόριθμοι λειτουργούν υπό την υπόθεση ότι οι εγκεφαλικές πηγές είναι ισοδύναμες με δίπολα εντός του εγκεφάλου. Όταν ένα δίπολο τοποθετείται μέσα σε ένα αγώγιμο μέσο, η ένταση του ρεύματος στο μέσο είναι αντιστρόφως ανάλογη της απόστασης μεταξύ του διπόλου και του σημείου μέτρησης υψωμένης στο τετράγωνο. Φυσικά, υπάρχει και ένας όρος που σχετίζεται με την αγωγιμότητα του μέσου που περιπλέκει αρκετά το πρόβλημα. Στον όγκο του κεφαλιού υπάρχουν διάφορα είδη βιολογικού ιστού με διαφορετική αγωγιμότητα [2]. Λόγω των διαφορών στην αγωγιμότητα, το σήμα που προέρχεται από τις

εγκεφαλικές πηγές διαστρεβλώνεται και εξασθενεί μέχρι να φτάσει στο σημείο μέτρησης, δηλαδή τα ηλεκτρόδια στο κρανίο. Ωστόσο, για τον ίδιο λόγο, δηλαδή την αγωγιμότητα του εγκεφάλου, τα σήματα που παράγονται εντός του εγκεφάλου φτάνουν σε όλα τα ηλεκτρόδια, απλά με διαφορετικούς βαθμούς διαστρέβλωσης και εξασθένησης. Ως εκ τούτου, είναι δυνατό να θεωρηθεί ότι οι μετρήσεις των ηλεκτροδίων είναι ένα γραμμικός συνδυασμός της δραστηριότητας των διαφόρων πηγών στον εγκέφαλο.

To forward problem στο πλαίσιο της ανάλυσης EEG αναφέρεται στον υπολογισμό των χρονοσειρών των ηλεκτροδίων, όταν είναι γνωστές οι χρονοσειρές των πηγών / διπόλων (source activations) καθώς και η “ένταση προβολής” των πηγών στα ηλεκτρόδια (leadfield matrix).

To leadfield matrix είναι ένας πίνακας διαστάσεων ($N \times L \times K$) όπου N ο αριθμός των ηλεκτροδίων, L οι δυνατοί προσανατολισμοί ανά δίπολο και K ο αριθμός των διπόλων εντός του εγκεφάλου. Εάν θεωρήσουμε τη δραστηριότητα των διπόλων ως ένα πίνακα με διαστάσεις ($K \times n$) όπου n ο αριθμός των παρατηρήσεων, τότε ο πολλαπλασιασμός των δύο πινάκων μας δίνει την συνολική προβολή της δραστηριότητας των διπόλων στα ηλεκτρόδια.

Το αντίστροφο πρόβλημα (inverse problem), δηλαδή ο μετασχηματισμός από το πεδίο των ηλεκτροδίων στο πεδίο των πηγών είναι αρκετά πιο περίπλοκο και δεν έχει μοναδική λύση. Το πρόβλημα λύνεται από αλγορίθμους όπως οι low resolution brain electromagnetic tomography (loreta), local autoregressive average (LAURA) και block sparse Bayesian learning (bsbl), οι οποίοι βασίζονται στην υπόθεση ότι οι μετρήσεις EEG είναι γραμμική υπέρθεση των υποκείμενων εγκεφαλικών πηγών και κάνουν χρήση ενός προϋπολογισμένου leadfield matrix για τη μοντελοποίηση της αγωγιμότητας.

Όπως αναφέρθηκε, η ανακατασκευή των πηγών από τις μετρήσεις δεν έχει μοναδική λύση και τα αποτελέσματα των παραπάνω αλγορίθμων θεωρούνται απλές προσεγγίσεις.

1.7 Αυτοπαλίνδρομα μοντέλα - Autoregressive (AR) models

Στο πλαίσιο των πειραμάτων EEG πολλές φορές είναι χρήσιμη η ανάλυση σε προσομοιωτικά δεδομένα. Ο κύριος λόγος για αυτό είναι ότι τα δεδομένα από πραγματικά πειράματα EEG είναι εξαρχής μολυσμένα με θόρυβο και τα αντίστοιχα καθαρά δεδομένα (ground truth) μας είναι άγνωστα. Ως εκ τούτου είναι αδύνατη για παράδειγμα η αξιολόγηση μεθόδων καθαρισμού, καθώς δεν

μπορούμε να γνωρίζουμε πόσο κοντά στο ground truth είναι τα καθαρισμένα δεδομένα. Μία μέθοδος παραγωγής προσομοιωτικών δεδομένων είναι μέσω των μοντέλων αυτοπαλινδρόμησης (AR models)

Τα AR μοντέλα είναι ένας τρόπος μοντελοποίησης στοχαστικών διαδικασιών που έχει εφαρμοστεί για την ανάλυση χρονοσειρών σε αρκετές επιστήμες. Η βασική υπόθεση αυτού του τύπου μοντελοποίησης είναι ότι η παρατηρούμενη μεταβλητή εξαρτάται γραμμικά από προηγούμενες τιμές της. Η μέγιστη υστέρηση στην οποία παρατηρείται αυτή η εξάρτηση λέγεται τάξη του μοντέλου. Έτσι, ένα AR μοντέλο τάξης p εκφράζεται μαθηματικά από τον τύπο

$$X_t = c + \sum_{i=1}^p u_i X_{t-i} + e_t \quad (1.12)$$

όπου c ένας σταθερός όρος, u_i οι παράμετροι του μοντέλου και e_t λευκός όρος. Το παραπάνω μοντέλο συμβολίζεται ως AR(p).

Τα παραπάνω μπορούν να επεκταθούν και στην μοντελοποίηση συστήματων με περισσότερες από μία μεταβλητές. Αυτή η γενίκευση γίνεται με τα Vector autoregressive (VAR) models. Ένα VAR(p) μοντέλο εκφράζεται μαθηματικά από τον παρακάτω τύπο.

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^p \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{e}_t \quad (1.13)$$

όπου εάν το μοντέλο μοντελοποιεί k μεταβλητές, το \mathbf{c} είναι διάνυσμα k στοιχείων, οι \mathbf{A}_i είναι πίνακες συντελεστών διαστάσεων ($k \times k$) και το \mathbf{e}_t είναι διάνυσμα λευκού όροβου k στοιχείων.

Καθώς τα συγκεκριμένα μοντέλα είναι ικανά να μοντελοποιήσουν σχέσεις αλληλεπίδρασης μεταξύ πολλών μεταβλητών, έχουν βρει εφαρμογή σε πεδία όπου είναι χρήσιμη η ανάλυση συνδεσμώντητας. Ένας τρόπος με τον οποίο χρησιμοποιούνται είναι η δημιουργία προσομοιωτικών δεδομένων με ελεγχόμενες αλληλεπιδράσεις. Ένας δεύτερος τρόπος είναι η ανάλυση συνδεσμώντητας μέσω των πινάκων συντελεστών. Συγχριμένα, η προσαρμογή δεδομένων σε μοντέλα VAR μπορεί να προσφέρει πληροφορίες για τις σχέσεις μεταξύ των μεταβλητών οι οποίες εκφράζονται στους πίνακες \mathbf{A}_i .

1.8 Μετρικές αιτιότητας

Ο εντοπισμός και η ανάλυση των αλληλεπιδράσεων των μεταβλητών ενός συστήματος παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς οι αλληλεπιδράσεις μπορούν να περιγράψουν τη δυναμική του συστήματος και να συμβάλουν στο ορισμό των

σχέσεων μεταξύ των μεταβλητών του. Για παράδειγμα, στο πλαίσιο του EEG, η γνώση των σημάτων των πηγών της δραστηριότητας θα μπορούσε να βοηθήσει στην καλύτερη κατανόηση του νευρικού συστήματος [8]. Οι δυναμικές σχέσεις είναι δυνατόν να απεικονιστούν με διαγράμματα διαδρομής (γράφους), στα οποία οι μεταβλητές αντιπροσωπεύονται από κόμβους, και οι σχέσεις μεταξύ των κόμβων αντιπροσωπεύονται από ακμές. Η περιγραφή της δυναμικής του συστήματος και των σχέσεων αιτιότητας στις πολυμεταβλητές χρονοσειρές μπορεί να πραγματοποιηθεί, εκτός των άλλων προσεγγίσεων όπως παρουσιάζεται από την ανασκόπηση στο θέμα από τους Hlaváčková-Schindler et al. (2007) [9] και με τη χρήση κάποιας μετρικής αιτιότητας [6]. Οι μετρικές αιτιότητας δεν εκφράζουν απλή συσχέτιση μεταξύ δύο μεταβλητών, αλλά κατά πόσο η μία μεταβλητή οδηγεί την άλλη. Στη συνέχεια αναφέρονται μερικές μετρικές αιτιότητας που χρησιμοποιούνται συχνά στην ανάλυση EEG.

1.8.1 (υπό συνθήκη) Αιτιότητα κατά Granger / (conditional) Granger causality

Η αιτιότητα κατά Granger (Granger causality) είναι μία μετρική που βασίζεται στην ικανότητα πρόβλεψης μίας μεταβλητής. Σύμφωνα με την αιτιότητα κατά Granger, εάν ένα σήμα X αιτιάζει (G-causes) σε ένα σήμα Y , τότε οι παρελθούσες τιμές του X περιέχουν πληροφορία η οποία θα μπορούσε να συμβάλει στην πρόβλεψη του Y . [10]

Η αιτιότητα κατά Granger ορίζεται ως κυρίως στο πλαίσιο των γραμμικών διανυσματικών μοντέλων παλινδρόμησης. Θεωρώντας ένα τέτοιο διανυσματικό μοντέλο αυτοπαλινδρόμησης (VAR) δύο μεταβλητών X_1 και X_2

$$X_1(t) = \sum_{j=1}^p A_{11,j} X_1(t-j) + \sum_{j=1}^p A_{12,j} X_2(t-j) + E_1(t)$$

$$X_2(t) = \sum_{j=1}^p A_{21,j} X_1(t-j) + \sum_{j=1}^p A_{22,j} X_2(t-j) + E_2(t)$$

όπου p η μέγιστη υστέρηση ή τάξη του συστήματος, A ο πίνακας που περιέχει τους συντελεστές του μοντέλου, δηλαδή την συμβολή της κάθε (προηγούμενης) παρατήρησης στις τιμές των $X_1(t)$ και $X_2(t)$ και E_1 και E_2 τα σφάλματα πρόβλεψης του μοντέλου. Τότε εάν τα σφάλματα E_1 και E_2 μειώνονται με την εισαγωγή του όρου X_2 (ή X_1) στην σχέση της άλλης μεταβλητής, λέγεται ότι το X_2 αιτιάζει κατά Granger / G-causes το X_1 (ή αντίστροφα). Το παραπάνω μπορεί να εκφραστεί και μέσω της συνθήκη $A_{12} \neq 0$

Τα παραπάνω μπορούν να επεκταθούν και σε περιπτώσεις όπου ο αριθμός των μεταβλητών είναι μεγαλύτερος από 2 με τη μετρική της υπό συνθήκης αιτιότητας κατά Granger / conditional Granger causality (CGC). Σε αυτήν την περίπτωση λέμε ότι το X_2 αιτιάζει κατά Granger το X_1 εάν οι παρατηρήσεις με υστέρηση της X_2 βοηθούν στην πρόβλεψη της X_1 όταν έχουμε λάβει υπόψη και τις παρατηρήσεις με υστέρηση των υπόλοιπων μεταβλητών X_3, \dots, X_n . Η ποσοτικοποίηση της υπό συνθήκης αιτιότητας κατά Granger ονομάζεται conditional Granger causality index (CGCI) και ορίζεται για κάθε πιθανό κατευθυνόμενο ζεύγος κόμβων $X_1 \rightarrow X_2$. Όλες αυτές οι τιμές μπορούν να συγκεντρωθούν σε έναν τετραγωνικό, μη συμμετρικό πίνακα, ώστε να αναπαριστώνται όλες οι αιτιακές σχέσεις μεταξύ όλων των κόμβων ενός συστήματος σε μία βάση δεδομένων.

1.8.2 Επεκτάσεις του (Conditional) Granger causality

Όπως αναφέρθηκε, η μετρική του Granger causality βασίζεται στην προσαρμογή ενός VAR μοντέλου στα δεδομένα. Ωστόσο, εάν το σύστημα που μοντελοποιείται είναι επαρκώς περίπλοκο και μη γραμμικό ή εάν ο αριθμός των παρατηρήσεων είναι πολύ μικρός, η προσαρμογή γραμμικών μοντέλων όπως τα VAR αποτυγχάνει λόγω του πλήθους των συντελεστών που πρέπει να προσαρμοστούν και του σχετικά μικρού όγκου δεδομένων [11]. Καθώς δεν είναι δυνατή η αξιόπιστη προσαρμογή VAR μοντέλων, μετρικές όπως το CGCI λειτουργούν ελαττωματικά.

Ένας τρόπος να αντιμετωπιστεί αυτό το πρόβλημα είναι μέσω μείωσης συνιστωσών του μοντέλου VAR. Η μετρική Restricted conditional Granger causality (rCGCI) ακολουθεί αυτή την προσέγγιση χρησιμοποιώντας τον αλγόριθμο Backwards in time selection (BTS) για την επιλογή συντελεστών, την προσαρμογή ενός VAR μοντέλου μειωμένων διαστάσεων και τον υπολογισμό του rCGCI από αυτό.

Μία άλλη επέκταση της αιτιότητας κατά Granger είναι η μετρική Partial directed coherence (PDC). Η PDC αναπτύχθηκε από τους Baccala et al. [12] και είναι βασισμένη στην partial coherence function, μία υπό συνθήκη μετρική αλληλεπίδρασης που υπολογίζεται στο χώρο της συχνότητας, σε συνδυασμό με την έννοια της αιτιότητας κατά Granger. Όπως και πολλά άλλα μεγέθη που υπολογίζονται στο πεδίο της συχνότητας, η PDC υπολογίζεται σε διαφορετικά παράμυρα συχνοτήτων. Αυτό έχει ιδιαίτερο νόημα στο πλαίσιο της ανάλυσης EEG καθώς η εγκεφαλική δραστηριότητα συχνά μπορεί να κατηγοριοποιηθεί

ανάλογα με τη συχνότητα της.

1.9 Θεωρία γράφων

Οι γράφοι είναι μαθηματικές δομές που χρησιμοποιούνται για τη μοντελοποίηση και αναπαράσταση συστημάτων αντικειμένων με αλληλεπίδραση. Ένας γράφος ορίζεται από ένα σύνολο κόμβων ή κορυφών $V = \{v_1, v_2, \dots, v_n\}$, όπου n ο αριθμός των κόμβων και ένα σύνολο ακμών $E = \{z_1, z_2, \dots, z_m\}$, όπου m ο αριθμός των ακμών. Οι κόμβοι μπορούν να θεωρηθούν ως τα διακριτά αντικείμενα του συστήματος και οι ακμές ως η σύνδεση μεταξύ ζευγών κόμβων που αλληλεπιδρούν με κάποιο τρόπο. Ο αντίστοιχος γράφος συμβολίζεται ως $G = (V, E)$. Οι γράφοι μπορούν να είναι κατευθυνόμενοι που σημαίνει ότι οι ακμές έχουν προσανατολισμό ή μη κατευθυνόμενοι στην αντίθετη περίπτωση. Η επιλογή κατευθυνόμενου ή μη κατευθυνόμενου γράφου έχει να κάνει με το είδος της αλληλεπίδρασης που μοντελοποιείται. Εάν για παράδειγμα εξετάζονται αιτιακές αλληλεπιδράσεις, δηλαδή περιπτώσεις όπου μία μεταβλητή ενός συστήματος οδηγεί μία άλλη (χωρίς να συμβαίνει απαραίτητα και το αντίστροφο), η μοντελοποίηση θα γινόταν με κατευθυνόμενο γράφο.

Μία πολύ σημαντική έννοια από το πεδίο της θεωρίας γράφων είναι αυτή του πίνακα γειτνίασης (Adjacency matrix). Ο πίνακας γειτνίασης είναι ένας $n \times n$ λογικός πίνακας στον οποίο ορίζονται όλα τα ζεύγη κόμβων ενός γράφου. Τα στοιχεία $A_{i,j}$ του πίνακα γειτνίασης μπορούν να είναι είτε 1, που υποδεικνύει την ύπαρξη ακμής μεταξύ των κόμβων i και j , είτε 0 που υποδεικνύει την απουσία ακμής. Ο πίνακας γειτνίασης ενός μη κατευθυνόμενου γράφου είναι εξ ορισμού συμμετρικός.

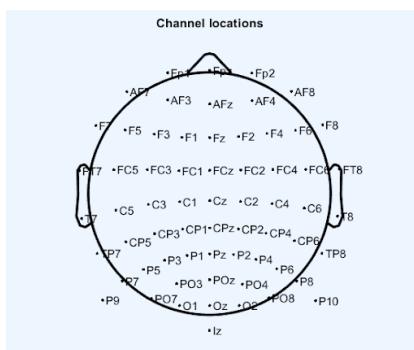
Στο πλαίσιο της συγκεκριμένης εργασίας, η έννοια του πίνακα γειτνίασης προσφέρει έναν πολύ απλό τρόπο μετατροπής κάποιας μετρικής αλληλεπίδρασης σε μία αναπαράσταση με γράφους. Για παράδειγμα, εάν οι μεταβλητές ενός συστήματος θεωρηθούν ως κόμβοι ενός γράφου, τότε οι ακμές μεταξύ των κόμβων μπορούν να οριστούν υπολογίζοντας κάποια μετρική αλληλεπίδρασης για κάθε ζεύγος κόμβων του γράφου. Οι υπολογισμένες τιμές, που στην ουσία είναι η ποσοτικοποίηση της αλληλεπίδρασης μεταξύ των κόμβων, σχηματίζουν ένα τετραγωνικό πίνακα και μπορούν να μετατραπούν σε λογικές τιμές θέτωντας κάποιο κατώφλι, πάνω από το οποίο θα υποδεικνύουν την ύπαρξη ακμής μεταξύ των κόμβων. Ο τελικός λογικός τετραγωνικός πίνακας θεωρείται ο πίνακας γειτνίασης του γράφου που μοντελοποιεί το σύστημα των μεταβλητών.

Κεφάλαιο 2

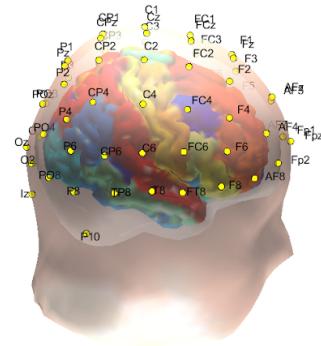
Υλικό και μέθοδοι

2.1 Τοποθέτηση ηλεκτροδίων

Για τα πειράματα θεωρήσαμε 64 ηλεκτρόδια με average reference, τοποθετημένα σύμφωνα με το international 10-20 system, όπως φαίνεται και στα σχήματα 2.1 και 2.2



Σχήμα 2.1: 2D τοποθέτηση ηλεκτροδίων



Σχήμα 2.2: 3D τοποθέτηση ηλεκτροδίων

Τα ηλεκτρόδια που δεν είναι ορατά στην κάτοψη του μοντέλου του κεφαλιού τοποθετούνται εκτός της περιμέτρου του κεφαλιού στην 2D αναπαράσταση της τοποθέτησης των ηλεκτροδίων.

Το international 10-20 system είναι μία παγκοσμίως αναγνωρισμένη μέθοδος για την επιλογή των θέσεων των ηλεκτροδίων.

2.2 Λύση των forward και inverse problems

Για το μετασχηματισμό από το πεδίο των πηγών στο πεδίο των ηλεκτροδίων και το αντίστροφο χρησιμοποιήθηκε το distributed EEG source imaging (dsi) toolbox για το EEGLAB plug-in στο προγραμματιστικό περιβάλλον matlab.

Η περίπτωση του μετασχηματισμού από το πεδίο των πηγών στο πεδίο των ηλεκτροδίων είναι αρκετά απλή, καθώς χρειάζεται μόνο ένα leadfield matrix, το οποίο παράγεται από το dsi toolbox. Το dsi toolbox θεωρεί 8003 δίπολα εντός του εγκεφάλου και ένα προσανατολισμό (κάθετο), επομένως το leadfield matrix που χρησιμοποιήθηκε ήταν διαστάσεων ($64 \times 1 \times 8003$).

Αναφέρεται επίσης ότι η λύση του forward problem προϋπονθέτει τη γνώση ενός μοντέλου του κεφαλιού του ασθενή που περιλαμβάνει πληροφορίες για το σχήμα, τον όγκο και την αγωγιμότητα. Σε πραγματικά πειράματα, ο υπολογισμός αυτού του μοντέλου γίνεται μέσω ενός MRI scan. Καθώς τα πειράματα γίνονται με προσομοιωτικά δεδομένα και δεν καταγράφονται από ασθενή, χρησιμοποιείται το μοντέλο Colin 27, το οποίο είναι ένα στάνταρ μοντέλο του ανθρώπινου εγκέφαλου.

Όσον αφορά το αντίστροφο πρόβλημα, το dsi toolbox προσφέρει τους αλγόριθμους loretal και bsbl. Στην παρούσα εργασία χρησιμοποιήθηκε ο αλγόριθμος bsbl.

Τα 8003 δίπολα που θεωρεί το dsi toolbox είναι ταξινομημένα σε 68 εγκεφαλικές περιοχές. Η δραστηριότητα των περιοχών μπορεί να υπολογιστεί με διάφορους τρόπους. Στην παρούσα μελέτη, υπολογίζεται το άθροισμα των δραστηριοτήτων των διπόλων ανά περιοχή και μετά διαιρείται με τον αριθμό των ενεργών διπόλων μέσα σε αυτήν, ώστε να πάρουμε το μέσο όρο της δραστηριότητας ανά περιοχή.

Αναφέρεται ότι η ανάλυση συνδεσμότητας έγινε μεταξύ των εγκεφαλικών περιοχών, αντί για μεταξύ των διπόλων. Ο λόγος για αυτή την επιλογή είναι ο τεράστιος όγκος δεδομένων για τα δίπολα ($8003 \text{ δίπολα} \times 5000 \text{ παρατηρήσεις}$), καθώς και ότι το averaging με το οποίο υπολογίζεται η δραστηριότητα ανά περιοχή ενισχύει το signal to noise ratio του σήματος.

2.3 Προσομοιωτικά δεδομένα πηγών

Η μελέτη έγινε χρησιμοποιώντας προσομοιωτικά δεδομένα που παρήχθησαν από vector autoregressive μοντέλο τάξης 1 (VAR(1)). Ο λόγος για τη χρήση προσομοιωτικών δεδομένων είναι ότι για τη συγκεκριμένη ανάλυση είναι αναγκαία η γνώση της ground truth, δηλαδή της ακριβής κατάστασης του συστήματος πριν την προσθήκη του θορύβου.

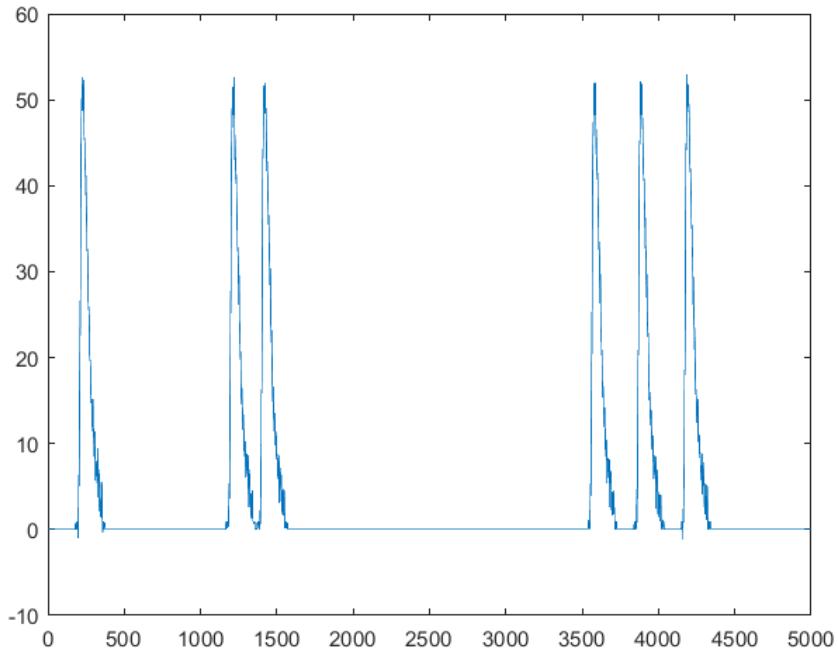
Οι χρονοσειρές που παρήχθησαν από το μοντέλο VAR(1) θεωρούνται ως οι χρονοσειρές της ηλεκτρικής δραστηριότητας των περιοχών εντός του εγκεφάλου (όπως αυτές ορίζονται από το dsi toolbox). Ο αριθμός των ενεργών περιοχών του εγκεφάλου που θεωρήσαμε είναι 7 και ο αριθμός παρατηρήσεων είναι 5000.

Επιπλέον, προστέθηκε ροζ θόρυβος στις υπόλοιπες 61 εγκεφαλικές περιοχές. Ο ροζ θόρυβος είναι ένας τύπος θορύβου που έχει επίπεδη συχνότητα στην λογαριθμική κλίμακα (όπως ο λευκός θόρυβος έχει επίπεδη συχνότητα σε γραμμική κλίμακα). Επιλέχθηκε ροζ θόρυβος επειδή η συγκεκριμένη μορφή συναντάται συχνά σαν “background” εγκεφαλική δραστηριότητα. Σημειώνεται ότι διαφορετικές εκτελέσεις ροζ θορύβου είναι ασυχέτιστες μεταξύ τους και επομένως θεωρείται ότι η μελέτη συνδεσμότητας δεν επηρεάζεται από την προσθήκη αυτού του είδους θορύβου. Η προσθήκη αυτού του είδους θορύβου αποσκοπεί μονάχα στην καλύτερη προσομοίωση της δραστηριότητας του εγκεφάλου και δε θα χρησιμοποιηθεί για την αξιολόγηση κάποιας μεθόδου καθαρισμού.

2.4 Θόρυβος

Ως θόρυβος χρησιμοποιήθηκε μία χρονοσειρά που προσομοιώνει το θόρυβο από το ανοιγοκλείσιμο του ματιού. Για να κατασκευαστεί αυτή η χρονοσειρά, αρχικά εφαρμόστηκε ο αλγόριθμος FastICA σε πραγματικά δεδομένα EEG που προέρχονταν από υγιή ασθενή σε κατάσταση χαλάρωσης, με ανοιχτά τα μάτια. Από τα ICs που υπολογίστηκαν από την εφαρμογή FastICA, επιλέχθηκε το IC που εμφανίζει τα χαρακτηριστικά spikes που προκαλούνται από την ηλεκτρική δραστηριότητα κατά το ανοιγοκλείσιμο του ματιού. Στη συνέχεια η παραπάνω χρονοσειρά κεντραρίστηκε και απομονώθηκαν τα samples σε ένα από αυτά τα spikes. Το κομμάτι που απομονώθηκε προστέθηκε σε μία μηδενική χρονοσειρά με συνολικά 10000 παρατηρήσεις σε σημεία που ορίστηκαν από κατανομή Poisson. Τέλος, έγινε downsampling για να μειωθεί ο αριθμός των παρατηρήσεων σε 5000. Η τελική χρονοσειρά φαίνεται στο σχήμα 2.3.

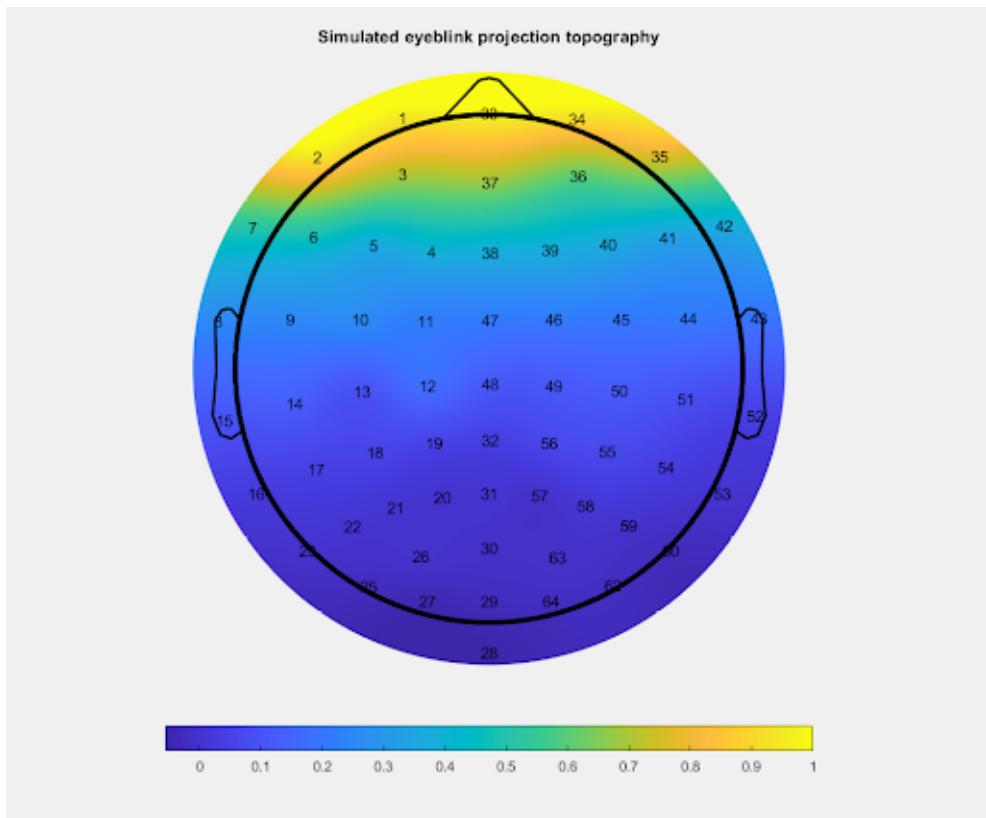
Ο θόρυβος αυτός προστέθηκε κατευθείαν στις χρονοσειρές των ηλεκτροδίων αντί να θεωρηθεί ως πηγή για την οποία θα λυνόταν το forward problem μέχρη του dsi toolbox. Δηλαδή το forward problem λύθηκε θεωρώντας μόνο την δραστηριότητα του ροζ θορύβου και του VAR(1) μοντέλου (που αναφέρθηκαν στην προηγούμενη ενότητα) για τον υπολογισμό του καθαρού EEG.



Σχήμα 2.3: Προσομοιωτική χρονοσειρά θορύβου

Στη συνέχεια προστέθηκε σε αυτό η χρονοσειρά του θορύβου. Καθώς χρησιμοποιούμε 64 ηλεκτρόδια, χρειάζεται και ένα διάνυσμα κανονικοποίησης 64 στοιχείων με το οποίο να υπολογίζεται η επίδραση / ένταση της προβολής του θορύβου σε κάθε ηλεκτρόδιο. Για τον υπολογισμό αυτού του διανύσματος χρησιμοποιήθηκαν πραγματικά δεδομένα EEG από 8 πειράματα με υγιείς ασθενείς σε κατάσταση χαλάρωσης με ανοιχτά τα μάτια. Όλα τα πειράματα χρησιμοποιούσαν την ίδια τοποθέτηση ηλεκτροδίων με την παρούσα εργασία. Για κάθε πείραμα και για κάθε ηλεκτρόδιο υπολογίστηκε το μέσο peak to peak distance (όπως ορίστηκε στην ενότητα 1.5.3) των spikes που προκαλούνται από ανοιγοκλείσιμο του ματιού. Στη συνέχεια για κάθε πείραμα έγινε κανονικοποίηση των μέσων αποστάσεων όλων των ηλεκτροδίων με διαίρεση με τη μέγιστη τιμή και τέλος υπολογίστηκε ο μέσος όρος των 8 κανονικοποιημένων διανυσμάτων 64 στοιχείων.

Η τοπογραφία του τελικού διανύσματος φαίνεται στο σχήμα 2.4.



Σχήμα 2.4: Τοπογραφία θορύβου

2.5 Καθαρισμός EEG

Συνολικά, εφαρμόστηκαν τέσσερεις μέθοδοι καθαρισμού δεδομένων EEG. Αυτοί είναι:

- Καθαρισμός με FastICA
- Καθαρισμός με wavelet enhanced ICA (wICA)
- Καθαρισμός με stripped ICA (sICA)
- Καθαρισμός με noise mixing approximation (NMA)

Ο θόρυβος που αφαιρέθηκε με αυτές τις μεθόδους ήταν ο θόρυβος από το ανοιγοκλείσιμο του ματιού. Όπως αναφέρθηκε προηγουμένως, η προσθήκη του ροζ θορύβου έγινε με μοναδικό σκοπό την καλύτερη προσομοίωση της εγκεφαλικής δραστηριότητας και δεν επηρεάζει σημαντικά τη συνδεσιμότητα των

δεδομένων, επομένως δεν έγινε καμία απόπειρα καθαρισμού του.

Η επιλογή των IC που αντιστοιχούν σε θόρυβο έγινε αυτόματα με τον υπολογισμό (για κάθε IC) του συντελεστή συσχέτισης με τη (γνωστή) χρονοσειρά του θορύβου και τη σύγκρισή του με ένα κατώφλι.

Αναφέρεται ότι συχνά εμφανίζονται περισσότερα από ένα ICs που να εμφανίζουν τουλάχιστον ένα από τα χαρακτηριστικά spikes που προκαλεί το ανοιγοκλείσιμο του ματιού. Ωστόσο με κάθε IC που αφαιρείται χάνεται και χρήσιμη πληροφορία που σχετίζεται με εγκεφαλική δραστηριότητα. Επομένως, σε πραγματικά πειράματα η επιλογή των ICs που θα αφαιρεθούν γίνεται με μεγάλη προσοχή στο λόγο neural / artifactual δραστηριότητας του IC και ανάλογα με τις ανάγκες της εκάστοτε ανάλυσης.

Όσον αφορά την επιλογή του IC που αντιστοιχεί σε θόρυβο κατά τον καθαρισμό με stripped ICA, γίνεται με τον ίδιο τρόπο που αναφέρθηκε και προηγουμένως. Ωστόσο, αυτή τη φορά υπάρχει η πιθανότητα να μην εμφανιστεί κάποιο IC του οποίου ο συντελεστής συσχέτισης με τη χρονοσειρά να είναι μεγαλύτερος από το κατώφλι. Σε αυτή την περίπτωση, επιλέγεται το IC με το μεγαλύτερο συντελεστή συσχέτισης.

Η επιλογή αυτή έγινε διότι σε αυτό το στάδιο της ανάλυσης είναι χρήσιμη η αξιολόγηση του καθαρισμού με stripped ICA σε όλα τα ηλεκτρόδια.

Όσον αφορά τη μέθοδο noise mixing approximation, η κύρια δυσκολία είναι η ανακατασκευή του διανύσματος προβολής του θορύβου στα ηλεκτρόδια. Ο τρόπος με τον οποίο επιτυγχάνεται αυτό έχει αναλυθεί στην ενότητα 1.5.3 και βασίζεται στον υπολογισμό της μέσης διακύμανσης που εισάγει ο θόρυβος ανά ηλεκτρόδιο. Ωστόσο, στη συγκεκριμένη προσομοίωση, το ανοιγοκλείσιμο του ματιού συμβαίνει μόλις 6 φορές που σημαίνει ότι ο μέσος όρος υπολογίζεται από μόνο 6 στοιχεία για το κάθε ηλεκτρόδιο.

Παρ' όλ' αυτά συνεχίζει να έχει νόημα η σύγκριση της μεθόδου καθαρισμού με τον καθαρισμό με ICA. Αναφέρεται επίσης ότι τα πραγματικά πειράματα EEG συνήθως διαρκούν από αρκετά λεπτά μέχρι και μερικές ώρες και ως εκ τούτου τα περισσότερα είδη θορύβου εμφανίζονται αρκετές εκατοντάδες φορές σε κάθε καταγραφή.

Κεφάλαιο 3

Μεθοδολογία πειραμάτων

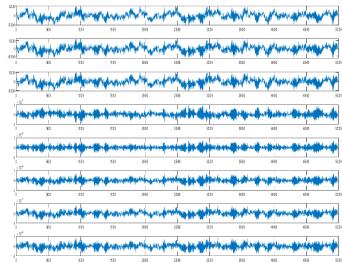
3.1 Περιγραφή πειράματος

Σε αυτό το κεφάλαιο θα αναλυθεί ο τρόπος με τον οποίο έγιναν τα πειράματα. Το πρώτο βήμα έγκειται στη δημιουργία προσομοιωτικών δεδομένων. Για αυτό το σκοπό, αρχικά παράγουμε 7 χρονοσειρές με 5000 παρατηρήσεις από ένα VAR(1) μοντέλο με γνωστό πίνακα γειτνίασης. Οι χρονοσειρές που παράγονται θεωρούνται ως η δραστηριότητα σε 7 περιοχές του εγκεφάλου. Στις υπόλοιπες 61 περιοχές του εγκεφάλου που θεωρεί το dsi toolbox, προστίθεται ροζ θόρυβος.

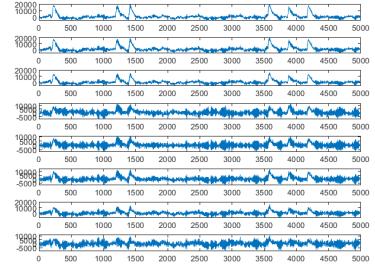
Ο τρόπος με τον οποίο προστίθεται μία χρονοσειρά σε μία εγκεφαλική περιοχή είναι ο εξής: όπως αναφέρθηκε, η λύση του forward problem γίνεται με τον πολλαπλασιασμό των δεδομένων των πηγών με το leadfield matrix το οποίο στην περίπτωσή μας είναι διαστάσεων ($64 \times 1 \times 8003$). Τα δεδομένα των πηγών είναι ένας αρχικά μηδενικός πίνακας διαστάσεων (8003×5000). Για να προσθέσουμε μία από τις χρονοσειρές του VAR μοντέλου σε μία περιοχή, διαλέγουμε 10 τυχαία δίπολα που ανήκουν στην περιοχή και προσθέτουμε τη χρονοσειρά στις αντίστοιχες γραμμές του πίνακα δεδομένων πηγών. Για να προσθέσουμε ροζ θόρυβο σε μία περιοχή, διαλέγουμε 5 τυχαία δίπολα που ανήκουν στην περιοχή και προσθέτουμε το ροζ θόρυβο στις αντίστοιχες γραμμές. Οι υπόλοιπες γραμμές των δεδομένων πηγών μένουν μηδενικές.

Από τα παραπάνω δεδομένα πηγών υπολογίζεται το ισοδύναμο (καθαρό) EEG με πολλαπλασιασμό με τον προϋπολογισμένο leadfield matrix. Στο σχήμα 3.1 φαίνονται οι χρονοσειρές των ηλεκτροδίων FP1, AF7, AF3, F1, F3, F5, F7 και FT7 για τα καθαρά δεδομένα EEG από ένα τυχαίο πείραμα. Το θορυβώδες EEG υπολογίζεται προσθέτοντας τη χρονοσειρά του θορύβου πολλαπλασιασμένη με το διάνυσμα προβολής θορύβου, τα οποία έχουν προϋπολογιστεί όπως περιγράφεται στην ενότητα 2.4. Οι χρονοσειρές των προαναφερθέντων ηλεκτροδίων

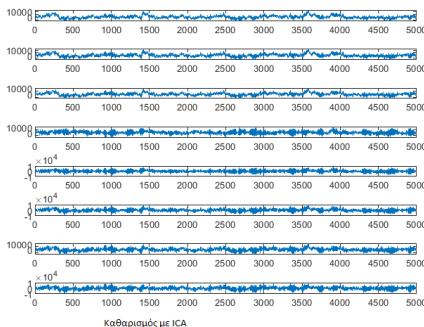
του θορυβώδους EEG δίνονται στο σχήμα 3.2. Στη συνέχεια υπολογίζονται 4 καθαρισμένα EEG εφαρμόζοντας κάθε μία από τις 4 μεθόδους καθαρισμού που αναφέρθηκαν στο κεφάλαιο 2.5. Οι χρονοσειρές των 8 καθαρισμένων ηλεκτροδίων δίνονται στο σχήμα 3.3. Για κάθε ένα από τα 6 αυτά EEG, λύνεται το inverse problem χρησιμοποιώντας τον αλγόριθμο bsbl του dsi toolbox.



Σχήμα 3.1: Καθαρά δεδομένα, 8 ηλεκτρόδια



Σχήμα 3.2: Θορυβώδη δεδομένα, 8 ηλεκτρόδια



Σχήμα 3.3: Καθαρισμένα δεδομένα, 8 ηλεκτρόδια

3.2 Περιγραφή μετρήσεων

Το δεύτερο βήμα περιλαμβάνει τις μετρήσεις που έγιναν σε κάθε έναν από τους 6 πίνακες δεδομένων EEG (καθαρό EEG, θορυβώδες EEG και EEG καθαρισμένο με μία από τις 4 μεθόδους). Οι μετρήσεις έγιναν σε τρία επίπεδα:

- Στο επίπεδο των 64 ηλεκτροδίων (χωρίς να είναι απαραίτητη η λύση του inverse problem)
- Στο επίπεδο όλων των 68 εγκεφαλικών περιοχών στις οποίες ταξινομείται η δραστηριότητα με τη λύση του inverse problem
- Στο επίπεδο των 7 περιοχών στις οποίες είχαν προστεθεί οι χρονοσειρές VAR

Οι μετρικές που υπολογίστηκαν σε κάθε επίπεδο είναι οι εξής:

64 ηλεκτρόδια	7 περιοχές	68 περιοχές
μέσο τετραγωνικό σφάλμα (mse) rCGCI localization error	CGCI PDC	rCGCI

Πίνακας 3.1: Επίπεδα μετρήσεων

Για τα επίπεδα των 64 ηλεκτροδίων και 68 περιοχών, ο μεγάλος αριθμός μεταβλητών κάνει την προσαρμογή μοντέλου VAR αναξιόπιστη και ως εκ τούτου τις μετρικές CGCI και PDC ακατάλληλες. Για αυτό το λόγο στα συγκεκριμένα επίπεδα μετράται μόνο η μετρική rCGCI.

Επιπλέον, όλες οι μετρήσεις έγιναν σε τρία παράθυρα / περιόδους δειγμάτων:

- Σε όλα τα 5000 δείγματα
- Στα δείγματα 2000 με 3000 (τα οποία δεν περιέχουν spikes από ανοιγοκλείσιμο του ματιού)
- Στα δείγματα 3500 με 4350 (τα οποία περιέχουν τρία διαδοχικά spikes από ανοιγοκλείσιμο του ματιού)

Η ερμηνεία του mse είναι αρκετά απλή. Τα υπόλοιπα μεγέθη είναι μετρικές αιτιότητας, δηλαδή μετράνε κατά πόσο μία χρονοσειρά “οδηγεί” μία άλλη. Θα πρέπει επομένως να οριστεί ένα κατώφλι, πάνω από το οποίο θα θεωρείται ότι μία τιμή του εκάστοτε μεγέθους υποδηλώνει την ύπαρξη αιτιακής σχέσης μεταξύ των χρονοσειρών. Αυτά επιλέχθηκαν ως εξής:

rCGCI	0.15
CGCI	0.05
PDC	0.15

Η επιλογή των κατωφλίων έγινε έτσι ώστε να επιτυγχάνονται ικανοποιητικές μετρικές συνδεσιμότητας στα καθαρά δεδομένα.

Για το κατώφλι του CGCI στο επίπεδο των 7 περιοχών η επιλογή ήταν αρκετά εύκολη, καθώς στα καθαρά δεδομένα, η μικρότερη μέτρηση CGCI που υπολογίστηκε σε ζεύγος κόμβων με αιτιακή σύνδεση ήταν 0.1829, ενώ η μεγαλύτερη μέτρηση CGCI σε ζεύγος κόμβων χωρίς αιτιακή σύνδεση ήταν 0.0168. Ως εκ τούτου, οποιαδήποτε τιμή μεταξύ των δύο θα οδηγούσε σε 100% accuracy.

Για τις μετρικές του rCGCI και PDC υπήρχε overlap στο εύρος τιμών που μετρούνται σε ζεύγη κόμβων με και χωρίς αιτιακή σχέση. Αυτό σημαίνει ότι δυστυχώς είναι αδύνατο να πετύχουμε 100% accuracy. Επιπλέον αναφέρεται ότι θα ήταν ακατάλληλη η επιλογή του κατωφλίου με σκοπό τη μεγιστοποίηση του accuracy καθώς στα 4556 ζεύγη κόμβων του συστήματος υπάρχουν μόλις 10 αιτιακές συνδέσεις. Αυτό ουσιαστικά σημαίνει ότι το κατώφλι που μεγιστοποιεί το accuracy θα είναι πιθανότατα πολύ υψηλό με αποτέλεσμα να αναγνωρίζονται πολύ λίγες αιτιακές συνδέσεις στο σύστημα. Έτσι η επιλογή αυτών των κατωφλίων έγινε με σκοπό να επιτευχθεί μία ικανοποιητική ισορροπία μεταξύ sensitivity και specificity.

Συγκεκριμένα για το PDC, αναφέρεται ότι ο αλγόριθμος που χρησιμοποιήθηκε επιστρέφει τιμές PDC σε 50 ομοιόμορφα παράθυρα συχνοτήτων (frequency bins) από 0 Hz μέχρι και τη συχνότητα Nyquist. Καθώς τα δεδομένα μας είναι προσομοιωτικά, η συχνότητα δειγματοληψίας και ως εκ τούτου η συχνότητα Nyquist είναι αυθαίρετες. Για τον μετασχηματισμό σε μία τιμή ανά ζεύγος κόμβων και τη σύγχριση με το κατώφλι, υπολογίστηκε ο μέσος όρος των πρώτων 18 frequency bins.

Εφαρμόζοντας τα παραπάνω κατώφλια, παράγεται από κάθε μέτρηση αιτιότητας των μετρήσεων ο αντίστοιχος πίνακας γειτνίασης ενός γράφου δικτύου που απεικονίζει τις αιτιακές σχέσεις μεταξύ των περιοχών του εγκεφάλου, σύμφωνα με το εκάστοτε μέτρο αιτιότητας. Ο πίνακας γειτνίασης είναι μη συμμετρικός και τετραγωνικός λογικός πίνακας με μέγεθος ανάλογο με το επίπεδο στο οποίο έγιναν οι μετρήσεις. Για παράδειγμα, στο επίπεδο των 68 πηγών, ο πίνακας γειτνίασης θα είναι διαστάσεων (68 × 68) όπου οι τιμές των στοιχείων θα είναι 0 όταν η μετρική προβλέπει την έλλειψη αιτιακής σχέσης μεταξύ των πηγών και 1 όταν η μετρική προβλέπει την ύπαρξη αιτιακής σχέσης.

Για να ερμηνευθούν τα δίκτυα αυτά, πρέπει να συγχριθούν με το ground truth των καθαρών δεδομένων. Αυτό για τις μετρικές που υπολογίστηκαν σε χρονοσειρές εγκεφαλικών περιοχών είναι γνωστό - στο επίπεδο των 7 περιοχών

είναι ο πίνακας γειτνίασης του μοντέλου VAR(1) με το οποία παρήχθησαν τα δεδομένα των πηγών και στο επίπεδο των 68 περιοχών είναι ο ίδιος πίνακας γειτνίασης μετά από επέκταση σε διαστάσεις (68×68) (όπου όλα τα νέα ζεύγη κόμβων θεωρούνται ως μη (αιτιακά) συνδεμένα, άρα τα αντίστοιχα στοιχεία του πίνακα γειτνίασης θέτονται να είναι 0). Ωστόσο για τις μετρικές που υπολογίστηκαν σε δεδομένα ηλεκτροδίων το ground truth μας είναι άγνωστο. Για αυτό το λόγο, τα συγκεκριμένα μέτρα συγχρίνηκαν με τον αντίστοιχο πίνακα γειτνίασης που υπολογίστηκε από τα καθαρά δεδομένα, πριν την προσθήκη του θορύβου από το ανοιγοκλείσιμο του ματιού.

Τα παραπάνω μπορούν να θεωρηθούν ισοδύναμα με τη λύση ενός προβλήματος ταξινόμησης (classification) με δύο κλάσεις οι οποίες υποδηλώνουν την ύπαρξη ή μη αιτιακής σχέσης μεταξύ δύο κόμβων. Μετά το σχηματισμό των δικτύων και τη σύγχρισή τους με το ground truth, όπως αναλύθηκε προηγουμένως, σχηματίζονται τα confusion matrix που συνοψίζουν τα αποτελέσματα της ταξινόμησης. Η μορφή ενός confusion matrix δίνεται στο σχήμα 3.4.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Σχήμα 3.4: Confusion matrix από binary classification

Στο πλαίσιο των πειραμάτων μας, ως κλάση positive θεωρείται η ύπαρξη αιτιακής σχέσης και ως κλάση negative θεωρείται η έλλειψη αιτιακής σχέσης. True class είναι η πραγματική κλάση ενός ζεύγους κόμβων, όπως ορίζεται από τον πίνακα γειτνίασης του μοντέλου VAR στις μετρήσεις στο επίπεδο των περιοχών και από τον πίνακα γειτνίασης που υπολογίζεται στα καθαρά δεδομένα στις μετρήσεις στο επίπεδο των ηλεκτροδίων. Predicted Class είναι η τάξη που προβλέπει η εκάστοτε μετρική.

True positive (TP) λέγεται η περίπτωση που το predicted class είναι positive και το true class είναι επίσης positive.

True negative (TN) λέγεται η περίπτωση που το predicted class είναι negative και το true class είναι επίσης negative.

False positive (FP) λέγεται η περίπτωση που το predicted class είναι positive και το true class είναι negative.

False negative (FN) λέγεται η περίπτωση που το predicted class είναι negative και το true class είναι positive.

Στο confusion matrix απεικονίζονται το πλήθος των TP, FP, FN και TN. Από αυτά μπορούν να οριστούν επίσης:

Accuracy : ο λόγος του αριθμού των σωστών προβλέψεων προς το συνολικό αριθμό των προβλέψεων = $(TP+TN) / (TP+FP+FN+TN)$

Sensitivity : Ο λόγος του αριθμού των true positive προβλέψεων προς το πλήθος των στοιχείων τάξης positive = $TP / (TP + FN)$

Specificity : Ο λόγος του αριθμού των true negative προβλέψεων προς το πλήθος των στοιχείων τάξης negative = $TN / (TN + FP)$

Τα παραπάνω υπολογίζονται κάθε φορά που μετριέται κάποια μετρική αιτιότητας (CGCI, rCGCI ή PDC).

Μία επιπλέον ποσότητα που μετράται είναι το σφάλμα τοπικοποίησης (Localization error) των διαφορετικών μεθόδων καθαρισμού. Ως σφάλμα τοπικοποίησης ορίζεται το σφάλμα της εκτίμησης της σχετικής έντασης προβολής του θορύβου στο κάθε ηλεκτρόδιο. Για να υπολογιστεί χρειάζονται δύο στοιχεία: μία κανονικοποιημένη έχφραση της τοπικοποίησης του IC που θεωρείται θόρυβος και καθαρίζεται από την εκάστοτε μέθοδο καθαρισμού και την πραγματική τοπικοποίηση του θορύβου. Η πραγματική τοπικοποίηση του θορύβου μας είναι γνωστή: είναι το διάνυσμα προβολής θορύβου όπως ορίστηκε στην ενότητα 2.4. Η τοπικοποίηση του IC που θεωρείται θόρυβος μπορεί να γίνει με διάφορους τρόπους. Ένας σχετικά απλός τρόπος είναι να υπολογιστεί ο θόρυβος στο επίπεδο των ηλεκτροδίων, να υπολογιστεί η τυπική απόκλιση για το κάθε ηλεκτρόδιο σε ένα παράθυρο παρατηρήσεων κατά το οποίο υπάρχει ανοιγοκλείσιμο του ματιού και τέλος να γίνει κανονικοποίηση με διαίρεση με τη μέγιστη τιμή. Ωστόσο αυτή η μέθοδος δεν διατηρεί το πρόσημο της προβολής του θορύβου. Για αυτό το λόγο, αντί για την τυπική απόκλιση χρησιμοποιήθηκε η ποσότητα

$$s_3 = \sqrt[3]{\frac{\sum(x_i - m)^3}{N}} \quad (3.1)$$

Ως σφάλμα τοπικοποίησης ορίζεται η διαφορά των δύο (χανονικοποιημένων) εκφράσεων που αναφέρθηκαν παραπάνω. Το σφάλμα τοπικοποίησης υπολογίζεται για τις μεθόδους καθαρισμού ICA, stripped ICA και noise mixing approximation. Ο καθαρισμός με wICA χρησιμοποιεί ακριβώς την ίδια τοπικοποίηση θορύβου με την ICA και ως εκ τούτου δεν θα είχε νόημα να υπολογιστεί το σφάλμα τοπικοποίησης και για τις δύο μεθόδους.

Ο λόγος που μετράται η παραπάνω ποσότητα είναι απλός. Όπως αναφέρθηκε προηγουμένως, η χρήση ICA γίνεται με την παραδοχή ότι θα καθαριστεί μόνο ένα ποσοστό του θορύβου. Υπολογίζοντας το σφάλμα τοπικοποίησης μπορούμε να ελέγξουμε κατά πόσο αυτό το ποσοστό είναι παρόμοιο για όλα τα ηλεκτρόδια, ή κατά πόσο υπάρχουν ανομοιογένειες για παράδειγμα σε περίπτωση που η ICA λειτουργεί ικανοποιητικά για τα μπροστινά ηλεκτρόδια αλλά όχι για όλα τα υπόλοιπα.

Συνοψίζοντας, το βάσικο πείραμα περιλαμβάνει την παραγωγή 6 EEG: αρχικό καθαρό EEG που παράγεται θεωρώντας πηγές των οποίων η δραστηριότητα είτε προέρχεται από το μοντέλο VAR είτε είναι ροζ θόρυβος και λύνοντας το forward problem, θορυβώδες EEG που παράγεται προσθέτοντας το θόρυβο στο καθαρό EEG και EEG καθαρισμένο με μία από τις 4 μεθόδους που αναφέρθηκαν. Για κάθε EEG (συμπεριλαμβανομένου του καθαρού) γίνεται και η ανακατασκευή των πηγών και εκτελούνται οι μετρήσεις που αναφέρθηκαν. Το παραπάνω πείραμα εκτελέστηκε 100 φορές για τη συλλογή αρκετών μετρήσεων.

Σε κάθε εκτέλεση χρησιμοποιήθηκε διαφορετικό random seed ώστε να διαφοροποιούνται τα δεδομένα πηγών και ροζ θορύβου.

Επιπλέον, σε κάθε εκτέλεση τα 64 στοιχεία του διανύσματος προβολής θορύβου πολλαπλασιάστηκαν με 64 τυχαίες τιμές από ομοιόμορφη κατανομή με εύρος τιμών [0.975 1.025]

Ο σκοπός των δύο αυτών βημάτων είναι να εισάγουμε κάποιες διαφοροποιήσεις μεταξύ των πειραμάτων.

Όλες οι υπόλοιπες παράμετροι ήταν κοινές για όλα τα πειράματα. Ο πίνακας γειτνίασης του VAR(1) μοντέλου που παρήγαγε τη δραστηριότητα των πηγών δίνεται στον πίνακα 3.2. Οι 7 εγκεφαλικές περιοχές στις οποίες προστίθεται η δραστηριότητα του VAR(1) μοντέλου είναι οι:

- left bank of the superior temporal sulcus (l.bankssts)
- right bank of the superior temporal sulcus (r.bankssts)
- left caudal middle frontal gyrus

0	1	0	0	0	0	0
0	0	1	0	0	0	0
0	1	0	1	0	0	0
0	0	1	0	1	0	0
0	0	0	1	0	1	0
0	0	0	0	1	0	0
0	0	0	0	0	1	0

Πίνακας 3.2: Πίνακας γειτνίασης μοντέλων VAR

- right caudal middle frontal gyrus
- left cuneus
- left endorhinal cortex
- left paracentral lobule

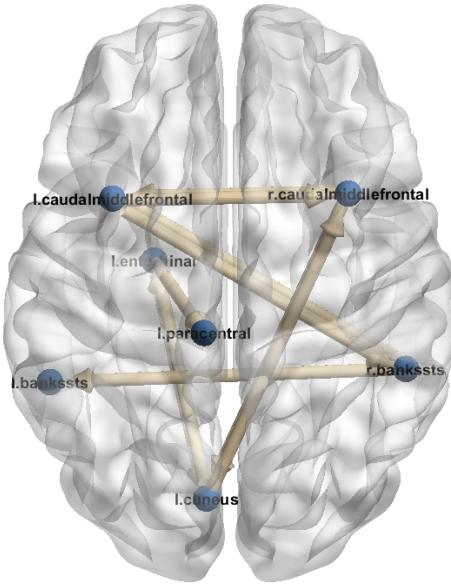
3.3 Παράδειγμα ανάλυσης πειράματος

Για τη διευκόλυνση κατανόησης της διαδικασίας, δίνονται αναλυτικά τα αποτελέσματα ενός πειράματος.

Όπως έχει αναφερθεί, τα αρχικά καθαρά δεδομένα παράγονται δημιουργώντας 7 χρονοσειρές από μοντέλο VAR(1) και 61 χρονοσειρές ροζ θορύβου, θεωρώντας αυτές τις χρονοσειρές ως τη δραστηριότητα των 68 διακριτών περιοχών του εγκεφάλου που υπολογίζονται στη συνδεσιμότητα και με σύγκριση με τον πίνακα γειτνίασης του VAR(1) μοντέλου, πετυχαίνουμε 100% accuracy στο επίπεδο των 7 περιοχών και σε όλα τα παράθυρα μετρήσεων με τη μετρική CGCI και με κατώφλι 0.05. Η τοπολογία του δικτύου που παράγεται από την παραπάνω μέτρηση δίνεται στο σχήμα 3.5.

Η αντίστοιχη τοπολογία του δικτύου που παράγεται από τη μετρική PDC σε όλες τις 5000 παρατηρήσεις δίνεται στο σχήμα 3.6.

Σε αυτό το δίκτυο υπάρχουν 2 misclassifications σε σχέση με τον πίνακα γειτνίασης του μοντέλου VAR(1). Και τα δύο λάθη αφορούν συνδέσεις που εντοπίστηκαν ενώ δεν υπήρχαν στα πραγματικά δεδομένα, επομένως το sensitivity της μέτρησης είναι 100%, ωστόσο το specificity είναι 94.87% ενώ το συνολικό accuracy είναι 95.92%. Η ίδια μέτρηση στα δύο μικρότερα παράθυρα παρατηρήσεων παρήγαγε δίκτυα με 5 και 7 λάθος συνδέσεις. Πάλι, το sensitivity ήταν 100%, επομένως τα λάθη αφορούσαν συνδέσεις που εντοπίστηκαν

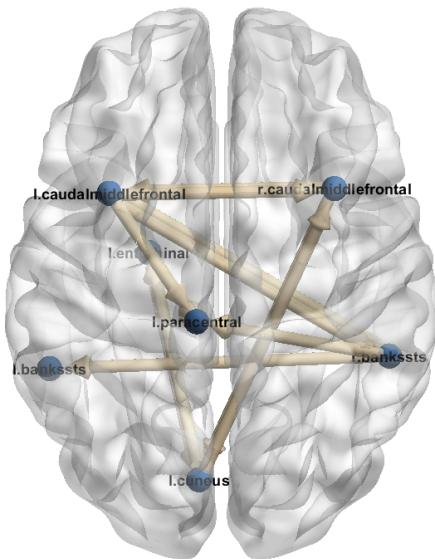


Σχήμα 3.5: Δίκτυο CGCI, καθαρά δεδομένα 7 περιοχών

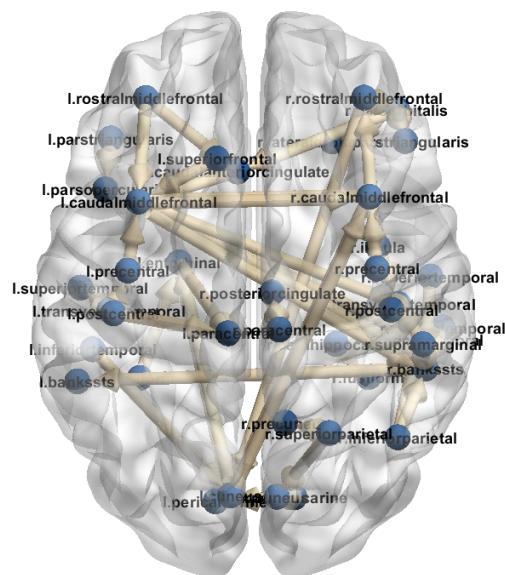
ενώ δεν υπάρχουν. Η μειωμένη αποδοτικότητα στα μικρότερα παράθυρα ήταν αναμενόμενη καθώς αυτά περιέχουν μόλις 1000 και 850 παρατηρήσεις.

Στο επίπεδο των 68 περιοχών, το δίκτυο που παράγεται από τη μετρική rCGCI έχει 35, 48 και 54 misclassifications στα παράθυρα 5000, 1000 και 850 παρατηρήσεων αντίστοιχα. Στα πρώτα δύο παράθυρα παρατηρήσεων το sensitivity είναι 100%. Στο τρίτο παράθυρο, ένα από τα λάθη αφορά μία σύνδεση που υπήρχε, αλλά δεν εντοπίστηκε, με αποτέλεσμα το sensitivity να είναι 90%. Το δίκτυο που παράγεται από το παράθυρο των 5000 παρατηρήσεων φαίνεται στο σχήμα 3.7. Στο σχήμα αναπαραριστούνται ως κόμβοι μόνο οι περιοχές οι οποίες συμμετέχουν σε κάποιο ζέυγος με αιτιακή σύνδεση

Αναφέρεται ότι ο αριθμός των misclassifications ήταν αναμενόμενος, καθώς η δραστηριότητα ανά περιοχή υπολογίζεται χρησιμοποιώντας όλα τα δίπολα στην ευρύτερη περιοχή αλλά και καθώς η ανακατασκευή των πηγών με τον αλγόριθμο bsbl είναι απλά μία προσέγγιση και είναι αναμενόμενο να εισάγει κάποιο σφάλμα, ειδικά σε περιοχές που είναι γειτονικές σε κάποια πηγή δραστηριότητας. Παρ' όλα αυτά, η ακρίβεια του δικτύου θεωρείται ικανοποιητική, καθώς το sensitivity ήταν 100% σε δύο από τα τρία παράθυρα μετρήσεων. Επίπλεον αναφέρεται ότι το specificity ήταν 99.24%, 98.96% και 98.85% για τα 3 παράθυρα μετρήσεων. Παρακάτω δίνονται τα confusion matrix των μετρήσεων rCGCI στις 5000, 1000 και 850 παρατηρήσεις αντίστοιχα.



Σχήμα 3.6: Δίκτυο PDC, καθαρά δεδομένα 7 περιοχών



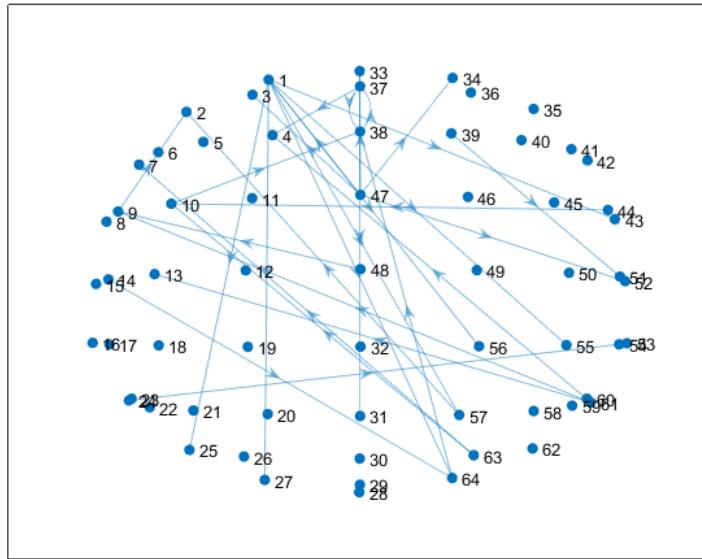
Σχήμα 3.7: Δίκτυο rCGCI, καθαρά δεδομένα 68 περιοχών

		Real value		Total
Prediction	p	n		
	p	10	35	10 + 35
n	0	4579		4579
Total	10	35 + 4579		

		Real value		Total
Prediction	p	n		
	p	48		10 + 48
	n	4566		4566
Total	10	48 + 4566		

		Real value		Total
Prediction	p	n		
	p	53		9 + 53
	n	4561		4561 + 1
Total	9 + 1	53 + 4561		

Όσον αφορά το επίπεδο των ηλεκτροδίων, η μετρική rCGCI εντοπίζει 30, 79 και 72 αιτιακές συνδέσεις στα παράθυρο των 5000, 1000 και 850 παρατηρήσεων αντίστοιχα. Ο γράφος του δικτύου από τις μετρήσεις στις 5000 παρατηρήσεις φαίνεται στο σχήμα 3.8.



Σχήμα 3.8: Δίκτυο rCGCI, καθαρά δεδομένα 64 ηλεκτροδίων

Όπως έχει αναφερθεί, η ερμηνεία αυτών των δικτύων είναι πιο δύσκολη, καθώς δεν υπάρχει κάποιος γνωστός πίνακας γειτνίασης με τον οποίο να μπορούμε να τα συγχρίνουμε.

Τα ψηρυβώδη δεδομένα υπολογίζονται προσθέτοντας τη χρονοσειρά του ψο-

ρύβου πολλαπλασιασμένη με ένα συντελεστή σε κάθε ένα από τα 64 ηλεκτρόδια του καθαρού EEG. Στη συνέχεια λύνεται και το inverse problem για την ανάλυση συνδεσιμότητας στο επίπεδο των εγκεφαλικών περιοχών.

Για τα δεδομένα με θόρυβο, στο επίπεδο των 7 πηγών η μετρική του CGCI πετυχαίνει πάλι 100% accuracy στα παράθυρα των 5000 και 1000 παρατηρήσεων. Στο τελευταίο παράθυρο, δηλαδή στο παράθυρο με τις 850 παρατηρήσεις κατά το οποίο εμφανίζονται τρία spikes από ανοιγοκλείσιμο του ματιού, υπάρχει ένα false negative λάθος, με αποτέλεσμα το sensitivity να είναι 90% για αυτή τη μέτρηση.

Στό ίδιο επίπεδο, η μετρική PDC κάνει 7,10 και 7 misclassifications στα τρία παράθυρα μετρήσεων. Σε όλες τις περιπτώσεις υπάρχει μία πραγματική σύνδεση που δεν εντοπίστηκε ενώ όλα τα υπόλοιπα λάθη είναι false negatives. Επομένως το sensitivity είναι 90%.

Στο επίπεδο των 68 περιοχών, η μετρική rCGCI κάνει 25, 54 και 40 misclassifications στα τρία παράθυρα μετρήσεων. Στον πίνακα 3.3 δίνονται τα confusion matrices για τις μετρήσεις στις 5000, 1000 και 850 παρατηρήσεις αντίστοιχα.

	p	n		p	n		p	n
p	9	24	p	10	54	p	10	40
n	1	4590	n	0	4560	n	0	4574

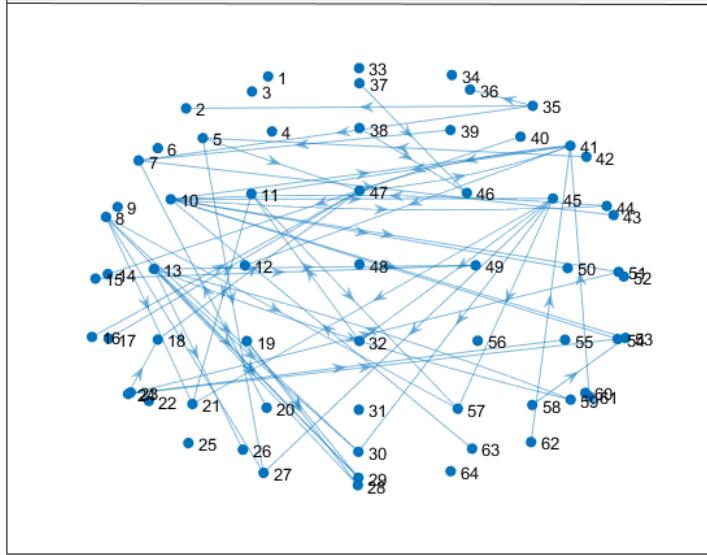
3.3: Confusion matrices rCGCI, θορυβώδη δεδομένα 68 περιοχών

Και σε αυτή την περίπτωση, φαίνεται ότι το sensitivity είναι μάλλον πιο κατάλληλη μετρική για την αξιολόγηση των δικτύων από το accuracy. Συγκεχριμένα, βλέπουμε ότι ο αριθμός των misclassifications ήταν μικρότερος μετά την προσθήκη του θορύβου. Ωστόσο, το sensitivity στις μετρήσεις στις 5000 παρατηρήσεις ήταν χαμηλότερο.

Στο επίπεδο των ηλεκτροδίων, τα αποτελέσματα της ανάλυσης είναι αρκετά διαφορετικά μετά την προσθήκη του θορύβου. Στο σχήμα 3.9 φαίνεται ο γράφος του δικτύου που υπολογίζεται από το rCGCI στις 5000 παρατηρήσεις.

Ο παραπάνω γράφος μπορεί να συγχριθεί με τον αντίστοιχο γράφο που είχε υπολογιστεί στα καθαρά δεδομένα για να αξιολογήσουμε κατά πόσο η προσθήκη του θορύβου επηρέασε την ανάλυση συνδεσιμότητας στο επίπεδο των ηλεκτροδίων. Από αυτή τη σύγκριση παράγονται τα παρακάτω confusion matrices για τα τρία παράθυρα παρατηρήσεων.

Είναι προφανές ότι ο θόρυβος είχε πολύ ισχυρή επίδραση στην ανάλυση συνδεσιμότητας στο επίπεδο των ηλεκτροδίων. Αυτό φαίνεται καλύτερα από το sensitivity το οποίο έχει πέσει στο 10% και 8% στα παράθυρα 5000 και 850



Σχήμα 3.9: Δίκτυο rCGCI, θορυβώδη δεδομένα 64 ηλεκτροδίων

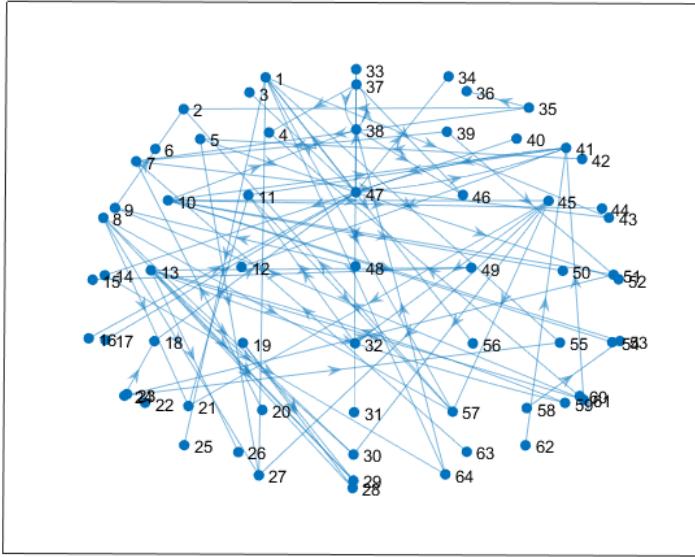
	p	n		p	n		p	n
p	3	56	p	79	0	p	6	99
n	27	4010	n	0	4017	n	66	3925

3.4: Confusion matrices rCGCI, θορυβώδη δεδομένα 64 ηλεκτροδίων

παρατηρήσεων αντίστοιχα. Σημειώνεται ότι στο παράθυρο των 1000 παρατηρήσεων δεν έχει προστεθεί θόρυβος, επομένως τα αποτελέσματα της ανάλυσης είναι ίδια με αυτά της ανάλυσης στα καθαρά δεδομένα. Όπως έχει αναφερθεί η ερμηνεία αυτών των αποτελεσμάτων είναι αρκετά πολύπλοκη, ωστόσο μία παρατήρηση που θα μπορούσε να γίνει είναι ότι ο συνολικός αριθμός των ηλεκτροδίων που συμμετέχουν σε κάποια αιτιακή σχέση φαίνεται να έχει αυξηθεί αρκετά με την προσθήκη του θορύβου, καθώς και ότι αυτές οι συνδέσεις εμφανίζονται σε όλή την έκταση του εγκεφάλου.

Τα παραπάνω μπορούν να οπτικοποιηθούν και σχηματίζοντας το δίκτυο misclassifications. Αυτό παράγεται από το γράφο που έχει ως κόμβους τα ηλεκτρόδια και ως ακμές τα ζεύγη ηλεκτροδίων που συμμετείχαν σε κάποιο misclassification. Στο συγκεκριμένο παράδειγμα το δίκτυο αυτό φαίνεται στο σχήμα 3.10.

Στη συνέχεια δίνονται τα αποτελέσματα για τις 4 μεθόδους καθαρισμού.



Σχήμα 3.10: Misclassifications rCGCI, θιρυβώδη δεδομένα 64 ηλεκτροδίων

Στους πίνακες η μέθοδος stripped ICA αναφέρεται ως sICA και η μέθοδος noise mixing approximation ως NMA.

Αρχικά στον πίνακα 3.5 δίνονται τα αποτελέσματα μετρήσεων στο επίπεδο των 7 περιοχών και στο παράθυρο των 5000 παρατηρήσεων. Τα αποτελέσματα στα δύο μικρότερα παράθυρα μετρήσεων είναι παρόμοια.

CGCI	ICA	wICA	sICA	NMA
Misclassifications	0	0	0	0
Accuracy	100%	100%	100%	100%
Specificity	100%	100%	100%	100%
Sensitivity	100%	100%	100%	100%
PDC	ICA	wICA	sICA	NMA
Misclassifications	10	9	7	2
Accuracy	79.5%	81.6%	85.7%	95.9%
Specificity	76.9%	79.4%	84.6%	94.8%
Sensitivity	90%	90%	90%	100%

3.5: Μετρικές ταξινόμησης, 7 περιοχές, 5000 παρατηρήσεις

Στο επίπεδο των 68 περιοχών τα αντίστοιχα αποτελέσματα για τη μετρική rCGCI δίνονται στον πίνακα 3.6

rCGCI	ICA	wICA	sICA	NMA
Misclassifications	24	26	30	44
Accuracy	99.4%	99.4%	99.3%	99%
Specificity	99.5%	99.4%	99.3%	99%
Sensitivity	90%	100%	80%	100%

3.6: Μετρικές ταξινόμησης, 68 περιοχές, 5000 παρατηρήσεις

Τέλος, η αντίστοιχη ανάλυση στο επίπεδο των ηλεκτροδίων δίνεται στον πίνακα 3.7

rCGCI	ICA	wICA	sICA	NMA
Misclassifications	79	75	35	40
Accuracy	98%	98.1%	99.1%	99%
Specificity	98.7%	98.8%	99.7%	99.7%
Sensitivity	6.6%	6.6%	13.3%	3.3%

3.7: Μετρικές ταξινόμησης, 64 ηλεκτρόδια, 5000 παρατηρήσεις

Επιπλέον από την ανάλυση στο επίπεδο των ηλεκτροδίων δίνεται στον πίνακα 3.8 το μέσο τετραγωνικό σφάλμα στα ηλεκτρόδια Fp1, Fpz, Fp2 (τα οποία βρίσκονται στο μπροστινό μέρος του κεφαλιού) και C1, Cz, C2 (τα οποία βρίσκονται στο κέντρο του κεφαλιού).

MSE	Noisy	ICA	wICA	sICA	NMA
FP1	14124080	4012402	3066677	47071	46699
FPz	15494627	4396964	3346145	51628	51628
FP2	13528237	3831911	2890923	44529	45135
C1	477775	137135	96183	6707	17120
Cz	203835	83367	47657	3801	678
C2	194354	11429	61203	4610	965

3.8: Μέσο τετραγωνικό σφάλμα 6 ηλεκτροδίων, 5000 παρατηρήσεις

Για το συγκεκριμένο παράδειγμα αναφέρεται ότι για την περίπτωση ανάλυσης συνδεσμότητας στο επίπεδο των 68 περιοχών, όπου ο καθαρισμός με ICA προκαλεί τα λιγότερα misclassifications, οι μέθοδοι wICA και NMA ενδεχομένως

να είναι καταλληλότερες καθώς πετυχαίνουν καλύτερο sensitivity. Αναφέρεται επίσης ότι στο επίπεδο των 64 ηλεκτροδίων, καμία από τις μεθόδους καθαρισμού δεν καταφέρνει να επαναφέρει τη συνδεσιμότητα των καθαρών δεδομένων. Ακόμα και όταν ο αριθμός των misclassifications είναι σχετικά χαμηλός, όπως στην περίπτωση καθαρισμού με sICA ή NMA, το sensitivity είναι πολύ χαμηλό. Όσον αφορά τη συνδεσιμότητα στο επίπεδο των 7 πηγών, δε φαίνεται να επηρεάστηκε σημαντικά ούτε από την προσθήκη θορύβου, ούτε από τον καθαρισμό με οποιαδήποτε από τις 4 μεθόδους όταν η μετρική συνδεσιμότητας που χρησιμοποιείτε είναι το CGCI. Η μετρική PDC εμφανίζει σημαντικές διαφορές μεταξύ καθαρών και θορυβώδων δεδομένων, ωστόσο η συγκεκριμένη μετρική φαίνεται να είναι γενικά λιγότερο αποτελεσμάτικη στο πλαίσιο αυτού του πειράματος, καθώς δεν πετυχαίνει 100% accuracy ακόμα και στα αρχικά καθαρά δεδομένα.

Κεφάλαιο 4

Αποτελέσματα και συζήτηση

4.1 Αποτελέσματα στο επίπεδο των 7 περιοχών

Από τα 100 πειράματα, ο μέσος όρος των μετρικών που υπολογίστηκαν στο παράθυρο των 5000, 1000 και 850 παρατηρήσεων δίνονται στους πίνακες 4.1, 4.2 και 4.3 αντίστοιχα. Επίσης στα θηχογράμματα 4.1 και 4.2 δίνεται η κατανομή των misclassifications για τις 2 μετρικές συνδεσιμότητας στο παράθυρο 850 παρατηρήσεων.

CGCI	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	0	0	0.02	0	0.02	0
Accuracy	100%	100%	99.9%	100%	99.9%	100%
Specificity	100%	100%	100%	100%	100%	100%
Sensitivity	100%	100%	99.8%	100%	99.8%	100%
PDC	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	3.38	5.28	6.62	6.54	3.76	3.23
Accuracy	93.1%	89.2%	86.4%	86.6%	92.3%	93.4%
Specificity	91.4%	88.2%	85.9%	85.7%	90.8%	91.8%
Sensitivity	99.5%	93%	88.7%	90.2%	98.2%	99.5%

4.1: Μέσος όρος μετρικών ταξινόμησης, 7 περιοχές, 5000 παρατηρήσεις

Οπως φάνηκε και από το παράδειγμα που δόθηκε προηγουμένως, στα παράθυρα 5000 και 1000 παρατηρήσεων η μετρική CGCI καταφέρνει να υπολογίσει τη συνδεσιμότητα με 100% accuracy ακόμα και όταν προστίθεται θόρυβος. Μετά από καθαρισμό, υπάρχουν λίγα misclassifications με τις μεθόδους ICA και sICA, ωστόσο αυτά θα μπορούσαν να θεωρηθούν outliers καθώς το συνολικό

CGCI	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	0	0	0.18	0.03	0.02	0
Accuracy	100%	100%	99.6%	99.9%	99.9%	100%
Specificity	100%	100%	99.7%	99.9%	99.9%	100%
Sensitivity	100%	100%	99.3%	100%	99.9%	100%
PDC	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	5.62	6.25	8.55	9.22	6.9	5.7
Accuracy	88.5%	87.2%	82.5%	81.8%	85.9%	88.3%
Specificity	85.9%	84.1%	80.5%	78%	83%	85.5%
Sensitivity	98.8%	99.3%	90.2%	93.6%	97.3%	99.2%

4.2: Μέσος όρος μετρικών ταξινόμησης, 7 περιοχές, 1000 παρατηρήσεις

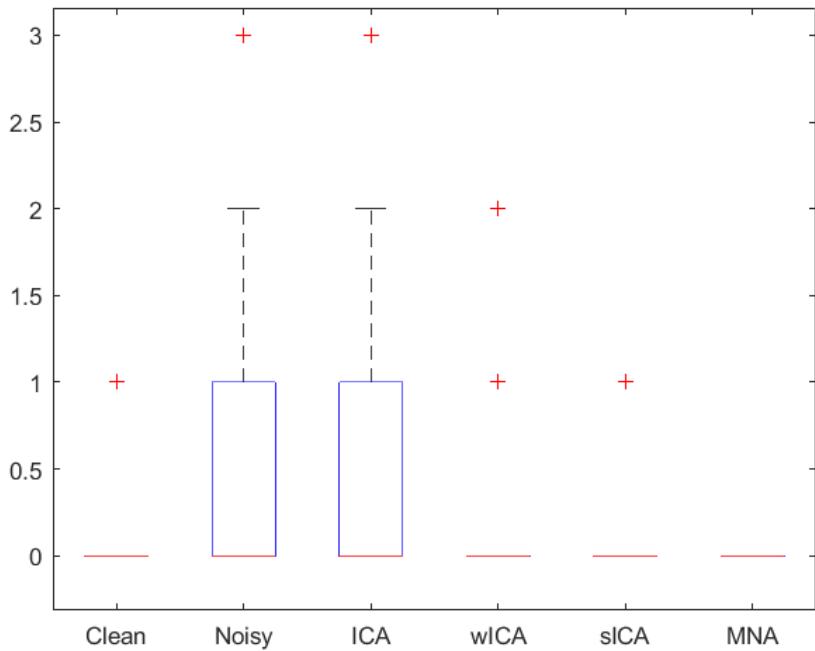
CGCI	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	0.01	0.62	0.54	0.23	0.03	0
Accuracy	99.9%	98.7%	98.9%	99.5%	99.9%	100%
Specificity	99.9%	99.4%	99.2%	99.7%	99.9%	100%
Sensitivity	100%	95.8%	97.5%	98.8%	100%	100%
PDC	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	5.78	9.81	9.87	10.23	7.58	6
Accuracy	88.2%	79.9%	79.8%	79.1%	84.5%	87.7%
Specificity	85.5%	77.8%	78%	76.5%	81.3%	84.8%
Sensitivity	98.7%	88.4%	86.8%	89%	97.1%	99%

4.3: Μέσος όρος μετρικών ταξινόμησης, 7 περιοχές, 850 παρατηρήσεις

accuracy είναι υψηλότερο από 99%.

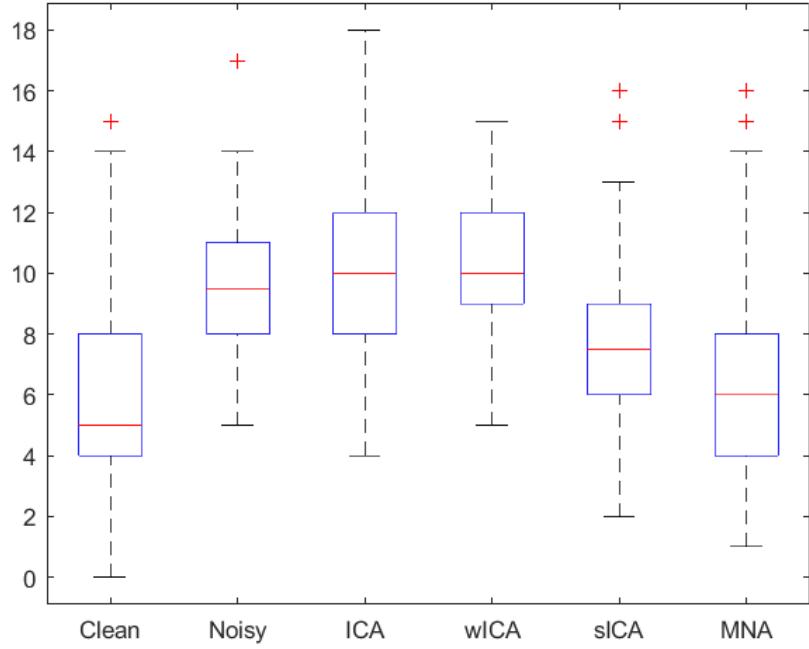
Η μετρική PDC φαίνεται λιγότερο αποτελεσματική καθώς εμφανίζονται misclassifications ακόμη και στα καθαρά δεδομένα, ωστόσο από αυτές τις μετρήσεις φαίνεται καλύτερα η επιρροή του θορύβου και του καθαρισμού στη συνδεσιμότητα. Συγκεκριμένα, φαίνεται ότι το sensitivity (και το specificity) πέφτει όταν προστίθεται θόρυβος. Επιπλέον φαίνεται ότι ο καθαρισμός με ICA ή wICA πολλές φορές οδηγεί σε περισσότερα σφάλματα από τον θόρυβο. Αυτό το φαινόμενο είναι πιο προφανές στο παράθυρο των 1000 παρατηρήσεων. Υπενθυμίζεται ότι σε αυτό το παράθυρο δεν έχει προστεθεί θόρυβος και ως εκ τούτου οποιαδήποτε μέθοδος καθαρισμού προκαλεί αλλαγές θα εισάγει εξ ορισμού θόρυβο. Υπενθυμίζεται ακόμη ότι η wICA είναι μία μέθοδος που αναπτύχθηκε συγκεκριμένα για την ελαχιστοποίηση των διαφοροποιήσεων του καθαρισμού δεδομένων σε περιόδους παρατηρήσεων στις οποίες δεν υπάρχει

θόρυβος. Από αυτήν την άποψη είναι πιο αποτελεσματική από την ICA (93.6% έναντι 90.2% sensitivity στο παράθυρο μετρήσεων χωρίς θόρυβο), ωστόσο δε συγκρίνεται καλά με τις μεθόδους sICA και NMA. Επιπλέον, μόνο η μέθοδος NMA πετυχαίνει το sensitivity που θα είχαμε από την ανάλυση θορυβώδων δεδομένων.



Σχήμα 4.1: Κατανομή misclassifications CGCI, 7 περιοχές, 850 παρατηρήσεις

Όπως ήταν αναμενόμενο, το παράθυρο των 850 μετρήσεων είναι αυτό στο οποίο υπάρχουν οι πιο αισθητές διαφορές στη συνδεσιμότητα μεταξύ καθαρών και θορυβώδων ή καθαρισμένων δεδομένων. Σε αυτό το παράθυρο οι διαφορές φαίνονται και στη μετρική CGCI, καθώς το sensitivity πέφτει στο 95.8%. Ο καθαρισμός με ICA ή wICA βελτιώνει την ανάλυση, ωστόσο τα καλύτερα αποτελέσματα επιτυγχάνονται από τις μεθόδους sICA και NMA με τις οποίες το sensitivity είναι 100%. Για τη μετρική PDC τα συμπεράσματα είναι παρόμοια. Στο θηκόγραμμα 4.2 φαίνεται ότι σε ορισμένες εκτελέσεις του πειράματος οι μέθοδοι sICA και NMA έχουν όσα misclassifications έχουν και οι άλλες μέθοδοι. Ωστόσο όπως φαίνεται από το θηκόγραμμα 4.3, ειδικά για τη μέθοδο NMA, πολύ λίγα από αυτά τα σφάλματα είναι false negative και επομένως το sensitivity παραμένει υψηλό.



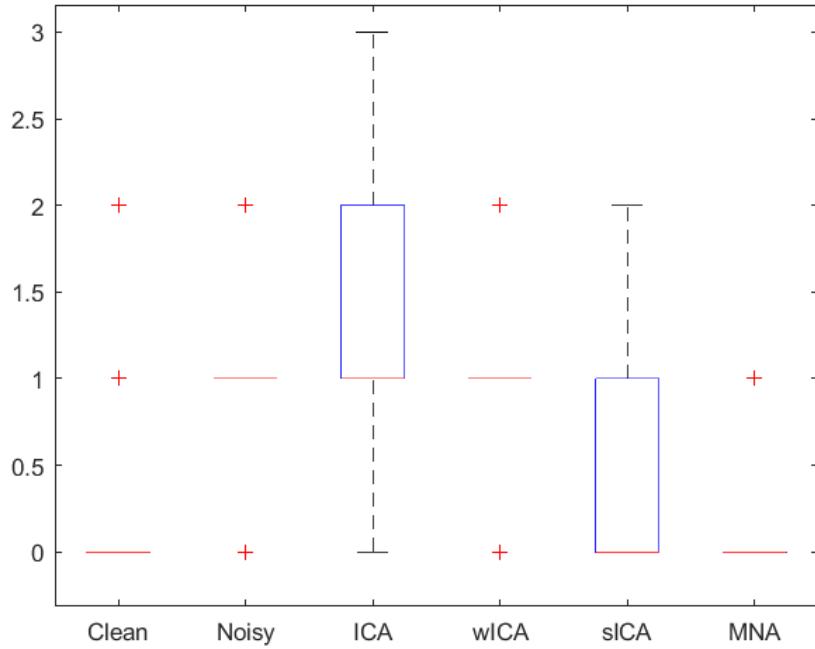
Σχήμα 4.2: Κατανομή misclassifications PDC, 7 περιοχές, 850 παρατηρήσεις

Ένα προφανές συμπέρασμα της παραπάνω ανάλυσης είναι ότι η μετρική PDC δεν φαίνεται να είναι όσο ακριβής είναι η CGCI. Ένας λόγος για αυτό ενδεχομένως να είναι ο τρόπος με τον οποίο δημιουργήθηκαν τα προσομοιωτικά δεδομένα, δηλαδή από ένα μοντέλο VAR(1) χωρίς να δωθεί προσοχή στις συχνότητες των σημάτων που παράγονται. Εάν οι πηγές προσομοιώνονταν με τρόπο ώστε να έχουν διαφορετικά χαρακτηριστικά στο πεδίο της συχνότητας, ενδεχομένως η PDC να είχε καλύτερα αποτελέσματα.

4.2 Αποτελέσματα στο επίπεδο των 68 περιοχών

Όπως και για το επίπεδο 7 περιοχών, αρχικά δίνονται οι μέσοι όροι των μετρικών ταξινόμησης στους πίνακες 4.4, 4.5 και 4.6.

Στα παράθυρα των 5000 και 1000 παρατηρήσεων, η προσθήκη του θορύβου φαίνεται να έχει ελάχιστη επιρροή στη μελέτη συνδεσιμότητας. Μάλιστα, όπως φάνηκε και στο επίπεδο των 7 περιοχών όταν χρησιμοποιούνταν η μετρική PDC, ο καθαρισμός των δεδομένων φαίνεται να δυσχεραίνει την ανάλυση συνδεσιμότη-



Σχήμα 4.3: Κατανομή false negative PDC, 7 περιοχές, 850 παρατηρήσεις

RCGCI	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	32.14	24.66	23.49	25.31	23.91	30.72
Accuracy	99.3%	99.4%	99.4%	99.4%	99.4%	99.3%
Specificity	99.3%	99.4%	99.5%	99.4%	99.5%	99.3%
Sensitivity	97.5%	96%	84.5%	94.9%	88.7%	93.5%

4.4: Μέσος όρος μετρικών ταξινόμησης, 68 περιοχές, 5000 παρατηρήσεις

RCGCI	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	49.46	51.94	35.31	44.92	38.2	47.19
Accuracy	98.9%	98.8%	99.2%	99%	99.1%	98.9%
Specificity	98.9%	98.8%	99.2%	99%	99.1%	98.9%
Sensitivity	99.1%	98.8%	90.5%	98.8%	96.4%	98.2%

4.5: Μέσος όρος μετρικών ταξινόμησης, 68 περιοχές, 1000 παρατηρήσεις

τας. Ειδικά με τη μέθοδο καθαρισμού ICA, το sensitivity πέφτει σε 84.5% και

RCGCI	Clean	Noisy	ICA	wICA	sICA	NMA
Misclassifications	52.41	47.19	38.15	43.1	38.36	50.18
Accuracy	98.8%	98.9%	99.1%	99%	99.1%	98.9%
Specificity	98.8%	98.9%	99.2%	99%	99.1%	98.9%
Sensitivity	98.6%	93.2%	89.9%	95.4%	96.8%	97.7%

4.6: Μέσος όρος μετρικών ταξινόμησης, 68 περιοχές, 850 παρατηρήσεις

90.5%, αρκετά χαμηλότερο από το sensitivity που επιτυγχάνονταν στα θορυβώδη δεδομένα. Μία διαφοροποίηση σε σχέση με την ανάλυση στο επίπεδο των 7 περιοχών, είναι ότι η wICA φαίνεται να πετυχαίνει να αφήσει σχετικά ανεπηρέαστες τις περιόδους παρατηρήσεων στις οποίες δεν υπάρχει θόρυβος. Αυτό φαίνεται από τις μετρικές στις 1000 παρατηρήσεις, στις οποίες η μέθοδος πετυχαίνει το ίδιο sensitivity (και καλύτερο από τις υπόλοιπες μεθόδους καθαρισμού) με τα θορυβώδη δεδομένα (τα οποία στην πραγματικότητα είναι καθαρά, καθώς δεν έχει προστεθεί θόρυβος σε αυτό το παράθυρο παρατηρήσεων) με χαμηλότερο αριθμό συνολικών misclassifications.

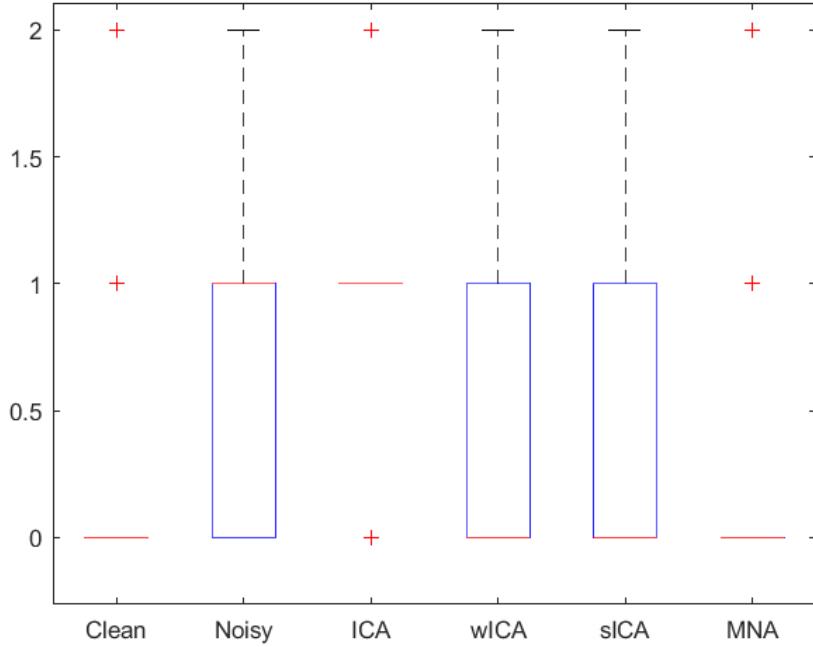
Όπως ήταν αναμενόμενο, το παράθυρο των 850 παρατηρήσεων είναι αυτό με τις πιο αισθητές διαφορές μεταξύ καθαρών και θορυβώδων δεδομένων και άρα το παράθυρο με το μεγαλύτερο χώρο για βελτίωση από τον καθαρισμό. Ωστόσο ακόμα και σε αυτή την περίπτωση ο καθαρισμός με ICA οδηγεί σε χαμηλότερο sensitivity από την ανάλυση στα θορυβώδη δεδομένα. Σε αυτό το παράθυρο το καλύτερο sensitivity επιτυγχάνεται από τη μέθοδο NMA, ωστόσο πρέπει να αναφερθεί ότι οι μέθοδοι wICA και sICA έχουν επίσης ικανοποιητικό sensitivity ενώ κρατάνε το συνολικό αριθμό misclassifications χαμηλότερο.

Στα θηκογράμματα 4.4 και 4.5 φαίνονται οι κατανομές των false negative και false positive αντίστοιχα για το παράθυρο 850 παρατηρήσεων.

Επίσης, στους πίνακες 4.7, 4.8 και 4.9 δίνονται τα αποτελέσματα της άνθροισης των 100 confusion matrices για κάθε μία από τις μεθόδους καθαρισμού.

ICA	p	n	wICA	p	n	sICA	p	n	NMA	p	n
p	845	2194	p	949	2480	p	887	2278	p	935	3007
n	155	459206	n	51	458920	n	113	459122	n	65	458393

4.7: Ανθροισμένα confusion matrices rCGCI, 68 περιοχές, 5000 παρατηρήσεις



Σχήμα 4.4: Κατανομή false negative RCGCI, 68 περιοχές, 850 παρατηρήσεις

ICA		wICA		sICA		NMA	
p	n	p	n	p	n	p	n
905	3436	988	4480	964	3789	982	4701
95	457964	12	456920	36	457611	18	456699

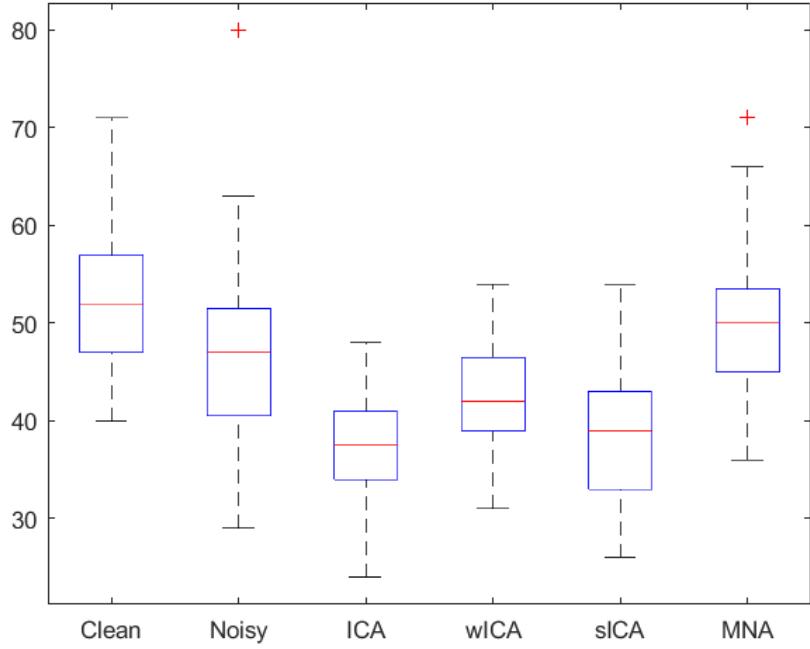
4.8: Αθροισμένα confusion matrices rCGCI, 68 περιοχές, 1000 παρατηρήσεις

ICA		wICA		sICA		NMA	
p	n	p	n	p	n	p	n
899	3714	954	4264	968	3831	977	4995
101	457686	46	457136	32	457569	23	456405

4.9: Αθροισμένα confusion matrices rCGCI, 68 περιοχές, 850 παρατηρήσεις

4.3 Αποτελέσματα στο επίπεδο των 64 ηλεκτροδίων

Το επίπεδο των ηλεκτροδίων είναι αυτό που παρουσιάζει τις πιο δραστικές διαφοροποιήσεις στην ανάλυση συνδεσιμότητας μετά την προσθήκη όρου υψηλής σταθερότητας. Όπως και για τα προηγούμενα δύο επίπεδα, αρχικά δίνονται οι μέσοι όροι των μετρικών ταξινόμησης για τα δίκτυα που σχη-



Σχήμα 4.5: Κατανομή false positive RCGCI, 68 περιοχές, 850 παρατηρήσεις

ματίζονται από τη μετρική RCGCI στους πίνακες 4.10, 4.11 και 4.12. Καθώς οι συγκρίσεις των δικτύων γίνονται με το δίκτυο που υπολογίστηκε στα καθαρά δεδομένα (και όχι με κάποιο γνωστό πίνακα γειτνίασης) τα αποτελέσματα για τα καθαρά δεδομένα δεν αναφέρονται.

RCGCI	Noisy	ICA	wICA	sICA	NMA
Misclassifications	80.8	81.56	80.3	39.38	41.68
Accuracy	98%	98%	98%	99%	98.9%
Specificity	98.7%	98.6%	98.7%	99.7%	99.6%
Sensitivity	7.7%	7.4%	8.3%	10.6%	14.4%

4.10: Μέσος όρος μετρικών ταξινόμησης, 64 ηλεκτρόδια, 5000 παρατηρήσεις

Ένα κομμάτι της ανάλυσης συνδεσμότητας ήταν ο εντοπισμός των 50 misclassifications που εμφανίστηκαν τις περισσότερες φορές για κάθε μία από τις μεθόδους καθαρισμού. Για τα θορυβώδη δεδομένα, τα 50 πιο συχνά σφάλματα σχεδιάζονται στο σχήμα 4.6. Αναφέρεται ότι 41 από τα 50 αυτά σφάλματα εμφανίστηκαν στην αντίστοιχη λίστα των πιο συχνών σφαλμάτων στα δίκτυα μετά

RCGCI	Noisy	ICA	wICA	sICA	NMA
Misclassifications	0	146.87	135.34	86.77	83.64
Accuracy	100%	96.4%	96.7%	97.8%	97.9%
Specificity	100%	97.9%	98.2%	99.2%	99.1%
Sensitivity	100%	12.3%	16.4%	25.3%	34%

4.11: Μέσος όρος μετρικών ταξινόμησης, 64 ηλεκτρόδια, 1000 παρατηρήσεις

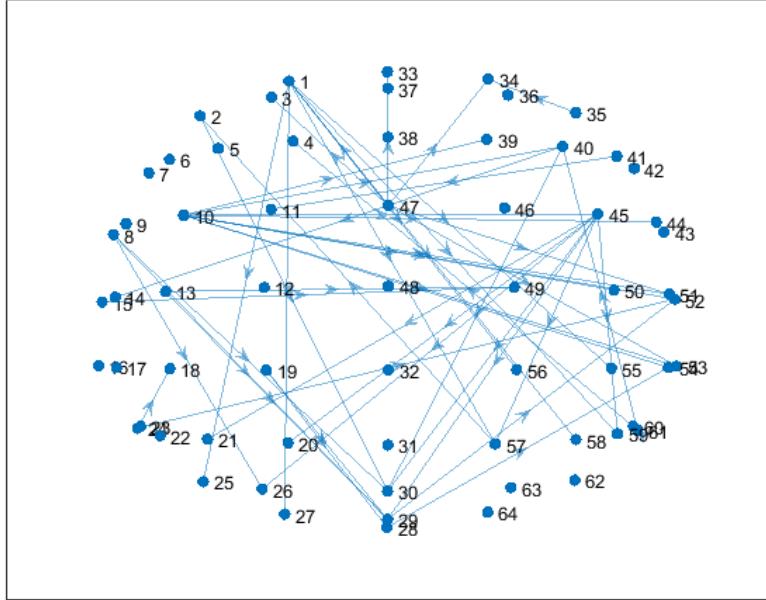
RCGCI	Noisy	ICA	wICA	sICA	NMA
Misclassifications	171.2	162.39	170.55	97.73	99.43
Accuracy	95.8%	96%	95.8%	97.6%	97.5%
Specificity	97.5%	97.7%	97.5%	99%	98.9%
Sensitivity	9.3%	10.4%	10.9%	25.1%	29.6%

4.12: Μέσος όρος μετρικών ταξινόμησης, 64 ηλεκτρόδια, 850 παρατηρήσεις

από καθαρισμό με ICA και wICA. Για τις μεθόδους sICA και NMA αυτός ο αριθμός ήταν 23 και 22 αντίστοιχα. Επίσης, 45 από τα 50 πιο συχνά σφάλματα μετά από καθαρισμό με ICA εμφανίστηκαν στην αντίστοιχη λίστα σφαλμάτων της wICA. Τα παραπάνω υποδεικνύουν ότι ο καθαρισμός δεν βελτιώνει ιδιαίτερα την ανάλυση συνδεσμότητας στο επίπεδο των ηλεκτροδίων, καθώς φαίνεται ότι τα misclassifications συνήθως δεν διορθώνονται.

Όσον αφορά το μέσο τετραγωνικό σφάλμα, υπολογίστηκε ο μέσος όρος των 100 πειραμάτων και στη συνέχεια ο μέσος όρος των ηλεκτροδίων ανά strip. Υπενθυμίζεται ότι τα ηλεκτρόδια χωρίστηκαν σε 5 strips ανάλογα με την απόστασή τους από την πηγή του θορύβου ως εξής

- strip a: FP1, FPz, FP2, AF3, AF4, AF7, AF8, AFz
- strip b: F1, F3, F5, F7, FT7, FC5, FC3, FC1, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, FCz
- strip c: C1, C3, C5, T7, TP7, CP5, CP3, CP1, CPz, Cz, C2, C4, C6, T8, TP8, CP6, CP4, CP2
- strip d: P1, P3, P5, P7, P9, PO7, PO3, POz, Pz, P2, P4, P6, P8, P10, PO8, PO4
- strip e: O1, Iz, Oz, O2



Σχήμα 4.6: 50 πιο συχνά misclassifications, 64 ηλεκτρόδια, 5000 παρατηρήσεις

όπου το strip a είναι το πλησιέστερο στην πηγή του όμορύβου. Στον πίνακα 4.13 δίνεται το μέσο τετραγωνικό σφάλμα που υπολογίστηκε από όλες τις 5000 παρατηρήσεις ανά strip.

mse	Noisy	ICA	wICA	sICA	NMA
strip a	1.0663e+07	4.6652e+06	3.7481e+06	8.1763e+04	8.2560e+04
strip b	1.6408e+06	9.3503e+05	7.4438e+05	8.7505e+04	9.3618e+04
strip c	1.1555e+05	2.4498e+05	1.8112e+05	5.9692e+04	6.7702e+04
strip d	3.5645e+03	2.3145e+05	1.6385e+05	5.4602e+04	8.2525e+04
strip e	2.2604e+04	2.0020e+05	1.4438e+05	2.9274e+05	6.8447e+04

4.13: Μέσο τετραγωνικό σφάλμα, 5000 παρατηρήσεις

Το αντίστοιχο μέσο τετραγωνικό σφάλμα για τα παράθυρα 1000 και 850 παρατηρήσεων δίνονται στους πίνακες 4.14 και 4.15. Μία ενδιαφέρον παρατήρηση είναι ότι για τα τελευταία δύο strips το σφάλμα των καθαρισμένων δεδομένων είναι μεγαλύτερο από το σφάλμα των όμορυβώδων δεδομένων (ανεξάρτητα της μεθόδου καθαρισμού).

mse	Noisy	ICA	wICA	sICA	NMA
strip a	0	1.6289e+06	5.3382e+05	7.3863e+04	7.3785e+04
strip b	0	3.1959e+05	1.0267e+05	2.8616e+04	9.9886e+03
strip c	0	9.2913e+04	2.8112e+04	3.1191e+03	848
strip d	0	9.4828e+04	2.8245e+04	1.2989e+03	461
strip e	0	7.8938e+04	2.3658e+04	1.6537e+05	285

4.14: Μέσο τετραγωνικό σφάλμα, 1000 παρατηρήσεις

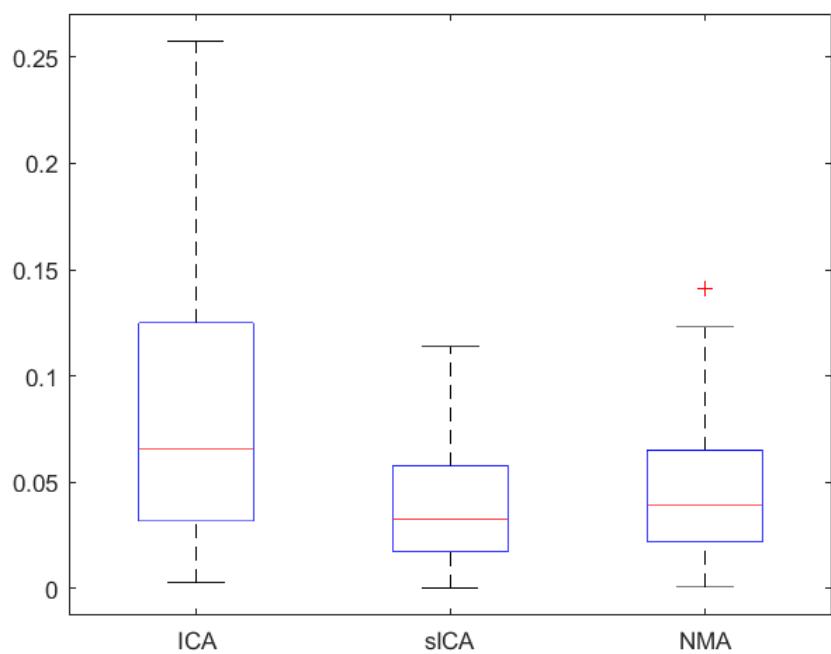
mse	Noisy	ICA	wICA	sICA	NMA
strip a	3.1328e+07	1.0613e+07	1.0052e+07	9.5253e+04	9.7698e+04
strip b	4.8210e+06	2.0821e+06	1.9439e+06	2.0477e+05	2.5482e+05
strip c	3.3952e+05	4.8827e+05	4.2596e+05	1.6918e+05	1.9690e+05
strip d	1.0473e+04	4.2970e+05	3.5954e+05	1.5801e+05	2.4116e+05
strip e	6.6414e+04	3.8224e+05	3.2499e+05	5.2907e+05	2.0026e+05

4.15: Μέσο τετραγωνικό σφάλμα, 850 παρατηρήσεις

	ICA	sICA	NMA
strip a	0.0169	0.0074	0.0089
strip b	0.0948	0.0472	0.0533
strip c	0.0885	0.0425	0.0492
strip d	0.0942	0.0404	0.0521
strip e	0.0830	0.0563	0.0459

4.16: Mean absolute localization error

Τέλος, στον πίνακα 4.16 δίνεται το μέσο απόλυτο localization error ανά strip ενώ στο θηρόγγραμμα 4.7 δίνεται η κατανομή του μέσου απόλυτου localization error στα ηλεκτρόδια για κάθε μία από τις 3 μεθόδους.



Σχήμα 4.7: Κατανομή απόλυτου localization error, 64 ηλεκτρόδια

Κεφάλαιο 5

Πειράματα σε πραγματικά δεδομένα

Σε αυτό το κεφάλαιο θα αναλυθούν τα αποτελέσματα του βασικού πειράματος που περιγράφηκε στις προηγούμενες ενότητες σε πραγματικά δεδομένα. Όπως έχει αναφερθεί, η ερμήνευση αποτελεσμάτων από πραγματικά δεδομένα είναι αρκετά πιο πολύπλοκη καθώς δεν υπάρχει γνώση του ground truth. Παρ' όλα αυτά, τα πειράματα σε πραγματικά δεδομένα ενδεχομένως να βοηθήσουν στην απόρριψη ή επαλήθευση συμπερασμάτων της ανάλυσης σε προσομοιωτικά δεδομένα. Στο πλαίσιο της παρούσας εργασίας πραγματοποιήθηκαν δύο πειράματα σε πραγματικά δεδομένα.

Το πρώτο πείραμα χρησιμοποιεί πραγματικά δεδομένα από ασθενή σε κατάσταση χαλάρωσης με κλειστά τα μάτια που καταγράφηκαν με συχνότητα δειγματοληψίας 1024 Hz η οποία μέσω downsampling μειώνεται σε 256 Hz. Η τοποθέτηση των ηλεκτροδίων είναι ίδια με αυτήν που θεωρήθηκε στα πειράματα σε προσομοιωτικά δεδομένα. Σε αυτά τα δεδομένα προστίθεται θόρυβος που προσομοιώνει το ανοιγοκλείσιμο του ματιού. Στη συνέχεια πραγματοποείται καθαρισμός με τις 4 μεθόδους και γίνονται οι μετρήσεις. Η προσθήκη και ο καθαρισμός του θορύβου γίνονται με τον ίδιο τρόπο όπως και στα πειράματα σε προσομοιωτικά δεδομένα. Όσον αφορά τις μετρήσεις, αυτές γίνονται μόνο στο επίπεδο των 68 εγκεφαλικών περιοχών και των 64 ηλεκτροδίων. Η ανάλυση στο επίπεδο των 7 εγκεφαλικών περιοχών δεν είναι εφικτή καθώς δεν γνωρίζουμε ποιες εγκεφαλικές περιοχές είναι ενεργές ώστε να απομονώσουμε μόνον αυτές. Τα δίκτυα που υπολογίζονται από τις μετρήσεις αιτιότητας συγκρίνονται με τα δίκτυα που υπολογίστηκαν στα καθαρά δεδομένα, όπως έγινε και στα προηγούμενα πειράματα στο επίπεδο των 64 ηλεκτροδίων. Ως καθαρά δεδομένα ορίζονται τα αρχικά πραγματικά δεδομένα EEG. Φυσικά, αυτά τα δεδομένα περιέχουν δραστηριότητα η οποία θα θεωρούνται θόρυβος σε πραγ-

ματικά πειράματα. Ωστόσο στο πλαίσιο αυτού του πειράματος εξετάζεται η ικανότητα των μεθόδων καθαρισμού να απομωνόσουν και να αφαιρέσουν μόνο τη δραστηριότητα που προέρχεται από το ανοιγοκλείσιμο του ματιού.

Το δεύτερο πείραμα χρησιμοποιεί πραγματικά δεδομένα από ασθενή σε κατάσταση χαλάρωσης με ανοιχτά τα μάτια. Η τοποθέτηση των ηλεκτροδίων και η συχνότητα δειγματολογίας είναι ίδια με αυτή του προηγούμενου πειράματος. Σε αυτά τα δεδομένα υπάρχει ήδη θόρυβος από ανοιγοκλείσιμο του ματιού, ο οποίος καθαρίζεται με τις 4 μεθόδους. Στο πλαίσιο αυτού του πειράματος, τα αρχικά καθαρά δεδομένα μας είναι άγνωστα και επομένως είναι αδύνατη η αξιολόγηση της ικανότητας των μεθόδων καθαρισμού να απομωνόσουν και να αφαιρέσουν το θόρυβο.

Αναφέρεται ότι ο αλγόριθμος ανακατασκευής πηγών bsbI κάνει χρήση του leadfield matrix για τον υπολογισμό της δραστηριότητας των πηγών / περιοχών εντός του εγκεφάλου. Φυσικά το leadfield matrix εξαρτάται από την αγωγιμότητα ολόκληρου του όγκου του εγκεφάλου. Υπενθυμίζεται ότι στα πειράματα σε προσομοιωτικά δεδομένα, για να μοντελοποιηθεί η αγωγιμότητα αυτή χρησιμοποιήθηκε το μοντέλο Colin 27. Στα συγκεκριμένα πειράματα χρησιμοποιήθηκε το ίδιο leadfield matrix, ωστόσο αναφέρεται ότι σε πραγματικά πειράματα η επιστημονικά ορθή διαδικασία θα περιλάμβανε την πραγματοποίηση ενός MRI scan για την ακριβή μοντελοποίηση του κεφαλιού και της αγωγιμότητάς του.

5.1 Πραγματικά δεδομένα με προσομοιωτικό θόρυβο

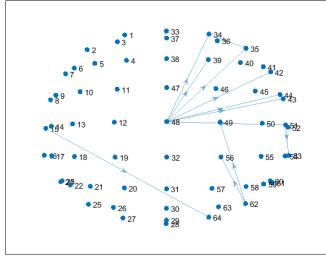
Για αυτό το πείραμα χρησιμοποιήθηκαν πραγματικά δεδομένα EEG 64 ηλεκτροδίων και 38401 παρατηρήσεων [18], το οποίο με συχνότητα δειγματοληψίας 256 Hz αντιστοιχεί σε καταγραφή 150 δευτερολέπτων. Όπως αναφέρθηκε, η διαδικασία προσθήκης θορύβου είναι η ίδια με αυτήν που περιγράφηκε στην ενότητα 2.4, με τη μόνη διαφορά ότι η χρονοσειρά του θορύβου πρέπει πρώτα να επεκταθεί από τις 5000 στις 38401 παρατηρήσεις. Αυτό γίνεται επαναλαμβάνοντας την ίδια χρονοσειρά, όσες φορές χρειαστεί. Όπως και στα προηγούμενα πειράματα, παράγονται συνολικά 6 EEG: αρχικά καθαρά δεδομένα, δεδομένα με θόρυβο και δεδομένα καθαρισμένα με μία από τις 4 μεθόδους καθαρισμού που εξετάζονται. Για κάθε ένα από αυτά υπολογίζεται και η ανακατασκευή των πηγών. Όπως και στα προηγούμενα πειράματα, όλες οι μετρήσεις γίνονται σε τρία παράθυρα παρατηρήσεων. Πρώτον σε ένα παράθυρο 1000 παρατηρήσεων στο οποίο δεν προστίθεται θόρυβος, δεύτερον σε ένα παράθυρο 850 παρατηρήσεων στο οποίο

υπάρχουν τρία spikes από ανοιγοκλείσιμο του ματιού και τέλος στο παράθυρο παρατηρήσεων 5000-10000. Όσον αφορά το τρίτο παράθυρο, η επιλογή να χρησιμοποιήσουμε 5000 απομωνομένες παρατηρήσεις αντί για το σύνολο της καταγραφής έγινε για δύο λόγους. Ο πρώτος είναι η πολυπλοκότητα των αλγορίθμων που κάνουν τις μετρήσεις σε όλες τις 38401 παρατηρήσεις υπερβολικά υπολογιστικά βαριές. Ο δεύτερος είναι ότι, σε αντίθεση με τα προσομοιωτικά δεδομένα που χρησιμοποιήθηκαν στα προηγούμενα κεφάλαια, η δυναμική των μεταβλητών στα πραγματικά δεδομένα EEG αλλάζει στο χρόνο. Αυτό σημαίνει ότι οι αιτιακές σχέσεις που εξετάζουμε ενδεχομένως να μην παραμένουν στάσιμες καθ' όλη τη διάρκεια του πειράματος. Ως εκ τούτου, όσο μεγαλύτερο είναι το παράθυρο παρατηρήσεων στο οποίο κάνουμε τις μετρήσεις, τόσο μεγαλύτερη είναι και η πιθανότητα να αλλάζει η δυναμική του εγκεφάλου κατά την αντίστοιχη περίοδο του πειράματος, καθιστώντας τις μετρήσεις αιτιότητας αναξιόπιστες.

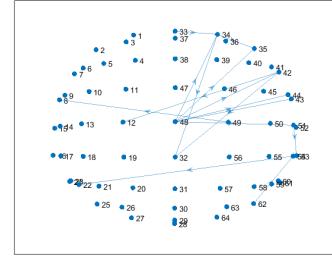
Μία διαφοροποίηση σε σχέση με τα πειράματα σε προσομοιωτικά δεδομένα είναι το κατώφλι που εφαρμόζεται για την κατασκευή των δικτύων από τις μετρήσεις αιτιότητας. Οι μετρήσεις αιτιότητας στα αρχικά καθαρά δεδομένα ηλεκτροδίων επιστρέφουν μόλις 2 αιτιακές συνδέσεις (στο παράθυρο 5000 παρατηρήσεων), υποδεικνύοντας ότι το κατώφλι 0.15 για τη μετρική RCGCI ίσως είναι πολύ υψηλό. Για αυτό το λόγο το κατώφλι μειώνεται σε 0.05 για το επίπεδο των ηλεκτροδίων. Αντιθέτως, στο επίπεδο των 68 περιοχών, οι μετρήσεις με κατώφλι 0.15 εντοπίζουν 107, 132 και 126 συνδέσεις στα 3 παράθυρα μετρήσεων. Αν και αυτά τα αποτελέσματα ενδεχομένως να είναι θεμιτά, στο πλαίσιο του συγκεκριμένου πειράματος θα επικεντρωθούμε μόνο στις πιο ισχυρές και εύκολα εντοπίσιμες συνδέσεις. Για αυτό το λόγο το κατώφλι αυξάνεται σε 0.25. Με αυτές τις παραμέτρους, οι μετρήσεις συνδεσμότητας στα αρχικά καθαρά δεδομένα εντοπίζουν 12, 48 και 47 αιτιακές σχέσεις στα παράθυρα 5000, 1000 και 850 παρατηρήσεων αντίστοιχα στο επίπεδο των ηλεκτροδίων, ενώ στο επίπεδο των πηγών οι αντίστοιχοι αριθμοί συνδέσεων είναι 61, 64 και 65.

Τα δίκτυα που σχηματίζονται από τις μετρήσεις 5000 παρατηρήσεων στα καθαρά και θορυβώδη δεδομένα ηλεκτροδίων δίνονται στα σχήματα 5.1 και 5.2.

Όπως και στα προσομοιωτικά δεδομένα, φαίνεται ότι ο αριθμός των ηλεκτροδίων που συμμετέχουν σε αιτιακή σχέση αυξάνεται με την προσθήκη του θορύβου. Ωστόσο αυτό το φαινόμενο δεν είναι τόσο έντονο. Στα δίκτυα από τις μετρήσεις στις 850 παρατηρήσεις τα οποία δίνονται στα σχήματα 5.3 και 5.4, τα παραπάνω γίνονται πιο προφανή. Συγκεκριμένα, φαίνεται ότι ενώ ο συνολικός αριθμός των συνδέσεων που εντοπίζονται πέφτει από 47 σε 39, αυτές οι συνδέσεις είναι πιο ομοιόμορφα κατανεμημένες σε όλο τον όγκο του κεφαλιού. Στα σχήματα 5.5 και 5.6 φαίνονται τα αντίστοιχα δίκτυα συνδεσμότητας από τις μετρήσεις στις 850 παρατηρήσεις στο επίπεδο των εγκεφαλικών περιοχών.

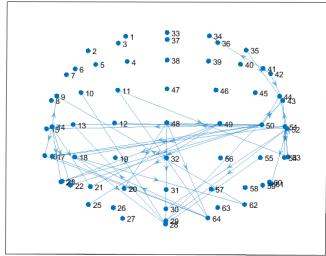


Σχήμα 5.1: Δίκτυο RCGCI και φάρα δεδομένα 64 ηλεκτροδίων, 5000 παρατηρήσεις

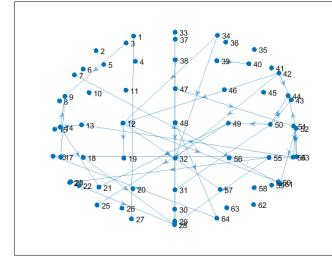


Σχήμα 5.2: Δίκτυο RCGCI θορυβώδη δεδομένα 64 ηλεκτροδίων, 5000 παρατηρήσεις

Για να διευκολυνθεί η ανάγνωση, έχουν αφαιρεθεί τα labels των εγκεφαλικών περιοχών. Αναφέρεται ότι μετά την προσθήκη θορύβου, ο συνολικός αριθμός συνδέσεων που εντοπίζονται αυξάνεται από 65 σε 73 και ο αριθμός των κόμβων που συμμετέχουν σε κάποια σύνδεση αυξάνεται από 31 σε 35.

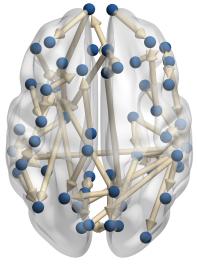


Σχήμα 5.3: Δίκτυο RCGCI και φάρα δεδομένα 64 ηλεκτροδίων, 850 παρατηρήσεις

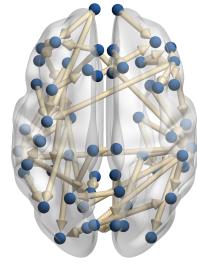


Σχήμα 5.4: Δίκτυο RCGCI θορυβώδη δεδομένα 64 ηλεκτροδίων, 850 παρατηρήσεις

Όσον αφορά τον καθαρισμό, αυτός τελικά εκτελέστηκε με 3 μεθόδους: ICA, sICA και NMA, καθώς η μέθοδος wICA απέτυχε να διορθώσει το artifactual IC. Κατά τον καθαρισμό με ICA, το artifactual IC αντικαταστήθηκε με το μέσο όρο του αντί να μηδενιστεί. Αυτή η επιλογή έγινε διότι το artifactual IC ήταν μη μηδενικό αλλά σταθερό ακόμα και σε περιόδους κατά τις οποίες δεν είχε προστεθεί θόρυβος. Ως εκ τούτου, ο μηδενισμός του θα οδηγούσε και στην αφαίρεση ενός σταθερού όρου (ο οποίος ήταν αρκετά μεγαλύτερος στα μπροστινά ηλεκτρόδια από ότι στα οπίσθια). Καθώς πρόκειται για σταθερό όρο αυτή η αφαίρεση δεν θα επηρέαζε την ανάλυση συνδεσμότητας, ωστόσο μπορούμε να πετύχουμε πολύ χαμηλότερο μέσο τετραγωνικό σφάλμα εάν την



Σχήμα 5.5: Δίκτυο RCGCI καθαρά δεδομένα 68 περιοχών, 850 παρατηρήσεις



Σχήμα 5.6: Δίκτυο RCGCI θορυβώδη δεδομένα 68 περιοχών, 850 παρατηρήσεις

αποφύγουμε. Στη συνέχεια δίνονται τα confusion matrices που προκύπτουν και οι μετρικές ταξινόμησης για όλες τις μεθόδους καθαρισμού που χρησιμοποιήθηκαν, πρώτα για τα δίκτυα στο επίπεδο των ηλεκτροδίων και μετά για τα δίκτυα στο επίπεδο των εγκεφαλικών περιοχών.

ICA	p	n	sICA	p	n	NMA	p	n
p	5	2	p	6	34	p	6	4
n	7	4082	n	6	4050	n	6	4080

5.1: Confusion matrices rCGCI, 64 ηλεκτρόδια, 5000 παρατηρήσεις

ICA	p	n	sICA	p	n	NMA	p	n
p	21	24	p	14	45	p	19	26
n	27	4024	n	34	4003	n	29	4022

5.2: Confusion matrices rCGCI, 64 ηλεκτρόδια, 1000 παρατηρήσεις

ICA	p	n	sICA	p	n	NMA	p	n
p	16	32	p	19	54	p	27	30
n	31	4017	n	28	3995	n	20	4019

5.3: Confusion matrices rCGCI, 64 ηλεκτρόδια, 850 παρατηρήσεις

Η ανάλυση συνδεσιμότητας φαίνεται να επαληθεύει αρκετά από τα συμπεράσματα των πειραμάτων σε προσομοιωτικά δεδομένα. Αρχικά φαίνεται ότι η συνδεσιμότητα επηρεάζεται πολύ περισσότερο στο επίπεδο των ηλεκτροδίων από ότι στο επίπεδο των εγκεφαλικών περιοχών. Επιπλέον φαίνεται ότι όλοι οι

RCGCI	Noisy	ICA	sICA	NMA
Misclassifications	11	9	40	10
Accuracy	99.7%	99.7%	99%	99.7%
Specificity	99.8%	99.9%	99.1%	99.9%
Sensitivity	66.6%	41.6%	50%	50%

5.4: Μετρικές ταξινόμησης, 64 ηλεκτρόδια, 5000 παρατηρήσεις

RCGCI	Noisy	ICA	sICA	NMA
Misclassifications	0	51	79	55
Accuracy	100%	98.7%	98%	98.6%
Specificity	100%	99.4%	98.8%	99.3%
Sensitivity	100%	43.7%	29.1%	39.5%

5.5: Μετρικές ταξινόμησης, 64 ηλεκτρόδια, 1000 παρατηρήσεις

RCGCI	Noisy	ICA	sICA	NMA
Misclassifications	54	63	82	50
Accuracy	98.6%	98.4%	98%	98.7%
Specificity	99.4%	99.2%	98.6%	99.2%
Sensitivity	34%	34%	40.4%	57.4%

5.6: Μετρικές ταξινόμησης, 64 ηλεκτρόδια, 850 παρατηρήσεις

ICA	p	n	sICA	p	n	NMA	p	n
p	55	5	p	49	9	p	56	3
n	6	4558	n	12	4554	n	5	4560

5.7: Confusion matrices rCGCI, 68 περιοχές, 5000 παρατηρήσεις

ICA	p	n	sICA	p	n	NMA	p	n
p	55	4	p	55	17	p	55	6
n	9	4556	n	9	4543	n	9	4554

5.8: Confusion matrices rCGCI, 68 περιοχές, 1000 παρατηρήσεις

μέθοδοι καθαρισμού προκαλούν σημαντικές διαφοροποιήσεις και σε περιόδους στις οποίες δεν έχει προστεθεί θόρυβος.

Μία διαφοροποίηση σε σχέση με τα πειράματα σε προσομοιωτικά δεδομένα είναι ότι η μέθοδος sICA δεν φαίνεται να καθαρίζει αποτελεσματικά τα δεδομένα.

ICA	p	n	sICA	p	n	NMA	p	n
p	57	15	p	49	16	p	53	12
n	8	4544	n	16	4543	n	12	4547

5.9: Confusion matrices rCGCI, 68 περιοχές, 850 παρατηρήσεις

RCGCI	Noisy	ICA	sICA	NMA
Misclassifications	13	11	21	8
Accuracy	99.7%	99.7%	99.5%	99.8%
Specificity	99.8%	99.9%	99.8%	99.9%
Sensitivity	88.5%	90.1%	80.3%	91.8%

5.10: Μετρικές ταξινόμησης, 68 περιοχές, 5000 παρατηρήσεις

RCGCI	Noisy	ICA	sICA	NMA
Misclassifications	0	13	26	15
Accuracy	100%	99.7%	99.4%	99.6%
Specificity	100%	99.9%	99.6%	99.8%
Sensitivity	100%	85.9%	85.9%	85.9%

5.11: Μετρικές ταξινόμησης, 68 περιοχές, 1000 παρατηρήσεις

RCGCI	Noisy	ICA	sICA	NMA
Misclassifications	32	23	32	24
Accuracy	99.3%	99.5%	99.3%	99.5%
Specificity	99.5%	99.6%	99.6%	99.7%
Sensitivity	81.5%	87.7%	75.4%	81.5%

5.12: Μετρικές ταξινόμησης, 68 περιοχές, 850 παρατηρήσεις

Ακόμα και σε περιπτώσεις που πετυχαίνει υψηλότερο sensitivity από τις υπόλοιπες μεθόδους, ο αριθμός των misclassifications είναι πολύ μεγάλος για να δικαιολογείται η χρήση της.

Από το μέσο τετραγωνικό σφάλμα που δίνεται ανά strip στους πίνακες 5.13, 5.14 και 5.15, φαίνεται ότι η sICA έχει απόδοση συγχρίσιμη με τις υπόλοιπες μεθόδους για τα μπροστινά strips αλλά αρκετά χειρότερη για τα οπίσθια strips. Από αυτό μπορούμε να συμπεράνουμε ότι ενδεχομένως να είχε νόημα η χρήση διαφορετικών μεθόδων καθαρισμού για διαφορετικά σετ ηλεκτροδίων.

Τέλος, στο θηκόγραμμα 5.7 φαίνεται η κατανομή του απόλυτου localization

mse	Noisy	ICA	sICA	NMA
strip a	4.1729e+09	5.5760e+08	8.1620e+07	7.3721e+07
strip b	6.4124e+08	8.5875e+07	9.4720e+07	1.3005e+07
strip c	4.5178e+07	7.0683e+06	3.1121e+06	1.1418e+06
strip d	1.3940e+06	4.3648e+05	3.5480e+05	3.1550e+04
strip e	8.8266e+06	1.2872e+06	4.4835e+06	4.6193e+05

5.13: Μέσο τετραγωνικό σφάλμα, 5000 παρατηρήσεις

mse	Noisy	ICA	sICA	NMA
strip a	0	2.8495e+08	6.3055e+07	5.5868e+07
strip b	0	4.4096e+07	1.5833e+08	7.8133e+06
strip c	0	2.9221e+06	8.1936e+05	5.7869e+05
strip d	0	8.7492e+04	1.7358e+05	2.1684e+04
strip e	0	6.9187e+05	4.9155e+06	1.9509e+05

5.14: Μέσο τετραγωνικό σφάλμα, 1000 παρατηρήσεις

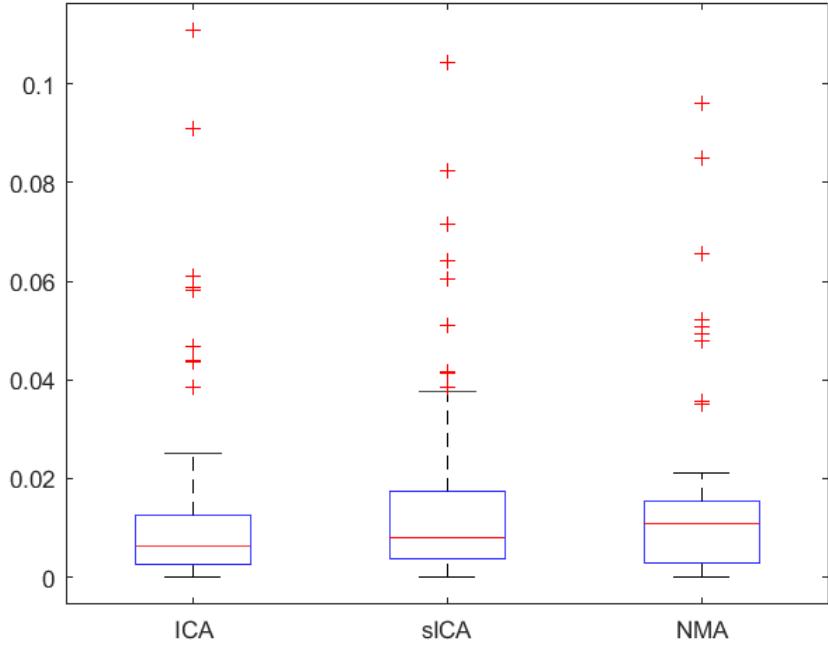
mse	Noisy	ICA	sICA	NMA
strip a	1.2551e+10	1.3771e+09	6.4671e+07	2.4236e+08
strip b	1.9293e+09	2.1094e+08	5.9441e+07	3.5982e+07
strip c	1.3592e+08	1.8417e+07	6.2234e+06	3.1607e+06
strip d	4.1942e+06	1.1867e+06	6.3472e+05	1.1314e+05
strip e	2.6556e+07	2.8647e+06	5.8999e+06	1.5944e+06

5.15: Μέσο τετραγωνικό σφάλμα, 850 παρατηρήσεις

error για τις 3 μεθόδους καθαρισμού. Από αυτό φαίνεται ότι η μέθοδος NMA έχει υψηλότερο σφάλμα κατά μέσο όρο από την ICA αλλά μικρότερη διασπορά σφάλματος.

5.2 Πραγματικά δεδομένα

Το τελευταίο πείραμα στο πλαίσιο της παρούσας μελέτης συνίσταται στην εφαρμογή των τριών μεθόδων καθαρισμού (ICA, sICA, NMA) σε πραγματικά δεδομένα από ασθενή σε κατάσταση χαλάρωσης με ανοιχτά τα μάτια [18] για τον καθαρισμό του artifact από το ανοιγοκλείσιμο των ματιών. Καθώς σε αυτό το πείραμα η κατάσταση των καθαρών δεδομένων μας είναι εντελώς άγνωστη,



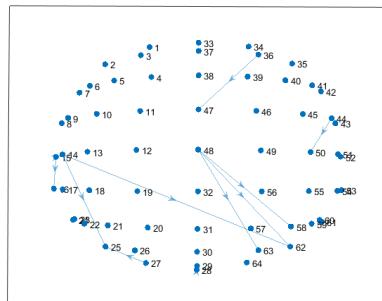
Σχήμα 5.7: Κατανομή απόλυτου localization error, 64 ηλεκτρόδια

δεν είναι δυνατή η σύγκριση και ερμηνεία των αποτελεσμάτων. Επιπλέον, όπως και στο προηγούμενο πείραμα, μας λείπει ένα ακριβές μοντέλο του κεφαλιού του ασθενή. Αντί αυτού χρησιμοποιείται το μοντέλο Colin 27 και ως εκ τούτου ο αλγόριθμος ανακατασκευής πηγών δεν μπορεί να θεωρηθεί αξιόπιστος. Το νόημα του συγκεκριμένου πειράματος είναι απλά να εξεταστεί κατά πόσο οι τρεις μέθοδοι καθαρισμού λειτουργούν με τον ίδιο τρόπο όπως και στα προηγούμενα πειράματα.

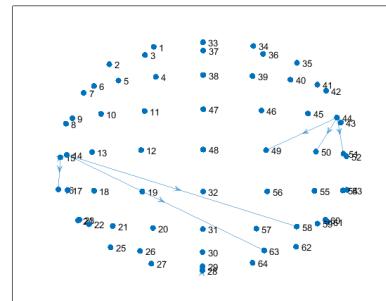
Συνολικά εξετάζονται 4 EEG: αρχικά ύφορυβώδη δεδομένα και δεδομένα καθαρισμένα με μία από τις τρεις μεθόδους. Οι μετρήσεις γίνονται στο επίπεδο των 64 ηλεκτροδίων και στο επίπεδο των 68 εγκεφαλικών περιοχών σε ένα παράθυρο 5000 παρατηρήσεων. Στα δύο αυτά επίπεδα μετριέται η μετρική rCGCI και για τον μετασχητισμό των μετρήσεων σε δίκτυα εφαρμόστηκαν τα κατώφλια 0.25 για το επίπεδο των περιοχών και 0.05 για το επίπεδο των ηλεκτροδίων. Με τη μέθοδο καθαρισμού sICA καθαρίζονται μόνο τα δύο μπροστινά σετ ηλεκτροδίων.

Στα σχήματα 5.8 με 5.11 φαίνονται τα δίκτυα που υπολογίζονται στο επίπεδο

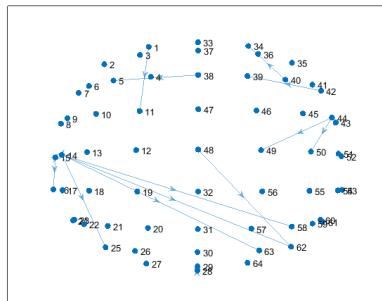
των ηλεκτροδίων. Στις συγκεκριμένες μετρήσεις, φαίνεται ότι η μέθοδος καθαρισμού που προκαλεί τη μεγαλύτερη διαφοροποίηση στη συνδεσιμότητα είναι η sICA ενώ η μέθοδος που προκαλεί τη μικρότερη είναι η NMA. Όσον αφορά τη συνδεσιμότητα στο επίπεδο των εγκεφαλικών περιοχών, στα σχήματα 5.13 με 5.15 δίνονται οι διαφοροποιήσεις των δεδομένων καθαρισμένα από κάποια μέθοδο σε σχέση με το δίκτυο των θορυβώδων δεδομένων. Ο λόγος που δίνονται οι διαφοροποιήσεις αντί για τα ίδια τα δίκτυα είναι ότι ο αριθμός των συνδέσεων είναι σχετικά μεγάλος, και η απεικόνιση των δικτύων δεν θα ήταν εύκολο να ερμηνευθεί. Για αναφορά, το δίκτυο συνδεσιμότητας εγκεφαλικών περιοχών στα θορυβώδη δεδομένα δίνεται στο σχήμα 5.12.



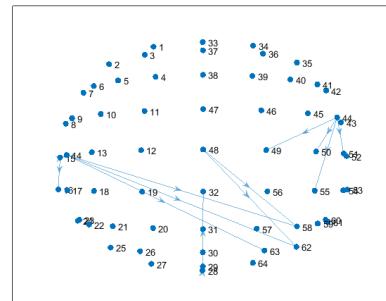
Σχήμα 5.8: Δίκτυο rCGCI, επίπεδο ηλεκτροδίων: θορυβώδη δεδομένα



Σχήμα 5.9: Δίκτυο rCGCI, επίπεδο ηλεκτροδίων: καθαρισμός με ICA

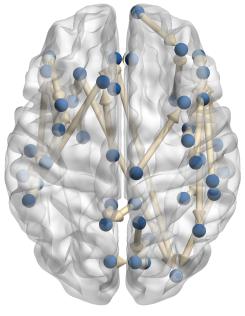


Σχήμα 5.10: Δίκτυο rCGCI, επίπεδο ηλεκτροδίων: καθαρισμός με sICA

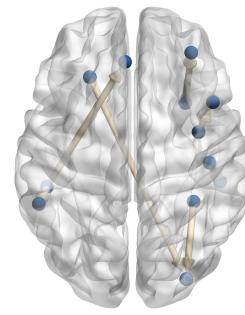


Σχήμα 5.11: Δίκτυο rCGCI, επίπεδο ηλεκτροδίων: καθαρισμός με NMA

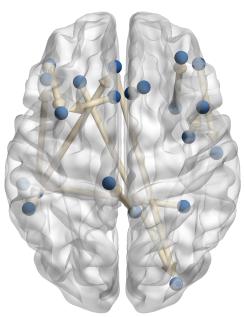
Οι τρεις μέθοδοι προκαλούν 7, 15 και 6 διαφοροποιήσεις στη συνδεσιμότητα εγκεφαλικών περιοχών σε σχέση με τα θορυβώδη δεδομένα. Επομένως φαίνεται



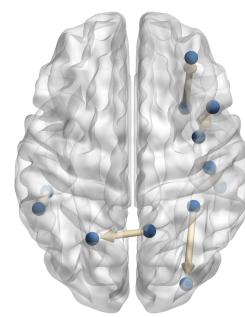
Σχήμα 5.12: Δίκτυο rCGCI, επίπεδο περιοχών: θορυβώδη δεδομένα



Σχήμα 5.13: Διαφοροποιήσεις δικτύου rCGCI, επίπεδο περιοχών: καθαρισμός με ICA



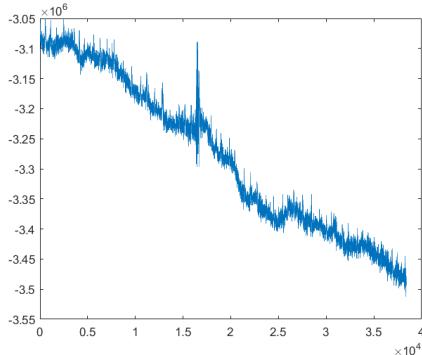
Σχήμα 5.14: Διαφοροποιήσεις δικτύου rCGCI, επίπεδο περιοχών: καθαρισμός με sICA



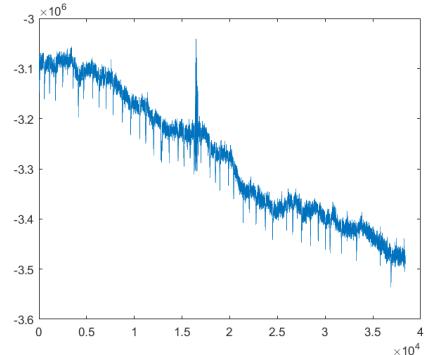
Σχήμα 5.15: Διαφοροποιήσεις δικτύου rCGCI, επίπεδο περιοχών: καθαρισμός με NMA

και σε αυτό το επίπεδο ότι η μέθοδος sICA επηρεάζει τη συνδεσιμότητα περισσότερο από τις υπόλοιπες. Μεταξύ των μεθόδων ICA και NMA υπάρχουν 4 κοινές διαφοροποιήσεις.

Συγκεκριμένα για τη μέθοδο sICA γίνεται αρκετά προφανές ότι αποτυγχάνει να καθαρίσει τα δεδομένα. Αυτό φαίνεται καλύτερα από τις χρονοσειρές ορισμένων καθαρισμένων ηλεκτροδίων. Για παράδειγμα στα σχήματα 5.16 και 5.17 δίνονται οι χρονοσειρές του ηλεκτροδίου FC1 (το οποίο ανήκει σε σετ που καθαρίστηκε από την sICA) πριν και μετά από καθαρισμό με sICA. Αναφέρεται ότι το συγκεκριμένο ηλεκτρόδιο έχει τάση (drift) η οποία σε κανονικά πειράματα θα αφαιρούνταν κατά το pre-processing. Ωστόσο η παρουσία αυτής της τάσης δεν επηρεάζει σημαντικά τη λειτουργία της sICA. Αναφέρεται επίσης ότι αυτού του είδους τα προβλήματα εμφανίζονται πιο έντονα στα ηλεκτρόδια του δεύτερου σετ που καθαρίστηκε, δηλαδή με τη μεγαλύτερη απόσταση από την πηγή του θορύβου. Όσον αφορά τις άλλες δύο μεθόδους, είναι δυστυχώς αδύνατο να κρίνουμε ποια από τις δύο λειτουργεί καλύτερα.



Σχήμα 5.16: Ηλεκτρόδιο FC1, θορύβωδη δεδομένα



Σχήμα 5.17: Ηλεκτρόδιο FC1, καθαρισμός με sICA

Κεφάλαιο 6

Συμπεράσματα και μελλοντικές επεκτάσεις

6.1 Περίληψη συμπερασμάτων

Στην παρούσα εργασία εξετάστηκε η ικανότητα της ICA να καθαρίσει το θόρυβο που προκαλείται από ανοιγοκλείσιμο του ματιού σε δεδομένα EEG. Επιπλέον, η ICA συγκρίθηκε με 3 άλλες μεθόδους καθαρισμού: wICA, sICA και NMA.

Ένα από τα πιο ξεκάθαρα συμπεράσματα είναι ότι όσον αφορά τη μέθοδο sICA, αν και τα αποτελέσματα στα προσομοιωτικά δεδομένα ήταν ενθαρρυντικά, η ανάλυση σε πραγματικά δεδομένα δείχνει ότι υστερεί σε σχέση με τις άλλες μεθόδους. Παρ' όλα αυτά, ίσως να είχε νόημα σε περιπτώσεις που χρησιμοποιούνται διαφορετικές μέθοδοι καθαρισμού για τον καθαρισμό διαφορετικών ηλεκτροδίων.

Ένα επιπλέον συμπέρασμα είναι ότι καμία από τις μεθόδους καθαρισμού δεν πέτυχε ικανοποιητικά αποτελέσματα όταν η ανάλυση συνδεσμούτητας έγινε στο επίπεδο των ηλεκτροδίων, τουλάχιστον στην περίπτωση που η απόδοση κρίνεται από το sensitivity. Αυτό εν μέρη οφείλεται στο ότι ο θόρυβος επηρεάζει σε πάρα πολύ μεγάλο βαθμό τη συνδεσμούτητα (υπενθυμίζεται ότι το sensitivity της ανάλυσης στα θορυβώδη δεδομένα έπεφτε στο 7.7% και 9.3% στα προσομοιωτικά δεδομένα στα παράθυρα 5000 και 850 παρατηρήσεων, ενώ στο πρώτο πείραμα σε πραγματικά δεδομένα το sensitivity έπεφτε στο 66.6% και 34% για τα ίδια παράθυρα μετρήσεων). Μάλιστα, φαίνεται ότι σε πολλές περιπτώσεις ο καθαρισμός κάνει την ανάλυση συνδεσμούτητας ακόμα πιο ανακριβή (αν και αυτό ισχύει κυρίως για το sensitivity).

Για το επίπεδο των εγκεφαλικών περιοχών, αναφέρεται ότι όταν απομο-

νώνονταν οι ενεργές εγκεφαλικές περιοχές τα αποτελέσματα της ανάλυσης ήταν πάντα ικανοποιητικά και η ανάλυση σε καθαρισμένα δεδομένα ήταν πάντα πιο ακριβής από την ανάλυση στα θορυβώδη δεδομένα (με εξαίρεση το παράθυρο παρατηρήσεων στις οποίες δεν έχει προστεθεί θόρυβος), τουλάχιστον όταν χρησιμοποιούνταν η μετρική CGCI. Ωστόσο, σε πραγματικά δεδομένα αυτού του τύπου η ανάλυση είναι αδύνατη καθώς δεν μπορούμε να γνωρίζουμε ακριβώς ποιες εγκεφαλικές περιοχές πρέπει να συμπεριληφθούν.

Για το επίπεδο των 68 εγκεφαλικών περιοχών η ερμηνεία των αποτελεσμάτων είναι λίγο δυσκολότερη. Ο καθαρισμός με οποιαδήποτε από τις 4 μεθόδους φαίνεται να βελτιώνει το συνολικό accuracy, ωστόσο φαίνεται επίσης ότι ο καθαρισμός με ICA μερικές φορές οδηγεί σε χαμηλότερο sensitivity από την ανάλυση σε θορυβώδη δεδομένα, τουλάχιστον στα προσομοιωτικά δεδομένα.

Ένα αξιοσημείωτο αποτέλεσμα είναι ότι η ανάλυση στο παράθυρο 5000 παρατηρήσεων οδηγεί σε χαμηλότερο sensitivity αλλά υψηλότερο accuracy για όλες τις μεθόδους καθαρισμού εκτός από την ICA. Αυτό συμβαίνει παρά το ότι οι αιτιακές σχέσεις των δεδομένων έμεναν στάσιμες στο χρόνο στα προσομοιωτικά δεδομένα. Παραδόξως, αυτό το φαινόμενο δεν παρατηρήθηκε στο πρώτο πείραμα σε πραγματικά δεδομένα, παρότι σε αυτήν την περίπτωση δεν έχουμε λόγο να πιστεύουμε ότι οι αιτιακές σχέσεις μένουν στάσιμες στο χρόνο.

Όσον αφορά το μέσο τετραγωνικό σφάλμα το οποίο είναι και η πιο εύκολα ερμηνεύσιμη μετρική, τα αποτελέσματα δείχνουν ότι η ICA αδυνατεί να καθαρίσει όλο το θόρυβο. Από τις 4 μεθόδους η πιο αποτελεσματική φαίνεται να είναι η NMA. Υπενθυμίζεται επίσης ότι συγκεκριμένα για το μπροστινό σετ ηλεκτροδίων, τα αποτελέσματα της sICA ήταν συγκρίσιμα με αυτά της NMA.

Τέλος, αναφέρεται ότι το ποσοστό του θορύβου που παραμένει μετά τον καθαρισμό είναι αρκετά διαφορετικό από ηλεκτρόδιο σε ηλεκτρόδιο, όπως φαίνεται από την κατανομή του localization error.

6.2 Μελλοντικές επεκτάσεις

Μελλοντικές θα μπορούσαν να πειραματιστούν με το συνδυασμό των μεθόδων NMA και sICA με τη μέθοδο wICA. Επιπλέον, θα είχε νόημα να δοκιμαστεί ο καθαρισμός των δεδομένων πηγών μετά την ανακατασκευή πηγών με κάποιο αλγόριθμο όπως ο bsbl αντί για τον καθαρισμό των δεδομένων ηλεκτροδίων. Τέλος, ένας από τους απλούστερους τρόπους βελτίωσης του καθαρισμού φαίνεται να είναι η εφαρμογή κάποιου είδους windowing που θα απέτρεπε τις διαφοροποιήσεις σε περιόδους κατά τις οποίες δεν υπάρχει ενεργοποίηση της πηγής θορύβου.

Κεφάλαιο 7

Βιβλιογραφία

- [1] Castellanos, Nazareth Makarov, Valeri, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis" Journal of neuroscience methods, vol. 158, no. 2, pp. 300-312, (2007), 10.1016/j.jneumeth.2006.05.033.
- [2] Zohreh Zakeri, "Optimised use of independent component analysis for EEG signal processing" (2017),
<https://etheses.bham.ac.uk/id/eprint/7430/>
- [3] Alam, Md Erfanul, "Electroencephalographic Signal Processing and Classification Techniques for Noninvasive Motor Imagery Based Brain Computer Interface" (2017), Electronic Theses and Dissertations. 1591.
<https://digitalcommons.georgiasouthern.edu/etd/1591>
- [4] Dustin Stansbury "The Statistical Whitening Transform"
theclevermachine (2015)
<https://theclevermachine.wordpress.com/2013/03/30/the-statistical-whitening-transform/>
- [5] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis" IEEE Transactions on Neural Networks, vol. 10, no. 3, pp. 626-634, (1999), doi: 10.1109/72.761722.
- [6] Χανδακας Ευαγγελος. "Εκτίμηση αιτιότητας κατά Granger με χρήση ανάλυσης ανεξαρτήτων μεταβλητών (ICA)", (2016), doi : 10.26262/heal.auth.ir.286832
- [7] Wikipedia contributors, "FastICA," Wikipedia, The Free Encyclopedia,
<https://en.wikipedia.org/w/index.php?title=FastICA&oldid=1068070537>.
- [8] Eichler, Michael "Comment on: Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance", (2012), doi :
<https://doi.org/10.48550/arXiv.1210.7125>
- [9] Katerina Hlaváčková-Schindler, Milan Paluš, Martin Vejmelka, Joydeep

- Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis", Physics Reports, vol. 441, Issue 1, Pages 1-46, ISSN 0370-1573, (2007),
<https://doi.org/10.1016/j.physrep.2006.12.004>.
- [10] Anil Seth Granger causality. Scholarpedia, 2(7):1667, (2007), doi::10.4249/scholarpedia.1667
http://www.scholarpedia.org/article/Granger_causality
- [11] E. Siggiridou and D. Kugiumtzis, "Granger Causality in Multivariate Time Series Using a Time-Ordered Restricted Vector Autoregressive Model," IEEE Transactions on Signal Processing, vol. 64, no. 7, pp. 1759-1773, (2016), doi: 10.1109/TSP.2015.2500893.
- [12] Baccala, Luiz Sameshima, Koichi. "Partial directed coherence: A new concept in neural structure determination", Biological Cybernetics, vol. 84, pp. 463-474. (2001) 10.1007/PL00007990.
- [13] Siggiridou, Elsa, Christos Koutlis, Alkiviadis Tsimpiris, and Dimitris Kugiumtzis. "Evaluation of Granger Causality Measures for Constructing Networks from Multivariate Time Series" Entropy, vol. 21, no. 11: 1080, (2019). <https://doi.org/10.3390/e21111080>
- [14] Delorme A Makeig S "EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics", Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9-21, (2004), doi: 10.1016/j.jneumeth.2003.10.009. PMID: 15102499.
- [15] Alejandro Ojeda, Marius Klug, Kenneth Kreutz-Delgado, Klaus Gramann, Jyoti Mishra bioRxiv 559450; (2019) doi:
<https://doi.org/10.1101/559450>
- [16] Xia M, Wang J, He Y "BrainNet Viewer: A Network Visualization Tool for Human Brain", Connectomics, vol. 8 (2013) PLoS ONE 8: e68910
- [17] A. Omidvarnia, M. Mesbah, J. M. O'Toole et al., "Analysis of the time-varying cortical neural connectivity in the newborn EEG: A time-frequency approach, in Systems, Signal Processing and their Applications (WoSSPA)", 2011 7th International Workshop on Systems, Signal Processing and their Applications, WoSSPA 2011, pp. 179-182, (2011) doi: 10.1109/WOSSPA.2011.5931445.
- [18] M. Torkamani-Azar, S. D. Kanik, S. Aydin and M. Cetin, "Prediction of Reaction Time and Vigilance Variability From Spatio-Spectral Features of Resting-State EEG in a Long Sustained Attention Task," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 9, pp. 2550-2558, Sept. 2020, doi: 10.1109/JBHI.2020.2980056.
<https://ieeexplore.ieee.org/document/9034192>