



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA BIOMEDICÍNSKÉHO INŽENÝRSTVÍ
Katedra biomedicínské techniky

**Automatická extrakce nezávislých komponent
ze spánkových EEG dat**

**Automatic extraction of independent components
from sleep EEG dataset**

Diplomová práce

Studijní program: Biomedicínská a klinická technika

Studijní obor: Biomedicínský inženýr

Vedoucí práce: Ing. Marek Piorecký

Bc. Eva Černá

Kladno 2020

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Černá** Jméno: **Eva** Osobní číslo: **456642**
Fakulta: **Fakulta biomedicínského inženýrství**
Garantující katedra: **Katedra biomedicínské techniky**
Studijní program: **Biomedicínská a klinická technika**
Studijní obor: **Biomedicínský inženýr**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Automatická extrakce nezávislých komponent ze spánkových EEG dat

Název diplomové práce anglicky:

Automatic extraction of independent components from sleep EEG dataset

Pokyny pro vypracování:

Ze spánkových EEG záznamů extrahujte nezávislé komponenty pomocí 3 vybraných variant analýzy nezávislých zdrojů (ICA), standardně využívaných ke zpracování biologických signálů. Extrahované komponenty promítněte do 2D prostoru pomocí nelineární metody redukce dimenze. Využijte automatický klasifikační algoritmus pro nalezení struktury v extrahovaných datech. Porovnejte charakter komponent pro rozdílné metody potlačení artefaktů.

Seznam doporučené literatury:

- [1] Maximilien Chaumona, Dorothy V.M. Bishopc, Niko A. Busch, A practical guide to the selection of independent components of the electroencephalogram for artifact correction, Journal of Neuroscience Methods, ročník 250, číslo 1, 2015
- [2] Vlastimil Koudelka, Jan Štrobl, Marek Piorecký, Martin Brunovský, Vladimír Krajča, Nonlinear Dimensionality Reduction and Feature Analysis for Artifact Component Identification in hdEEG Datasets, World Congress on Medical Physics and Biomedical Engineering 2018 , ročník 68, číslo 2, 2018
- [3] J. Birjandtalab, M. Baran Pouyan, M. Nourani, Nonlinear dimension reduction for EEG-based epileptic seizure detection, Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference on, ročník 21, číslo 15937139, 2016, 10.1109/BHI.2016.7455968

Jméno a příjmení vedoucí(ho) diplomové práce:


Ing. Marek Piorecký

Jméno a příjmení konzultanta(ky) diplomové práce:

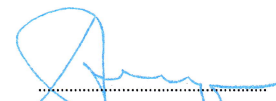
Ing. Jan Štrobl, Koudelka, Vlastimil, Ing., Ph.D.

Datum zadání diplomové práce: **10.02.2020**

Platnost zadání diplomové práce: **19.09.2021**



prof. Ing. Peter Kneppo, DrSc., dr.h.c.
podpis vedoucí(ho) katedry



prof. MUDr. Ivan Dylevský, DrSc.
podpis děkana(ky)

Prohlášení

Prohlašuji, že jsem diplomovou práci s názvem „Automatická extrakce nezávislých komponent ze spánkových EEG dat“ vypracovala samostatně a použila k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k diplomové práci.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů.

V Kladně dne

.....

Bc. Eva Černá

Poděkování

Děkuji mému vedoucímu, panu Ing. Marku Pioreckému, za velkou podporu a příkladné vedení mé diplomové práce. Také děkuji celému "BRAIN Teamu" za pomoc a věcné připomínky. Děkuji své partnerce, své rodině a svým pracovním kolegyním za podporu po celou dobu studia.

ABSTRAKT

Automatická extrakce nezávislých komponent ze spánkových EEG dat:

Analýza nezávislých komponent (ICA) je jedním z významných nástrojů používaných k analýze elektroencefalografického signálu (EEG) a k potlačování artefaktů. Existuje několik variant jak tuto metodu aplikovat, přičemž se v současnosti všechny varianty považují za rovnocenné. Ve své práci jsem na reálných EEG záznamech porovnála tři nejpoužívanější varianty ICA – Infomax, FastICA a SOBI. Cílem práce bylo ověřit podobnost nezávislých komponent (IC) napříč těmito metodami a zhodnotit tak efektivitu pro potlačení různých artefaktů a vliv případného řetězení různých metod za sebou. Pro další porovnání jsem využila 19dimenzionální prostor topografických map, který jsem redukovala nelineární metodou t-SNE a klasifikovala shlukovacím algoritmem DBSCAN. Byly porovnávány IC obsahující artefakty napříč prostorem, časem a frekvenční oblastí. Z výsledků práce vyplývá, že testované varianty ICA se nedají považovat za rovnocenné. Získané nezávislé komponenty se napříč metodami liší.

Klíčová slova

DBSCAN, EEG, FieldTrip, ICA, MATLAB, redukce dimenze, t-SNE

ABSTRACT

Automatic extraction of independent components from sleep EEG dataset:

Independent component analysis (ICA) is one of the important tools used for electroencephalographic signal (EEG) analysis and artifact suppression. There are several variants of how to apply this method, and currently all variants are considered equivalent. In this master thesis three most used ICA variants on real EEG records - Infomax, FastICA and SOBI were examined. The aim of the work was to verify the similarity of independent components (IC) across these methods and thus evaluate the effectiveness for the suppression of various artifacts and the effect of possible chaining of various methods in a row. For further comparison, a 19-dimensional space of topographic maps was used, which I reduced by the nonlinear t-SNE method and classified it using the DBSCAN clustering algorithm. ICs containing artifacts across space, time, and frequency domains were compared. The results of the work show that the tested variants of ICA cannot be considered equivalent. The independent components obtained vary across the methods.

Key words

DBSCAN, EEG, FieldTrip, ICA, MATLAB, dimension reduction, t-SNE

Obsah

Seznam zkratk	8
1 Úvod	9
2 Přehled současného stavu	13
3 Cíle práce	16
4 Metody	17
4.1 Dataset	17
4.2 Detekce a odstranění vzdálených hodnot a artefaktů	17
4.3 Analýza nezávislých komponent	18
4.3.1 Definice ICA	18
4.3.2 Předpoklady a vlastnosti ICA	19
4.3.3 Předzpracování dat pro ICA	21
4.3.4 Information-Maximization	22
4.3.5 Fast Fixed-Point	28
4.3.6 Second-Order Blind Identification	30
4.4 Pearsonův korelační koeficient	32
4.5 Dvouvýběrový Wilcoxonův test	32
4.6 t-Distributed Stochastic Neighbor Embedding	33
4.7 Density-Based Spatial Clustering of Applications with Noise	35
5 Výsledky	37
5.1 Předzpracování dat	37
5.2 Analýza nezávislých komponent	40
5.2.1 Analýza nezávislých komponent paralelních pokusů	41
5.2.2 Analýza nezávislých komponent ukázkového záznamu	46
5.2.3 Analýza nezávislých komponent všech záznamů	77
5.3 t-SNE a DBSCAN	84
5.4 Charakter komponent pro rozdílné metody potlačení artefaktů	86
6 Diskuse	120
6.1 Předzpracování dat	120
6.2 Analýza nezávislých komponent paralelních pokusů	120
6.3 Analýza nezávislých komponent ukázkového záznamu	121
6.4 Analýza nezávislých komponent všech záznamů	123
6.5 t-SNE a DBSCAN	125
6.6 Charakter komponent pro rozdílné metody potlačení artefaktů	127
6.7 Shrnutí	128

7 Závěr	130
Seznam použité literatury	131

Seznam zkratek

Zkratka	Význam
2D	2dimenzionální
cdf	distribuční funkce
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
ECG, EKG	elektrokardiograf
EEG	elektroencefalograf
EMG	elektromyograf
EOG	elektrookulograf
FastICA	Fast Fixed-Point
FIR	finite impulse response, filtr s konečnou impulzní odezvou
IC	nezávislá komponenta
ICA	analýza nezávislých komponent
Infomax	Information-Maximization
NREM, NONREM	spánková fáze s absencí rychlých očních pohybů
pdf	hustota pravděpodobnosti
REM	spánková fáze s rychlými očními pohyby
SOBI	Second-Order Blind Identification

1 Úvod

Elektroencefalogram (EEG) představuje záznam elektrické aktivity mozku snímá-
ný pomocí elektrod na povrchu skalpu. Často se využívá k diagnostice různých
mozkových stavů, jako je určení typu a umístění epileptické aktivity, k analýze
poruch spánku, k detekci neurologických dysfunkcí např. encefalopatie či demen-
ce. Na pokožce hlavy nedokážeme změřit aktivitu jediného neuronu, ale dokážeme
změřit signál, který vzniká mícháním elektrických potenciálů velkého množství neu-
ronů v různých oblastech mozku. EEG signál je náhodný, nestacionární signál s ne-
gaussovským rozložením a nelineární povahou. Má široký spektrální rozsah a kromě
užitečného signálu obsahuje také množství artefaktů. [1, 2]

EEG má frekvenční rozsah od 0,01 Hz do 100 Hz a jeho amplituda se pohy-
buje od jednotek mikrovoltů do 100 μV . Pomalé vlny do 0,5 Hz nejsou v praxi
obvykle zaznamenávány, respektive jsou filtrovány. Energie EEG je více koncent-
rována v nižším rozsahu frekvenčního spektra. Amplitudy a frekvence EEG signálu
se u zdravých dospělých lidí mění při přechodu mezi bděním a spánkem. Charakte-
ristiky vln se také mění s věkem. [1, 3]

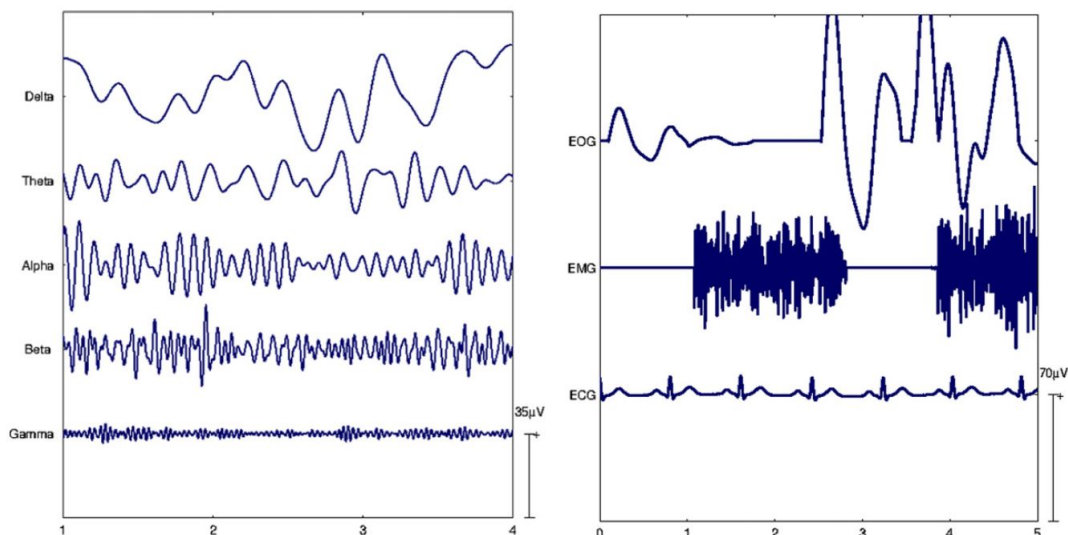
Existuje pět hlavních mozkových vln odlišujících se různými kmitočtovými roz-
sahy: delta (0,5–4 Hz), theta (4–8 Hz), alfa (8–13 Hz), beta (13–30 Hz) a gama
(> 30 Hz, ale obvykle < 100 Hz), viz obrázek 1.1 vlevo. [1]

Delta vlny leží v rozmezí 0,5–4 Hz. Tyto vlny jsou primárně spojeny s hlu-
bokým spánkem. ve stavu bdění se dají převážně nalézt u kojenců do 1 roku. [3,
2]

Theta vlny mají frekvenční rozsah 4 Hz až 8 Hz. U zdravých dětí a dospělých se
vyskytují při stavu ospalosti a spánku. Přítomnost vysoké theta aktivity u bdělých
dospělých naznačuje abnormální až patologický stav. [2, 3]

Alfa vlny mají frekvenční rozsah 8–13 Hz a amplitudu většinou menší než 50
 μV . Obvykle jsou vidět v zadních oblastech hlavy na obou stranách, přičemž vyšší
amplitudu mají na dominantní straně. Alfa rytmus je hlavní rytmus pozorovatelných
u zdravých uvolněných dospělých se zavřenýma očima. Alfa aktivita je snížena nebo
eliminována otevřením očí, nasloucháním neznámým zvukům, úzkostí nebo mentální
koncentrací či pozorností. Zajímavostí je, že Albert Einstein mohl řešit složité ma-
tematické problémy, zatímco jeho alfa aktivita nebyla potlačována, a byla přítomna
i beta a theta aktivita. [3, 2]

Beta vlna je elektrická aktivita mozku, kolísající v rozmezí 14–26 Hz (dle [2]
13–30 Hz), s normální amplitudou pod 30 μV . Je to obvyklý rytmus mozku, který se
nachází v bdělém stavu, a je spojený s aktivním myšlením, aktivní pozorností nebo
řešením konkrétních problémů u zdravých dospělých lidí. Rytmičká beta aktivita



Obrázek 1.1: Vlevo je zobrazeno pět normálních mozkových vln. Od nízkých po vysoké frekvence se jedná o vlny delta, theta, alfa, beta a gamma. Vpravo jsou zobrazeny tři nejčastější druhy biologických artefaktů: oční (EOG), svalové (EMG) a srdeční (ECG). [1]

se vyskytuje hlavně v přední a centrální oblasti. Může být blokována motorickou aktivitou nebo taktilní stimulací. Vysoká amplituda beta vlny může být přítomna u lidí nacházejících se v úzkostném a panickém stavu. [2, 3]

Frekvence nad 30 Hz (hlavně do 45 Hz) odpovídají **gama vlnám**. Ačkoliv amplitudy těchto rytmů jsou velmi nízké a jejich výskyt je vzácný, detekcí těchto rytmů lze potvrdit výskyt určitých mozkových onemocnění. Oblasti generující vysoké frekvence EEG a mající nejvyšší hladiny mozkového průtoku krve jsou umístěny ve fronto-centrální oblasti. [3]

Artefakty jsou nežádoucí elektrické potenciály, jejichž zdrojem není mozková činnost. Malá amplituda EEG signálu je velmi citlivá na rušení. Kvůli zlepšení interpretace a dalšímu zpracování EEG musí být artefakty v signálu detekovány a odstraněny. Na obrázku 1.1 vpravo jsou uvedeny tři nejběžnější fyziologické artefakty vyskytující se v EEG signálu. Jedná se o oční artefakty (EOG), svalové (EMG) a srdeční artefakty (ECG). [1, 2]

Oční artefakty

Mezi tyto artefakty patří pohyby očí a mrkání. Elektrookulogram měří elektrickou aktivitu způsobenou pohybem očí. Tato aktivita je obvykle natolik silná, že je možné ji zaznamenat spolu s EEG. Oční pohyby jsou primárně zachycovány čelními elektrodami, i když se stává, že se šíří mnohem dále. Síla interference EOG a EEG signálu závisí na blízkosti elektrod u očí a na směru, kterým se oko pohybuje. Mrkání způsobuje znečištění záznamu EEG, ve kterém se projeví jako prudká

změna. To způsobuje výskyt vyšší frekvence v EEG a amplituda mrkání je také mnohem vyšší než amplituda EEG aktivity. Z praktického hlediska je vhodné zaznamenávat v průběhu měření EEG i průběh EOG, usnadní to odstranění očních artefaktů. Naopak EOG záznam může být někdy rušen mozkovou aktivitou. [1]

Svalové artefakty

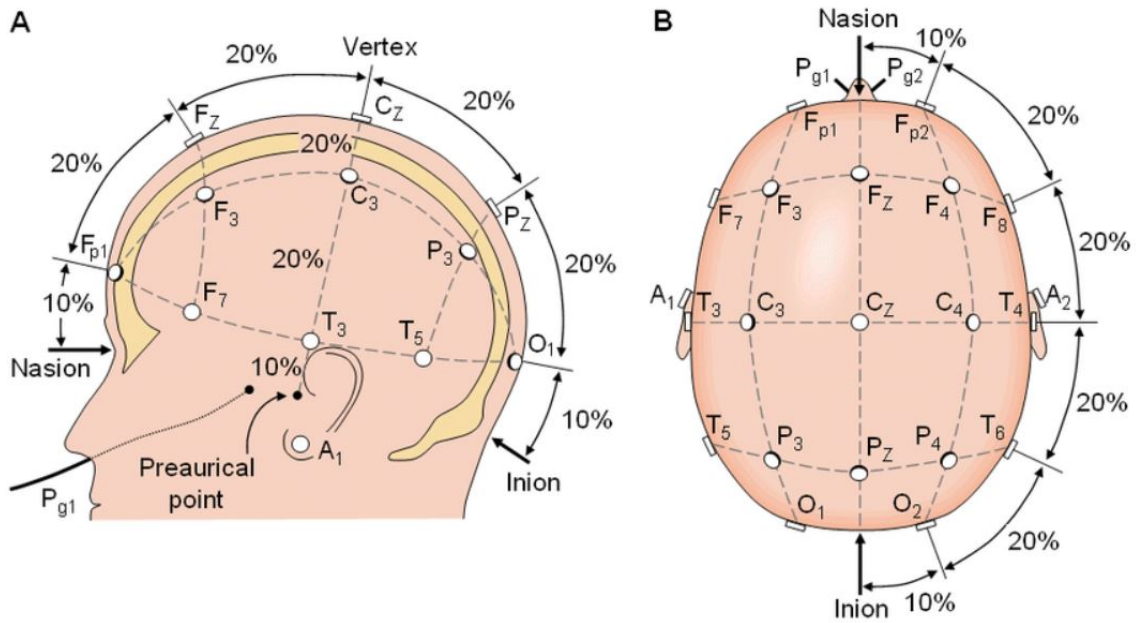
Elektromyogram měří na povrchu těla elektrickou aktivitu způsobenou aktivitou svalů. Tento artefakt je typický pro bdělé pacienty. Tvary a amplitudy rušení závisí na stupni svalové kontrakce a na typu aktivního svalu. EMG má široký spektrální rozsah a značně se překrývá s beta aktivitou v rozsahu 15–30 Hz. Lze ho často detekovat na celém skalpu v důsledku objemového vedení myogenní aktivity nezávisle generované svaly hlavy, obličeje a krku. [1]

Srdeční artefakty

Mezi srdeční artefakty patří pulzní artefakty a elektrický potenciál generovaný srdeční aktivitou. Elektrokardiogram (EKG) měří elektrickou aktivitu srdce. Amplituda srdeční aktivity na pokožce hlavy je obvykle nízká, to však do značné míry závisí na poloze elektrod. EKG má velmi charakteristický opakující se pravidelný vzorec, který se bohužel někdy může zaměnit například za epileptiformní aktivitu. EKG je rutinně měřeno spolu s mozkovou aktivitou, což usnadňuje odstranění tohoto artefaktu, protože obvykle je k dispozici referenční tvar vlny. K pulzním artefaktům dochází, když je elektroda EEG umístěna nad pulzující cévou. Pulzní artefakty vytváří pomalé periodické vlny, které se mohou podobat aktivitě EEG. Je tedy mnohem obtížnější tyto artefakty detekovat, protože můžou být časově a frekvenčně podobné měřenému EEG. Pulzní artefakty se však vyskytují pouze na jedné elektrodě a lze je tedy minimalizovat jejím správným umístěním. Existuje přímý vztah mezi EKG a pulzní aktivitou, pulzní vlny předcházejí EKG v konstantním intervalu. [1]

V roce 1958 představil Herbert H. Jasper mezinárodní standard rozmístění elektrod při měření EEG, který se používá dodnes. Předozadní rozmístění jsou založena na vzdálenosti mezi antropometrickými body nasion a inion přes vrchol v sagitální rovině. Boční rozmístění jsou založena na vzdálenosti mezi pravým a levým preaurikulárním bodem přes vrchol v koronární rovině. Na pomyslných spojnicích těchto antropometrických bodů jsou rozloženy elektrody vždy ve vzdálenosti 10 - 20 - 20 - 20 - 20 - 10 % celkové délky spojnic od referenčních bodů. Odtud název Systém 10 - 20, viz obrázek 1.2. [4]

V roce 1968 skupina vědců pod vedením Allana Rechtschaffena a Anthonyho Kalese vyvinula a publikovala první manuál pro skórování spánku u normálních lidských subjektů. Komise navrhla používání určitých EEG kanálů, délku epoch spánku ke skórování a pravidla pro rozdělení těchto epoch do 5 různých stádií spánku



Obrázek 1.2: Systém 10 – 20. Elektrody jsou označeny písmenem a číslicí. Písmena představují předozadní lokalizaci podle oblastí: Fp - frontopolární, F - frontální, P - parietální, T - temporální, O - okcipitální, C - centrální. Lichými čísly se označují elektrody umístěné nad levou hemisférou, sudými čísly nad pravou hemisférou. Z označuje nepárové (vertexové) elektrody. A1 a A2 jsou referenční elektrody [5]

a bdění. podle různých kritérií, např. podle frekvence EEG vln, jejich amplitudy či přítomnosti rychlých očních pohybů, rozdělili spánek na REM fázi a NONREM fáze 1, 2, 3 a 4. Tento skórovací systém se používá dodnes. [6]

2 Přehled současného stavu

Mozkové neurony při své aktivitě mění elektrický potenciál svých membrán. Nejsme schopni zachytit potenciály jednotlivých buněk, ale můžeme měřit sumační elektrický signál ze skalpu. Časový záznam tohoto signálu je elektroencefalogram (EEG). Kromě samotné informace o činnosti mozku se v signálu vyskytují artefakty. Artefakt je rušivý signál, jehož původ nesouvisí s činností mozku. Artefakty mohou být technické, např. rušení z elektrické sítě, nebo biologické, např. elektrický signál vznikající činností svalů. Abychom mohli získat užitečné informace z EEG signálu je nutné artefakty odstranit a signál analyzovat. k oběma účelům, kromě dalších metod, slouží analýza hlavních komponent (ICA).

ICA patří mezi metody nazývané slepá separace zdrojů (BSS). Tyto metody nepotřebují ke své práci referenční vlnu požadovaného signálu či artefaktu a využívají informace ze všech EEG elektrod. ICA zahrnuje několik metod pro oddělování lineárně smíšených nezávislých signálů pouze z časových záznamů těchto signálů, a to zavedením statistické nezávislosti zdrojů. [1]

Jednou z metod, jak aplikovat algoritmus ICA na EEG signál, je Infomax. Infomax je algoritmus vyvinutý A.J. Bellem, T.J. Sejnowskim v roce 1995. Jedná se o gradientní algoritmus, který maximalizuje entropii výstupních dat. To ve výsledku snižuje redundanci výstupních dat a zajišťuje maximální statistickou nezávislost výstupních komponent. [7]

Dalším velmi oblíbeným algoritmem pro metodu ICA je FastICA. Tento algoritmus byl vyvinut A. Hyvärinenem v roce 1999 na Helsinské technické univerzitě. FastICA maximalizuje negentropii, což je míra non-Gaussianity, výsledných komponent. Algoritmus pracuje sekvenčně, komponenty jsou tvořeny ortonormálně, čímž je zajištěno, že výsledná matice rozkladu je jedinečná a plná. Hlavní výhodou tohoto algoritmu je, že místo náročného výpočtu přesné negentropie počítá její aproximaci. Tím je zajištěno snížení výpočetní složitosti a také zrychlení chodu algoritmu. [7]

V roce 1997 představil A. Belouchrani a jeho kolegové algoritmus SOBI. SOBI je slepý identifikační algoritmus druhého řádu. Provádí diagonalizaci korelačních matic vypočítaných z původních dat s různým časovým zpožděním. Platí, že pro nezávislé zdroje musí být časově závislé kovarianční matice diagonální. [8]

Existuje dalších nejméně deset možných metod pro aplikaci ICA. Nicméně nejvíce používané pro analýzu EEG signálu jsou tři výše uvedené metody. Proto se ve své práci budu zabývat hlavně jimi. [9, 10, 11]

Algoritmy ICA se staly velice oblíbenými při analýze EEG a odstraňování artefaktů. Autor práce [12] navrhl algoritmus automatické klasifikace spánkových fází u pacientů trpících obstrukčním syndromem spánkové apnoe. ICA použil k systema-

tickému extraktu komponent spánkového EEG a k eliminaci rušení. Mezi nejčastější artefakty, kterými je rušeno nejen spánkové EEG, patří svalové artefakty (EMG), artefakty vyvolané srdeční aktivitou (ECG) a pohybem očí (EOG).

ICA se také využívá k redukcím redundantních informací či funkcí získaných při analýze EEG. v práci [13] autoři pomocí diskrétní vlnkové transformace (DWT) extrahovali příznaky vhodné k detekci a klasifikaci spánkových poruch. Množství získaných příznaků redukovali ICA, čímž odstranili redundantní příznaky.

Ve studii [14], zabývající se behaviorálním mikrosnávkem, bylo použito ICA a výkonová analýza k detekování aktivity mozku během mikrosnávkem, k nalezení zdrojů této aktivity a k jejich charakterizaci. Nejprve byly pomocí ICA odstraněny artefakty vyvolané pohybem očí. Poté metodou source-space ICA byly odděleny zdroje mikrosnávkem, rekonstruován jejich časový průběh a lokalizován jejich zdroj v mozku.

Autoři v práci [15] ukázali, že různé metody navržené pro analýzu nezávislých komponent vedou ke stejnému účinnému iteračnímu algoritmu vhodnému pro slepé oddělení smíšených nezávislých zdrojů. z teoretického hlediska všechny algoritmy ICA přibližně maximalizují nezávislost extrahovaných komponent.

Ve studii [16] bylo porovnáno několik metod ICA. z velké části různé algoritmy vedly k podobným výsledkům, navzdory rozdílům v jejich principech. Simulace ukázaly, že selektivita různých metod byla zhruba ekvivalentní, naproti tomu jejich aplikace pro reálná data neuronálních souborů zjistila, že algoritmus xICA (rozšířený Infomax) fungoval lépe než metody JADE and FastICA.

Autoři v práci [17] použili ICA na extrakci komponent z časových řad získaných funkční magnetickou rezonancí. Konkrétně se zabývali tím, do jaké míry závisí výsledky ICA na použité metodě, přičemž porovnávali metodu Infomax a FastICA. Oba algoritmy přinesly vysoce přesné výsledky, nicméně FastICA překonala Infomax z hlediska prostorové a časové přesnosti. a naopak, Infomax byl lepší z hlediska globálního odhadu modelu ICA a schopností redukovat šum.

Již v roce 2003 upozorňoval Arnaud Delorme a Scott Makeig na možné rozdíly v metodách ICA při použití na EEG datech. Míra, do jaké EEG data skutečně odpovídají předpokladům ICA, není známa. Metody Infomax, JADE a FastICA aplikovali na simulované, relativně nízkorozměrné datové soubory, pro které byly přesně splněny předpoklady ICA. Všechny algoritmy extrahovaly téměř ekvivalentní komponenty. Nicméně fyziologický význam jakýchkoliv rozdílů ve výsledcích stejných nebo různých metod ICA nebyl systematicky testován. Proto různé dekompozice ICA mohou poskytovat mírně odlišné výsledky.

Protože signály produkované mozkiem a zdroje signálů nacházejících se mimo mozek (tedy artefakty) nejsou pravděpodobně zcela nezávislé, a protože různé al-

goritmické přístupy vedoucí k maximalizaci nezávislosti komponent se liší, různé metody ICA mohou vracet poněkud odlišné výsledky, pokud se použijí na stejná EEG data. Na toto upozornili autoři práce [10]. Ve své studii porovnali 20 metod používaných k slepé separaci zdrojů, mezi nimi i několik metod ICA. Jejich dataset tvořil 71kanálové EEG záznamy od 13 subjektů, kteří během měření prováděli úkoly zaměřené na vizuální paměť. K hodnocení shodnosti extrahovaných nezávislých komponent použili tři míry: množství redukce vzájemné informace mezi obnovenými časovými průběhy IC vzhledem k zaznamenaným datovým kanálům, střední zbývající vzájemná informace mezi páry časových průběhů komponent a „dipolaritu“ rozkladu definovanou jako počet extrahovaných IC, jejichž topografické mapy lze přizpůsobit projekci skalpu jediného ekvivalentního dipólu s méně než stanoveným prahem chyby. Jako nejúčinnější metody ICA označili mimo jiné ty, které v algoritmu využívají přirozeného gradientu k maximalizaci entropie, tedy metodu Infomax.

Z uvedených zdrojů vyplývá, že analýza hlavních komponent je jedna z nej-používanějších technik analýzy EEG signálu. Také z nich vyplývá, že komponenty extrahované různými metodami ICA na stejných datech nemusí být shodné. Aby se uživatel těchto metod mohl spolehnout na získané výsledky, je třeba tuto shodnost ověřit.

3 Cíle práce

Jedním z cílů mé diplomové práce je ověření shodnosti výsledků analýzy hlavních komponent provedené 3 vybranými variantami. Můj dataset budou tvořit reálné spánkové EEG záznamy několika desítek subjektů. Existuje několik metod ICA, ve své práci se zaměřím na tři nejpoužívanější v oblasti zpracování EEG signálu. Shodnost získaných komponent budu ověřovat ze třech možných pohledů: podle časových řad, topografických map a výkonových spekter získaných komponent. K ověření shodnosti použiji metodu korelace. Komponenty extrahované různými metodami ICA budou označeny za shodné, pokud budou korelované. Pearsonovým korelačním koeficientem získám informace o počtech korelovaných a nekorelovaných (tedy neshodných) IC. Pokud se nebudou napříč metodami ICA počty lišit, budou metody ICA shodné. Tato část práce tedy bude ověřovat shodnost z hlediska matematické přesnosti.

Druhým cílem mé práce je multidimenzionální prostor topografických map nezávislých komponent redukovat a promítnout do dvoudimenzionálního prostoru. K redukcii použiji nelineární techniku. 2D prostor topografických map klasifikuji automatickým klasifikačním algoritmem pro nalezení shluků obsahujících stejné komponenty. Komponenty stejných shluků porovnáám napříč použitými metodami ICA. Druhá část práce bude ověřovat shodnost z hlediska klinického významu extrahovaných komponent.

Analýza hlavních komponent je hojně využívána k odstraňování artefaktů z EEG záznamů. Proto v poslední části své práce porovnáám charakter komponent pro rozdílné metody potlačení artefaktů. Opět budu porovnávat komponenty s artefakty v časové oblasti, v oblasti topografických map a výkonových spekter.

4 Metody

Spánek může být fyziologický i patologický. K identifikaci celé řady poruch spánku se používají různé testy např. noční oxymetrie, video-polysomnografie, psychomotorický test bdělosti nebo test mnohočetné latence usnutí (Multiple Sleep Latency Testing, MSLT). MSLT se skládá z 5 dvacetiminutových záznamů bdělosti a spánku. Během testu je monitorováno a zaznamenáváno EEG, svalová aktivita, srdeční aktivita a oční pohyby. [18]

4.1 Dataset

V této práci byly použity záznamy z MSLT pacientů s nadměrnou spavostí. Měření bylo provedeno na Ústavu spánkové medicíny v Národním ústavu duševního zdraví a bylo schváleno místní etickou komisí. K dispozici jsem měla záznamy 60 pacientů, celkem se jednalo o 213 záznamů. Pacienti byli muži a ženy středního věku. K záznamu EEG byl použit systém Brainscope s 21 elektrodami a vzorkovací frekvencí 1000 Hz. Na základě hodnocení lékařů byly v záznamech detekovány fáze bdění a fáze spánku - rychlé pohyby očí (rapid eye movement, REM) a absence rychlých pohybů očí (Non-rapid eye movement, NREM). NREM fáze se dále dělí na další tři podfáze: NREM1, NREM2 a NREM3, který se nazývá pomalovlnný spánek.

Ke zpracování dat jsem použila programové prostředí MATLAB R2019b a jeho další rozšíření (toolboxy) jako např. statistický, Fieldtrip atd. Záznamy byly podvzorkovány na frekvenci 250 Hz a filtrovány dvoucestným filtrem s konečnou impulzní odezvou (finite impulse response, FIR) řádu 1234. Řád filtru byl určen pomocí nástroje *Filter Designer* v programovacím prostředí MATLAB. Zájmová frekvenční oblast byla 0,5 - 30 Hz.

Po filtraci dat jsem rozdělila každý záznam dle indexů od lékařů na příslušné spánkové fáze a fázi bdění. Tyto fáze jsem dále rozdělila na 1 minutu dlouhé segmenty, které jsem po odstranění vzdálených hodnot a artefaktů podrobila analýze nezávislých komponent.

4.2 Detekce a odstranění vzdálených hodnot a artefaktů

K detekci vzdálených hodnot a artefaktů jsem naprogramovala jednoduchý detektor, který pracuje na principu interkvartilního rozsahu.

$$LDL_{Tukey} = Q_1 - \eta_{Tukey}(Q_3 - Q_1), UDL_{Tukey} = Q_3 + \eta_{Tukey}(Q_3 - Q_1), \quad (4.1)$$

kde LDL_{Tukey} značí dolní hranici detekčního limitu, UDL_{Tukey} je horní hranice detekce, Q_1 a Q_3 je dolní respektive horní kvartil a η_{Tukey} je malá nezáporná konstanta. [19]

Pokud se $\eta_{Tukey} = 1,5$, označují se detekované hodnoty jako "odlehle". V případě, že se $\eta_{Tukey} = 3$, označují se detekované hodnoty jako "daleké". [20]

Ve své práci jsem použila $\eta_{Tukey} = 3$.

4.3 Analýza nezávislých komponent

Analýza nezávislých komponent (Independent component analysis, ICA) je statistická a výpočetní technika, která umožňuje v pozorovaných datech odhalit jejich skryté zdroje. V modelu ICA předpokládáme, že pozorované nezávisle proměnné jsou lineární nebo nelineární směsí neznámých zdrojových nezávisle proměnných, a neznámý je i systém, ve kterém dochází k mísení signálů. Neznámé zdrojové signály jsou považovány za non-gaussovské a vzájemně nezávislé a nazývají se nezávislé komponenty zdrojových dat. Tyto nezávislé komponenty lze nalézt pomocí metod ICA. [21]

Nejvíce se metody ICA používají ke slepé separaci zdrojů, např. při zpracování biologických signálů, a extrakci příznaků, tedy vhodných reprezentací obrazových, zvukových nebo jiných dat, za účelem komprese nebo odstranění šumu. [22]

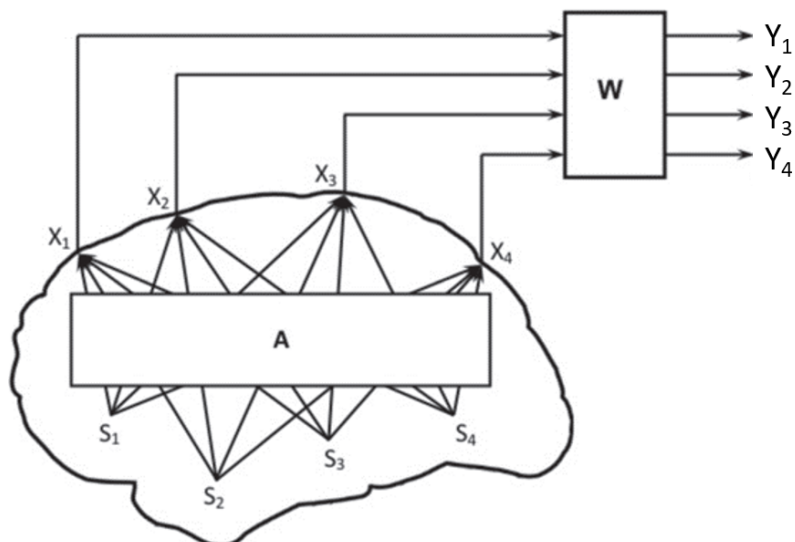
V následujících podkapitolách nejprve definuji lineární bezšumový model ICA, předpoklady a vlastnosti ICA a předzpracování dat nutné pro správnou analýzu. Dále podrobněji vysvětlím tři vybrané metody analýzy nezávislých komponent: Information-Maximization (Infomax), Fast Fixed-Point (FastICA) a Second-Order Blind Identification (SOBI). Tyto tři konkrétní metody jsem vybrala z důvodu jejich nejčastějšího použití v praxi při analýze biologických signálů, jak je uvedeno v [11, 9, 10]

4.3.1 Definice ICA

Základní lineární bezšumový model ICA předpokládá lineární kombinaci zdrojových signálů:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (4.2)$$

kde \mathbf{x} je směs zdrojových signálů, které pozorujeme, např. elektroencefalogram, \mathbf{A} je směšovací matice, která charakterizuje prostředí, skrze které zdrojové signály procházejí, a \mathbf{s} jsou zdrojové signály. [21, 23, 1]



Obrázek 4.1: Lineární mísení zdrojových signálů a slepá separace zdrojů v EEG. \mathbf{A} představuje prostředí, ve kterém probíhá lineární mísení zdrojových signálů \mathbf{s} . Na povrchu skalpu zaznameneáme směs zdrojových signálů \mathbf{x} , ze kterých se pomocí nesměšovací matice \mathbf{W} snažíme získat nezávislé komponenty \mathbf{y} , tedy původní zdrojové signály. Převzato z [1] a upraveno.

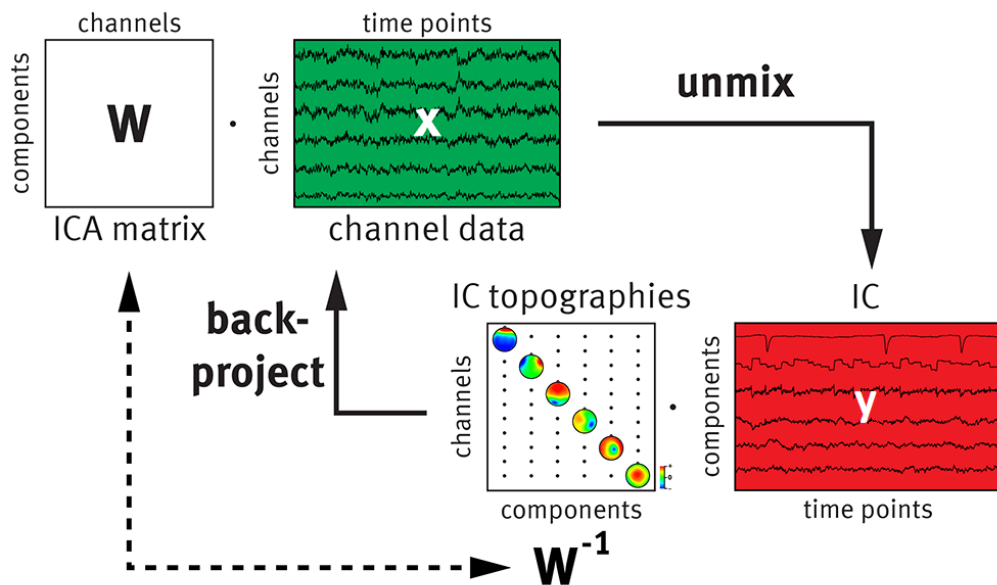
Na obrázku 4.1 je zobrazeno schéma představující základní lineární bezšumový model ICA. Nezávislé komponenty \mathbf{y} lze získat pomocí rovnice 4.3 [21].

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{x} = \mathbf{W}\mathbf{x} = \mathbf{y}, \quad (4.3)$$

Přestože efektivní počet statisticky nezávislých komponent přispívajících k naměřenému EEG na pokožce hlavy není znám, z maticového zobrazení je zřejmé, že vždy můžeme extrahovat maximálně tolik komponent, kolik je elektrod. ICA lze charakterizovat jako techniku rozkladu. To znamená, že proces míchání lze obrátit vynásobením IC s inverzní nesměšovací maticí \mathbf{W}^{-1} . Sloupce inverzní nesměšovací matice \mathbf{W}^{-1} obsahují relativní příspěvky a polaritu zpětných projekcí z každé komponenty na každou elektrodu, které lze vykreslit jako topografickou mapu. z tohoto důvodu je ICA někdy označována jako lineární prostorový filtr, protože umožňuje vyhodnocení, klasifikaci a výběr komponent na základě jejich časových řad (uložených v řádcích v matici \mathbf{y}) a topografií (uložených ve sloupcích) inverzní nesměšovací matice \mathbf{W}^{-1}). [24]

4.3.2 Předpoklady a vlastnosti ICA

Nemůžeme určit rozptyly (a tedy energie) jednotlivých nezávislých komponent a to z důvodu neznalosti zdrojových signálů \mathbf{s} i směšovací matice \mathbf{A} . Jakékoliv násobení



Obrázek 4.2: Schematický vývojový diagram ICA a zpětného promítání šesti EEG kanálů, které vede k šesti nezávislým komponentám, z nichž každá má specifickou časovou řadu a topografickou mapu. \mathbf{W} je hledaná nesměšovací matice, \mathbf{x} je směs zdrojových signálů, tedy EEG, \mathbf{y} představují nezávislé komponenty. Převzato z [24] a upraveno.

skalárem v jednom ze zdrojových signálů s_i lze eliminovat dělením odpovídajícího sloupce a_i matice \mathbf{A} stejným skalárem. V důsledku toho můžeme docela dobře určit a upravovat rozsah nezávislých komponent. Protože se jedná o náhodné proměnné, nejpřirozenějším způsobem, jak toho dosáhnout, je předpokládat, že každá z nich má jednotkový rozptyl: $E\{s_i^2\} = 1$. Stále je ponechána dvojznačnost znaménka: můžeme vynásobit nezávislou komponentu -1 , aniž bychom ovlivnili model. Tato nejednoznačnost je naštěstí ve většině aplikací zanedbatelná. [21]

Není definováno žádné pořadí mezi nezávislými komponentami. Komponenty, získané stejnou metodou aplikovanou na stejných datech, mohou mít odlišné pořadí v každém pokusu. Opět je důvodem neznalost zdrojových signálů \mathbf{s} a směšovací matice \mathbf{A} . Můžeme tedy volně měnit pořadí nezávislých komponent a kteroukoliv z nich nazvat jako první. [25]

Velmi důležitým praktickým hlediskem je, že spolehlivost jakéhokoliv rozkladu ICA nelze považovat za samozřejmost, na rozdíl od analýzy hlavních komponent (PCA), která při opakovaném použití na identických datech vrací vždy identické komponenty. Důvodem je to, že nesměšovací matice \mathbf{W} je získávána iteračně s použitím náhodně vybraných vzorků ze vstupních dat \mathbf{X} . Výsledkem může být to, že při opakovaném použití stejné metody na stejných vstupních datech můžeme získat odlišné komponenty. [26]

Omezení metod ICA spočívá v tom, že můžeme odhadnout pouze ne-gaussovské nezávislé komponenty, maximálně jedna z nezávislých komponent může mít Gaussovo (normální) rozložení. Žádným postupem nedokážeme oddělit dva nezávislé zdroje s Gaussovým rozložením, protože jejich součet bude mít opět Gaussovo rozložení. [25, 27]

Počet senzorů, snímajících směs zdrojových signálů \mathbf{x} , musí být větší nebo roven počtu zdrojových signálů \mathbf{s} . [15]

Při slepé separaci zdrojů jsou nezávislé komponenty nazývány zdrojovými signály. [25, 28, 27]

4.3.3 Předzpracování dat pro ICA

Vhodné předzpracování dat zlepšuje a zjednodušuje odhad nezávislých komponent. Ke standardnímu předzpracování dat pro ICA patří centrování a bělení.

Centrování

Základní předzpracování dat spočívá v odstranění vektoru středních hodnot $\mathbf{m} = E\{\mathbf{x}\}$ z matice \mathbf{x} . Z rovnice 4.2 vyplývá, že také matice \mathbf{s} bude mít nulové střední hodnoty. Toto předzpracování se provádí za účelem zjednodušení algoritmů ICA. Po odhadu směšovací matice \mathbf{A} můžeme odhad zdrojových signálů \mathbf{s} dokončit přičtením vektoru středních hodnot matice \mathbf{s} k centrované matici zdrojových signálů \mathbf{s} . Vektor středních hodnot matice \mathbf{s} je dán $\mathbf{A}^{-1}\mathbf{m}$, kde \mathbf{m} je vektor středních hodnot pozorovaných signálů \mathbf{x} . [22]

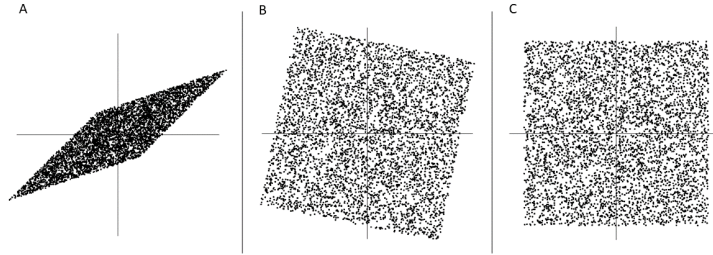
Bělení

Po centrování nastává bělení dat. Bělení dat znamená oddělení signálů od aditivního bílého šumu. To spočívá v lineární transformaci pozorovaných signálů \mathbf{x} v novou matici $\tilde{\mathbf{x}}$, jejíž vektory jsou nekorelované a mají jednotkový rozptyl. To znamená, že kovarianční matice $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\}$ je rovna jednotkové matici \mathbf{I} . [28, 22]

Tato transformace je vždy možná. Nejčastěji užívaným způsobem bělení je použití analýzy hlavních komponent (Principal component analysis, PCA). Bělení transformuje směšovací matici \mathbf{A} na novou $\tilde{\mathbf{A}}$. Rovnici 4.2 tak lze přepsat: [21, 22]

$$\tilde{\mathbf{x}} = \tilde{\mathbf{A}}\mathbf{s}. \quad (4.4)$$

Užitečnost této transformace spočívá v tom, že nová směšovací matice je ortogonální, což snižuje odhadovaný počet parametrů (prvků matice). Původní směšovací matice \mathbf{A} obsahuje n^2 prvků. Naproti tomu nová ortogonální směšovací matice $\tilde{\mathbf{A}}$ má $n(n-1)/2$ stupňů volnosti. Například ve dvou rozměrech je ortogonální



Obrázek 4.3: Ukázka významu bělení vstupních dat. **A.** Společná distribuce směsi pozorovaných signálů x_1 a x_2 . Horizontální osa je x_1 a vertikální x_2 . **B.** Společná distribuce vybělené směsi pozorovaných signálů \tilde{x}_1 a \tilde{x}_2 . Horizontální osa je \tilde{x}_1 a vertikální \tilde{x}_2 . **C.** Společná distribuce nezávislých komponent s_1 a s_2 s rovnoměrným rozdělením pravděpodobnosti. Horizontální osa je s_1 a vertikální s_2 . Převzato z [22] a upraveno.

transformace určena jedním parametrem a to parametrem úhlu otočení, jak vidíme na obrázku 4.3. [22]

Protože bělení je velmi jednoduchý a standardní postup, je vhodné tímto způsobem snížit složitost problému ICA. Současně je užitečné i snížit dimenzi vstupních dat. Díky použití metody PCA je možné odstranit hlavní komponenty, které nejsou nejméně informací. [21, 22]

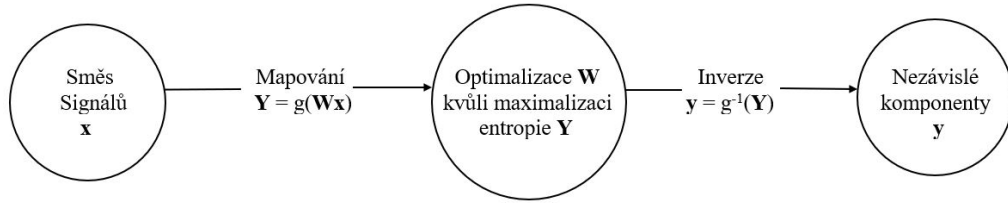
Toto předzpracování před spuštěním algoritmu vyžaduje metoda FastICA a SOBI. U Infomaxu není nutné, ale zvyšuje rychlost konvergence metody. U metod, které bělení striktně nevyžadují, je výhodné jeho použití i z důvodu snížení šumu ve vstupních datech. [22, 9]

V následujícím výkladu předpokládám centrovaná a bělená vstupní data.

4.3.4 Information-Maximization

V roce 1995 Anthony J. Bell a Terrence I. Sejnowski představili nový algoritmus analýzy hlavních komponent založený na maximalizaci entropie. Cílem každého algoritmu ICA je nalézt zdrojové signály (komponenty), které jsou vzájemně nezávislé. Nezávislost signálů měřit nelze, ale můžeme měřit jejich entropii. Infomax je založen na principu maximalizace entropie, což zajišťuje nezávislost zdrojových signálů. [28, 27]

Na Obrázku 4.4 je graficky zobrazena strategie Infomaxu. Entropie vstupní směsi signálů \mathbf{x} je konstantní. Abychom mohli entropii maximalizovat, musíme



Obrázek 4.4: Strategie Infomaxu. Převzato z [28] a upraveno.

provést mapování výstupních signálů $\mathbf{y} = \mathbf{W}\mathbf{x}$ s pomocí alternativních signálů $\mathbf{Y} = g(\mathbf{y}) = g(\mathbf{W}\mathbf{x})$. Entropie \mathbf{Y} je maximalizována optimalizací nesměšovací matice \mathbf{W} . Když je entropie maximální, výsledné signály jsou nezávislé. V dalším kroku provedeme inverzi $\mathbf{y} = g^{-1}(\mathbf{Y})$. Výsledkem jsou zdrojové komponenty \mathbf{y} , které jsou také nezávislé. [28]

Infomax řeší, jak maximalizovat vzájemné informace, které extrahované komponenty \mathbf{y} obsahují o vstupním smíšeném signálu \mathbf{x} . To je definováno jako

$$I(Y, X) = H(Y) - H(Y|X), \quad (4.5)$$

kde $I(Y, X)$ je vzájemná informace mezi výstupem a vstupem, $H(Y)$ je entropie výstupu a $H(Y|X)$ je podmíněná entropie. [27]

Uvažujeme gradient (parciální derivaci) rovnice 4.5 vzhledem k nějakému parametru w

$$\frac{\partial}{\partial w} I(Y, X) = \frac{\partial}{\partial w} H(Y), \quad (4.6)$$

kde vidíme, že $H(Y|X)$ nezávisí na w . Díky tomu lze vzájemnou informaci mezi vstupy a výstupy maximalizovat maximalizací entropie samotných výstupů. [27]

Entropii pro danou proměnnou definoval v roce 1948 Claude Elwood Shannon:

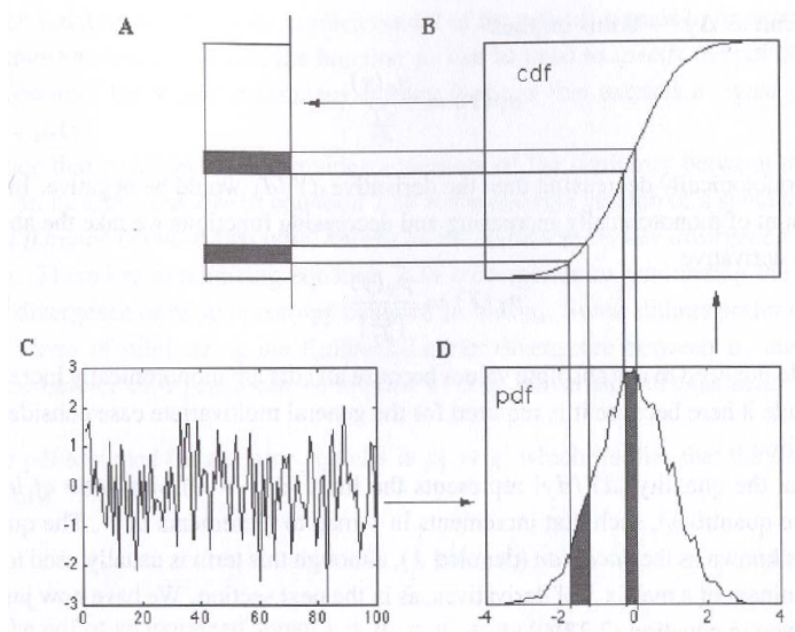
$$H(Y) = - \sum_y P(y) \log P(y), \quad (4.7)$$

kde $P(y)$ je pravděpodobnost, že Y je ve stavu y . [29]

James V. Stone definoval entropii jako míru rovnoměrnosti distribuce tak, že úplná rovnoměrnost se rovná maximální entropii. [23]

Základní vztah pro výpočet entropie mapovaného signálu Y je

$$H(Y) = - \frac{1}{N} \sum_t \ln p_Y(Y^t), \quad (4.8)$$



Obrázek 4.5: Transformace (mapování) z y na Y . Převzato z [23]

kde $Y = g(y)$, $y^t = y(t)$, horní index t značí, že y je funkce času. Funkce $g(y)$ je distribuční funkce (cumulative distribution function, cdf) požadovaného signálu y , často označovaná jako model cdf. $p_Y(Y^t)$ je hustota pravděpodobnosti (probability density function, pdf) mapovaného signálu $Y = g(y)$ a souvisí s hustotou pravděpodobnosti výstupního signálu y , $p_y(y^t)$, jak je znázorněno na obrázku 4.5 a v rovnici 4.10. [28]

Obrázek 4.5 ilustruje, jakým způsobem lze signál y , zobrazený v části C, použít pro aproximaci jeho hustoty pravděpodobnosti, v části D. Transformace této hustoty pravděpodobnosti do jeho distribuční funkce, uvedené v části B, vede k rovnoměrnému rozdělení, zobrazeného v části A. [23]

Rovnici 4.8 lze rozšířit na obecný případ, ve kterém je více než jedna náhodná proměnná

$$H(\mathbf{Y}) = -\frac{1}{N} \sum_t \ln p_Y(\mathbf{Y}^t), \quad (4.9)$$

kde mapovaný signál $\mathbf{Y} = g(\mathbf{y}) = g(\mathbf{W}\mathbf{x})$. [28]

Hustota pravděpodobnosti $p_Y(\mathbf{Y}^t)$ mapovaného signálu $\mathbf{Y} = g(\mathbf{y})$ je dle [28]

$$p_Y(\mathbf{Y}) = \frac{p_y(\mathbf{y})}{\left| \frac{\partial \mathbf{Y}}{\partial \mathbf{y}} \right|}. \quad (4.10)$$

Protože $\mathbf{Y} = g(\mathbf{y})$, kde $g(\mathbf{y})$ je model cdf zdrojového signálu, pak $\frac{\partial \mathbf{Y}}{\partial \mathbf{y}} = g'(\mathbf{y})$ a $g'(\mathbf{y})$ je pdf zdrojového signálu $p_s(\mathbf{y})$. Nahrazení tohoto výsledku do rovnice 4.10 získáme dle [28]

$$p_Y(\mathbf{Y}) = \frac{p_y(\mathbf{y})}{p_s(\mathbf{y})}. \quad (4.11)$$

Pro pdf extrahovaného signálu $p(\mathbf{y})$ vezmeme v úvahu vztah v rovnici 4.10, který platí pro jakoukoliv invertibilní funkci.

$$p_y(\mathbf{y}) = \frac{p_x(\mathbf{x})}{\left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right|}, \quad (4.12)$$

kde $p_y(\mathbf{y})$ je pdf extrahovaného signálu $\mathbf{y} = \mathbf{W}\mathbf{x}$. [28]

V rovnici 4.10 a 4.12 je ve jmenovateli Jakobián. Jakobián J je skalární hodnota determinantu Jakobiánovi matice \mathbf{J} parciálních derivací. Vztah mezi Jakobiánovou maticí \mathbf{J} a nesměšovací maticí \mathbf{W} je dle [28]

$$J = |\mathbf{J}| = |\mathbf{W}|. \quad (4.13)$$

Díky tomu lze přepsat rovnici 4.12 na tvar

$$p_y(\mathbf{y}) = \frac{p_x(\mathbf{x})}{|\mathbf{W}|}. \quad (4.14)$$

Substitucí rovnic 4.11 a 4.14 do výrazu pro entropii v rovnici 4.9 získáme

$$H(\mathbf{Y}) = -\frac{1}{N} \sum_{t=1}^N \ln \left(\frac{p_x(\mathbf{x}^t)}{|\mathbf{W}| p_s \mathbf{y}^t} \right), \quad (4.15)$$

kde t je vzorkovací čas, N je počet časových vzorků. [28]

Z vlastností logaritmů a znalosti rovnice 4.9 můžeme rovnici 4.15 přepsat:

$$H(\mathbf{Y}) = H(\mathbf{X}) + \frac{1}{N} \sum_{t=1}^N \ln p_s(\mathbf{y}^t) + \ln |\mathbf{W}|, \quad (4.16)$$

kde $H(\mathbf{X})$ je entropie původního vstupního signálu \mathbf{x} . [28]

Protože nesměšovací matice \mathbf{W} , která maximalizuje entropii $H(\mathbf{Y})$, neovlivňuje entropii $H(\mathbf{X})$, lze tento člen ignorovat. Rovnice 4.16 může být dále modifikována ignorováním uspořádání signálů M , čímž získáme základní rovnici používanou v al-

goritmu Infomax:

$$h(\mathbf{Y}) = \frac{1}{N} \sum_{i=1}^M \sum_{t=1}^N \ln p_s(y_i^t) + \ln |\mathbf{W}|, \quad (4.17)$$

kde \mathbf{W} maximalizuje entropii $H(\mathbf{Y})$, což znamená, že signály \mathbf{Y} jsou nezávislé. Protože \mathbf{y} je inverzní k \mathbf{Y} , znamená to, že komponenty \mathbf{y} jsou také nezávislé. [28]

Cílem Infomaxu je najít nesměšovací matici \mathbf{W} , která maximalizuje entropii \mathbf{Y} , nebo ekvivalentně maximalizuje $h(\mathbf{Y})$, kde $\mathbf{Y} = g(\mathbf{y}) = g(\mathbf{W}\mathbf{x})$. K optimalizaci nesměšovací matice \mathbf{W} používá metodu vzestupného gradientu. Vzestup gradientu je iterační proces, při kterém postupuje ve směru maximálního sklonu, dokud není dosaženo lokálního maxima. [28]

Gradient se zjistí pomocí parciální derivace $h(\mathbf{Y})$ vzhledem k \mathbf{W} , tedy

$$\frac{\partial h}{\partial W_{ij}} = E \left\{ \sum_{i=1}^M \frac{\partial \ln g'(\mathbf{y}_i)}{\partial W_{ij}} \right\} + \frac{\partial \ln |\mathbf{W}|}{\partial W_{ij}}, \quad (4.18)$$

kde je za účelem zjednodušení nejprve nalezen gradient s ohledem na jeden prvek \mathbf{W} , $\partial h / \partial \mathbf{W}$, a poté je rozšířen na každý prvek, $\partial h / \partial W_{ij}$. [28]

Řešením rovnice 4.18 pro jeden prvek \mathbf{W} je

$$\frac{\partial h}{\partial W_{ij}} = E \left\{ \sum_{i=1}^M \Psi(y_i) x_j \right\} + [\mathbf{W}^{-T}]_{ij}, \quad (4.19)$$

kde

$$\Psi(y_i) = \frac{g''(y_i)}{g'(y_i)}, \quad (4.20)$$

$[\mathbf{W}^{-T}]_{ij}$ je jeden element inverze transponované nesměšovací matice \mathbf{W} a $x_j = \partial y_i / \partial W_{ij}$. [28]

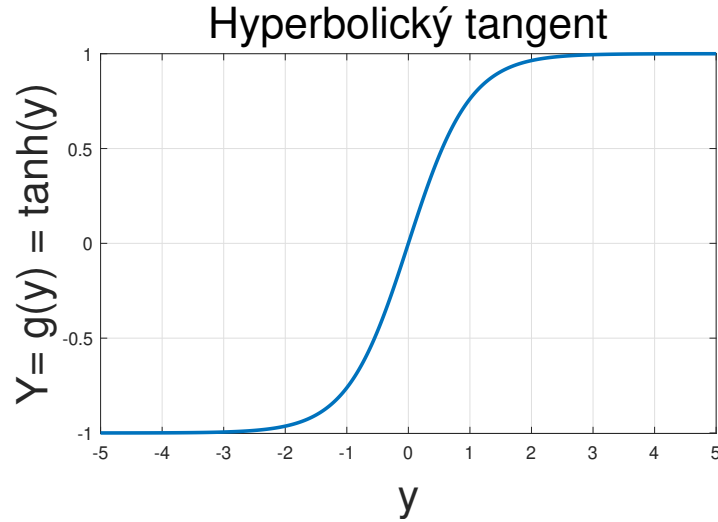
Gradient entropie ∇h pro všechny elementy nesměšovací matice \mathbf{W} za předpokladu, že jsou signály ergodické je dle [28]

$$\nabla h = \mathbf{W}^{-T} + \frac{1}{N} \sum_{t=1}^N \Psi(\mathbf{y}^t) [\mathbf{x}^t]^T. \quad (4.21)$$

Optimální nesměšovací matice \mathbf{W} je nalezena maximalizací entropie. Algoritmus Infomax iterativně postupuje směrem maximálního gradientu ∇h , dokud není dosaženo lokálního maxima:

$$\mathbf{W}_{nova} = \mathbf{W}_{predchozi} + \eta \nabla h, \quad (4.22)$$

kde η je malá konstanta. [28]



Obrázek 4.6: Graf hyperbolického tangentu

Vložením výrazu pro ∇h do rovnice 4.22 získáme rovnici pro optimalizaci nesměšovací matice \mathbf{W} pro maximalizaci entropie:

$$\mathbf{W}_{nova} = \mathbf{W}_{predchozi} + \eta \frac{1}{N} \sum_{t=1}^N \Psi(\mathbf{y}^t) [\mathbf{x}^t]^T. \quad (4.23)$$

Tato rovnice je obecná forma algoritmu Infomax používající vzestup gradientu pro optimalizaci nesměšovací matice \mathbf{W} . Nevýhodou tohoto postupu je, že algoritmus najde pouze lokální maximum, které nemusí být nutně globálním maximem funkce. To lze kompenzovat vícenásobným spuštěním algoritmu iniciovaným z různých počátečních bodů. [28]

Typickým modelem cfd zdrojového signálu \mathbf{y} , tedy funkcí $g\mathbf{y}$ používaným v algoritmu Infomax je hyperbolický tangent, jehož hodnoty vždy leží v intervalu $(-1, 1)$: [28],

$$\mathbf{Y} = g(\mathbf{y}) = \tanh(\mathbf{y}). \quad (4.24)$$

Graf hyperbolického tangentu je vidět na obrázku 4.6.

Metoda Infomax selhává v případech, kdy je směšovací matice \mathbf{A} téměř singulární. Potom je téměř singulární i nesměšovací matice \mathbf{W} a algoritmus se stává v blízkosti řešení nestabilní, jak je vidět z rovnice 4.13 a 4.14. [27]

4.3.5 Fast Fixed-Point

Metodu FastICA zavedli Aapo Hyvärinen a Erkki Oja v roce 1997. Algoritmus využívá vysoce účinné iterační schéma s pevným bodem pro nalezení lokálního extrému špičatosti (kurtózy) lineární kombinace pozorovaných proměnných. Nalezení lokálního extrému špičatosti je ekvivalentní odhadu non-Gaussových nezávislých komponent. [25]

Koeficient špičatosti pro náhodnou proměnnou x s nulovou střední hodnotou porovnává rozdělení x s normálním rozdělením pravděpodobnosti: [25]

$$\text{kurt}(x) = E \{x^4\} - 3(E \{x^2\})^2. \quad (4.25)$$

Normální (Gaussovo) rozdělení má nulovou špičatost. Kladný koeficient špičatosti mají náhodné veličiny, jejichž většina hodnot leží blízko její střední hodnotě, v našem případě blízko nule. Křivka hustoty pravděpodobnosti takové proměnné je špičatější, než-li u normálního rozdělení. Záporná špičatost značí, že rozdělení pravděpodobnosti náhodné proměnné je rovnoměrnější a její křivka hustoty je plošší než u Gaussova rozdělení. [25]

Hledáme vektor \mathbf{w} takový, aby lineární kombinace $\mathbf{w}^T \mathbf{x}$ měla maximální non-Gaussovo rozdělení (maximální non-gaussalitu). Non-gaussalita může být měřena více způsoby. Jedním ze způsobů je použití koeficientu špičatosti tak, že budeme hledat maximální nebo minimální koeficient špičatosti. Předpokládejme, že norma $\|\mathbf{w}\| = 1$. Nezávislé komponenty s normálním rozdělením nelze tímto způsobem najít, protože jejich koeficient špičatosti je nulový. Úpravou rovnice 4.25 získáme: [25]

$$\text{kurt}(\mathbf{w}^T \mathbf{x}) = E \{(\mathbf{w}^T \mathbf{x})^4\} - 3 [E \{(\mathbf{w}^T \mathbf{x})^2\}]^2 = E \{(\mathbf{w}^T \mathbf{x})^4\} - 3\|\mathbf{w}\|^4. \quad (4.26)$$

Pro minimalizaci nebo maximalizaci špičatosti lze použít algoritmus založený na vzestupném nebo sestupném gradientu: [21]

$$\frac{\partial |\text{kurt}(\mathbf{w}^T \mathbf{x})|}{\partial \mathbf{w}} = 4 \text{sign}(\text{kurt}(\mathbf{w}^T \mathbf{x})) [E \{\mathbf{x}(\mathbf{w}^T \mathbf{x})^3\} - 3\mathbf{w}\|\mathbf{w}\|^2] \quad (4.27)$$

Iterační algoritmy s pevným bodem proces hledání extrému špičatosti zrychlují. Takový algoritmus získáme, když gradient kurtózy položíme roven \mathbf{w} . [21]

$$\mathbf{w} = E \{\mathbf{x}(\mathbf{w}^T \mathbf{x})^3\} - 3\mathbf{w}\|\mathbf{w}\|^2 \quad (4.28)$$

Rovnice 4.28 je základní rovnicí, která je použita v iteračním algoritmu FastICA, který je uveden dále.

Algoritmus FastICA

V následujícím vysvětlení algoritmu FastICA vycházím z [25], není-li uvedeno jinak.

Předpokládáme, že máme k dispozici předzpracovaný vektor \mathbf{x} náhodné proměnné, což je soubor lineární směsi nezávislých zdrojových signálů podle rovnice 4.4.

Algoritmus vedoucí k nalezení jedné nezávislé komponenty:

1. Vezmeme náhodný počáteční vektor $\mathbf{w}(0)$, jehož $\|\mathbf{w}\| = 1$, kde $\|\mathbf{w}\|$ může být téměř jakákoliv běžná maticová norma, např. sloupcová nebo řádková norma (ale ne Frobeniova norma), a $k = 1$. [25, 22]
2. Nechť $\mathbf{w}(k) = E\{\mathbf{x}(\mathbf{w}(k-1)^T \mathbf{x})^3\} - 3\mathbf{w}(k-1)$. Střední hodnotu lze odhadnout při počtu vzorků náhodné veličiny \mathbf{x} nad 1000.
3. $\frac{\mathbf{w}(k)}{\|\mathbf{w}(k)\|}$.
4. Pokud $|\mathbf{w}(k)^T \mathbf{w}(k-1)|$ není dostatečně blízko 1, pak $k = k + 1$ a zpět do kroku 2. Jinak jsme našli vektor $\mathbf{w}(k)$.

Konečný vektor $\mathbf{w}(k)$, získaný tímto algoritmem, se rovná jednomu ze sloupců ortogonální směšovací matice \mathbf{A} . To znamená, že $\mathbf{w}(k)$ separoval jeden z non-Gaussových zdrojových signálů (nezávislých komponent).

Abychom odhadli n nezávislých komponent, musíme výše uvedený algoritmus spustit n krát. Zajištění, že pokaždé odhadneme jinou nezávislou komponentu, provedeme pomocí ortogonalizační projekci. Směšovací matice \mathbf{A} je ortogonální, protože předpokládáme bělení před spuštěním algoritmu FastICA. Nezávislé komponenty poté můžeme odhadnout jednu po druhé promítnutím současného řešení $\mathbf{w}(k)$ na prostor ortogonální ke sloupcům směšovací matice \mathbf{A} , která byla dříve nalezena. Definujeme matici $\overline{\mathbf{A}}$ jako matici, jejíž sloupce jsou dříve nalezenými sloupci \mathbf{A} . Poté přidáme ortogonalizační projekci na začátek kroku 3:

$$\mathbf{w}(k) = \mathbf{w}(k) - \overline{\mathbf{A}}\overline{\mathbf{A}}^T \mathbf{w}(k), \quad (4.29)$$

$\mathbf{w}(k)$ vydělíme jeho normou $\|\mathbf{w}(k)\|$. Před zahájením iterací by měl být tímto způsobem promítnut i počáteční náhodný vektor $\mathbf{w}(0)$. Aby se zabránilo chybám odhadu v $\overline{\mathbf{A}}$ kvůli zhoršujícímu se odhadu $\mathbf{w}(k)$, může být krok ortogonalizační projekce vynechán po několika prvních iteracích. Jakmile řešení $\mathbf{w}(k)$ vstoupí do přitažlivosti jednoho z pevných bodů, bude konvergovat k tomuto pevnému bodu.

4.3.6 Second-Order Blind Identification

Pokud jsou nezávislé komponenty časovými signály, mohou obsahovat více informací o struktuře než jednoduché náhodné proměnné. Tyto dodatečné informace můžeme použít ke zlepšení odhadu modelu. Příkladem využití informací o struktuře může být použití autokovarianční funkce. [21]

Second-Order Blind Identification (SOBI) je jedna z metod využívajících k výpočtu nezávislých komponent časovou strukturu pozorované směsi zdrojových signálů. Tuto metodu navrhli v roce 1997 Adel Belouchrani, Karim Abed-Meraim, Jean-Francois Cardoso a Eric Moulines. [21, 30]

Není-li uvedeno jinak, v následujícím vysvětlení metody a algoritmu SOBI vycházím z [30], kde je možné nalézt další informace a asymptotickou analýzu algoritmu.

Kovarianční matice směsi zdrojových signálů $\mathbf{x}(t)$ má následující strukturu:

$$\mathbf{R}(0) = E [\mathbf{x}(t)\mathbf{x}^*(t)], \quad (4.30)$$

kde horní index $*$ nebo H značí Hermiteovský sdruženou matici (nebo vektor).

Kovarianční matice směsi zdrojových signálů $\mathbf{x}(t)$ při aplikaci nějakého nennulového časového zpoždění τ má následující strukturu:

$$\mathbf{R}(\tau) = E [\mathbf{x}(t + \tau)\mathbf{x}^*(t)] = \mathbf{A}\mathbf{R}_s(\tau)\mathbf{A}^*, \quad (4.31)$$

kde $\mathbf{R}_s(\tau)$ je kovarianční matice zdrojových signálů.

Máme směšovací matici \mathbf{A} a bělicí matici \mathbf{B} , potom $\mathbf{B}\mathbf{A} = \mathbf{U}$, kde \mathbf{U} je $n \times n$ unitární matice. v důsledku toho může být matice \mathbf{A} faktorována jako $\mathbf{A} = \mathbf{B}^+\mathbf{U}$, kde horní index $+$ značí Moore–Penroseovu pseudoinverzi.

Proces bělení se pak zapíše jako

$$\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t). \quad (4.32)$$

Pomocí rovnice 4.31 a 4.32 získáme klíčový vztah pro prostorové bělené kovarianční matice $\underline{R}(\tau)$. Tento komplex matic není nic jiného než kovarianční matice vybeleného signálu $\mathbf{z}(t)$.

$$\underline{R}(\tau) = \mathbf{U}\mathbf{R}_s(\tau)\mathbf{U}^*, \quad (4.33)$$

pro $\forall \tau \neq 0$. Protože \mathbf{U} je unitární matice a $\mathbf{R}_s(\tau)$ je diagonální, znamená to, že jakákoliv bělená kovarianční matice je diagonalizována unitární maticí \mathbf{U} .

Pomocí několika časových zpoždění τ chceme současně diagonalizovat všechny odpovídající zpožděné kovarianční matice $R(\tau)$ v komplexu matic $\mathbf{R}_s(\tau)$. Dosaženou míru diagonalizace můžeme měřit pomocí operátoru off , který je definován pro matici \mathbf{M} o velikosti $n \times n$ s prvky M_{ij} jako

$$\text{off}(\mathbf{M}) = \sum_{1 \leq i \neq j \leq n} |M_{ij}|^2, \quad (4.34)$$

a unitární diagonalizace matice \mathbf{M} je ekvivalentní nulování $\text{off}(\mathbf{V}^* \mathbf{M} \mathbf{V})$ nějakou unitární maticí \mathbf{V} . [21, 30, 31]

Pokud je matice ve formě $\mathbf{M} = \mathbf{U} \mathbf{D} \mathbf{U}^H$, kde \mathbf{U} je unitární a \mathbf{D} je diagonální s odlišnými diagonálními prvky, může být unitárně diagonalizována pouze unitárními maticemi, které jsou v podstatě stejné jako \mathbf{U} , tj. pokud $\text{off}(\mathbf{V}^* \mathbf{M} \mathbf{V}) = 0$, pak $\mathbf{V} \doteq \mathbf{U}$.

Máme sadu K matic $O = \{\mathbf{M}_1, \dots, \mathbf{M}_K\}$ o velikosti $n \times n$. Kritérium společné diagonality je pro každou matici \mathbf{V} velikosti $n \times n$ definováno jako následující nezáporná funkce \mathbf{V} :

$$C(O, \mathbf{V}) = \sum_{k=1}^K \text{off}(\mathbf{V}^* \mathbf{M}_k \mathbf{V}). \quad (4.35)$$

Unitární matice \mathbf{V} je označována jako společný diagonalizátor množiny matic O , pokud minimalizuje kritérium společné diagonality nad množinou všech unitárních matic.

Algoritmus SOBI

Algoritmus vedoucí k nalezení nezávislých komponent:

1. Odhadneme kovarianci vzorků $\mathbf{R}(0)$ z T datových vzorků. Označíme $\lambda_1, \dots, \lambda_n$ n největších vlastních čísel a $\mathbf{h}_1, \dots, \mathbf{h}_n$ odpovídající vlastní vektory matice $\mathbf{R}(0)$.
2. Za předpokladu bílého šumu je odhad rozptylu šumu σ^2 průměrem $m - n$ nejmenších vlastních čísel $\mathbf{R}(0)$. Bělené signály jsou $\mathbf{z}(t) = [z_1(t), \dots, z_n(t)]^T$, které se počítají jako $z_i(t) = (\lambda_i - \sigma^2)^{(-1/2)} \mathbf{h}_i^* \mathbf{x}(t)$ pro $1 \leq i \leq n$. To je ekvivalentní vytvoření bělicí matice $\mathbf{W} = [(\lambda_1 - \sigma^2)^{(-1/2)} \mathbf{h}_1, \dots, (\lambda_n - \sigma^2)^{(-1/2)} \mathbf{h}_n]^*$
3. Odhady $\underline{R}(\tau)$ vytvoříme výpočtem kovariančních matic $\mathbf{z}(t)$ pro pevnou sadu časových zpoždění $\tau \in \{\tau_j | j = 1, \dots, K\}$.
4. Unitární matice \mathbf{U} je poté získána jako společný diagonalizátor sady $\{\underline{\mathbf{R}}(\tau_j) | j = 1, \dots, K\}$.
5. Zdrojové signály jsou odhadovány jako $\mathbf{s}(t) = \mathbf{U}^* \mathbf{W} \mathbf{x}(t)$ a směšovací matice jako $\mathbf{A} = \mathbf{W}^+ \mathbf{U}$

4.4 Pearsonův korelační koeficient

K porovnání nezávislých komponent získaných výše uvedenými metodami jsem použila Pearsonův korelační koeficient.

Předpokládejme, že X a Y jsou dvě proměnné: Proměnná X obsahuje n pozorování x_1, x_2, \dots, x_n a proměnná Y obsahuje n pozorování y_1, y_2, \dots, y_n . Pearsonův korelační koeficient je definován: [32]

$$\rho(X, Y) = \frac{(N \sum x_i y_i - \sum x_i \sum y_i)}{\sqrt{N x_i^2 - (\sum x_i)^2} \sqrt{N y_i^2 - (\sum y_i)^2}}. \quad (4.36)$$

Hodnota $\rho(X, Y)$ je v intervalu $< -1, 1 >$. Když $\rho(X, Y) = 1$, proměnné X a Y spolu pozitivně lineárně korelují. Pokud Když $\rho(X, Y) = -1$, proměnné X a Y spolu negativně lineárně korelují. V případě, že $\rho(X, Y) = 0$, lineární korelace mezi proměnnými není zřejmá. [32]

Dle Evansovi příručky z roku 1996 se jako velmi silně korelované proměnné nazývají proměnné, které mají Pearsonův korelační koeficient v absolutní hodnotě mezi 0,80 a 1,00.

Korelace pod 0,5 nebo nad -0,5 se nebere v potaz. [33]

4.5 Dvouvýběrový Wilcoxonův test

Dvouvýběrový Wilcoxonův test (The Wilcoxon rank sum test) je neparametrický test sloužící k porovnání dvou nezávislých výběrů. Používáme jej v případě, kdy nelze předpokládat normální rozdělení hodnot. Testujeme jím hypotézu, že oba výběry pochází ze stejného rozdělení. [34, 35]

Nechť $\{X_1, X_2, \dots, X_n\}$ je náhodným výběrem z rozdělení s distribuční funkcí F_x a $\{Y_1, Y_2, \dots, Y_m\}$ je náhodným výběrem z rozdělení s distribuční funkcí G_y . Nulová hypotéza H_0 : distribuční funkce obou rozdělení si jsou rovny $F_x = G_y$. Alternativní hypotéza H_1 : distribuční funkce obou rozdělení si nejsou rovny $F_x \neq G_y$. [34, 35]

Oba výběry spojíme a prvky uspořádáme podle velikost, přičemž prvkům, které mají stejnou velikost přiřadíme průměr jejich pořadí. Dvouvýběrový Wilcoxonův test je ekvivalentní Mann-Whitneyovu U testu. Statistika testu pro Mann-Whitneyův U Test je menší z U_1 a U_2 definovaných níže. [34, 35]

$$U_1 = nm + \frac{n(n+1)}{2} - R_1, U_2 = nm + \frac{m(m+1)}{2} - R_2, \quad (4.37)$$

kde R_1 je součet pořadí prvků z prvního výběru a R_2 je součet pořadí prvků z druhého výběru. Nulovou hypotézu lze odmítnout na hladině významnosti α pokud

$\min U_1, U_2 \leq u(\alpha)$, kde $u(\alpha)$ je kritická hodnota, kterou nalezneme v tabulkách. [34, 35]

4.6 t-Distributed Stochastic Neighbor Embedding

Získané komponenty jsem pro vizualizaci a pro následné nalezení struktury pomocí klasifikačních algoritmů redukovala na dvoudimenzionální prostor pomocí nelineární metody t-Distributed Stochastic Neighbor Embedding (t-SNE).

Tuto techniku redukce dimenze používají autoři [36] k automatické detekci epileptických záchvatů. Autoři biometrického autentizačního systému založeného na EEG použili t-SNE k vizualizaci účinků navrhovaného přístupu v práci [37].

T-SNE je nelineární technika snižování rozměrů multidimenzionálních dat vyvinutá Laurenssem van der Maatenem a Geoffreyem Hintonem v roce 2008. Je vhodná zejména k vizualizaci vysokodimenzionálních dat v nízkodimenzionálním prostoru, typicky dvoudimenzionálním či třídimenzionálním. [38, 39]

Každý multidimenzionální objekt modeluje nízkodimenzionálním bodem tak, že podobné objekty jsou modelovány blízkými body a odlišné objekty jsou modelovány vzdálenými body s vysokou pravděpodobností. Tento algoritmus je schopen velmi dobře zachytit lokální strukturu multidimenzionálních dat a zároveň dokázat odhalit globální strukturu, jako je přítomnost shluků ve velkých měřících. [38]

Při redukci dimenze vstupních dat provádí algoritmus t-SNE několik kroků [40]:

1. Výpočet párových vzdáleností mezi vysokorozměrnými objekty.
2. Konstrukce matice společných distribucí pravděpodobnosti P definované vzorcem 4.38 u vysokodimenzionálních dat.
3. Vytvoření počáteční sady nízkorozměrných bodů a výpočet společných distribucí pravděpodobnosti Q u nízkodimenzionálních dat dle výrazu 4.39.
4. Minimalizace Kullback-Leiblerovy divergence mezi Gaussovou distribucí v multidimenzionálním prostoru a studentovou t-distribucí v nízkodimenzionálním prostoru dle rovnice 4.40

Hodnoty společných distribucí pravděpodobnosti P jsou definovány jako symetrické podmíněné pravděpodobnosti

$$p_{i,j} = \frac{p_{j|i} + p_{i|j}}{2 \cdot n}. \quad (4.38)$$

Podobnost objektu x_j k objektu x_i představuje podmíněná pravděpodobnost $p_{j|i}$ taková, že x_i si vybere x_j jako svého souseda, pokud sousedé byli vybráni v poměru

k jejich hustotě pravděpodobnosti pod Gaussovou křivkou soustředěnou nad x_i . Pro blízké objekty je $p_{j|i}$ relativně vysoká, zatímco pro velmi vzdálené objekty bude $p_{j|i}$ téměř nekonečně malá (pro standardní hodnoty rozptylu Gaussovy funkce, σ_i). [38, 40]

Konkrétní hodnoty σ_i určuje nepřímou uživatelským prostřednictvím parametru perplexity. Perplexita je způsob, jak vyvážit pozornost algoritmu mezi lokálními a globálními aspekty vstupních dat, protože se hustota dat v celém souboru bude pravděpodobně v různých oblastech lišit. Dá se říci, že perplexita je odhadem počtu blízkých sousedů, které má každý objekt. A čím vyšší počet sousedů, tím vyšší hustota shluku. Typické hodnoty perplexity jsou mezi 5 a 50. [38, 39]

U nízkodimenzionálních bodů y_i a y_j multidimenzionálních objektů x_i a x_j počítáme společnou distribuci pravděpodobnosti $q_{i,j}$

$$q_{i,j} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}}. \quad (4.39)$$

Metoda t-SNE používá studentovu t-distribuci. Díky ní nedochází k vytlačování velice vzdálených bodů z mapového znázornění společných distribucí pravděpodobnosti. [38]

Pokud nízkodimenzionální mapové body y_i a y_j správně modelují podobnosti mezi vysokodimenzionálními objekty x_i a x_j , pravděpodobnosti $p_{i,j}$ a $q_{i,j}$ budou stejné. t-SNE se snaží o nalezení takové nízkodimenzionální reprezentace dat, která minimalizuje nesoulad mezi $p_{i,j}$ a $q_{i,j}$. Přírodním způsobem jak dosáhnout souladu je minimalizace Kullback-Leiblerovi divergence mezi společným rozdělením pravděpodobnosti P , v multidimenzionálním prostoru a společným rozdělením pravděpodobnosti Q v nízkodimenzionálním prostoru. Minimální hodnotu Kullback-Leiblerovi divergence KL nalezneme pomocí gradientu: [38]

$$\frac{\delta KL}{\delta y_i} = 4 \cdot \sum_j (p_{i,j} - q_{i,j}) \cdot (y_i - y_j) \cdot (1 + \|y_i - y_j\|^2)^{-1}. \quad (4.40)$$

Algoritmus t-SNE přirozeně rozšiřuje husté shluky a uzavírá řídké. Nicméně vizualizace vzdáleností mezi shluky neodpovídá reálným vzdálenostem mezi shluky v multidimenzionálním prostoru. [39].

4.7 Density-Based Spatial Clustering of Applications with Noise

Shluková analýza je proces seskupování velkých datových souborů na základě jejich podobnosti. Jedná se o proces učení bez učitele, protože o vstupních datech nemáme předem žádné znalosti. Hustotně založené shlukovací algoritmy jsou založeny na myšlence, že objekty, které tvoří hustou oblast, by měly být seskupeny do jednoho shluku. k určení hustých oblastí používají pevnou prahovou hodnotu. Hledají oblasti s vysokou hustotou dat, které jsou odděleny regiony s nižší hustotou. Jednou z nevýhod těchto algoritmů je to, že zachycují pouze určité druhy šumových bodů, pokud v datech existují shluky s různými hodnotami hustoty. Kromě toho fungují dobře, pokud jsou shluky od sebe dostatečně vzdáleny, ale pokud spolu shluky sousedí, nejsou tyto algoritmy uspokojivé. Mezi takové algoritmy patří Density-Based Spatial Clustering of Applications with Noise (DBSCAN). [41]

V DBSCANu se hustota spojená s bodem získá počítáním počtu bodů v oblasti zadaného poloměru kolem bodu. Body s hustotou nad stanoveným prahem jsou označeny jako shluky. Algoritmus DBSCAN má schopnost objevovat shluky s libovolným tvarem (lineárním, konkávním nebo oválným) a nevyžaduje určení počtu shluků na vstupu. [41]

Autoři v práci [42] aplikovali metodu DBSCAN na reálná EEG data a porovnali ji se standardně využívanou metodou k-means. DBSCAN byl, narozdíl od k-means, schopen oddělit vnořené shluky. Tento hustotně založený algoritmus vykazoval dobré výsledky pro klinicky velmi důležité třídy epilepsie a fyziologické aktivity. Vytvořil nehomogennější třídy s výjimkou třídy elektromyografických artefaktů, které nebylo možné identifikovat.

DBSCAN použili autoři v práci [43] k odstranění artefaktů a zvýšení robustnosti metody pro včasnou diagnostiku poruch autistického spektra ze signálů EEG. DBSCANem detekovali a vyřadili odlehle hodnoty a zbývající data požili pro další analýzu. V důsledku odstranění artefaktů se snížila chybovost navrhované diagnostické metody.

DBSCAN je navržen tak, aby objevoval libovolné shluky v jakékoliv vstupní databázi a současně rozlišoval šumové body. Klíčovou myšlenkou DBSCANu je, že pro každý objekt shluku musí sousedství daného poloměru $Eps(\epsilon)$ obsahovat alespoň minimální počet objektů $MinPts$, to znamená, že kardinalita sousedství musí překročit určitou prahovou hodnotu. Eps-sousedství libovolného bodu p je definována jako

$$N_{Eps} = \{q \in D \mid dist(p, q) \leq Eps\}, \quad (4.41)$$

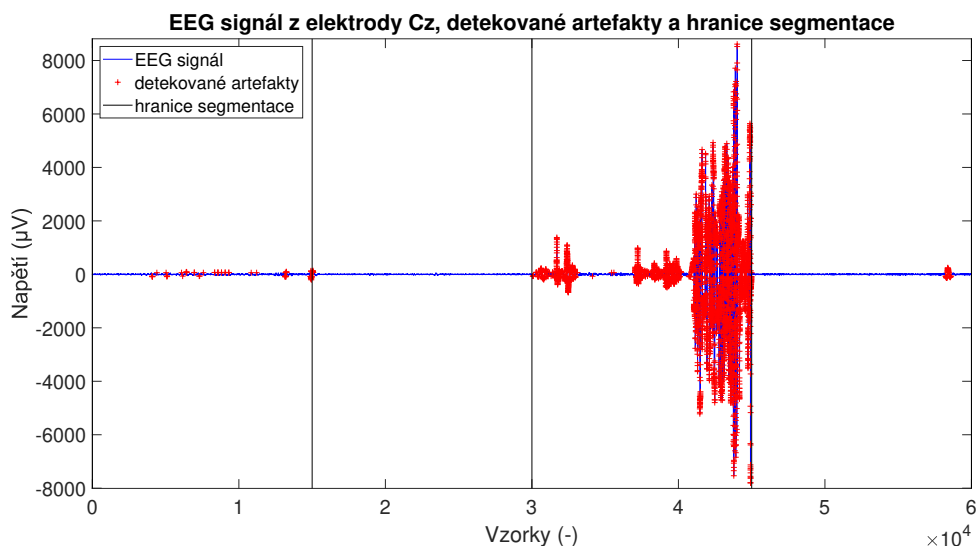
kde D je vstupní databáze objektů. [41]

Pokud Eps-sousedství bodu p obsahuje alespoň minimální počet bodů $MinPts$, pak se tento bod nazývá centrální bod. Centrální bod je definován jako

$$N_{Eps}(p) > MinPts, \quad (4.42)$$

kde Eps a $MinPts$ jsou parametry specifikované uživatelem, které znamenají poloměr sousedství a minimální počet bodů v sousedství centrálního bodu. Pokud tato podmínka není splněna, považuje se tento bod za nepodstatný bod. [44]

DBSCAN hledá shluky kontrolou Eps-sousedství každého objektu v datové sadě. Pokud Eps-sousedství objektu p obsahuje více než $MinPts$, vytvoří se nový shluk s p objektem jako základním objektem. Potom iterativně shromažďuje hustotně-dosažitelné objekty přímo z těchto základních objektů do stejného shluku. Proces končí, když žádný nový objekt nelze přidat do libovolného shluku. [44]



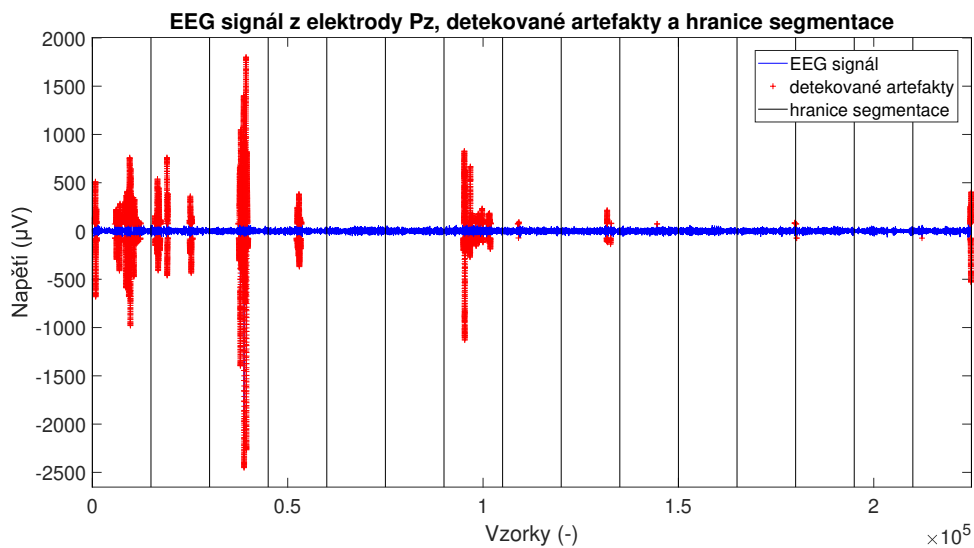
Obrázek 5.1: EEG signál z elektrody Cz, detekované artefakty a hranice segmentace.

5 Výsledky

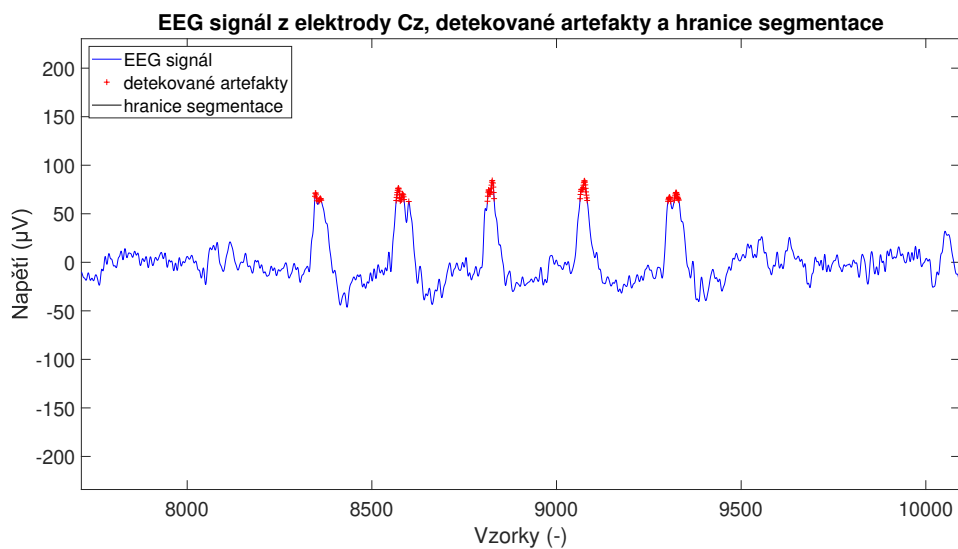
Kapitolu výsledky kvůli lepší orientaci rozdělím do několika částí. V první části představím výsledky předzpracování dat, zejména detekci a odstranění artefaktů. Ve druhé části se budu věnovat samotné analýze hlavních komponent. Ukáži výsledky ICA několika pokusů jednom záznamu. Poté na ukázkové segmentu ukáži a popíši časové řady, topografické mapy a výkonová spektra získaná všemi variantami ICA. V každé oblasti také uvedu jejich korelační koeficienty. Představím výsledky korelace na všech záznamech a otestuji jejich shodnost. Ve třetí části budou zobrazeny dvou-dimenzionální prostory topografických map všech variant klasifikované automatickým klasifikačním algoritmem. Ve čtvrté a poslední části Budou pro vybraný segment prezentovány nezávislé komponenty představující artefakty, jejich časové řady, topografické mapy i výkonová spektra.

5.1 Předzpracování dat

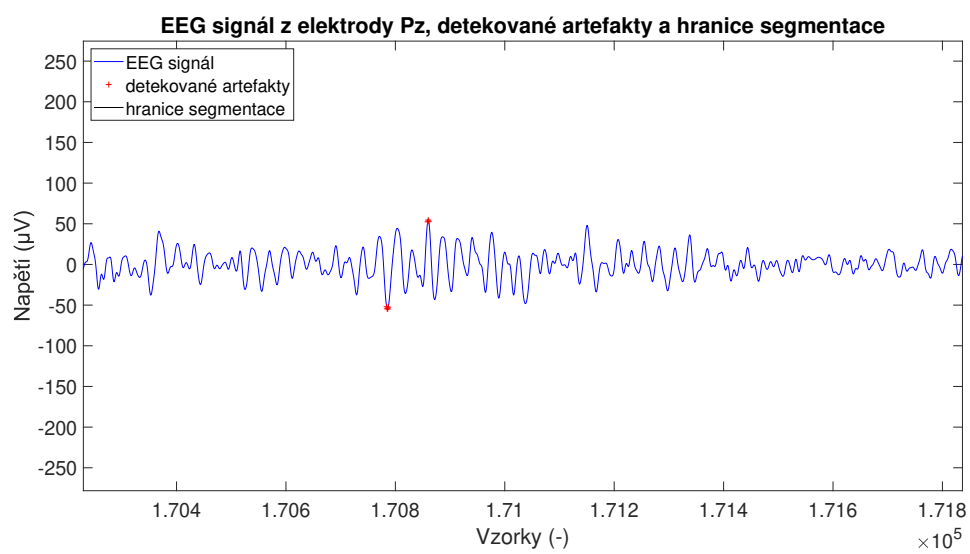
Po filtraci a segmentaci dat jsem provedla detekci a odstranění vzdálených hodnot a artefaktů. Aplikovaný postup je popsán v podkapitole 4.2. Ukázkový EEG signál s detekovanými vzdálenými hodnotami je vidět na obrázku 5.1 a 5.2. V detailu na obrázku 5.3 jsou vidět detekované oční artefakty. Nevýhoda tohoto jednoduchého odstranění artefaktů je vidět na obrázku 5.4, kde došlo k možnému chybnému označení signálu jako odlehlé hodnoty.



Obrázek 5.2: EEG signál z elektrody Pz, detekované artefakty a hranice segmentace.



Obrázek 5.3: Detail EEG signálu z elektrody Cz a detekované artefakty.



Obrázek 5.4: Detail EEG signálu z elektrody Pz a možná chybná detekce artefaktů.

5.2 Analýza nezávislých komponent

U každého záznamu jsem porovnála časovou řadu nezávislých komponent, topografické mapy nezávislých komponent a výkonová spektra ve čtyřech frekvenčních pásmech (alfa, beta, delta a theta) nezávislých komponent mezi třemi různými metodami ICA. K porovnání jsem použila Pearsonův korelační koeficient.

Nezávislé komponenty jsem rozdělila do třech skupin: IC s korelačním koeficientem v absolutní hodnotě větším než 0,8, které označuji jako silně korelované IC, IC s korelačním koeficientem v absolutní hodnotě větším než 0,5 ale menším než 0,8, které označuji jako středně korelované a komponenty, které nekorelovaly s žádnou další komponentou.

Metoda FastICA generovala počet nezávislých komponent variabilně, ne vždy vypočítala všech 19 IC. Rozdíly v počtu vypočítaných IC mezi metodou Infomax a FastICA a SOBI a FastICA jsou uvedeny v příslušných tabulkách v příloze v elektronické verzi této práce. Výsledky korelace časových řad IC a topografických map nezávislých komponent mezi různými metodami jsou uváděny ve tvaru $\bar{x} \pm \sigma$.

V podkapitole 4.3.2 jsem uvedla, že opakované použití totožných metod ICA na totožných datech může vést k rozdílným výsledkům. Pro ověření vlivu tohoto faktoru na samotné porovnání metod ICA mezi sebou, jsem použila každou metodu opakovaně desetkrát na jednom ukázkovém záznamu.

V první části uvedu výsledky opakovaných výpočtů IC všemi metodami na totožném záznamu. Ve druhé části budou výsledky porovnání IC vypočítaných třemi různými metodami ICA na jednom ukázkovém záznamu. Ukázkový záznam byl vybrán náhodně, jedinou podmínkou bylo, aby obsahoval fázi bdění i spánkové fáze REM, NONREM 1 nebo NONREM 2. Vzhledem k velkému rozsahu vstupních dat, jsou tabulky s výsledky pro všech 213 záznamů uvedeny v příloze v elektronické verzi této práce. Nakonec se budu zabývat výsledky korelace IC u všech záznamů.

5.2.1 Analýza nezávislých komponent paralelních pokusů

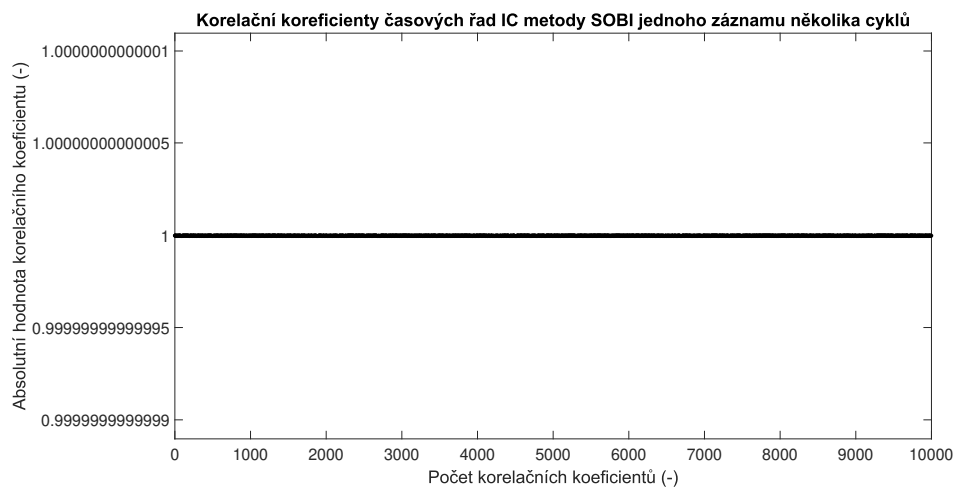
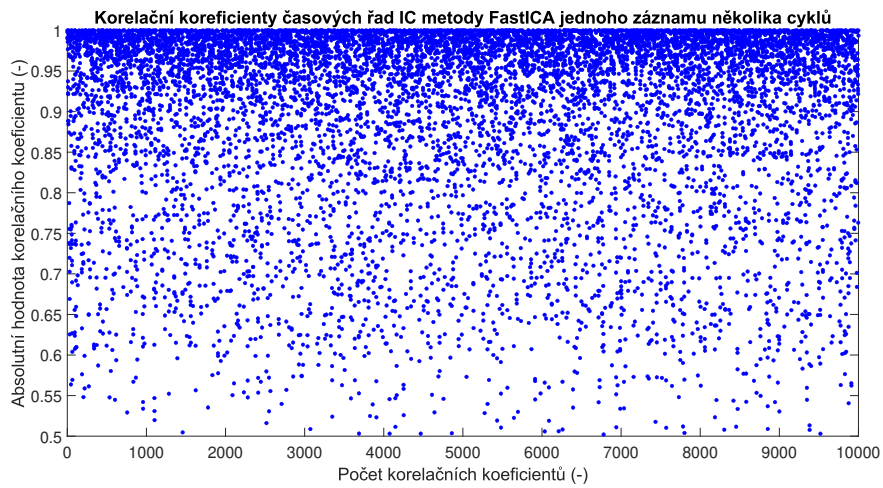
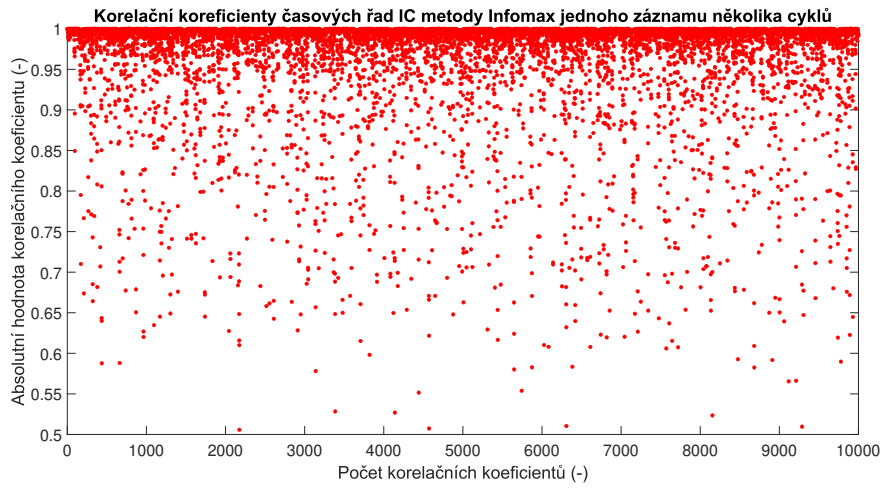
Porovnávala jsem, jak moc si budou časové řady, topografické mapy a výkonová spektra IC získaných stejnou metodou podobná. K tomuto porovnání jsem použila Pearsonův korelační koeficient. Vypočítala jsem korelační koeficienty všech pokusů mezi sebou. Zajímalo mě, zda korelační koeficient dosáhne hodnoty 1. V případě nižší hodnoty by bylo nutné touto hodnotou normovat výpočty pro výzkum rozdílů dílčích metod ICA.

Na obrázku 5.5 jsou vidět absolutní hodnoty korelačních koeficientů časových řad IC metod Infomax (nahore), FastICA (uprostřed) a SOBI (dole). Jedná se koeficienty prvních 10000 párů komponent v absolutní hodnotě větší než 0,5, jejichž komponenty považuji za korelované. Z obrázku je patrné, že maximální hodnota korelace je 1. U metody Infomax je vidět, že korelační koeficienty jsou nejvíce zastoupené v hodnotách mezi 0,95–1, naopak mezi hodnotami 0,5–0,55 je jen malé množství koeficientů. U metody FastICA je rozdělení koeficientů podobné, jako u metody Infomax, s tím rozdílem, že vysoké hodnoty dosáhlo méně párů komponent, zatímco ty nejnižší více. Část korelačních koeficientů, dá se říci, se z vyšších hodnot přesunula na nižší. U metody SOBI je jednoznačně vidět, že v každém pokusu vypočítala naprosto totožné komponenty. Proto je graf korelačních koeficientů metody SOBI v jiném měřítku.

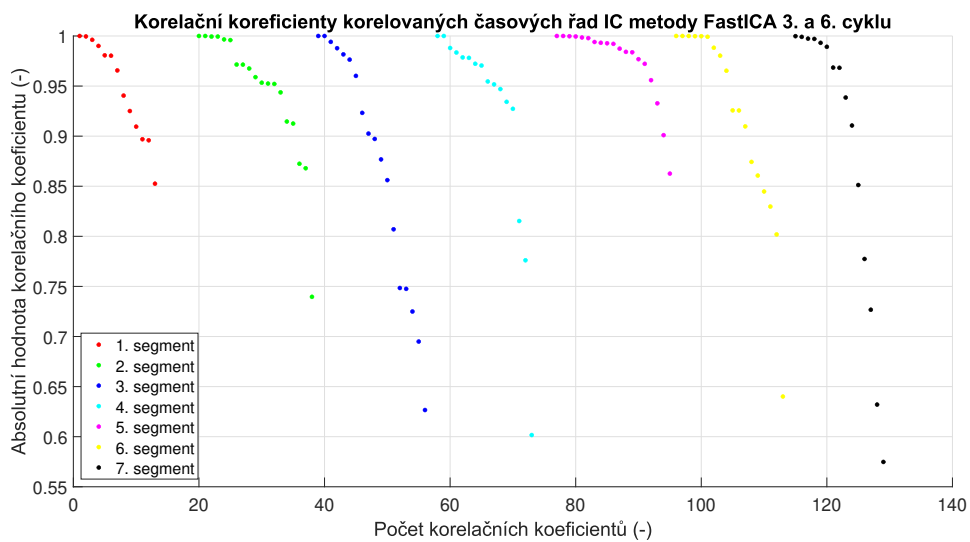
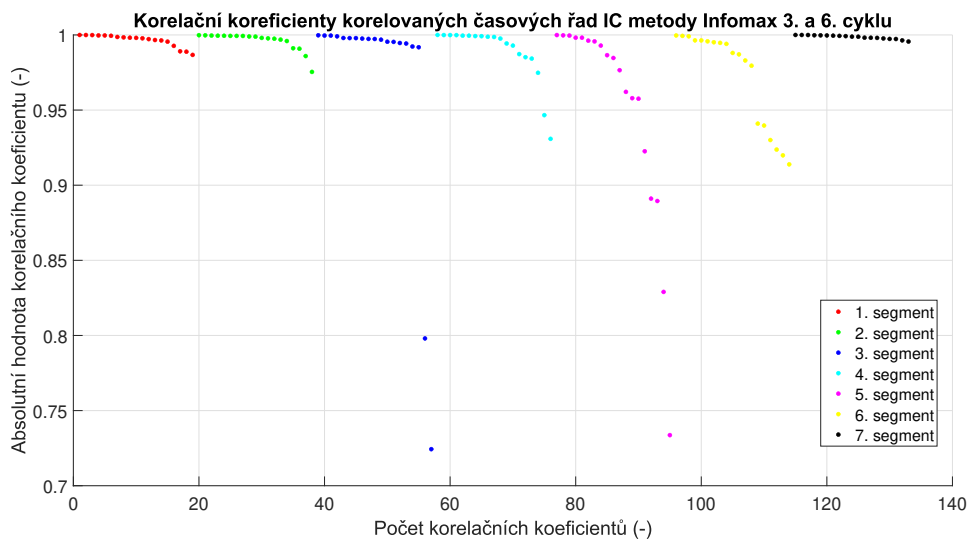
Rozdělení absolutních hodnot korelačních koeficientů topografických map IC a výkonových spekter IC bylo v případě metod Infomax a FastICA velmi podobné, v případě metody SOBI totožné, jako u korelace časových řad, proto nejsou grafy zobrazeny.

Zajímalo mě, jak vypadá rozdělení korelačních koeficientů časových řad IC u různých segmentů pocházejících ze dvou stejných pokusů. Výsledky jsou vidět na obrázku 5.6, kde jsou zobrazeny korelační koeficienty 3. a 6. pokusu. V grafech je jasně vidět, že metoda Infomax (na obrázku nahore) ve většině cyklů extrahovala téměř totožné komponenty. Také je patrné, že u metody FastICA (na obrázku dole) jsou korelační koeficienty rovnoměrněji rozděleny mezi hodnoty 0,5–1, extrahované komponenty si jsou méně podobné než v případě metody Infomax.

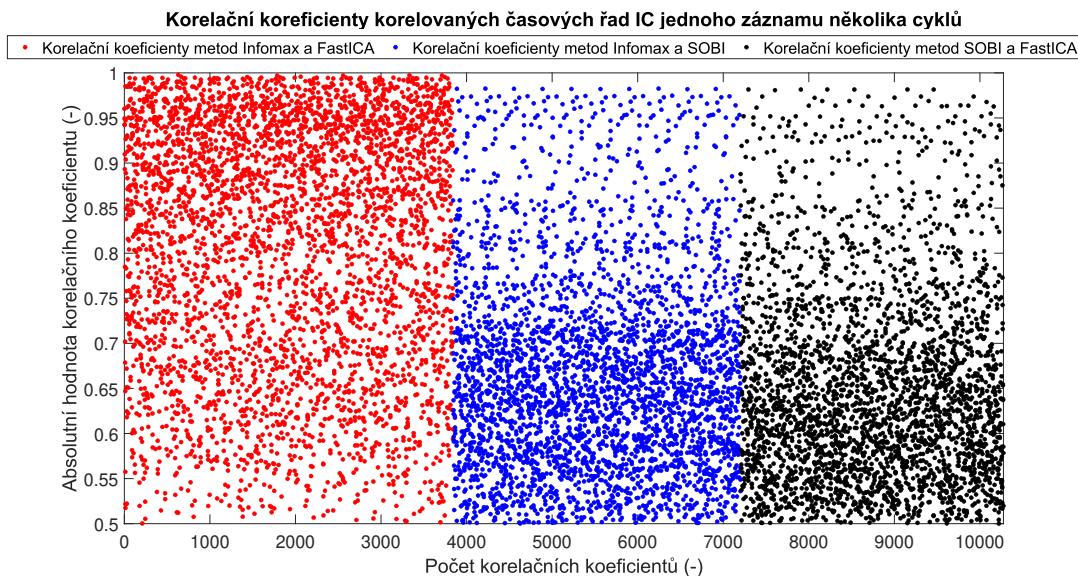
V posledním kroku jsem vypočítala korelaci nezávislých komponent všech pokusů mezi metodami Infomax a FastICA, Infomax a SOBI, SOBI a FastICA. Korelační koeficienty časových řad IC jsou vidět na obrázku 5.7. U metod Infomax–FastICA dosáhly koeficienty hodnoty 0,998, zatímco u metod Infomax–SOBI a SOBI–FastICA dosáhly nejvyšší hodnoty 0,983 respektive 0,982. Také u metod Infomax–SOBI a SOBI–FastICA dosahovalo více komponent nižších korelačních koeficientů než v případě metod Infomax–FastICA. Podle počtu korelačních koeficientů



Obrázek 5.5: Korelační koeficienty časových řad IC jednoho záznamu několika pokusů metody Infomax (nahore), metody FastICA (uprostřed) a SOBI (dole).



Obrázek 5.6: Korelační koeficienty korelovaných časových řad IC metod Infomax (nahore) a FastICA (dole) 3. a 6. cyklu ICA

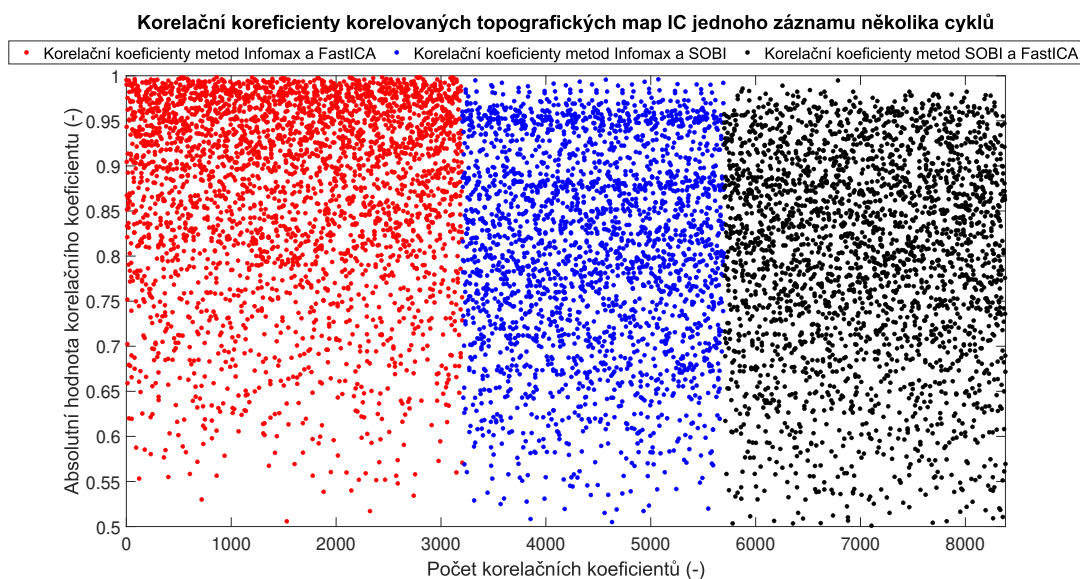


Obrázek 5.7: Korelační koeficienty korelovaných časových řad IC metod Infomax–FastICA, Infomax–SOBI a SOBI–FastICA

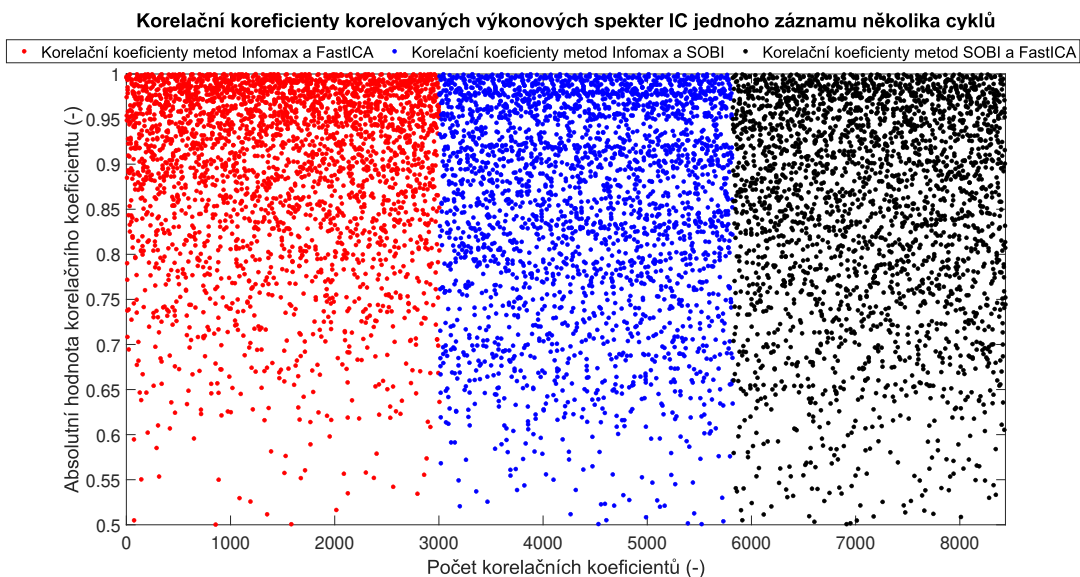
je také vidět, že největší počet korelovaných párů měly metody Infomax–FastICA a nejmenší Infomax–SOBI.

V případě korelačních koeficientů topografických map IC všech pokusů mezi všemi metodami došlo k přerozdělení koeficientů oproti časovým řadám, viz obrázek 5.8. Více párů komponent korelovalo více, zejména u metod Infomax–SOBI a SOBI–FastICA. Nicméně celkový počet korelovaných párů klesl (oproti korelaci časových řad). Nejmenšího počtu korelovaných párů v tomto případě dosáhly metody Infomax–SOBI. Metody Infomax–FastICA dosahovaly nejvyšší hodnoty korelačního koeficientu 0,999. Metody Infomax–SOBI 0,996 a metody SOBI–FastICA 0,995.

Korelační koeficienty výkonových spekter IC všech pokusů mezi všemi metodami jsou vidět na obrázku 5.9. U všech párů metod dosáhlo nejvyššího koeficientu více komponent než v předchozích případech. Největší počet korelovaných komponent měly opět metody Infomax–FastICA, i když už jejich převaha není tak výrazná. Více párů komponent mělo vyšší korelační koeficienty. Nízké korelace se vyskytovaly zřídka.



Obrázek 5.8: aKorelační koeficienty korelovaných topografických map IC metod Infomax-FastICA, Infomax-SOBI a SOBI-FastICA



Obrázek 5.9: Korelační koeficienty korelovaných výkonových spekter IC metod Infomax-FastICA, Infomax-SOBI a SOBI-FastICA

5.2.2 Analýza nezávislých komponent ukázkového záznamu

Ukázkový záznam obsahoval spánkové fáze REM a NONREM 2 a fázi bdění. Záznam ve fázi REM obsahoval 3 jednodominutové segmenty, ve fázi NONREM 2 9 segmentů a ve fázi bdění 5. Pro prezentaci nezávislých komponent spánku jsem vybrala fázi NONREM 2 a třetí segment, a to z toho důvodu, že metoda FastICA v tomto segmentu spočítala všech 19 IC, jako metody Infomax a SOBI.

Kvůli lepší orientaci uvedu na začátku této podkapitoly text a tabulky, který čtenáře provede výsledky. Na konci kapitoly jsou vloženy obrázky časových řad nezávislých komponent, jejich topografických map a výkonových spekter, kvůli možnosti optického porovnání výsledků.

Výsledky korelace metod Infomax - FastICA

V tabulce, 5.1 jsou uvedeny výsledky korelace časových řad, topografických map a výkonových spekter IC mezi metodami Infomax a FastICA. Celkem se v oblasti časových řad spárovalo 16 IC, z toho 3 mají korelační koeficient v absolutní hodnotě větší než 0,8. V oblasti topografických map silně korelovalo 9 párů, středně 5 párů. Pokud tyto výsledky srovnáme s korelacemi časových řad zjistíme, že 6 párů korelovalo společně v časové oblasti i v topografických mapách. Dva páry zanikly, a to infomax8 - fastica16 a infomax10 - fastica17, v časové řadě spolu korelovaly středně. Osm topografických map jedné metody korelovalo s jinými osmi mapami druhé metody než v případě časových řad. Jedná se např. o pár infomax1 - fastica 14, které v časové řadě korelovaly silně, nicméně v oblasti topografických map korelovaly s jinými mapami. Oproti časovým řadám ubyly 4 páry IC u korelace výkonového spektra. 6 párů zůstalo shodných a 5 IC jedné metody korelovalo s jinou komponentou druhé metody než v případě časových řad. Pouze dva páry IC korelovaly ve výkonovém spektru i v topografické mapě. U korelace výkonového spektra vznikly 2 páry nové, ale 4 zmizely oproti topografickým mapám.

Opticky porovnat ukázky časových řad nezávislých komponent můžete na obrázcích 5.10 a 5.11, kde jsou uvedeny IC metody Infomax, a na obrázcích 5.15 a 5.16, kde jsou vidět nezávislé komponenty vypočítané metodou FastICA. U některých časových řad je korelace patrná na první pohled, např. pár infomax1 - fastica14, u některých je korelace pouhým okem sotva postřehnutelná, např. jako u páru infomax8 - fastica16. Rozsah osy y není uveden, protože jak jsem ukázala v podkapitole 4.3.2, nemůžeme určit rozptyly jednotlivých nezávislých komponent a to z důvodu neznalosti zdrojových signálů \mathbf{s} i směšovací matice \mathbf{A} . V důsledku toho můžeme docela dobře určit a upravovat rozsah nezávislých komponent.

Topografické mapy nezávislých komponent vypočítaných metodou Infomax jsou uvedeny na obrázcích 5.12 a 5.13. U metody FastICA pak na obrázcích 5.17 a 5.18. Na obrázcích je vidět, že topografická mapa infomax2 koreluje nejen s mapou fastica11, ale i s mapou fastica4. Nicméně ve výsledcích uvádím pouze maximální korelační koeficienty, a to z toho důvodu, že hledám komponenty, které si jsou co nejvíce podobné, nebo které nejsou podobné s žádnou jinou. V dalším statistickém zpracování totiž použiji počty spárovaných komponent a počty těch nespárovaných. Některé korelace map jsou hned vidět, např. pár infomax3 - fastica 8, zatímco u páru infomax9 - fastica 19 není korelace ihned patrná.

Opticky porovnat výkonová spektra můžete na obrázku 5.14, kde jsou výkonová spektra nezávislých komponent metody Infomax, a na obrázku 5.19, kde jsou spektra IC metody FastICA. Na první pohled je zřejmá korelace páru infomax1 - fastica14, méně výrazná u páru infomax8 - fastica5.

Nespárované časové řady, topografické mapy a výkonové spektra obou metod jsou uvedeny v tabulce 5.2. V časové řadě byly nespárovány pouze 3 IC z každé metody. Jedinou nezávislou komponentou, která se nespárovala jak v časové řadě, v topografické mapě, tak i výkonovém spektru zůstala fastica7. Nespárovaných výkonových spekter bylo nejvíce, celkem 8 z každé metody.

Některé časové řady nezávislých komponent metody Infomax korelovaly s více než jednou časovou řadou IC metody FastICA a naopak. V tabulce 5.3 jsou dotyčné komponenty uvedeny. Vícenásobné korelace počítám pro komponenty jejichž korelační koeficient byl v absolutní hodnotě větší než 0,5. Průměr Pearsonových korelačních koeficientů počítám z jejich absolutních hodnot. U metody Infomax se jednalo o násobné korelace 2 komponent, u metody FastICA 4 IC.

Vícenásobně korelované topografické mapy IC metod Infomax a FastICA jsou uvedeny v tabulce 5.4. Oproti vícenásobné korelaci časových řad uvedených metod vzrostl počet násobně korelovaných IC u obou metod, vzrostl počet korelací (ve dvou případech až na 4) a také průměrné korelační koeficienty byly vyšší.

Tabulka 5.5 obsahuje soupis násobně korelovaných výkonových spekter IC metod Infomax a FastICA. Celkem korelovalo 17 spekter se dvěma nebo třemi spektry IC opačné metody. Zajímavé je, že trojnásobné korelace měli IC které neměli násobné korelace v časové řadě a topografických mapách.

Celkově pouze 2 páry nezávislých komponent korelovaly v časové řadě, topografické mapě i ve výkonovém spektru, a to pár infomax15 - fastica1 a pár infomax17 - fastica 4. Naopak jedinou nespárovanou komponentou ve všech oblastech byla fastica7.

Tabulka 5.1: Korelované páry časových řad, topografických map a výkonových spekter nezávislých komponent fáze NONREM 2 mezi metodami Infomax a FastICA a příslušné korelační koeficienty.

Korelované páry	Pearsonův korelační koeficient		
	Časové řady	Topografické mapy	Výkonová spektra
infomax1 - fastica14	-0,9396	–	0,9965
infomax1 - fastica18	–	0,8123	–
infomax2 - fastica11	0,7338	0,8458	–
infomax2 - fastica18	–	–	0,9369
infomax3 - fastica8	-0,5161	-0,9657	–
infomax3 - fastica16	–	–	0,7966
infomax4 - fastica13	-0,7009	–	–
infomax5 - fastica9	-0,6470	–	0,9267
infomax5 - fastica14	–	0,8886	–
infomax6 - fastica11	–	–	0,8836
infomax6 - fastica15	-0,5463	–	–
infomax7 - fastica3	0,6968	–	–
infomax7 - fastica12	–	0,7784	–
infomax8 - fastica5	–	–	0,6355
infomax8 - fastica16	-0,6355	–	–
infomax9 - fastica5	-0,5662	–	–
infomax9 - fastica19	–	-0,6916	–
infomax10 - fastica17	0,5812	–	0,9778
infomax11 - fastica6	-0,5371	-0,8700	–
infomax12 - fastica3	–	-0,8034	–
infomax13 - fastica10	-0,7581	-0,8280	–
infomax14 - fastica12	0,7556	–	–
infomax15 - fastica1	-0,9817	-0,9814	0,9844
infomax16 - fastica2	–	0,7752	–
infomax17 - fastica4	0,9848	0,9587	0,9952
infomax18 - fastica3	–	–	0,7972
infomax18 - fastica15	–	0,7794	–
infomax19 - fastica2	0,7239	–	0,9688
infomax19 - fastica13	–	0,7436	–

Tabulka 5.2: Nespárované nezávislé komponenty fáze NONREM 2 mezi metodami Infomax a FastICA.

Nespárované nezávislé komponenty metod Infomax a FastICA		
Časové řady	Topografické mapy	Výkonová spektra
–	infomax4	infomax4
–	infomax6	–
–	–	infomax7
–	infomax8	–
–	–	infomax9
–	infomax10	–
–	–	infomax11
infomax12	–	infomax12
–	–	infomax13
–	infomax14	infomax14
infomax16	–	infomax16
infomax18	–	–
–	fastica5	fastica5
–	–	fastica6
fastica7	fastica7	fastica7
–	–	fastica8
–	fastica9	–
–	–	fastica10
–	–	fastica12
–	–	fastica13
–	fastica16	–
–	fastica17	–
fastica18	–	–
fastica19	–	fastica19

Tabulka 5.3: Vícenásobné korelace časových řad nezávislých komponent fáze NONREM 2 mezi metodami Infomax a FastICA.

Vícenásobně korelované časové řady IC	Počet korelací	Průměrný Pearsonův korelační koeficient
infomax9	2	0,5643
infomax16	2	0,5312
fastica2	2	0,6401
fastica3	2	0,6719
fastica6	2	0,5369
fastica8	2	0,5393

Tabulka 5.4: Vícenásobné korelace topografických map nezávislých komponent fáze NONREM 2 mezi metodami Infomax a FastICA.

Vícenásobně korelované topografické mapy IC	Počet korelací	Průměrný Pearsonův korelační koeficient
infomax2	2	0,8398
infomax3	2	0,9058
infomax5	4	0,8794
infomax8	3	0,6830
infomax11	2	0,8122
infomax13	2	0,8035
infomax15	2	0,8486
fastica2	2	0,8106
fastica3	2	0,8409
fastica10	3	0,8173
fastica12	4	0,7381
fastica15	2	0,7663
fastica18	2	0,7885
fastica19	2	0,7095

Tabulka 5.5: Vícenásobné korelace výkonových spekter nezávislých komponent fáze NONREM 2 mezi metodami Infomax a FastICA.

Vícenásobně korelovaná výkonová spektra IC	Počet korelací	Průměrný Pearsonův korelační koeficient
infomax1	2	0,9016
infomax2	2	0,7263
infomax5	2	0,8654
infomax10	3	0,9324
infomax11	2	0,8696
infomax14	2	0,8620
infomax15	2	0,8295
infomax17	3	0,9468
infomax19	2	0,9146
fastica2	2	0,7905
fastica3	2	0,7417
fastica9	3	0,9236
fastica11	2	0,9023
fastica15	2	0,8163
fastica16	3	0,7054
fastica17	2	0,9497
fastica18	2	0,8968

Výsledky korelace metod Infomax - SOBI

Korelace časových řad, topografických map a výkonových spekter IC mezi metodami Infomax a SOBI je uvedena v tabulce 5.6. V časové řadě byl silně korelovaný pouze jeden pár, středně korelovalo 11 párů. Silně korelovaných párů topografických map bylo 5, středně korelovaných taktéž. Korelací map vznikly 3 nové páry, ale 5 jich zaniklo, narozdíl od korelace časových řad. 3 páry zůstaly shodné, a to infomax6 - sobi9, infomax15 - sobi8 a infomax17 - sobi5. 4 mapy jedné metody korelovalo s jinými 4 mapami druhé metody. Pokud porovnáme korelované páry časových řad a korelované páry výkonových spekter, zjistíme, že celkem 4 páry zůstaly shodné a 1 nový pár vznikl. Oproti korelaci topografických map vznikly u korelace výkonového spektra 3 nové páry, ale 2 původní zmizely. Pouze 1 pár zůstal shodný a to pár infomax17 - sobi5.

Časové řady IC vypočítaných metodou SOBI jsou uvedeny na obrázcích 5.20 a 5.21. Je vidět korelace páru infomax17 - sobi5, naopak hůře je vidět korelace u páru infomax13 - sobi13.

Opticky lze porovnat topografické mapy IC vypočítaných metodou SOBI na obrázcích 5.22 a 5.23. Některé korelace nejsou úplně zřejmé, např. infomax8 - sobi10, některé jsou patrné hned, např. infomax15 - sobi8.

Normovaná výkonová spektra IC metody SOBI jsou uvedena na obrázku 5.24. Některé korelace jsou vidět na první pohled, např. infomax10 - sobi8, některé nejsou tak patrné, např. jako u páru infomax1 - sobi10.

V tabulce 5.7 jsou uvedeny nespárované časové řady, topografické mapy a výkonová spektra IC metod Infomax a SOBI. Nespárované časové řady komponent infomax14, sobi11, sobi14 a sobi15 zůstaly nespárované i topografických mapách. U metody SOBI zůstala nespároována stejná výkonová spektra, kromě IC sobi3, jako u časových řad. U metody Infomax zůstaly shodně nespárované výkonová spektra IC infomax2 a infomax5, stejně jako u topografických map.

Vícenásobné korelace u některých časových řad IC metod Infomax a SOBI dosáhly počtu maximálně 2, jak je vidět v tabulce 5.8. U metody Infomax se jednalo o 2 IC (jiné než v případě korelace s metodou FastICA) u metody SOBI o 4 IC.

Topografická mapa infomax5 korelovala celkem se 6 topografickými mapami metody SOBI, jak je vidět v tabulce 5.9. Tato komponenta měla nejvíce korelací i s IC metody FastICA (celkem 4). Celkem 11 topografických map IC obou metod korelovalo s více než 1 mapou opačné metody, v případě korelace map metod Infomax–FastICA se jednalo o 14 vícenásobně korelovaných map. Téměř všechny průměrné koeficienty (kromě jednoho) jsou vyšší než u korelace časových řad stejných metod.

Tabulka 5.6: Korelované páry časových řad, topografických map a výkonových spekter nezávislých komponent fáze NONREM 2 mezi metodami Infomax a SOBI a příslušné korelační koeficienty.

Korelované páry	Pearsonův korelační koeficient		
	Časové řady	Topografické mapy	Výkonová spektra
infomax1 - sobi3	-0,7414	–	–
infomax1 - sobi10	–	–	0,8609
infomax2 - sobi2	0,5135	–	–
infomax3 - sobi1	–	0,9371	–
infomax4 - sobi10	-0,5415	–	–
infomax4 - sobi19	–	–	0,8857
infomax5 - sobi19	-0,6354	–	–
infomax6 - sobi9	-0,6512	-0,9510	–
infomax7 - sobi2	–	0,8742	–
infomax7 - sobi12	–	–	0,7111
infomax8 - sobi10	–	0,7106	–
infomax9 - sobi4	–	0,6751	–
infomax9 - sobi9	–	–	0,6660
infomax10 - sobi8	–	–	0,9866
infomax10 - sobi16	-0,6887	–	–
infomax11 - sobi12	-0,5144	–	–
infomax11 - sobi17	–	0,6982	–
infomax12 - sobi7	0,5220	–	0,9194
infomax13 - sobi13	-0,6333	–	0,7840
infomax14 - sobi1	–	–	0,9229
infomax15 - sobi2	–	–	0,9341
infomax15 - sobi8	-0,6125	-0,8228	–
infomax16 - sobi4	0,6526	–	0,6210
infomax17 - sobi5	-0,9680	-0,9418	0,9974
infomax18 - sobi18	–	-0,7928	–
infomax19 - sobi6	–	0,7691	–
infomax19 - sobi16	–	–	0,9614

Tabulka 5.7: Nespárované nezávislé komponenty fáze NONREM 2 mezi metodami Infomax a SOBI.

Nespárované nezávislé komponenty metod Infomax a SOBI		
Časové řady	Topografické mapy	Výkonová spektra
–	infomax1	–
–	infomax2	infomax2
infomax3	–	infomax3
–	infomax4	–
–	infomax5	infomax5
–	–	infomax6
infomax7	–	–
infomax8	–	infomax8
infomax9	–	–
–	infomax10	–
–	–	infomax11
–	infomax12	–
–	infomax13	–
infomax14	infomax14	–
–	infomax16	–
infomax18	–	infomax18
infomax19	–	–
sobi1	–	–
–	sobi3	sobi3
sobi6	–	sobi6
–	sobi7	–
sobi11	sobi11	sobi11
–	sobi12	–
–	sobi13	–
sobi14	sobi14	sobi14
sobi15	sobi15	sobi15
–	sobi16	–
sobi17	–	sobi17
sobi18	–	sobi18
–	sobi19	–

Tabulka 5.8: Vícenásobné korelace časových řad nezávislých komponent fáze NONREM 2 mezi metodami Infomax a SOBI.

Vícenásobně korelované časové řady IC	Počet korelací	Průměrný Pearsonův korelační koeficient
infomax1	2	0,6469
infomax6	2	0,6118
sobi4	2	0,5952
sobi7	2	0,5145
sobi8	2	0,5672
sobi10	2	0,5469

Tabulka 5.9: Vícenásobné korelace topografických map nezávislých komponent fáze NONREM 2 mezi metodami Infomax a SOBI.

Vícenásobně korelované topografické mapy IC	Počet korelací	Průměrný Pearsonův korelační koeficient
infomax2	3	0,7316
infomax3	3	0,7999
infomax5	6	0,8087
infomax6	3	0,9005
infomax13	2	0,6693
infomax18	2	0,7708
sobi1	4	0,8021
sobi2	6	0,7587
sobi4	2	0,7626
sobi8	3	0,7649
sobi18	2	0,8092

Tabulka 5.10: Vícenásobné korelace výkonových spekter nezávislých komponent fáze NONREM 2 mezi metodami Infomax a SOBI.

Vícenásobně korelovaná výkonová spektra IC	Počet korelací	Průměrný Pearsonův korelační koeficient
infomax1	2	0,7789
infomax5	2	0,7939
infomax10	3	0,9675
infomax11	3	0,8272
infomax17	5	0,8703
sobi1	2	0,9209
sobi2	2	0,8203
sobi5	2	0,9140
sobi7	3	0,8687
sobi10	2	0,9024
sobi12	2	0,7291
sobi16	3	0,8766
sobi19	3	0,8864

Všechna výkonová spektra IC uvedená v tabulce 5.10 metody Infomax korelovala taktéž s alespoň dvěma výkonovými spektrama IC metody FastICA. Výkonová spektra infomax1, sobi7 a sobi10 násobně korelovala i v časové oblasti. Infomax5, sobi1 a sobi2 násobně korelovala v oblasti topografických map.

Jediným párem IC, kterým koreloval v časové řadě, topografických mapách i výkonových spektrech, byl pár infomax17 - sobi5. Naopak ani v jedné oblasti se nespárovaly komponenty sobi11, sobi14 a sobi15.

Výsledky korelace metod SOBI - FastICA

Jediným silně korelovaným párem byl pár sobi5 - fastica4, jak je vidět v tabulce 5.11. Středně korelovaných párů bylo 11. Silně korelovaných párů topografických map bylo 7, středně korelovaných 5. Oproti korelaci časových řad vznikly 3 nové páry a jeden zmizel, 3 páry zůstaly shodné, sobi1 - fastica8, sobi4 - fastica5, sobi5 - fastica4. Silně korelovaných párů výkonových spekter bylo 7, středně korelovaných taktéž. Oproti korelaci časových řad vznikly 2 páry nové a 4 páry zůstaly shodné. 7 map IC jedné metody korelovalo s jinými 7 mapami druhé metody na rozdíl od korelace časových řad IC. Pokud porovnáme korelaci výkonových spekter a topografických map, dva páry zůstanou shodné, a to sobi4 - fastica5 a sobi5 - fastica4. Vzniknou 3 nové páry - sobi3 - fastica14, sobi9 - fastica16 a sobi19 - fastica17. Naopak jeden pár zmizí - sobi6 - fastica13. 9 výkonových spekter IC jedné metody koreluje s jinými spektry IC druhé metody než v případě korelace map.

Tabulka 5.11: Korelované páry časových řad, topografických map a výkonových spekter nezávislých komponent fáze NONREM 2 mezi metodami SOBI a FastICA a příslušné korelační koeficienty.

Korelované páry	Pearsonův korelační koeficient		
	Časové řady	Topografické mapy	Výkonová spektra
sobi1 - fastica8	-0,5787	-0,9270	–
sobi1 - fastica9	–	–	0,8582
sobi2 - fastica1	–	–	0,9437
sobi2 - fastica10	–	-0,8304	–
sobi3 - fastica14	0,6484	–	0,6954
sobi4 - fastica5	-0,7594	-0,7757	0,7961
sobi5 - fastica4	-0,9391	-0,8805	0,9861
sobi6 - fastica13	–	0,8315	–
sobi7 - fastica3	-0,5587	–	–
sobi7 - fastica18	–	–	0,8824
sobi8 - fastica1	0,6562	–	–
sobi8 - fastica2	–	0,9192	–
sobi9 - fastica16	–	–	0,5819
sobi10 - fastica18	–	-0,6684	–
sobi11 - fastica11	0,5165	–	–
sobi12 - fastica3	–	-0,8341	–
sobi12 - fastica15	-0,5353	–	0,7510
sobi13 - fastica3	–	–	0,5973
sobi13 - fastica9	–	0,7515	–
sobi13 - fastica17	0,5765	–	–
sobi14 - fastica11	–	–	0,9043
sobi14 - fastica15	–	0,7264	–
sobi16 - fastica2	–	–	0,9668
sobi16 - fastica10	0,5272	–	–
sobi16 - fastica11	–	-0,8568	–
sobi17 - fastica7	–	-0,7906	–
sobi17 - fastica19	–	–	0,7094
sobi18 - fastica8	–	–	0,7630
sobi19 - fastica9	0,5737	–	–
sobi19 - fastica17	–	–	0,9359

Tabulka 5.12: Nespárované nezávislé komponenty fáze NONREM 2 mezi metodami Infomax a SOBI.

Nespárované nezávislé komponenty metod Infomax a SOBI		
Časové řady	Topografické mapy	Výkonová spektra
sobi2	–	–
–	sobi3	–
–	–	sobi6
–	sobi7	–
–	–	sobi8
sobi9	sobi9	–
sobi10	–	sobi10
–	sobi11	sobi11
sobi14	–	–
sobi15	sobi15	sobi15
sobi17	–	–
sobi18	sobi18	–
–	sobi19	–
–	fastica1	–
fastica2	–	–
fastica6	fastica6	fastica6
fastica7	–	fastica7
–	–	fastica10
fastica12	fastica12	fastica12
–	–	fastica13
–	fastica14	–
fastica16	fastica16	–
–	fastica17	–
fastica18	–	–
fastica19	fastica19	–

Nespárované časové řady, topografické mapy a výkonová spektra metod SOBI a FastICA jsou vidět v tabulce 5.12. Nespárovaných časových řad IC obou metod zůstalo 7. U metody SOBI se nespárovaly 3 topografické mapy stejných IC (konkrétně sobi9, sobi15 a sobi18) jako v časových řadách, u metody FastICA 4 (konkrétně fastica6, fastica12, fastica16 a fastica19). U obou metod zůstalo nespárováno 5 výkonových spekter. Komponenty sobi11, sobi15, fastica6 a fastica12 se nespárovaly ve výkonových spektrech ani v topografických mapách.

Násobné korelace časových řad IC metod SOBI a FastICA jsou uvedeny v tabulce 5.13. U metody Infomax se jednalo pouze o 1 IC, u metody SOBI o 2, přičemž všechny korelovaly pouze se 2 IC opačné metody.

Vícenásobných korelací topografických map IC metod SOBI a FastICA dosáhlo celkem 12 IC. Komponenta sobi2 korelovala trojnásobně s IC metody FastICA, zatímco u metody Infomax se jednalo o šestnásobnou korelaci. Naproti tomu např.

Tabulka 5.13: Vícenásobné korelace časových řad nezávislých komponent fáze NONREM 2 mezi metodami SOBI a FastICA.

Vícenásobně korelované časové řady IC	Počet korelací	Průměrný Pearsonův korelační koeficient
sobi8	2	0,5948
sobi10	2	0,5290
fastica14	2	0,5906

Tabulka 5.14: Vícenásobné korelace topografických map nezávislých komponent fáze NONREM 2 mezi metodami SOBI a FastICA.

Vícenásobně korelované topografické mapy IC	Počet korelací	Průměrný Pearsonův korelační koeficient
sobi1	2	0,9162
sobi2	3	0,7315
sobi6	3	0,7441
sobi8	3	0,8671
sobi12	2	0,8263
sobi13	2	0,7243
sobi15	2	0,6024
fastica7	3	0,7904
fastica8	2	0,8849
fastica9	2	0,7777
fastica11	4	0,7903
fastica15	4	0,6980

komponenta fastica11 neměla násobné korelace s mapami metody Infomax, s mapami metody SOBI korelovala čtyřnásobně.

Celkem 20 výkonových spekter nezávislých komponent metod SOBI a FastICA násobně korelovalo s výkonovými spektry IC opačné metody. Spektrum sobi16 čtyřnásobně korelovalo se spektrem IC metody FastICA a trojnásobně se spektrem IC metody Infomax. V obou případech byl průměrný korelační koeficient v absolutní hodnotě větší než 0,8. IC fastica2 a fastica3 násobně korelovaly s IC metod Infomax v časové oblasti, topografických mapách a výkonovém spektru, s IC metody SOBI pouze ve výkonovém spektru. IC sobi8 násobně korelovala v časové oblasti, v topografických mapách a ve výkonovém spektru s IC metody FastICA, v časových řadách a topografických mapách s IC metody Infomax.

Celkově pouze 2 páry nezávislých komponent korelovaly v časové řadě, topografických mapách i ve výkonovém spektru, byly to páry sobi4 - fastica5 a sobi5 - fastica4 ve všech 3 oblastech zůstaly nespárované komponenty sobi15, fastica6 a fastica12.

Tabulka 5.15: Vícenásobné korelace výkonových spekter nezávislých komponent fáze NONREM 2 mezi metodami SOBI a FastICA.

Vícenásobně korelovaná výkonová spektra IC	Počet korelací	Průměrný Pearsonův korelační koeficient
sobi1	2	0,8375
sobi2	2	0,8390
sobi5	3	0,9416
sobi7	2	0,8409
sobi8	2	0,9630
sobi10	2	0,8883
sobi11	2	0,8640
sobi15	3	0,7737
sobi16	4	0,9138
fastica2	2	0,9651
fastica3	2	0,6983
fastica4	2	0,9559
fastica8	2	0,8472
fastica9	2	0,8571
fastica11	3	0,8942
fastica14	2	0,7731
fastica15	2	0,8180
fastica16	2	0,6581
fastica17	2	0,9492
fastica19	2	0,7833

Celkové výsledky korelace všech tří metod

Celkové výsledky korelace nezávislých komponent mezi všemi třemi metodami jsou uvedeny v tabulce 5.16. V následujících odstavcích popíšu nezávislé komponenty z hlediska klinického významu. Pokud nezávislá komponent obsahuje více druhů aktivity, budou vyjmenovány ty dominantní.

V časových řadách korelovaly IC infomax1–fastica14–sobi3, které představují srdeční artefakt, z časové řady dobře rozpoznatelný, viz obrázky 5.10, 5.16 a 5.20. Dále korelovaly IC infomax5–fastica9–sobi19, což jsou komponenty fyziologické delta aktivity zarušené očními pohyby. Komponenty infomax17–fastica4–sobi5 fyziologickou delta aktivitu zarušenou artefaktem ze špatné elektrody. Nespárovanou IC v časové oblasti metody Infomax byla IC číslo 18, ta představuje fyziologickou aktivitu rušenou srdečním artefaktem. V časové oblasti je hůře rozpoznatelný, lépe je vidět v topografických mapách, viz obrázek 5.13. Nespárované komponenty metody FastICA byly 3. Číslo 7 představuje fyziologickou aktivitu zarušenou artefaktem, číslo 18 a číslo 19 je fyziologická aktivita. Nespárovaná nezávislá komponenty sobi14 sobi 17 a sobi18 je fyziologická aktivita, sobi15 je artefakt ze špatné elektrody.

V oblasti topografických map korelovaly IC infomax3–fastica8–sobi1, to jsou artefakty z pohybu očí. IC infomax17–fastica4–sobi5 je fyziologická delta aktivita zarušená artefaktem ze špatné elektrody. Korelované IC infomax19–fastica13–sobi6 představují fyziologickou aktivitu zarušenou artefaktem a očními pohyby. Nespárovaná topografická mapa metody Infomax číslo 4 a číslo 14 je fyziologická aktivita, číslo 10 je fyziologická aktivita rušená pulzním artefaktem. IC fastica16 a fastica 17 je fyziologická aktivita rušená artefaktem ze špatné elektrody. Nespárovaná nezávislá komponenta metody SOBI číslo 3 je fyziologická aktivita rušená srdečním artefaktem, číslo 7 je fyziologická aktivita rušená artefaktem ze špatné elektrody, číslo 11 a číslo 19 je fyziologická aktivita.

V oblasti výkonového spektra korelovaly IC infomax15–fastica1–sobi2 představující oční artefakt a infomax19–fastica2–sobi16 představující fyziologickou aktivitu rušenou artefaktem a očními pohyby. Nespárované výkonové spektrum fastica6 je fyziologická aktivita rušená artefakty, fastica10 představuje artefakt, fastica12 je fyziologická aktivita.

Nezávislá komponenta metody Infomax číslo 2 je artefakt ze špatné elektrody, číslo 5 je fyziologická aktivita rušená očními pohyby, číslo 7 je fyziologická aktivita, číslo 8 představuje šumovou komponentu, číslo 9 a číslo 16 je fyziologická aktivita rušená svalovým artefaktem, číslo 11 a číslo 19 je fyziologická aktivita, číslo 12 fyziologická alfa aktivita.

Tabulka 5.16: Celkové výsledky korelace nezávislých komponent fáze NONREM 2 mezi všemi třemi metodami.

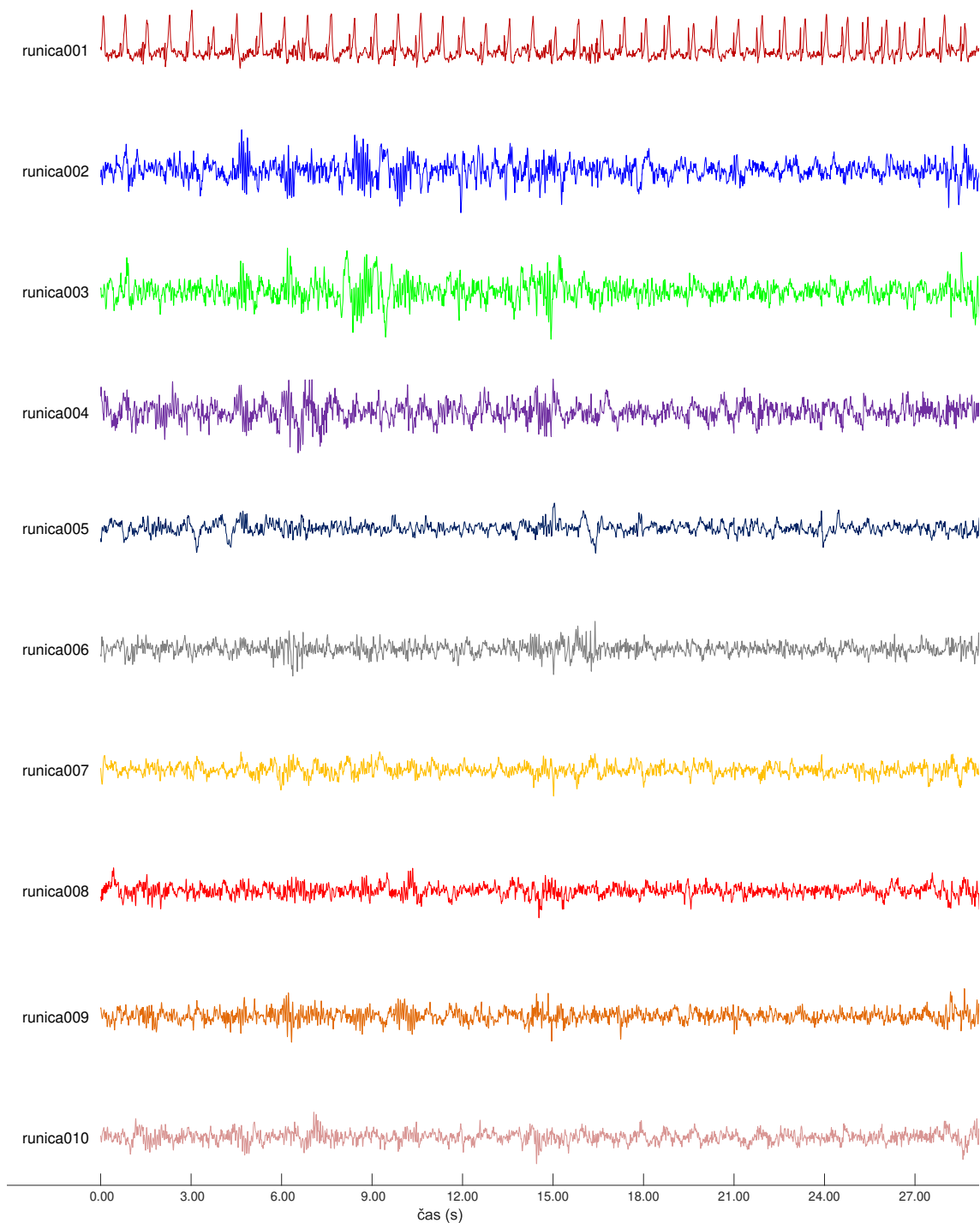
	Totožné IC	Nespárované IC
Časové řady IC	infomax1 - fastica14 - sobi3	Infomax: 18
	infomax5 - fastica9 - sobi19	FastICA: 7, 18, 19
	infomax17 - fastica4 - sobi5	SOBI: 14, 15, 17, 18
Topografické mapy IC	infomax3 - fastica8 - sobi1	Infomax: 4, 10, 14
	infomax17 - fastica4 - sobi5	FastICA: 16, 17
	infomax19 - fastica13 - sobi6	SOBI: 3, 7, 11, 15, 19
Výkonová spektra IC	infomax15 - fastica1 - sobi2	Infomax: 1
	infomax17 - fastica4 - sobi5	FastICA: 6, 7, 10, 12, 13
	infomax19 - fastica2 - sobi16	SOBI: 6, 11, 15

IC varianty Fastica číslo 3 a číslo 11 by mohly představovat fyziologickou aktivitu, číslo 5 fyziologickou aktivitu rušenou srdečním artefaktem, číslo 15 fyziologickou aktivitu rušenou očními pohyby.

Dle časových a spektrálních charakteristik předpokládám, že IC sobi4 představuje fyziologickou aktivitu rušenou srdečním artefaktem. IC metody SOBI číslo 8 představuje oční artefakty. Na základě charakteru IC číslo 9, číslo 12 a číslo 13 lze usuzovat, že se jedná o fyziologickou aktivitu, IC číslo 10 by mohla představovat artefakt.

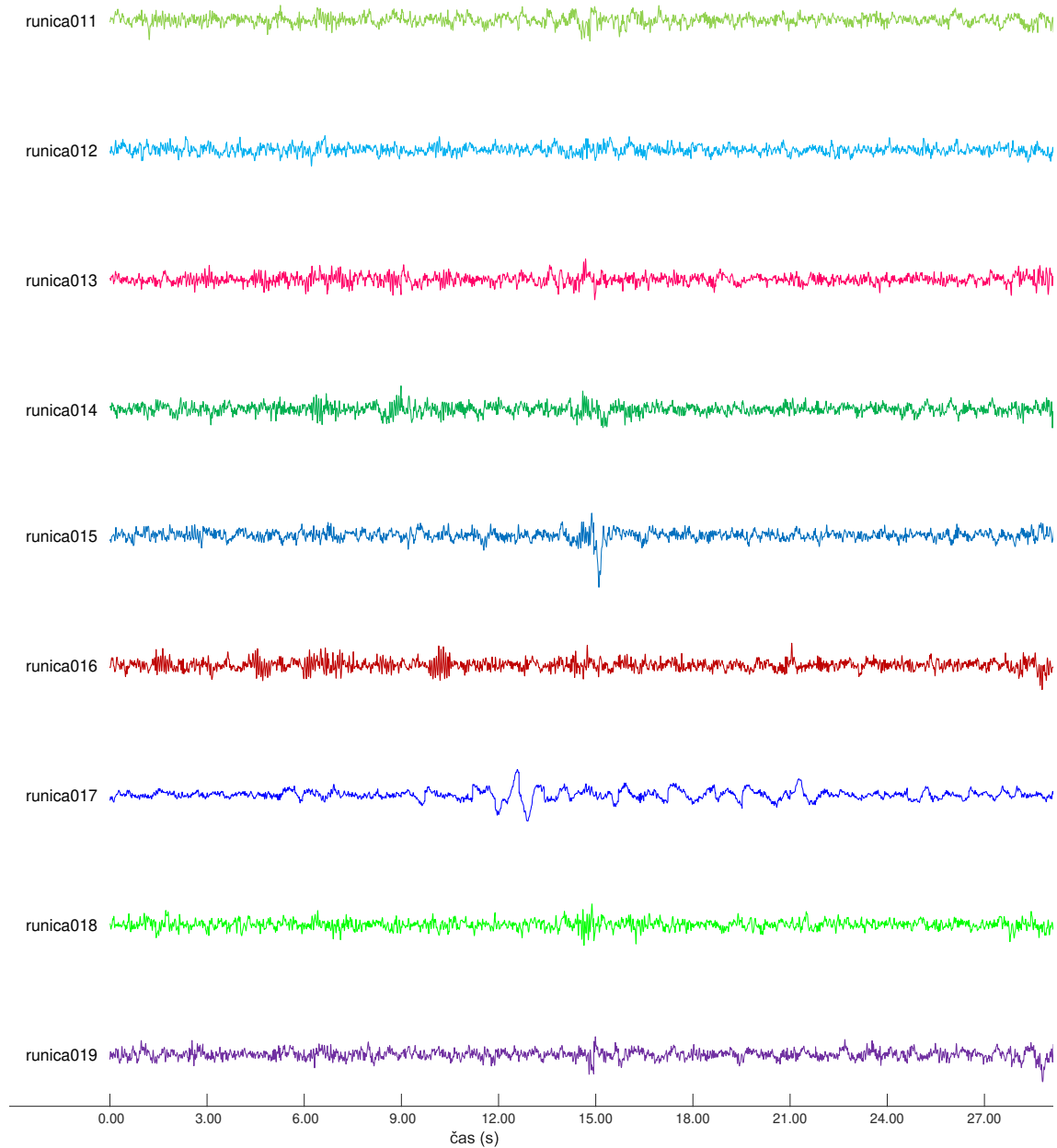
Páry nezávislých komponent, které korelovaly v časové řadě, topografické mapě i výkonovém spektru byly infomax15–fastica1, které tvořily oční artefakty. Dále infomax17–fastica4, infomax17–sobi5 a sobi5–fastica4, což je fyziologická delta aktivita rušená artefaktem ze špatné elektrody. Dle časových a spektrálních charakteristik usuzuji, že u páru sobi4–fastica5 se jedná o fyziologickou aktivitu rušenou srdečním artefaktem.

Nezávislé komponenty fáze NONREM 2 vypočítané metodou Infomax



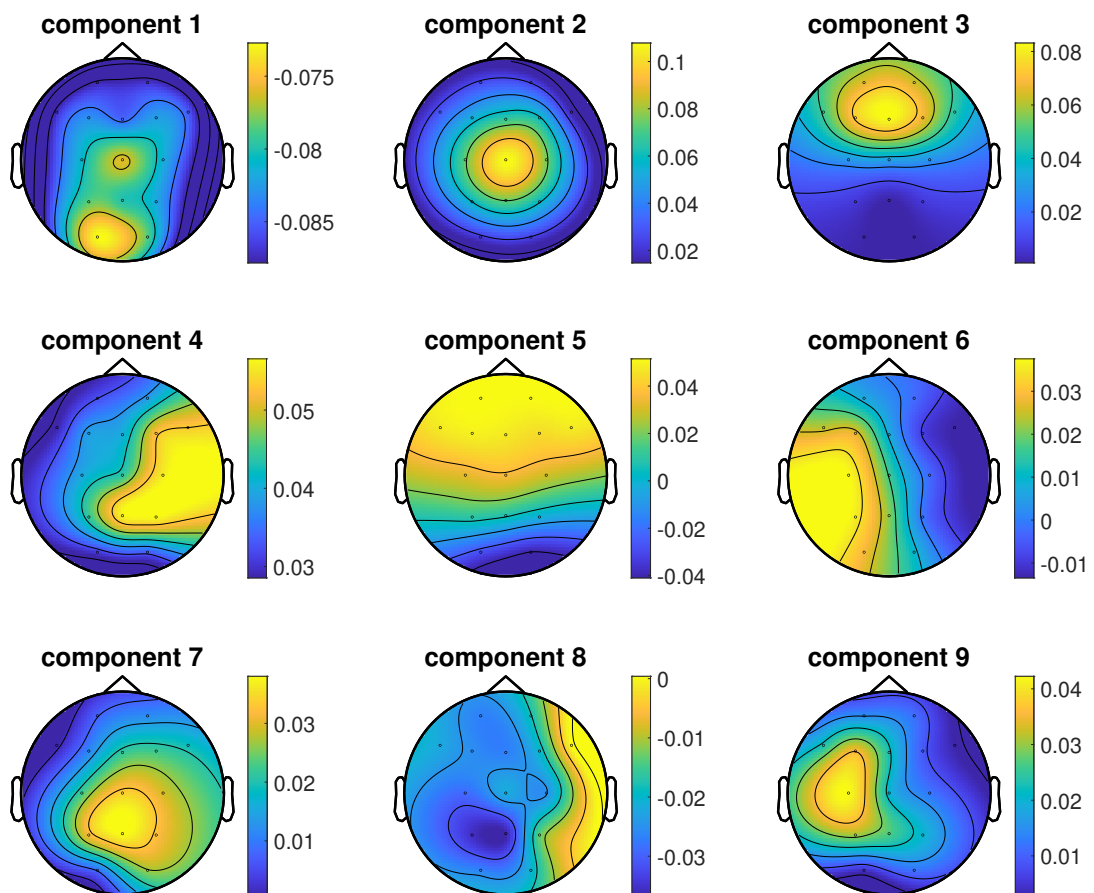
Obrázek 5.10: Časové řady nezávislých komponent číslo 1 až 10 fáze NONREM 2 vypočítané metodou Infomax, runica = infomax.

Nezávislé komponenty fáze NONREM 2 vypočítané metodou Infomax



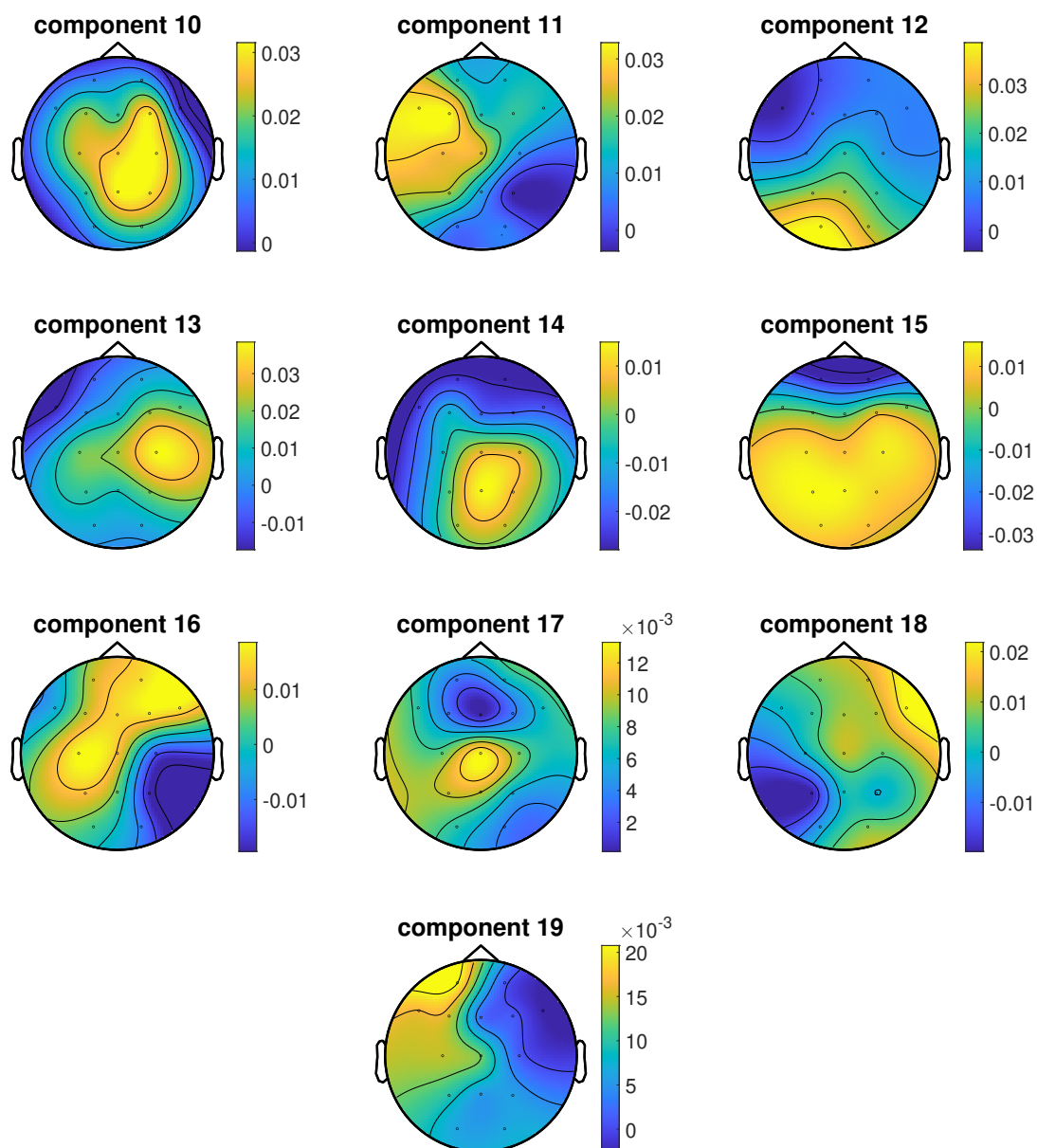
Obrázek 5.11: Časové řady nezávislých komponent číslo 11 až 19 fáze NONREM 2 vypočítané metodou Infomax, runica = infomax.

Topografické mapy IC fáze NONREM 2 vypočítané metodou Infomax



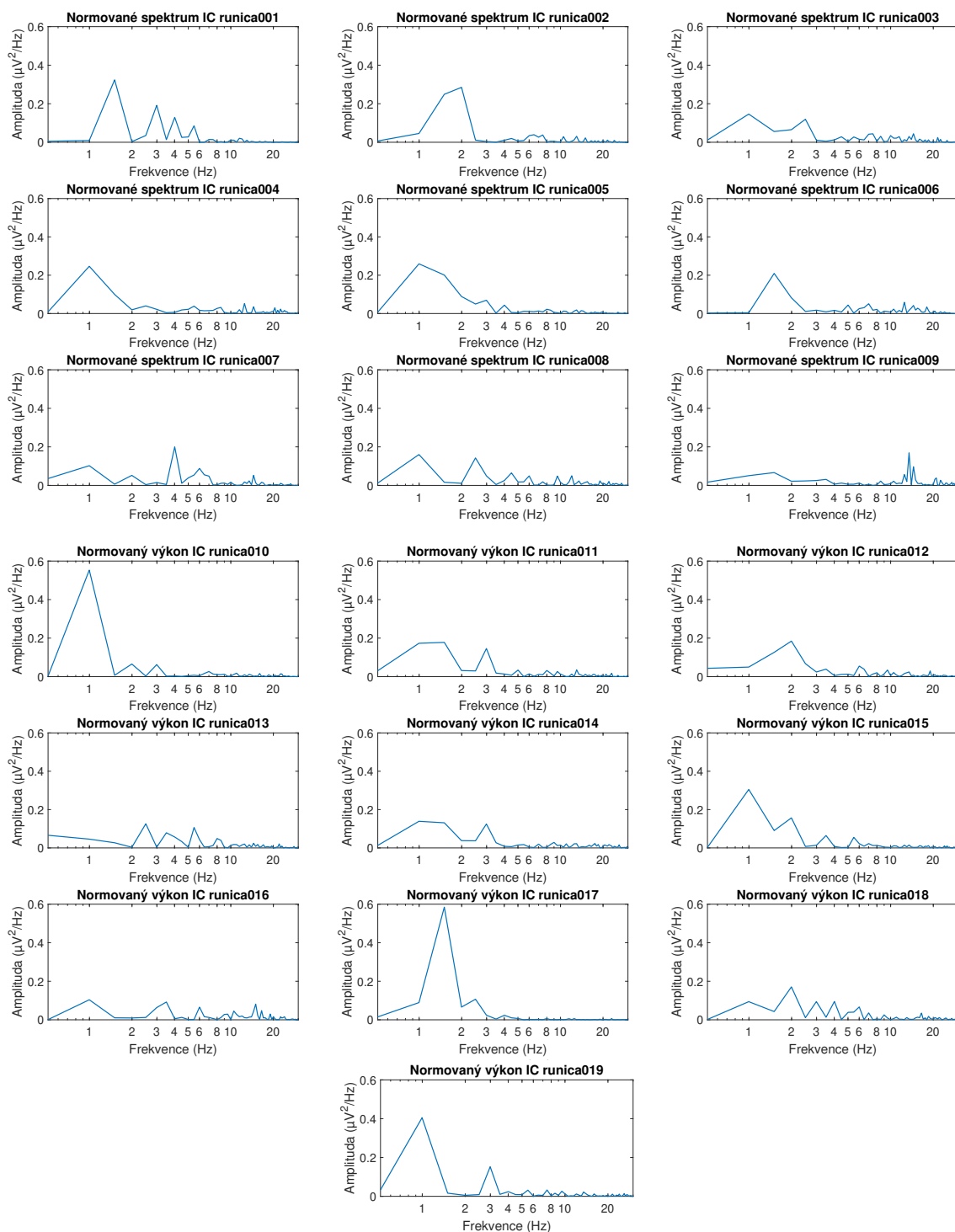
Obrázek 5.12: Topografické mapy nezávislých komponent číslo 1 až 9 fáze NONREM 2 vypočítané metodou Infomax. Barevná škála představuje napětí v μV , component = komponenta.

Topografické mapy IC fáze NONREM 2 vypočítané metodou Infomax



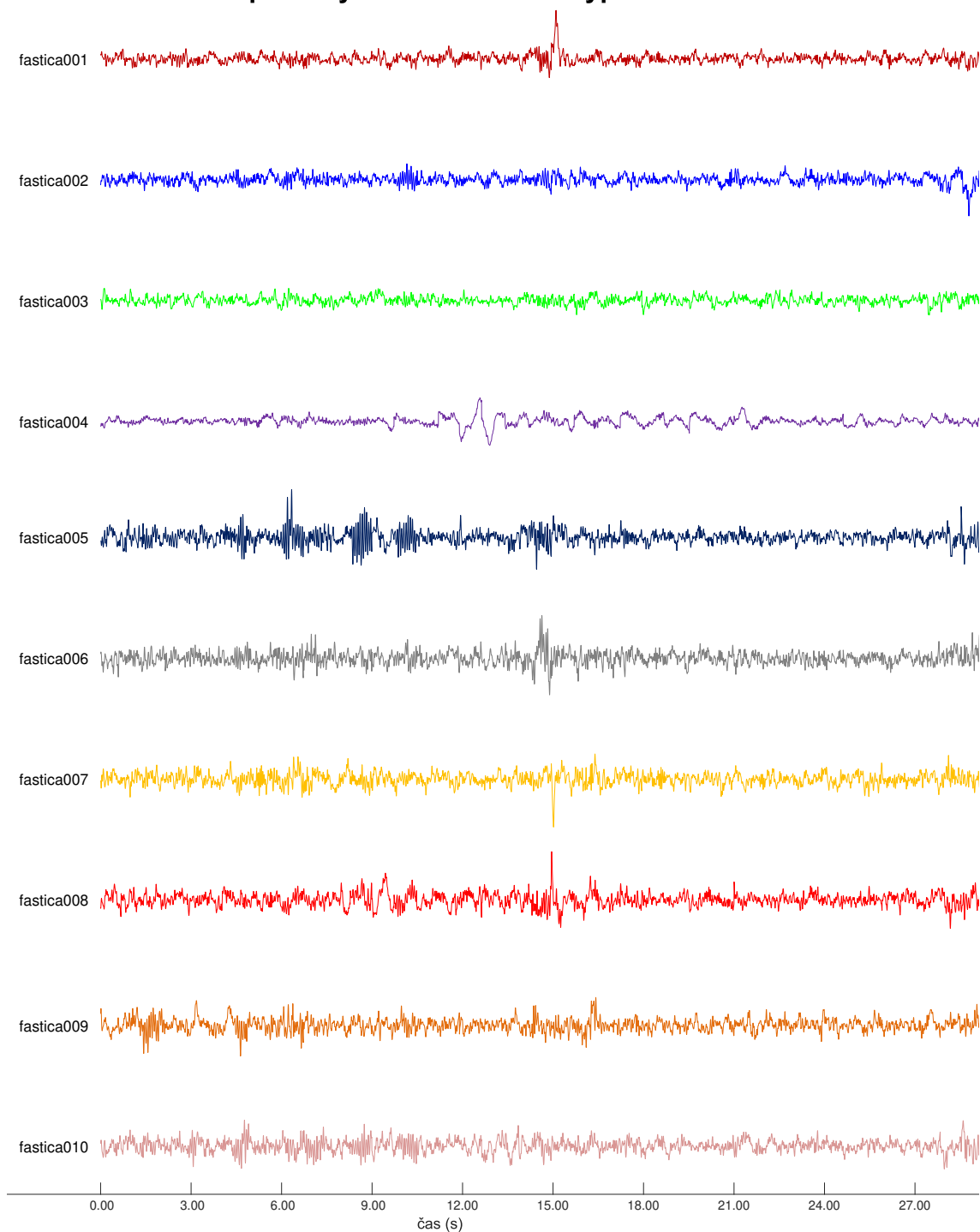
Obrázek 5.13: Topografické mapy nezávislých komponent číslo 10 až 19 fáze NONREM 2 vypočítané metodou Infomax. Barevná škála představuje napětí v μV , component = komponenta.

Normovaná výkonová spektra IC fáze NONREM 2 vypočítaných metodou Infomax



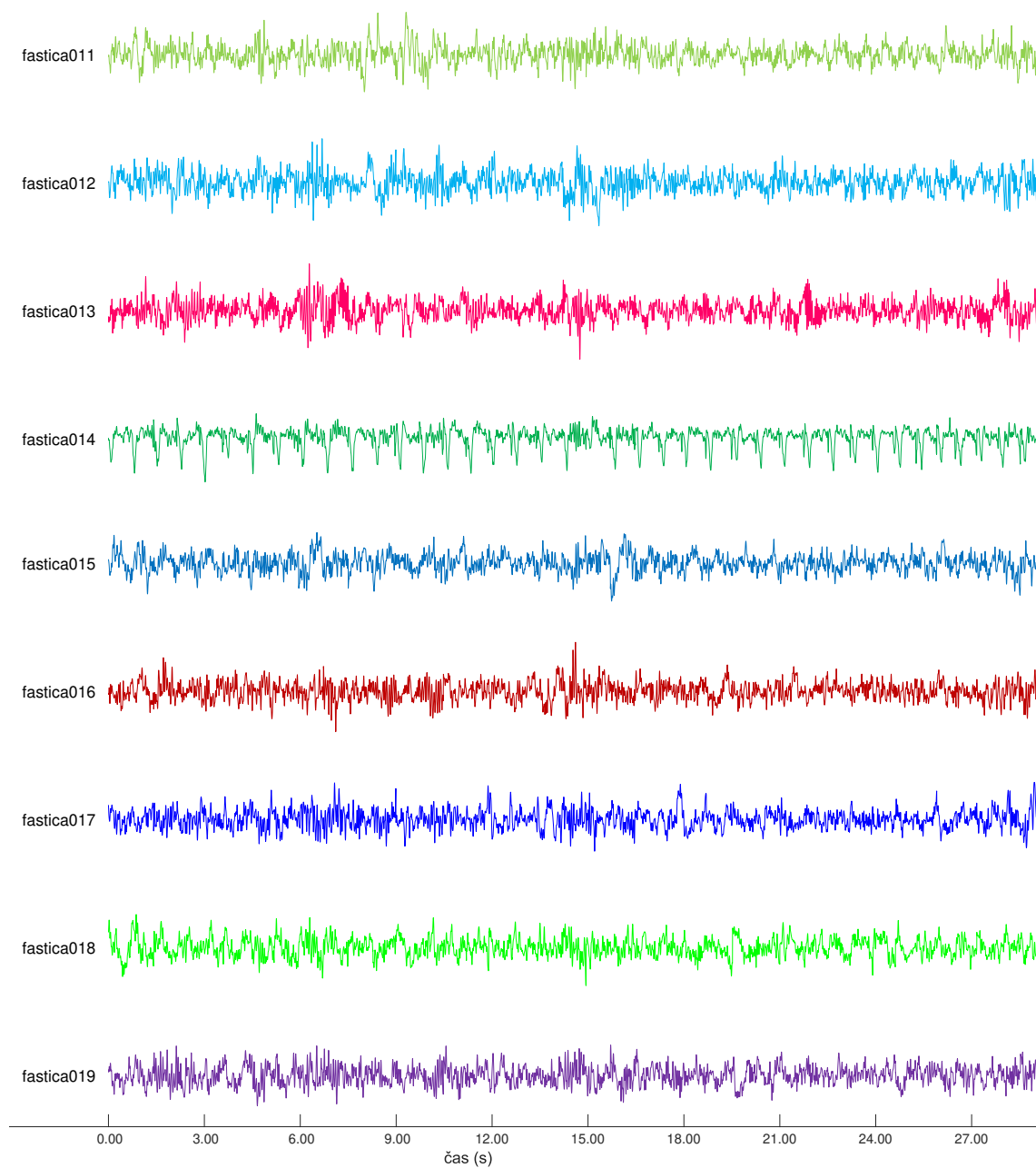
Obrázek 5.14: Normovaná výkonová spektra nezávislých komponent fáze NONREM 2 vypočítaných metodou Infomax. runica = infomax.

Nezávislé komponenty fáze NONREM 2 vypočítané metodou FastICA



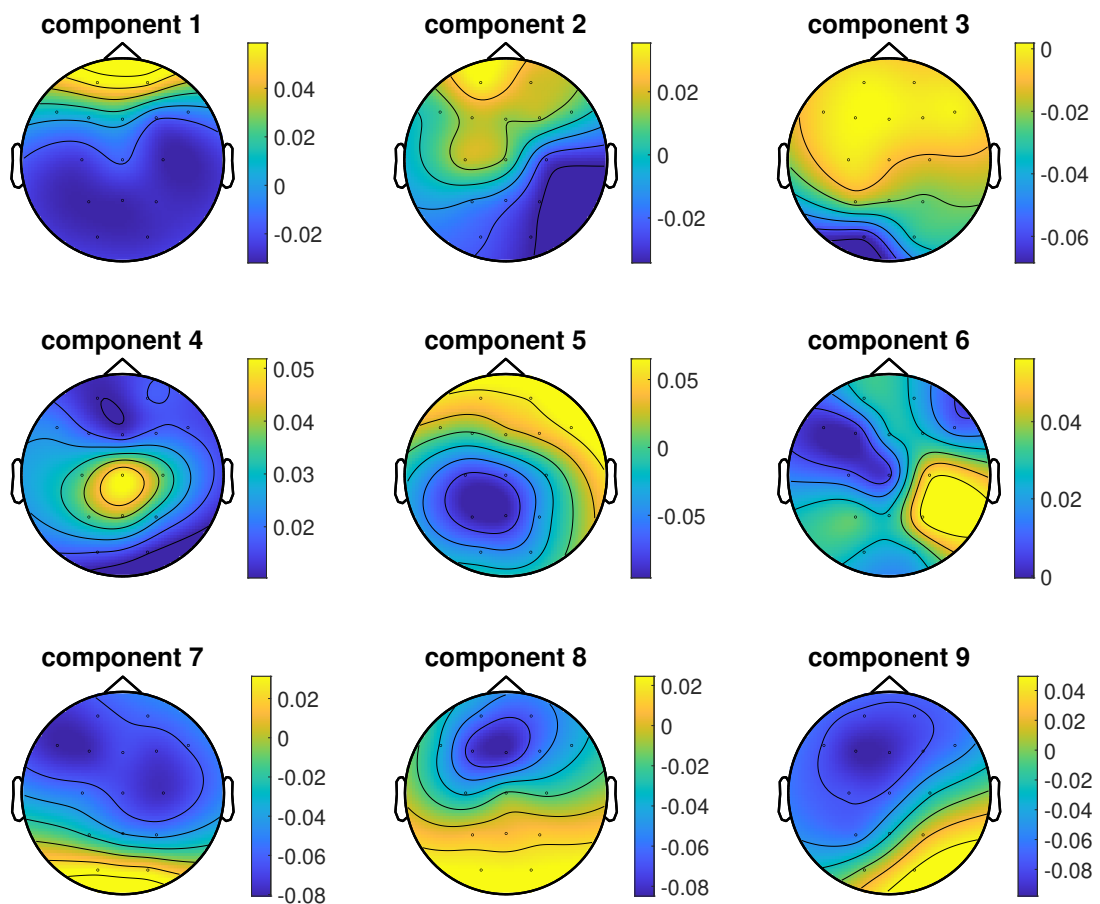
Obrázek 5.15: Časové řady nezávislých komponent číslo 1 až 10 fáze NONREM 2 vypočítané metodou FastICA.

Nezávislé komponenty fáze NONREM 2 vypočítané metodou FastICA



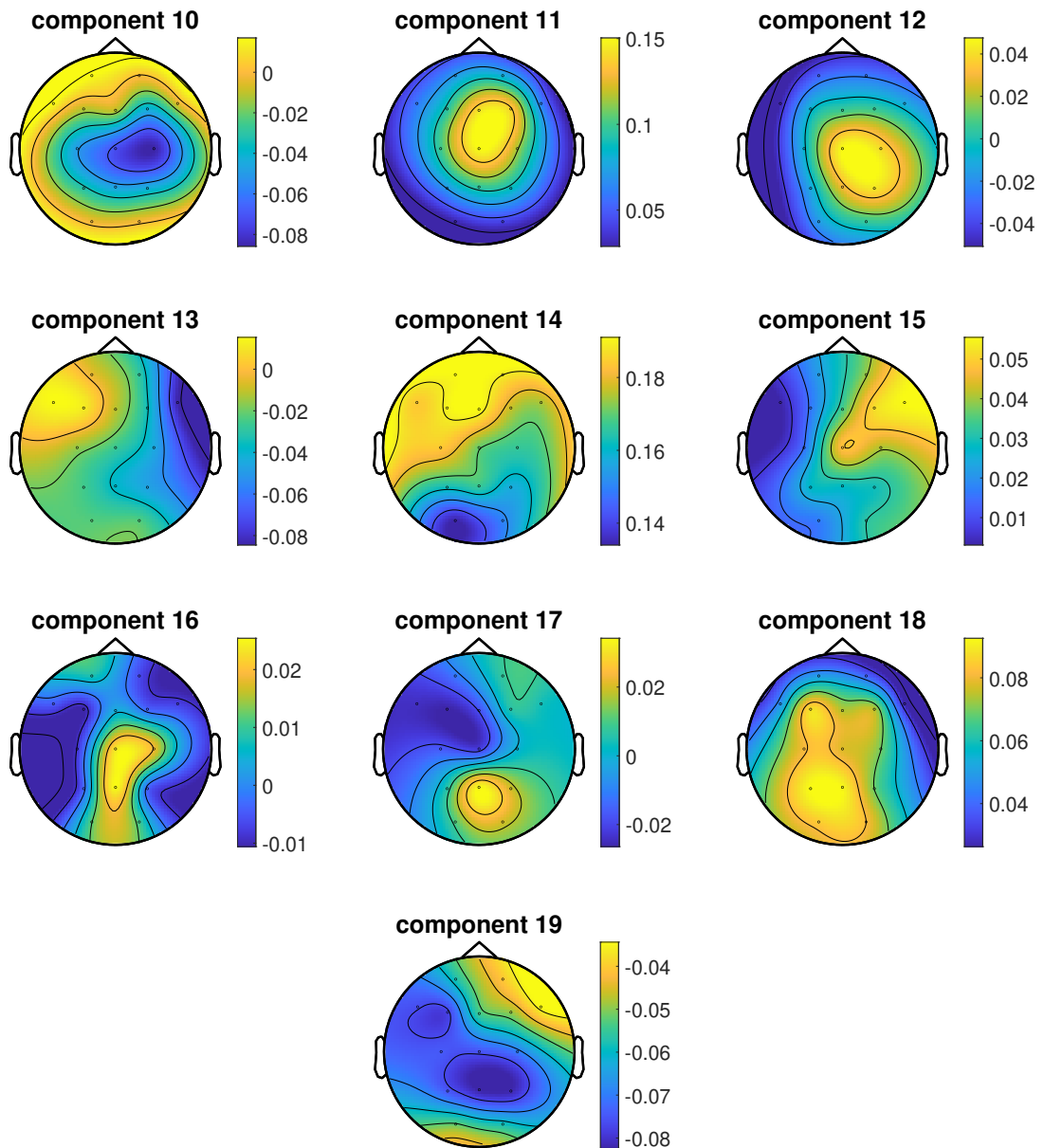
Obrázek 5.16: Časové řady nezávislých komponent číslo 11 až 19 fáze NONREM 2 vypočítané metodou FastICA.

Topografické mapy IC fáze NONREM 2 vypočítané metodou FastICA



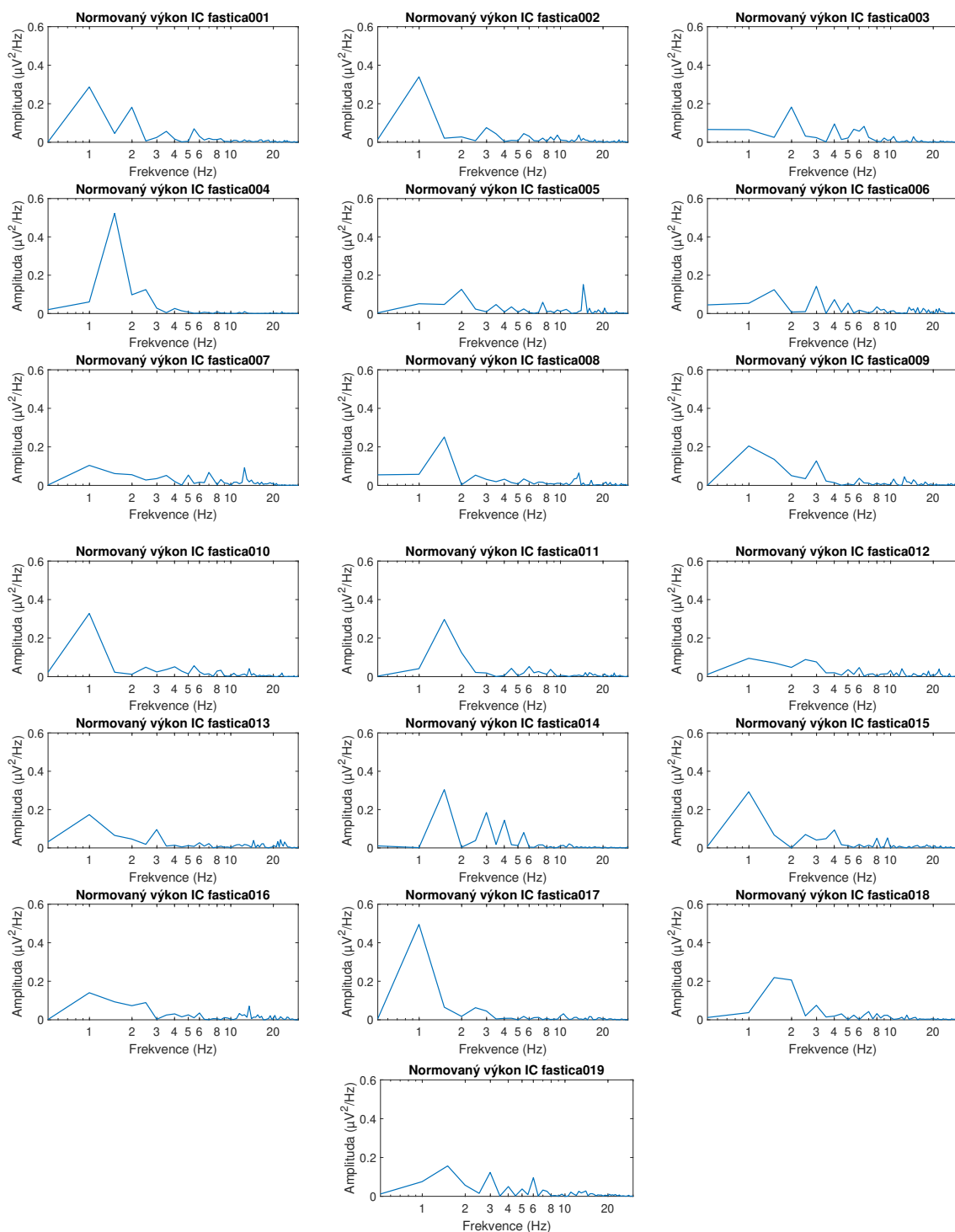
Obrázek 5.17: Topografické mapy nezávislých komponent číslo 1 až 9 fáze NONREM 2 vypočítané metodou FastICA. Barevná škála představuje napětí v μV , component = komponenta.

Topografické mapy IC fáze NONREM 2 vypočítané metodou FastICA



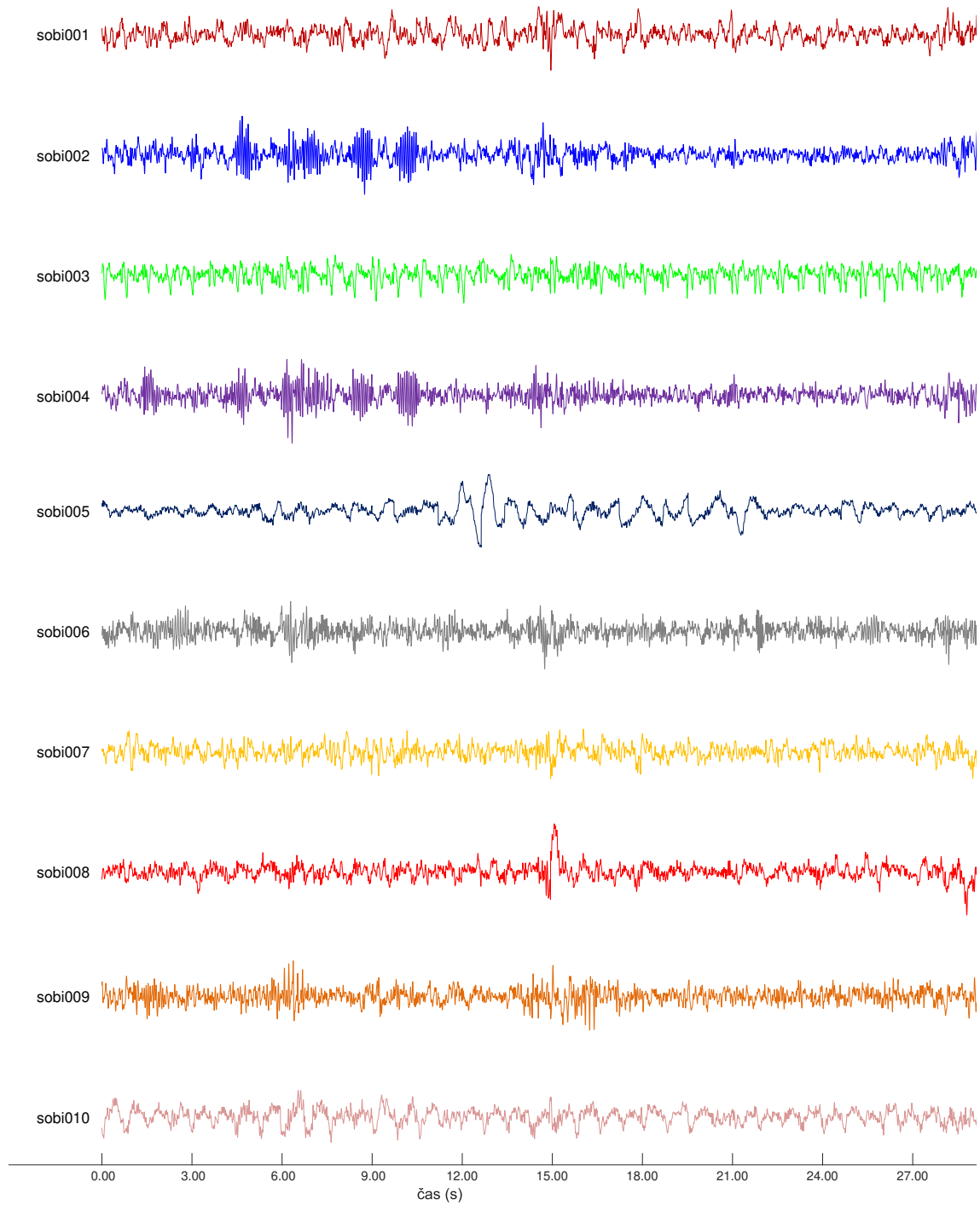
Obrázek 5.18: Topografické mapy nezávislých komponent číslo 10 až 19 fáze NONREM 2 vypočítané metodou FastICA. Barevná škála představuje napětí v μV , component = komponenta.

Normovaná výkonová spektra IC fáze NONREM 2 vypočítaných metodou FastICA



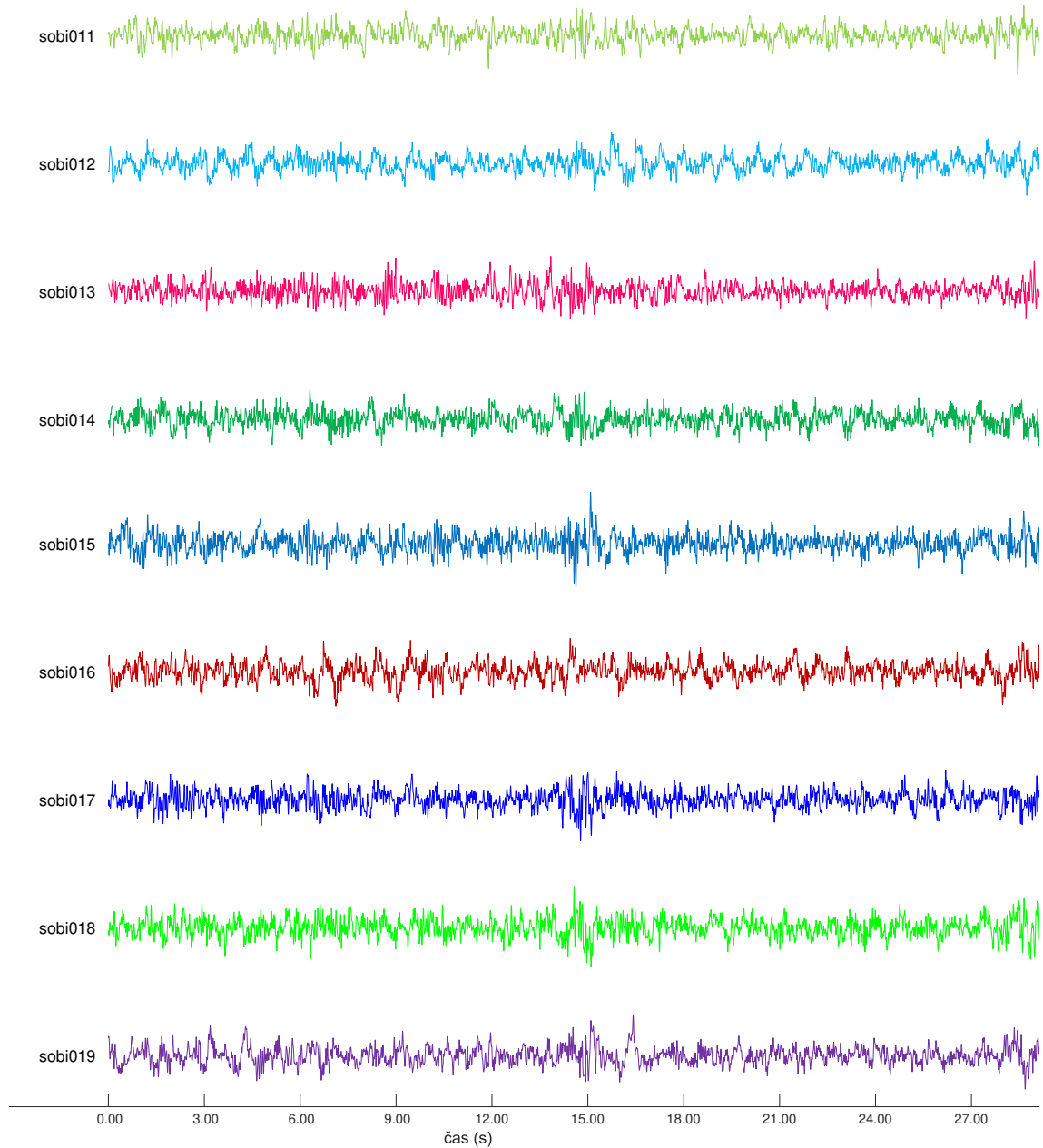
Obrázek 5.19: Normovaná výkonová spektra nezávislých komponent fáze NONREM 2 vypočítaných metodou FastICA.

Nezávislé komponenty fáze NONREM 2 vypočítané metodou SOBI



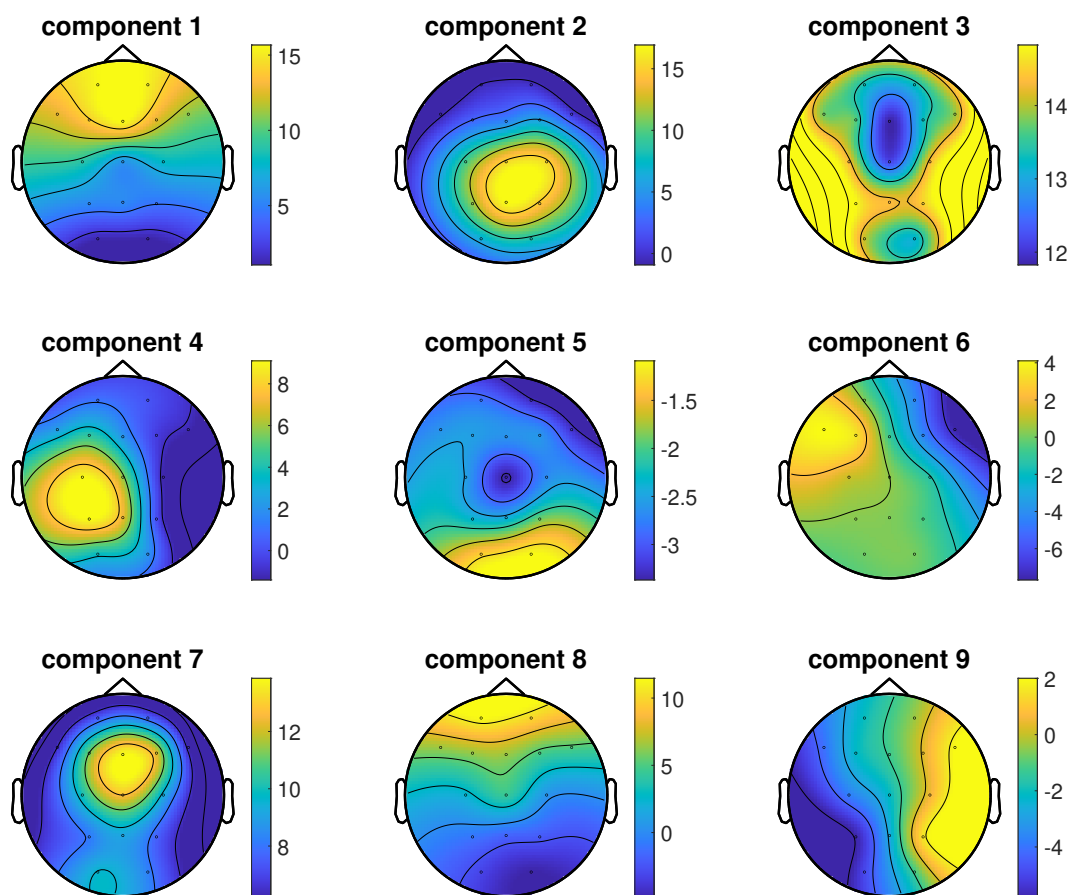
Obrázek 5.20: Časové řady nezávislých komponent číslo 1 až 10 fáze NONREM 2 vypočítané metodou SOBI.

Nezávislé komponenty fáze NONREM 2 vypočítané metodou SOBI



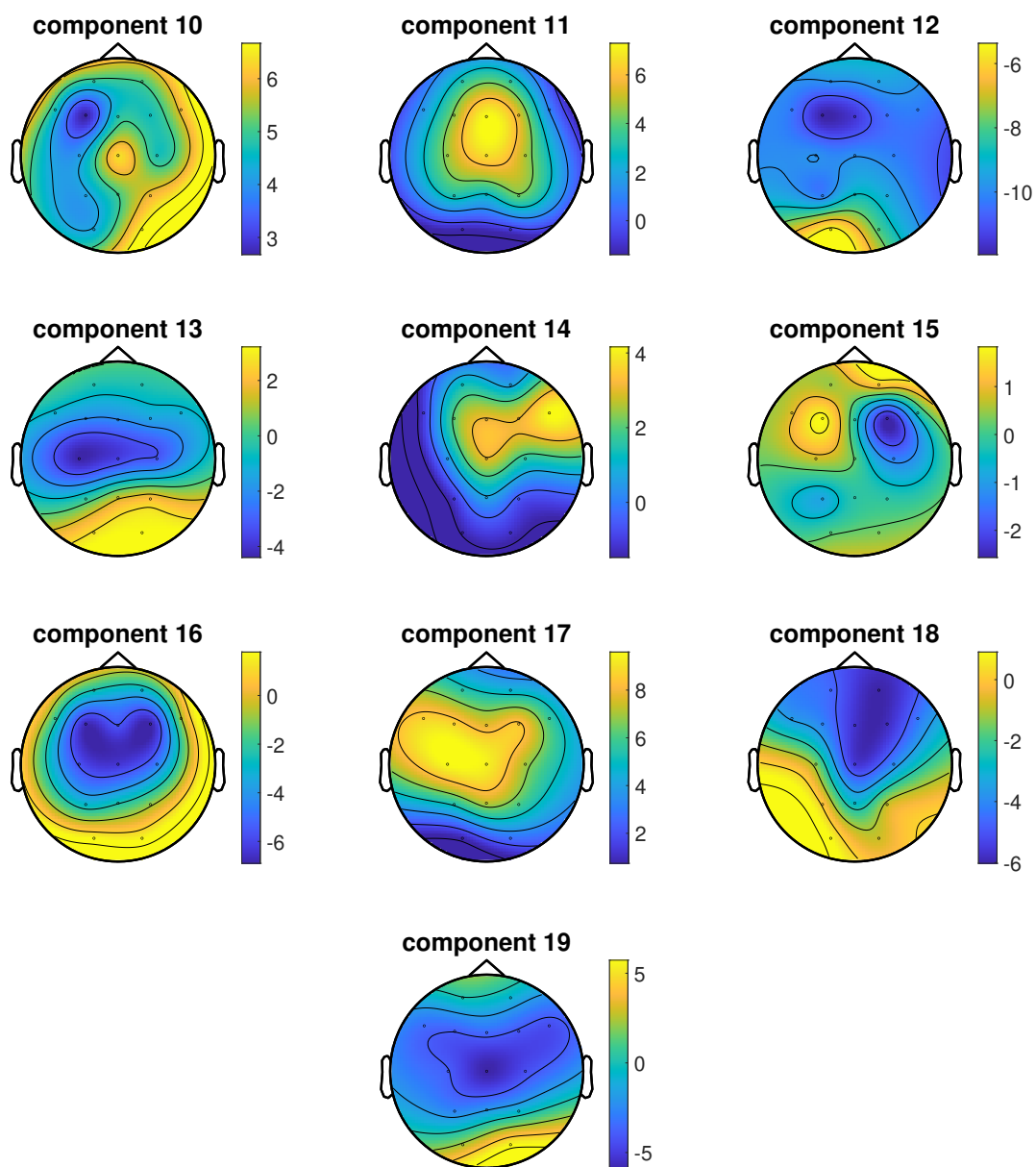
Obrázek 5.21: Časové řady nezávislých komponent číslo 11 až 19 fáze NONREM 2 vypočítané metodou SOBI.

Topografické mapy IC fáze NONREM 2 vypočítané metodou SOBI



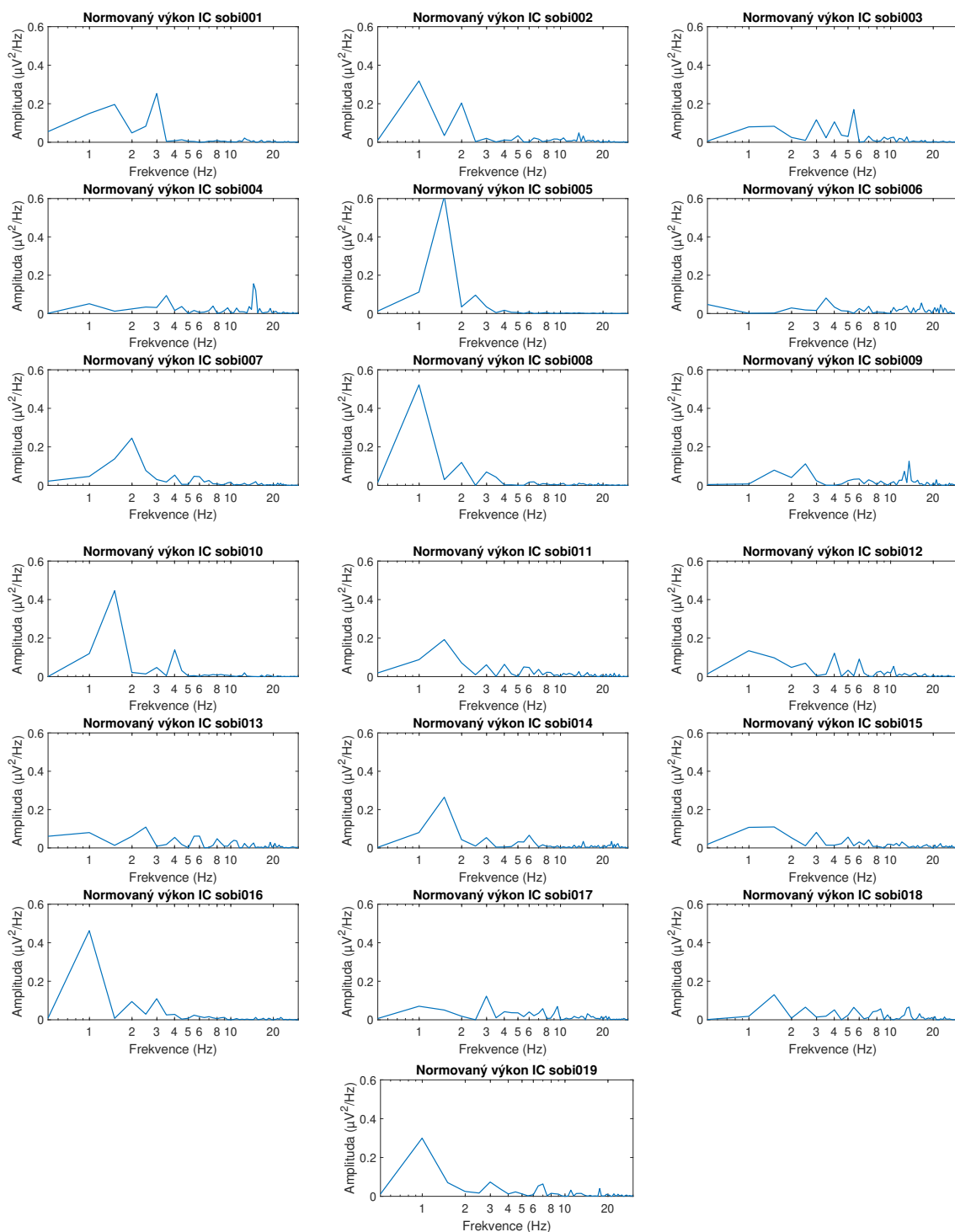
Obrázek 5.22: Topografické mapy nezávislých komponent číslo 1 až 9 fáze NONREM 2 vypočítané metodou SOBI. Barevná škála představuje napětí v μV , component = komponenta.

Topografické mapy IC fáze NONREM 2 vypočítané metodou SOBI



Obrázek 5.23: Topografické mapy nezávislých komponent číslo 10 až 19 fáze NONREM 2 vypočítané metodou SOBI. Barevná škála představuje napětí v μV , component = komponenta.

Normovaná výkonová spektra IC fáze NONREM 2 vypočítaných metodou SOBI



Obrázek 5.24: Normovaná výkonová spektra nezávislých komponent fáze NONREM 2 vypočítaných metodou SOBI.

Tabulka 5.17: Počet extrahovaných nezávislých komponent metodou FastICA u jednoho záznamu

Minimum	9
Maximum	19
Průměr	18,1
Směrodatná odchylka	1,5
Rozptyl	2,3
Medián	19,0
Modus	19,0

5.2.3 Analýza nezávislých komponent všech záznamů

Výše uvedená podkapitola představovala výsledky analýzy nezávislých komponent na jednom ukázkovém záznamu. V této podkapitole vyhodnotím výsledky ICA pro všechny záznamy z datasetu. Jak bylo uvedeno v předešlé kapitole, k porovnání nezávislých komponent mezi různými metodami jsem použila Pearsonův korelační koeficient. Porovnávala jsem mezi sebou časové řady, topografické mapy a výkonová spektra IC vypočítaných metodami Infomax, FastICA a SOBI. Kompletní tabulky s výsledky těchto srovnání jsou kvůli své velikosti uvedeny v přílohách v elektronické verzi této práce.

Jak bylo uvedeno v kapitole 4.1, měla jsem k dispozici 213 patientských záznamů, které celkově tvořili 4424 jednominutových segmentů. U každého segmentu mám informace o počtu IC, jejichž korelační koeficient je v absolutní hodnotě větší než 0.5. Porovnávala jsem metody Infomax–FastICA, Infomax–SOBI a SOBI–FastICA, tedy veškeré kombinace mezi testovanými metodami. Při porovnání metod mohlo vzniknout nejvýše 19 párů, tedy 38 korelovaných nezávislých komponent. A dále u každého segmentu vím, kolik zůstalo nespárovaných IC.

Maximální počet vypočítaných nezávislých komponent mohl být 19, protože dataset tvořil záznam z 19 EEG elektrod. Metoda FastICA ne vždy rozložila signál na 19 nezávislých komponent, jejich počet byl variabilní. Statistické zhodnocení počtu výsledných komponent získaných metodou FastICA je uveden v tabulce 5.17. Celkový počet segmentů, kde FastICA extrahovala 19 IC byl 3138, což tvoří 70,9 % všech segmentů.

Testovala jsem na hladině významnosti 5 % normalitu množství extrahovaných nezávislých komponent metodou FastICA. Dále jsem testovala, zda se počty extrahovaných IC metodou FastICA statisticky významně liší od množství extrahovaných IC zbylými metodami. Výsledky jsou uvedeny v tabulce 5.18. Protože data neměly normální rozdělení použila jsem pro testování shodnosti neparametrický dvouvýběrový Wilcoxonův test. Na hladině významnosti 5 % zamítám nulovou hy-

Tabulka 5.18: Výsledky Shapiro–Wilkova testu a dvouvýběrového Wilcoxonova testu pro množství extrahovaných nezávislých komponent metodou FastICA

Testy	p-hodnota	Výsledek
Shapiro–Wilkův test	$\ll 0,05$	Zamítám H_0
Dvouvýběrový Wilcoxonův test	$\ll 0,05$	Zamítám H_0

potézu, že oba výběry pocházejí z rozdělení se stejnou distribuční funkcí. Množství nezávislých komponent extrahovaných metodou FastICA se liší od počtů IC extrahovaných metodami Infomax a SOBI.

Časové řady nezávislých komponent

Výsledky korelace časových řad mezi všemi metodami jsou uvedeny v tabulce 5.19. Medián i modus metod Infomax – FastICA je 32 IC, tedy 16 párů. Tento počet korelovaných IC mělo 965 segmentů, což je 21,8 % všech segmentů. U metod Infomax – SOBI mělo medián i modus 903 korelovaných segmentů, to je 20,4 %. Metody SOBI – FastICA měly medián a modus rovný 26. Počet segmentů obsahujících 24 korelovaných časových řad IC byl 772 (17,5 %).

Tabulka 5.19: Počet korelovaných časových řad nezávislých komponent všech metod

Statistika	Infomax – FastICA	Infomax – SOBI	SOBI – FastICA
Minimum	14,0	10,0	4,0
Maximum	33,0	38,0	38,0
Průměr	31,1	27,1	24,8
Směrodatná odchylka	3,7	4,3	4,6
Rozptyl	13,8	18,1	21,5
Medián	32,0	28,0	26,0
Modus	32,0	28,0	26,0

Kromě informací o počtu korelovaných časových řadách IC jsem měla informace i o počtech nespárovaných časových řad IC, tedy takových, které nekorelovaly s žádnou další IC druhé metody. Výsledky jsou uvedeny v tabulce 5.20. Modus IC metod Infomax – FastICA byl 6 IC a vyskytoval se u 1018 segmentů, což je 23,0 % všech segmentů. Medián i modus u metod Infomax – SOBI byl taktéž shodný (10 IC) a vyskytoval se u 903 segmentů, tedy u 20,4 % všech segmentů. U metod SOBI – FastICA a byl medián i modus počtu nespárovaných IC 12 a vyskytoval se u 735 segmentů (16,6 %).

Tabulka 5.20: Počet nespárovaných časových řad nezávislých komponent všech metod

Statistika	Infomax – FastICA	Infomax – SOBI	SOBI – FastICA
Minimum	0,0	0,0	0,0
Maximum	18,0	28,0	31,0
Průměr	6,0	10,1	12,3
Směrodatná odchylka	3,0	4,3	4,3
Rozptyl	9,2	18,1	18,2
Medián	6,0	10,0	12,0
Modus	6,0	10,0	12,0

Normalitu dat jsem ověřila Shapiro–Wilkovým testem, jehož nulová hypotéza zní: Náhodný výběr má normální rozdělení s nespécifikovaným průměrem a rozptylem. Hladinu významnosti α jsem stanovila 5 %. Výsledky jsou uvedeny v tabulce 5.21. U všech metod jsem zamítla nulovou hypotézu na hladině významnosti 5 %. K ověření, zda jsou všechny tři metody stejné, použiji neparametrický test, konkrétně dvouvýběrový Wilcoxonův test.

Tabulka 5.21: Výsledky Shapiro–Wilkova testu korelovaných a nespárovaných časových řad nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Korelované IC Infomax – FastICA	$\ll 0,05$	Zamítám H_0
Korelované IC Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Korelované IC SOBI – FastICA	$\ll 0,05$	Zamítám H_0
Nespárované IC Infomax – FastICA	$\ll 0,05$	Zamítám H_0
Nespárované IC Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Nespárované IC SOBI – FastICA	$\ll 0,05$	Zamítám H_0

Pomocí Wilcoxonova testu jsem mezi sebou porovnála korelované časové řady IC všech tří párů metod mezi sebou a taktéž i nespárované časové řady IC. Výsledky korelovaných IC jsou uvedeny v tabulce 5.22 a výsledky nespárovaných IC v tabulce 5.23. Nulová hypotéza tohoto testu zní, že distribuční funkce obou rozdělení jsou shodné. Hladinu významnosti jsem zvolila 5%. Ve všech případech jsem zamítla nulovou hypotézu na hladině významnosti 5 %. To znamená, že se jednotlivé metody ICA liší v počtu korelovaných i nespárovaných párů časových řad extrahovaných IC.

Topografické mapy nezávislých komponent

Výsledky korelace topografických map nezávislých komponent vypočítaných všemi metodami je uveden v tabulce 5.24. Nejčastěji zastoupenou hodnotou ko-

Tabulka 5.22: Výsledky Wilcoxonova testu korelovaných časových řad nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Infomax – FastICA a Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Infomax – FastICA a SOBI – FastICA	$\ll 0,05$	Zamítám H_0
Infomax – SOBI a SOBI – FastICA	$\ll 0,05$	Zamítám H_0

Tabulka 5.23: Výsledky Wilcoxonova testu nespárovaných časových řad nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Infomax – FastICA a Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Infomax – FastICA a SOBI – FastICA	$\ll 0,05$	Zamítám H_0
Infomax – SOBI a SOBI – FastICA	$\ll 0,05$	Zamítám H_0

relace topografických map IC metod Infomax – FastICA bylo 26 korelovaných IC. Tento počet obsahovalo 994 segmentů, což činí 22,5 % všech segmentů. Modus korelovaných topografických map metod Infomax – SOBI mělo 1089 komponent, tedy 24,6 % celkového množství. Nejčastěji zastoupenou hodnotu korelovaných topografických map metod SOBI – FastICA mělo 1096 segmentů (24,8 %).

Tabulka 5.24: Počet korelovaných topografických map nezávislých komponent všech metod

Statistika	Infomax – FastICA	Infomax – SOBI	SOBI – FastICA
Minimum	14,0	6,0	12,0
Maximum	38,0	34,0	34,0
Průměr	26,0	22,4	23,1
Směrodatná odchylka	3,4	3,4	3,1
Rozptyl	11,2	11,8	9,4
Medián	26,0	22,0	24,0
Modus	26,0	24,0	24,0

Nejčastějším množstvím nespárovaných topografických map nezávislých komponent metod Infomax–FastICA bylo 10 IC, jak je vidět v tabulce 5.25. Tuto hodnotu obsahovalo 974 segmentů (22,0 % všech segmentů). Nejčastější počet nespárovaných topografických map metod Infomax–SOBI mělo 1089 segmentů, tedy 24,6 % všech segmentů. Modus nespárovaných map metod SOBI–FastICA byl 14 IC a toto množství obsahovalo 1087 segmentů (24,6 %).

V tabulce 5.26 jsou vidět výsledky ověření normality množství korelovaných a nespárovaných topografických map IC všech segmentů všech metod. Hladina vý-

Tabulka 5.25: Počet nespárovaných topografických map nezávislých komponent všech metod

Statistika	Infomax – FastICA	Infomax – SOBI	SOBI – FastICA
Minimum	0,0	4,0	4,0
Maximum	22,0	32,0	24,0
Průměr	11,1	15,2	14,0
Směrodatná odchylka	3,1	3,4	2,9
Rozptyl	9,9	11,8	8,3
Medián	11,0	16,0	14,0
Modus	10,0	14,0	14,0

znamnosti α byla 5 %. U všech veličin jsem zamítla nulovou hypotézu na hladině významnosti 0,05.

Tabulka 5.26: Výsledky Shapiro–Wilkova testu korelovaných a nespárovaných topografických map nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Korelované IC Infomax – FastICA	$\ll 0,05$	Zamítám H_0
Korelované IC Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Korelované IC SOBI – FastICA	$\ll 0,05$	Zamítám H_0
Nespárované IC Infomax – FastICA	$\ll 0,05$	Zamítám H_0
Nespárované IC Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Nespárované IC SOBI – FastICA	$\ll 0,05$	Zamítám H_0

Výsledky Wilcoxonova testu korelovaných topografických map jsou uvedeny v tabulce 5.27. Protože p-hodnota byla ve všech případech $\ll 0,05$, zamítám hypotézu H_0 na hladině významnosti α , kterou jsem stanovila 5 %. To znamená, že se jednotlivé metody ICA liší v počtu korelovaných i nespárovaných párů topografických map extrahovaných IC.

Tabulka 5.27: Výsledky Wilcoxonova testu korelovaných topografických map nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Infomax – FastICA a Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Infomax – FastICA a SOBI – FastICA	$\ll 0,05$	Zamítám H_0
Infomax – SOBI a SOBI – FastICA	$\ll 0,05$	Zamítám H_0

Další skupinou dat na které jsem použila Wilcoxonův test byly nespárované topografické mapy IC všech metod. V tabulce 5.28 je vidět, že ve všech případech je

p-hodnota výrazně nižší než stanovená hladina významnosti, tudíž zamítám nulovou hypotézu na hladině významnosti 5 %.

Tabulka 5.28: Výsledky Wilcoxonova testu nespárovaných topografických map nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Infomax – FastICA a Infomax – SOBI	$\ll 0,05$	Zamítám H ₀
Infomax – FastICA a SOBI – FastICA	$\ll 0,05$	Zamítám H ₀
Infomax – SOBI a SOBI – FastICA	$\ll 0,05$	Zamítám H ₀

Výkonové spektrum nezávislých komponent

Posledním způsobem, kterým jsem porovnála podobnost nezávislých komponent extrahovaných různými metodami, bylo porovnání výkonového spektra IC. Výsledky jsou vedeny v tabulce 5.29. Nejčastější počet korelovaných výkonů IC metod Infomax – FastICA bylo 26 IC. Tento počet měl 1034 segmentů, tedy 23,4 % z celkového počtu. U metod Infomax – SOBI bylo nejvíce segmentů majících 24 korelovaných IC, celkem 1139 segmentů (25,7 %). Metody SOBI – FastICA měly modus roven 24 korelovanými výkony. Tento modus mělo 1096 segmentů, což je 24,8 % z celkového počtu 4424 segmentů.

Statistické zhodnocení počtů nespárovaných výkonových spekter nezávislých komponent je uvedeno v tabulce 5.30. Počet nespárovaných výkonových spekter roven 12 mělo u metod Infomax – FastICA 987 segmentů (22,3 %). 14 nespárovaných výkonových spekter metod Infomax – SOBI mělo 1139 segmentů (25,7 %) a metod SOBI – FastICA 1068 segmentů (24,1 %).

Na základě výsledků Shapiro-Wilkova testu, které jsou vidět v tabulce 5.31, u všech výběrů zamítám hypotézu H₀, že výběry pochází z normálního rozdělení,

Tabulka 5.29: Počet korelovaných výkonových spekter nezávislých komponent všech metod

Statistika	Infomax – FastICA	Infomax – SOBI	SOBI – FastICA
Minimum	12,0	8,0	10,0
Maximum	36,0	34,0	34,0
Průměr	25,5	23,8	23,2
Směrodatná odchylka	3,4	3,1	3,3
Rozptyl	11,6	9,8	10,7
Medián	26,0	24,0	24,0
Modus	26,0	24,0	24,0

Tabulka 5.30: Počet nespárovaných výkonových spekter nezávislých komponent všech metod

Statistika	Infomax – FastICA	Infomax – SOBI	SOBI – FastICA
Minimum	2,0	4,0	3,0
Maximum	24,0	30,0	26,0
Průměr	11,6	14,2	13,9
Směrodatná odchylka	3,1	3,1	3,1
Rozptyl	9,3	9,8	9,3
Medián	12,0	14,0	14,0
Modus	12,0	14,0	14,0

Tabulka 5.31: Výsledky Shapiro–Wilkova testu korelovaných a nespárovaných výkonových spekter nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Korelované IC Infomax – FastICA	$\ll 0,05$	Zamítám H_0
Korelované IC Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Korelované IC SOBI – FastICA	$\ll 0,05$	Zamítám H_0
Nespárované IC Infomax – FastICA	$\ll 0,05$	Zamítám H_0
Nespárované IC Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Nespárované IC SOBI – FastICA	$\ll 0,05$	Zamítám H_0

na hladině významnosti 5 %. K porovnání opět použiji neparametrický dvouvýběrový Wilcoxonův test.

Výsledky Wilcoxonova testu pro korelovaná výkonová spektra IC jsou uvedeny v tabulce 5.32. U všech třech výběrů vyšla p-hodnota výrazně nižší než stanovená hladina významnosti. Proto na hladině významnosti 5 % zamítám nulovou hypotézu, že výběry pocházejí z rozdělení se stejnou distribuční funkcí. To znamená, že se jednotlivé metody ICA liší v počtu korelovaných i nespárovaných párů výkonových spekter extrahovaných IC.

Tabulka 5.32: Výsledky Wilcoxonova testu korelovaných výkonových spekter nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Infomax – FastICA a Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Infomax – FastICA a SOBI – FastICA	$\ll 0,05$	Zamítám H_0
Infomax – SOBI a SOBI – FastICA	$\ll 0,05$	Zamítám H_0

Tabulka 5.33: Výsledky Wilcoxonova testu nespárovaných topografických map nezávislých komponent všech metod

Metody	p-hodnota	Výsledek
Infomax – FastICA a Infomax – SOBI	$\ll 0,05$	Zamítám H_0
Infomax – FastICA a SOBI – FastICA	$\ll 0,05$	Zamítám H_0
Infomax – SOBI a SOBI – FastICA	$\ll 0,05$	Zamítám H_0

Pro nespárovaná výkonová spektra jsou výsledky Wilcoxonova testu uvedena v tabulce 5.33. U všech tří výběru opět zamítám H_0 na hladině významnosti $\alpha = 5\%$.

5.3 t-SNE a DBSCAN

Kvůli lepší přehlednosti a snazší orientaci v této podkapitole uvedu na začátku text s výsledky a na konci podkapitoly všechny uvedené obrázky.

19dimenzionální prostor všech topografických map získaných analýzou hlavních komponent jednotlivých metod jsem redukovala na 2dimenzionální (2D) prostor pomocí nelineární techniky t-Distributed Stochastic Neighbor Embedding. Při použití této techniky je vyžadováno nastavení perplexity (lze popsat jako odhad počtu blízkých sousedů každého objektu) uživatelem, přičemž autoři algoritmu doporučují nastavení 5-50, viz podkapitola 4.6. Provedla jsem sérii redukcí dimenze s nastavenou perplexitou 5 - 50 v kroku po 5. Nejlepších výsledků z hlediska separace shluků jsem dosáhla s perplexitou 40 a euklidovskou vzdálenostní metrikou.

Redukovaný 2D prostor jednotlivých metod jsem klasifikovala shlukovacím algoritmem Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Výsledky klasifikace metodou DBSCAN jsou částečně závislé na jeho uživateli. Uživatel na vstupu nastaví předpokládaný poloměr shluků $Eps(\varepsilon)$ a minimální počet objektů $MinPts$, který shluk musí obsahovat, aby byl klasifikován jako shluk obsahující stejné prvky. K nalezení optimálního nastavení $Eps(\varepsilon)$ a $MinPts$, jsem u každého nalezeného shluku vypočítala rozptyl topografických map, které shluk obsahoval. Nastavení, při kterém byl rozptyl malý, jsem použila ke klasifikaci a k následnému výpočtu průměrných topografických map reprezentujících každý shluk. Konečné nastavení bylo $Eps(\varepsilon) = 4,5$ a $MinPts = 630$.

Výsledek t-SNE a DBSCANu na prostoru topografických map nezávislých komponent vypočítaných metodou Infomax je vidět na obrázku 5.25. DBSCAN v 2D prostoru odhalil 16 shluků a jednu třídu představující šumové body, na obrázku je označena -1. Průměrné topografické mapy shluků jsou vidět na obrázcích 5.26 a 5.27.

U metody FastICA DBSCAN klasifikoval 2D prostor do 15 shluků a 1 třídu šumových komponent, viz obrázek 5.29. Průměrné topografické mapy jsou uvedeny na obrázcích 5.30 a 5.31. Některé topografické mapy metody FastICA byly podobné mapám metody Infomax.

2D prostor topografických map metody SOBI klasifikovaný DBSCANem je uveden na obrázku 5.33. U této varianty ICA bylo klasifikováno 19 shluků a 1 shluk šumových komponent. Rozložení a velikost shluků je mírně odlišná od předchozích metod. Na obrázcích 5.34 a 5.35 jsou uvedeny průměrné topografické mapy charakterizující každý shluk. I ty jsou v některých případech (např. mapa IC 11 či 15) odlišné.

Pro ukázkou časových řad a výkonových spekter jsem vybrala topografickou mapu, která byla podobná u všech variant. U metody Infomax se jednalo o mapu číslo 5, patřící stejnému shluku. Časové řady a výkonový spektra jsou vidět na obrázku 5.3. U metody FastICA to byla mapa číslo 8, ukázkou časových řad a spekter je možné vidět na obrázku 5.3. Na obrázku 5.3 jsou vidět časové řady a výkonová spektra klasifikovaná do shluku číslo 12 metody SOBI. Dle časových řad, topografických map i výkonových spekter se zřejmě jedná o fyziologickou alfa aktivitu a oční artefakty. Měřítko osy y není uvedeno, protože jak jsem uvedla v podkapitole 4.3.2 nemůžeme určit rozptyly jednotlivých nezávislých komponent a to z důvodu neznalosti zdrojových signálů \mathbf{s} i směšovací matice \mathbf{A} . V důsledku toho můžeme docela dobře určit a upravovat rozsahy nezávislých komponent.

Pomocí výše uvedeného nastavení shlukovacího algoritmu DBSCAN jsem neodhalila vnitřní charakter 2D prostoru topografických map ICA. Nalezené shluky nevykazovaly jednoznačné charakteristiky, který by byly typické pro každý shluk. Dále jsem u žádné varianty ICA nenalezla shluk s komponenty, který by další varianty neodhalily. Sama metoda t-SNE separovala u každé varianty ICA jistý počet malých shluků. Změnila jsem nastavení DBSCANu a provedla klasifikaci 2D prostoru všech variant ještě jednou.

Podruhé jsem nastavila $Eps(\varepsilon) = 2$ a $MinPts = 80$, abych mohla analyzovat malé shluky odtržené od zbytku topografických map. Výsledný klasifikovaný 2D prostor topografických map metody Infomax je uveden na obrázku 5.37. Metoda t-SNE v kombinaci s DBSCANem odhalila 18 shluků (-1 je klasifikován shluk představující šumové/odlehle komponenty). Taktéž u varianty FastICA bylo klasifikováno 18 shluků, viz obrázek 5.40. DBSCAN 2D prostoru topografických map metody SOBI je vidět na obrázku 5.43. V tomto případě bylo klasifikováno 38 shluků, shluk s označením -1 jsou odlehle komponenty.

Všechny shluky u všech metod jsem opět popsala jejich průměrnými topografickými mapami. Pro variantu Infomax jsou průměrné topografické mapy uvedeny

na obrázcích 5.38 a 5.39. Z map je patrné, že odlehlé shluky představují zejména artefakty ze špatné elektrody, konkrétně jde o elektrody P4 (shluk číslo 5), F7 (shluk číslo 6 a 7), Fp2 (shluk číslo 8, 13 a 14), P3 (shluk číslo 9 a 10), Pz (shluk číslo 12), F3 (shluk číslo 15) a F4 (shluk číslo 16 a 17).

Metody t-SNE a DBSCAN také odhalily artefakty ze špatné elektrody u varianty FastICA. Topografické mapy artefaktů ze špatné elektrody s označením Fp1 (shluk číslo 1), Fp2 (shluk číslo 8), F7 (shluk číslo 7), Fz (shluk číslo 12) a P3 (shluk číslo 9 a 10) jsou vidět na obrázku 5.41, pro elektrody Fp2 (shluk číslo 15), F4 (shluk číslo 17), P4 (shluk číslo 14) a částečně i Pz na obrázku 5.42. Shluk patřící elektrodě F3 u metody FastICA nebyla nalezen. Elektroda Fz naopak není vidět ve shlucích metody Infomax.

Redukce dimenze u prostoru topografických metody SOBI způsobila oddělení shluků, které patří artefaktům ze stejné elektrody. Jedná se o shluky 17 a 20, jejichž topografické mapy jsou uvedeny na obrázku 5.45. A shluky 25 a -1, jejichž topografické mapy jsou na obrázku 5.46. Další průměrné topografické mapy jsou uvedeny na obrázku 5.44. Artefaktů ze špatné elektrody s označením F3 patří shluk číslo 4 a 27, C4 shluk číslo 12 a 35, P4 shluk číslo 15, F7 shluk číslo 16, Fp2 shluk číslo 17, 20 a 21, C3 shluk číslo 22 a 23, P3 shluk číslo 24, 25 a -1, Fz shluk číslo 26, F4 shluk číslo 29 a 32, Pz shluk číslo 34. Metoda SOBI tedy navíc odhalila artefakty ze špatné elektrody s označením C3.

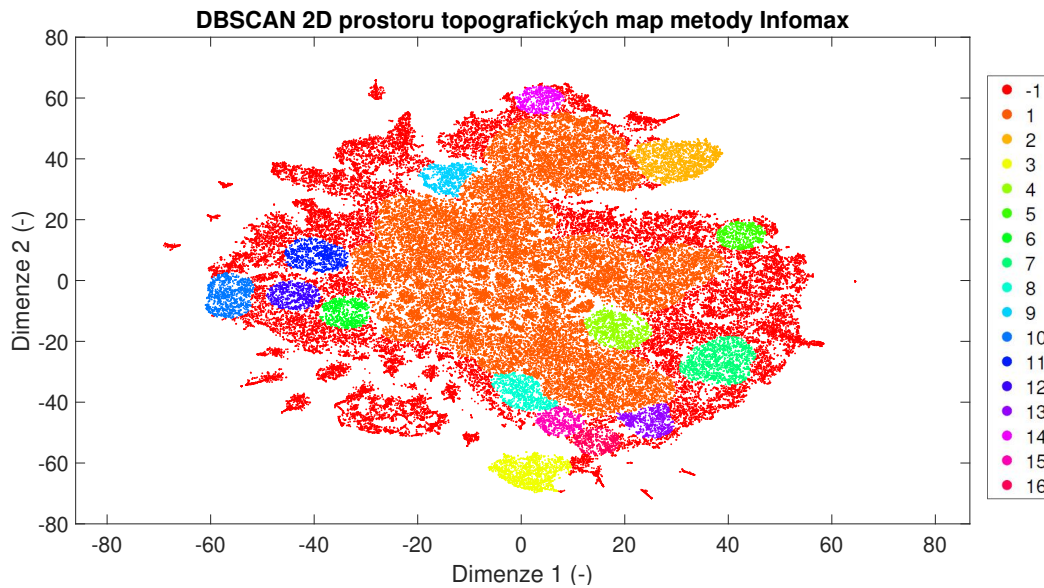
Ukázky časových řad a výkonových spekter vybraných shluků u artefaktů ze špatné elektrody neuvádím, protože mohou být velmi rozmanité. Ale k detekování tohoto artefaktu je důležitý vzhled topografické mapy.

5.4 Charakter komponent pro rozdílné metody potlačení artefaktů

Na obrázku 5.47 jsou vidět časové řady, topografické mapy a normovaná výkonová spektra nezávislých komponent vypočítaných metodou Infomax a FastICA a představujících elektrickou srdeční aktivitu. Dle vzhledu topografických map jsou do této komponenty zřejmě přimíchané i pulzní artefakty. Komponenty obou metod jsou téměř totožné.

Metoda SOBI nedokázala oddělit samostatnou komponentu srdeční aktivity. EKG je nejvíce v komponentě číslo 1 a 10 na obrázku 5.48. Kromě EKG je v IC číslo 1 patrná i fyziologická aktivita a dle topografické mapy i pulzní artefakt. V IC číslo 10 je taktéž vidět i fyziologická aktivita.

V bdělém stavu jsou významnou součástí EEG signálu i oční artefakty, mezi které patří pohyby očí a mrkání. Pro metodu Infomax jsou na obrázcích 5.49 a 5.50 dole



Obrázek 5.25: 2D prostor topografických map vypočítaných metodou Infomax redukováných metodou t-SNE a klasifikovaných metodou DBSCAN.

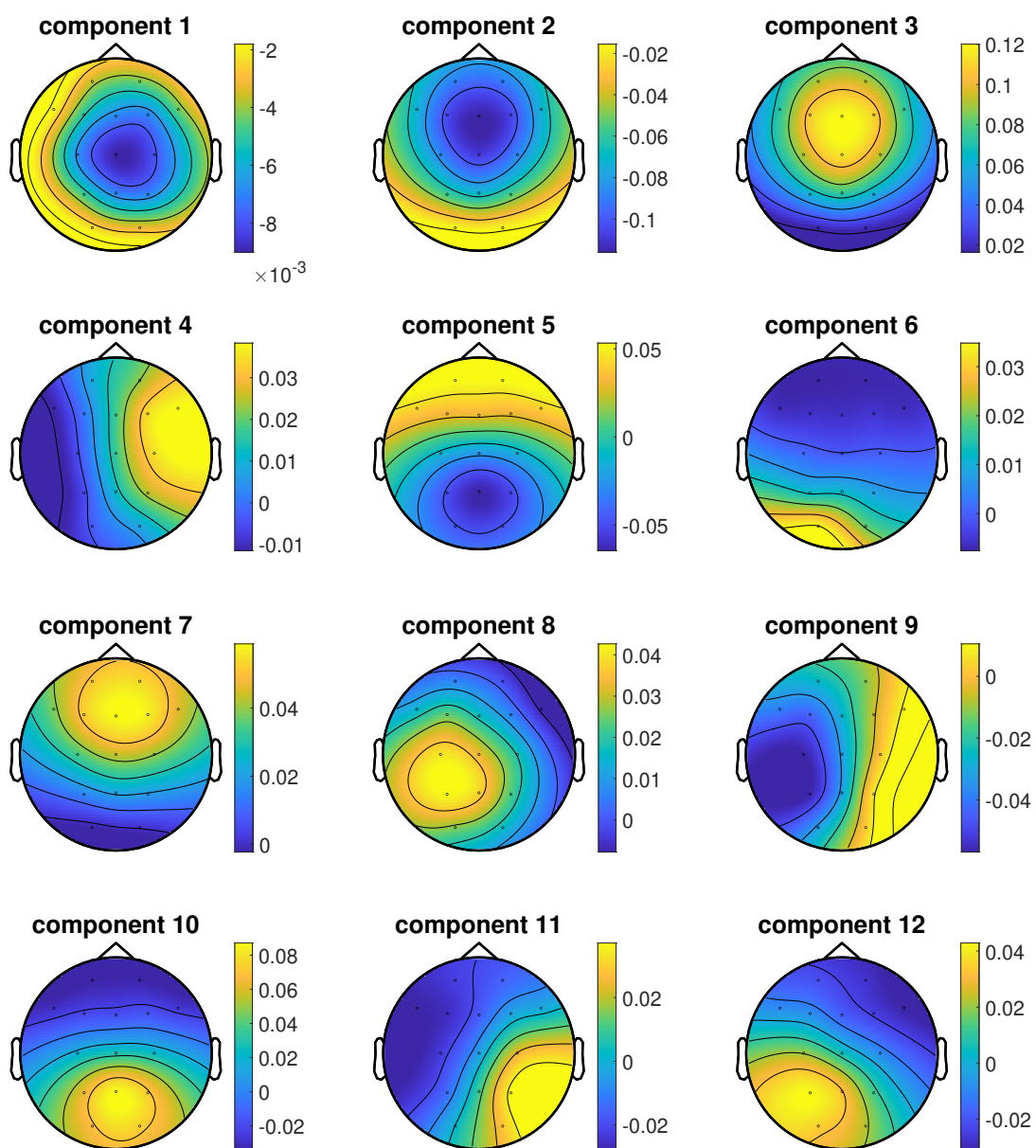
vidět komponenty představující mrkání, na obrázku 5.50 je komponenta představující pohyby očí. V topografické mapě jsou dobře rozpoznatelné jako zvýšená aktivita ve frontální oblasti. Největší výkon je patrný v nízkých frekvencích 1–2 Hz.

Přestože metoda FastICA vypočítala v tomto ukázkovém segmentu celkově 14 IC, "čistých" komponent představujících oční artefakty vypočítala 5. Zatímco metoda Infomax a SOBI 4 IC. Komponenty s artefakty způsobené mrkáním jsou zobrazeny na obrázcích 5.51 a 5.52. Na obrázku 5.53 je zobrazena komponenta, která kromě mrkání obsahuje i oční pohyby. Opět je vidět největší aktivita ve frontální oblasti a výkon ve frekvenční oblasti okolo 1–2 Hz.

Nezávislé komponenty s očními artefakty vypočítané metodou SOBI jsou vidět na obrázcích 5.54 a 5.55. IC číslo 4 a IC číslo 6 jsou si velice podobné v časové oblasti, topografické mapě i v rozložení výkonu a kromě očních artefaktů zřejmě obsahují i svalové artefakty. Komponenta číslo 13 obsahuje kromě očních artefaktů i fyziologickou aktivitu.

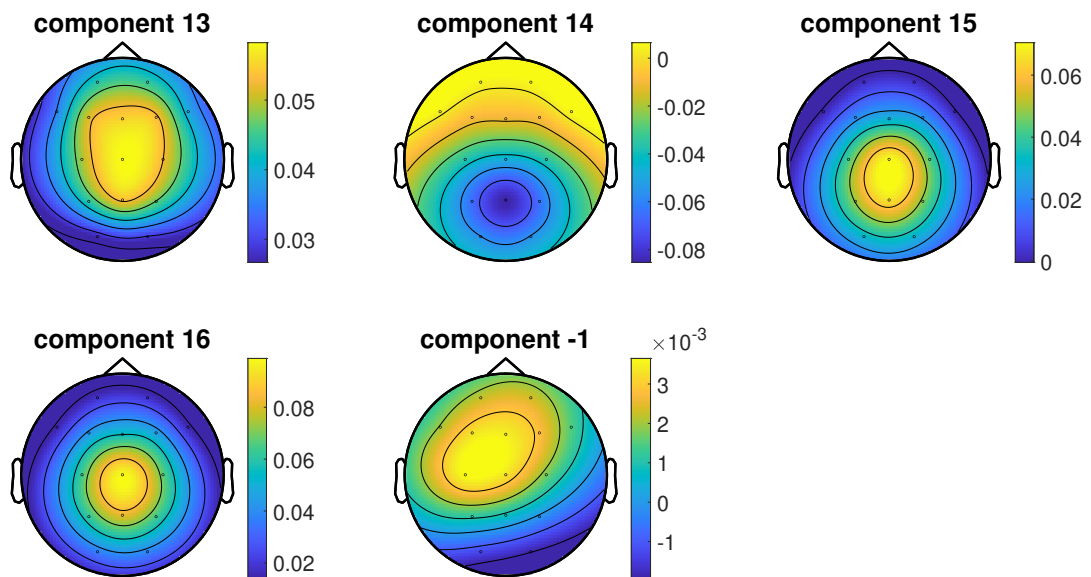
Artefakt ze špatné elektrody je v topografické mapě dobře rozpoznatelný, jedná se o vysokou aktivitu lokalizovanou na malé ploše pod odpovídající elektrodou. V časové a frekvenční oblasti tyto artefakty vypadají různě, v závislosti na tom, jaký šum okolí obsahuje (elektroda, která nemá pevný kontakt s povrchem skalpu funguje ve vzduchu jako anténa). Jak vypadají IC s artefakty ze špatné elektrody vypočítané metodou Infomax je vidět na obrázku 5.56. Pro metody FastICA a SOBI pak na obrázku 5.57.

Průměrné topografické mapy shluků metody Infomax



Obrázek 5.26: Průměrné topografické mapy nezávislých komponent vypočítaných metodou Infomax shluku číslo 1–12. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.

Průměrné topografické mapy shluků metody Infomax

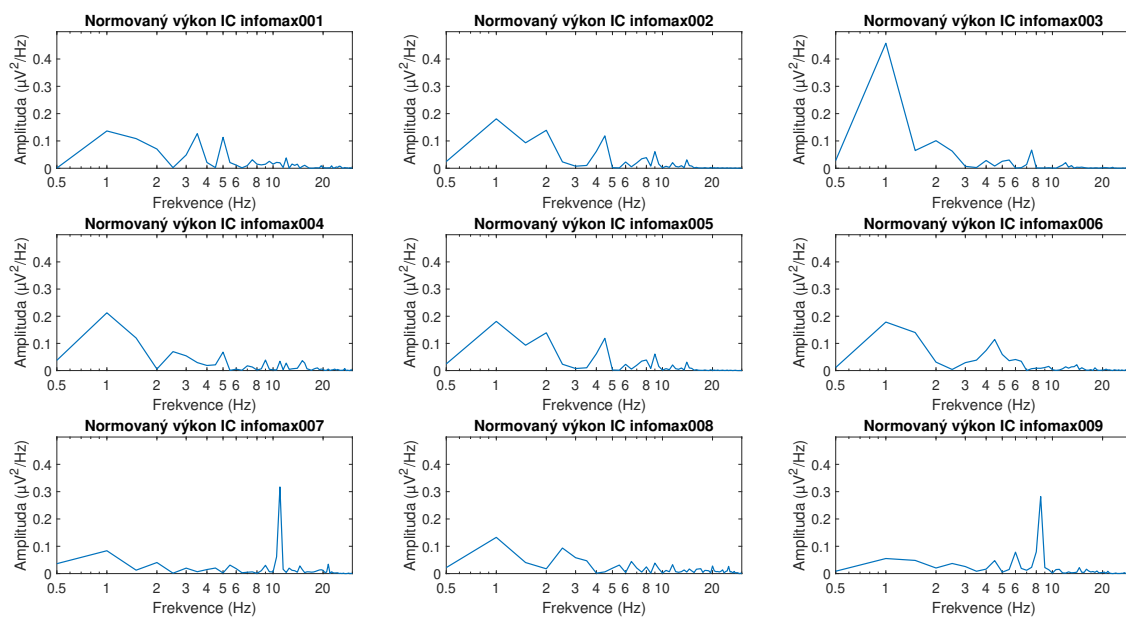
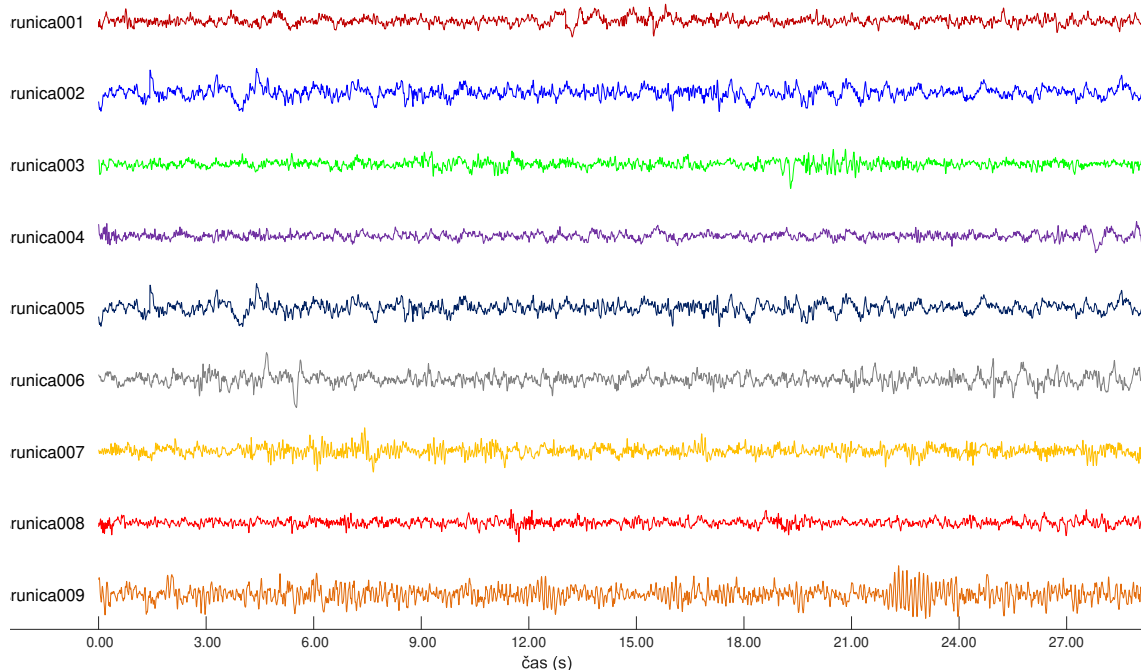


Obrázek 5.27: Průměrné topografické mapy nezávislých komponent vypočítaných metodou Infomax shluku číslo 13–16 a -1. Shluk s číslem -1 je skupina bodů, které DBSCAN definuje jako odlehlé. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.

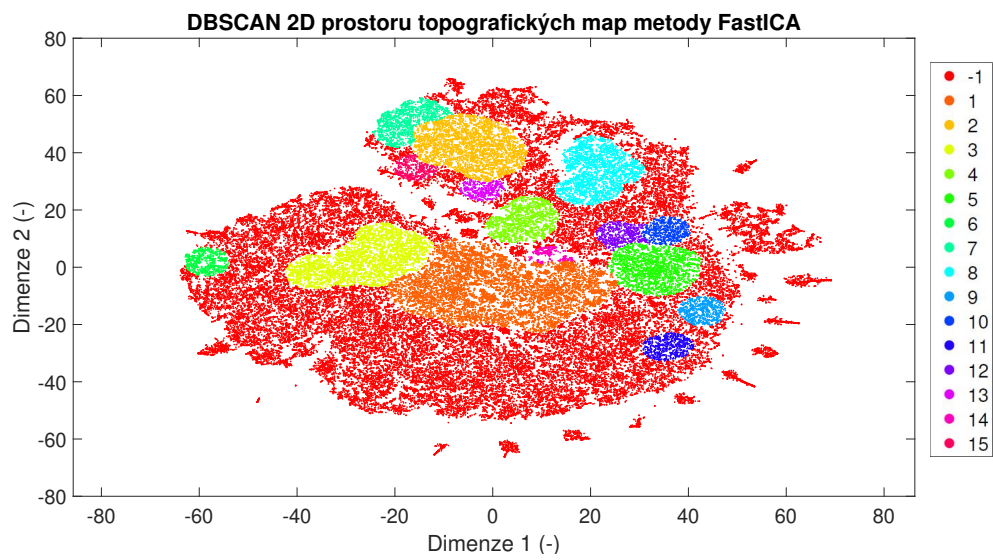
Ani jedna z metod nedokázala vypočítat komponentu, která by prezentovala pouze artefakty ze svalů. EMG artefakty byly rozprostřeny do více komponent, ve kterých kromě těchto artefaktů byla přítomna i např. EKG jako na obrázku 5.58 nahoře, či fyziologická aktivita jako na obrázku 5.58 dole. Ve frekvenčním pásmu je EMG nejvýkonnější od 20 Hz výš.

Kromě nezávislých komponent, ve kterých byla prezentována "čistá" EEG aktivita, nebo naopak "čisté" artefakty, byly vypočítány i komponenty, ve kterých EEG aktivita zůstala smíchaná se srdečním či očním artefaktem. A také vznikly IC, které prezentovaly směs signálů pocházejících z různých zdrojů, jak je vidět na obrázku 5.59. V tomto případě nezávislá komponenta obsahuje signál ze špatné elektrody, fyziologickou aktivitu a svalové artefakty.

Časové řady nezávislých komponent shluku číslo 5 metody Infomax

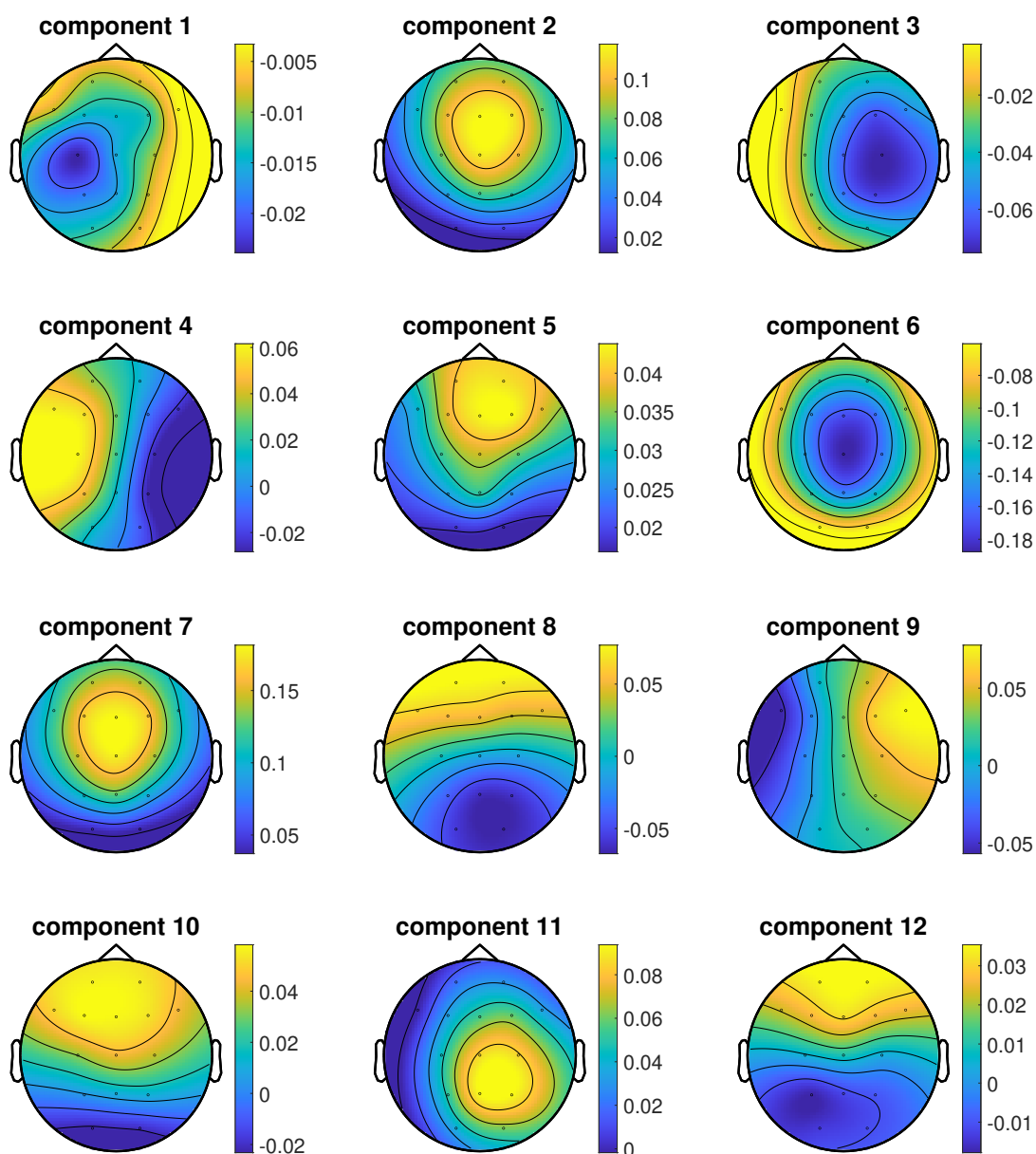


Obrázek 5.28: Ukázka časových řad (nahore) a normovaných výkonových spekter (dole) IC metody Infomax shluku číslo 5 získaného klasifikací DBSCAN.



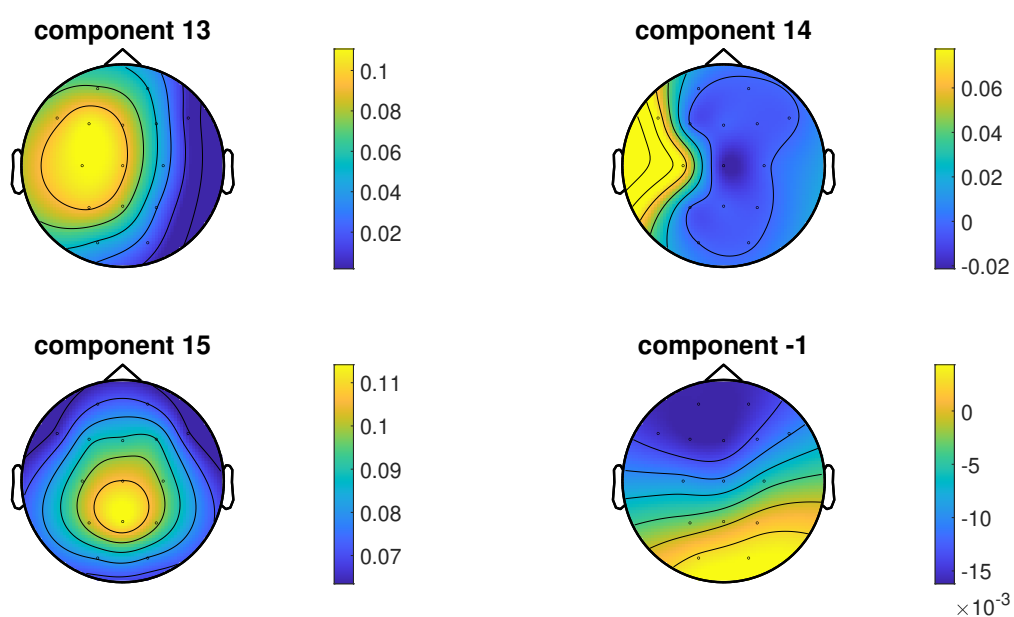
Obrázek 5.29: 2D prostor topografických map vypočítaných metodou FastICA redukováných metodou t-SNE a klasifikovaných metodou DBSCAN.

Průměrné topografické mapy shluků metody FastICA



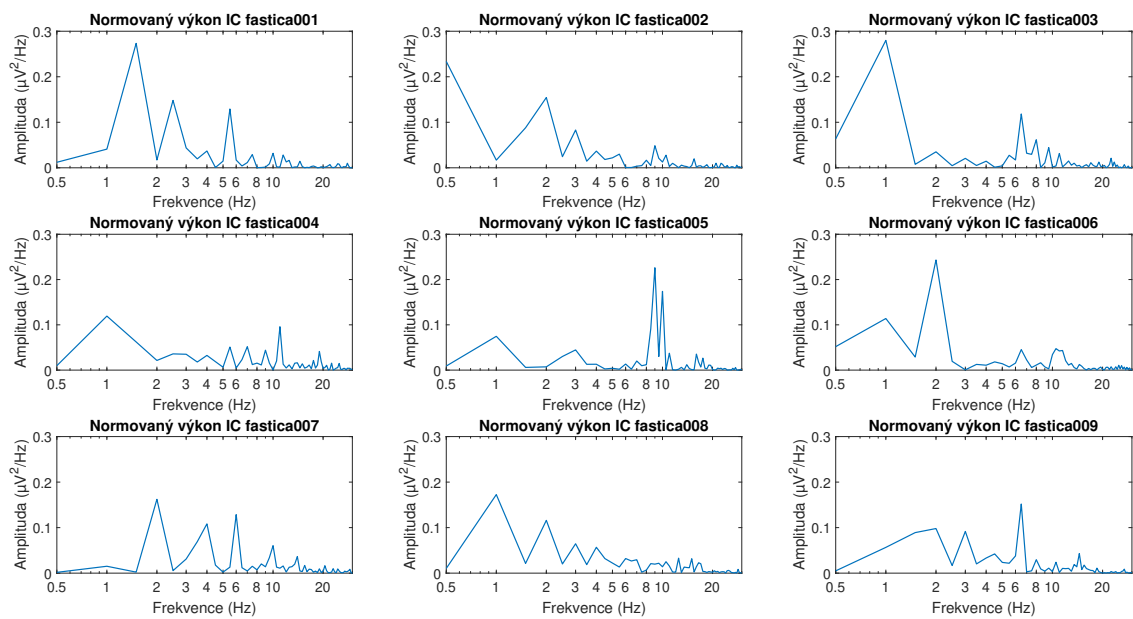
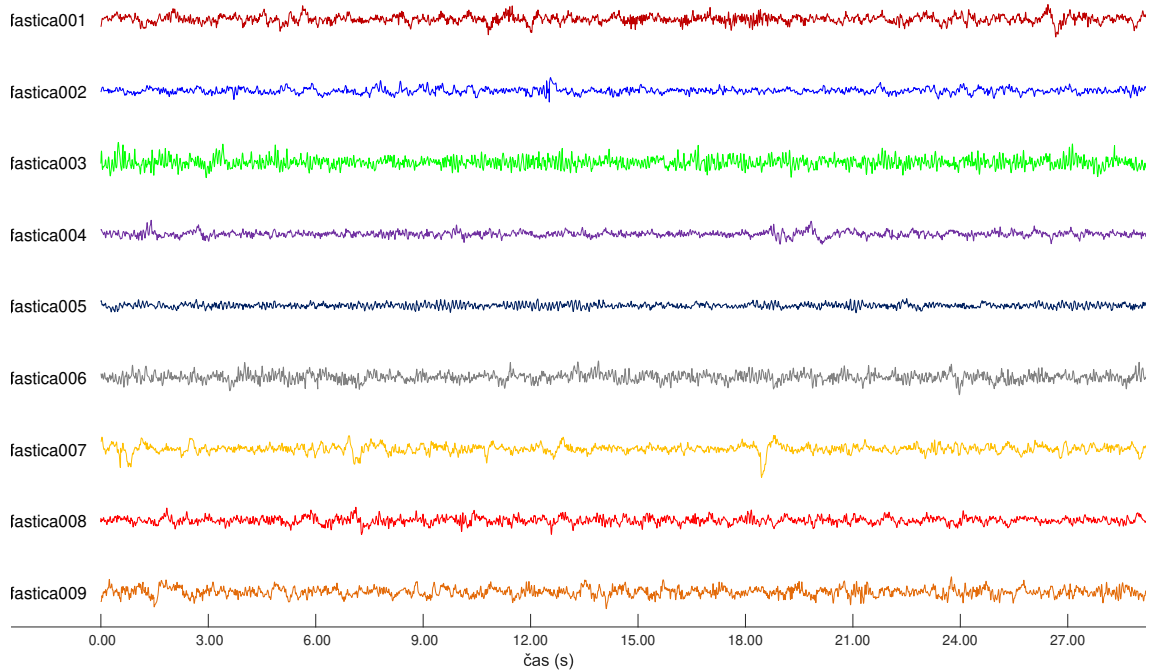
Obrázek 5.30: Průměrné topografické mapy nezávislých komponent vypočítaných metodou FastICA shluku číslo 1–12. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku

Průměrné topografické mapy shluků metody FastICA

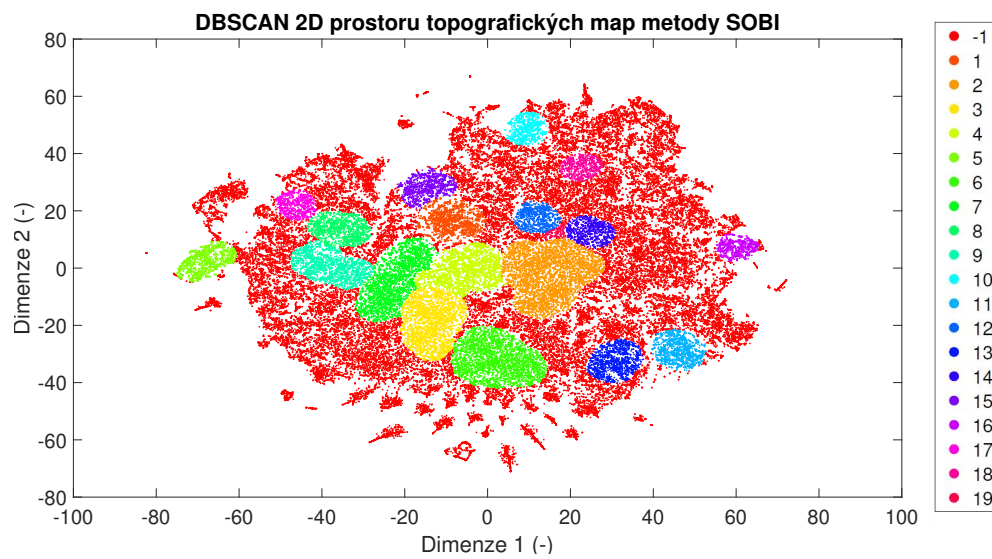


Obrázek 5.31: Průměrné topografické mapy nezávislých komponent vypočítaných metodou FastICA shluku číslo 12–15 a -1. Shluk s číslem -1 je skupina bodů, které DBSCAN definuje jako odlehlé. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku

Časové řady nezávislých komponent shluku číslo 8 metody FastICA

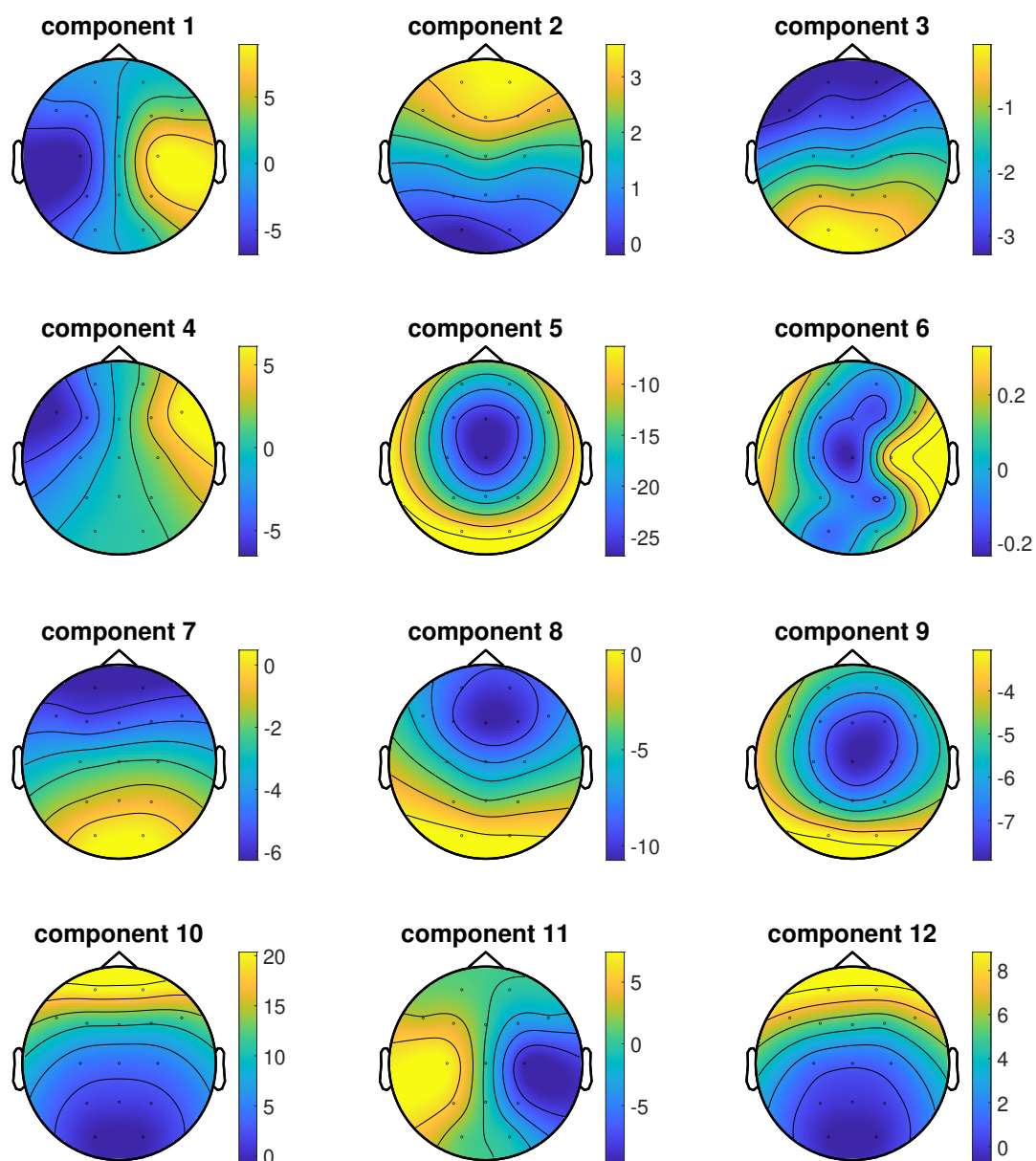


Obrázek 5.32: Ukázka časových řad (nahore) a normovaných výkonových spekter (dole) IC metody FastICA shluku číslo 8 získaného klasifikací DBSCAN.



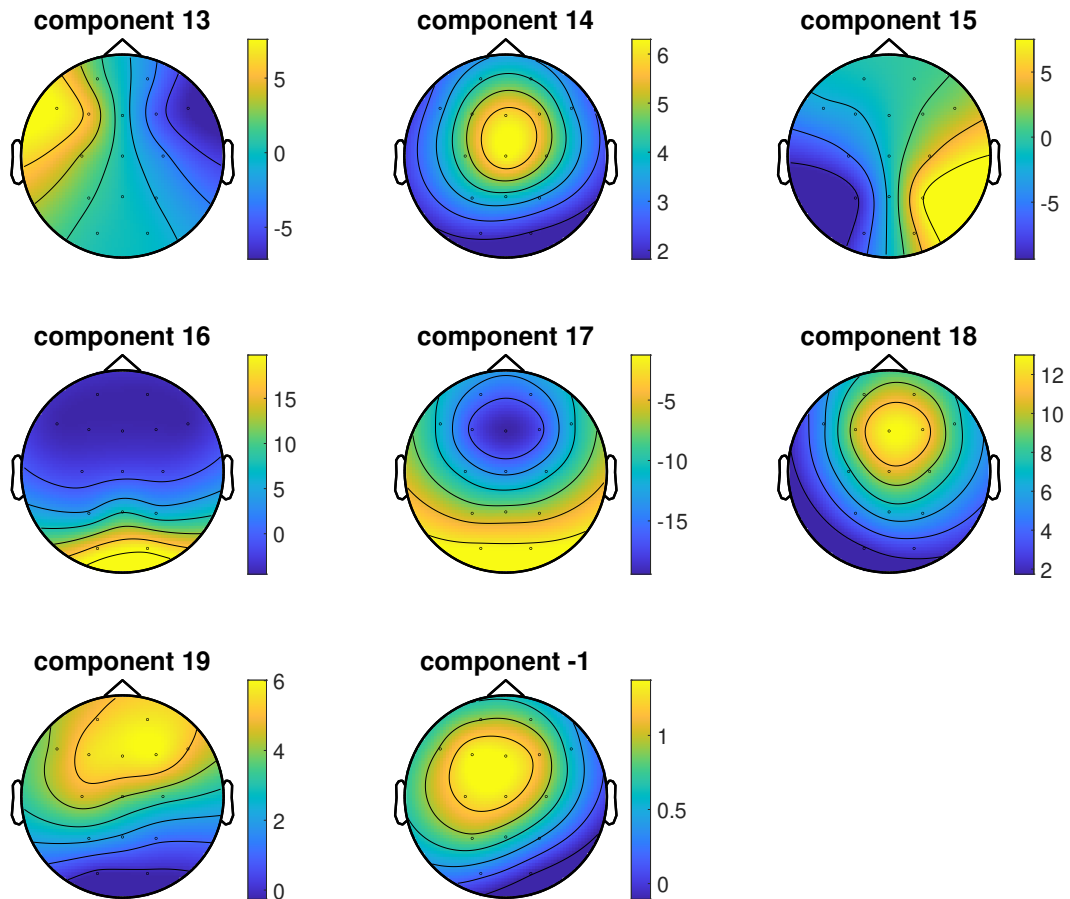
Obrázek 5.33: 2D prostor topografických map vypočítaných metodou SOBI redukováných metodou t-SNE a klasifikovaných metodou DBSCAN.

Průměrné topografické mapy shluků metody SOBI



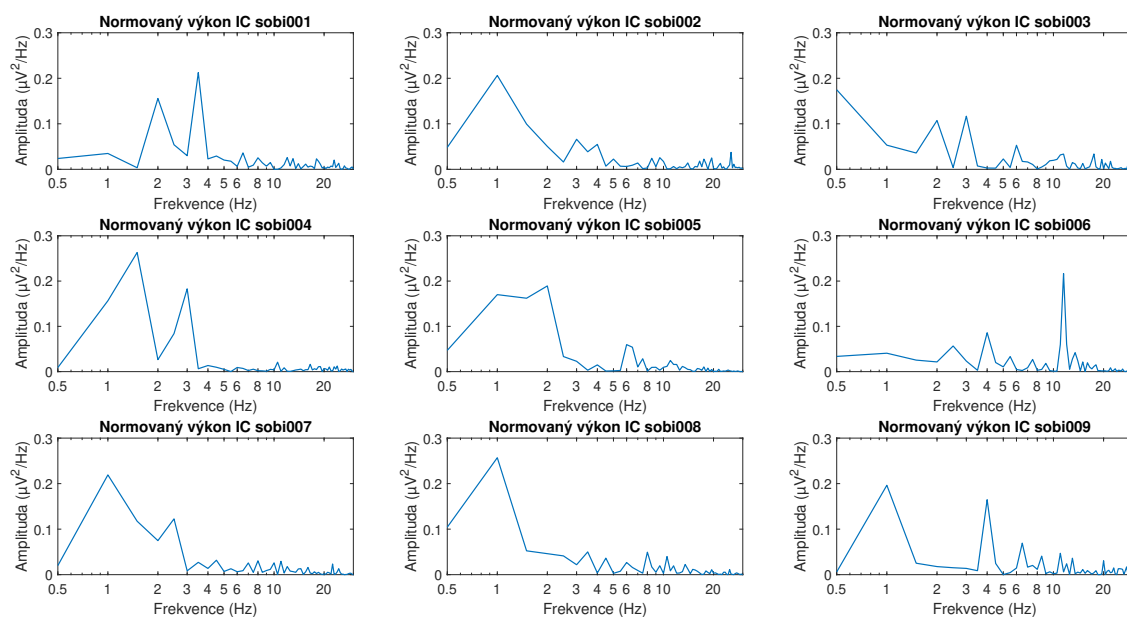
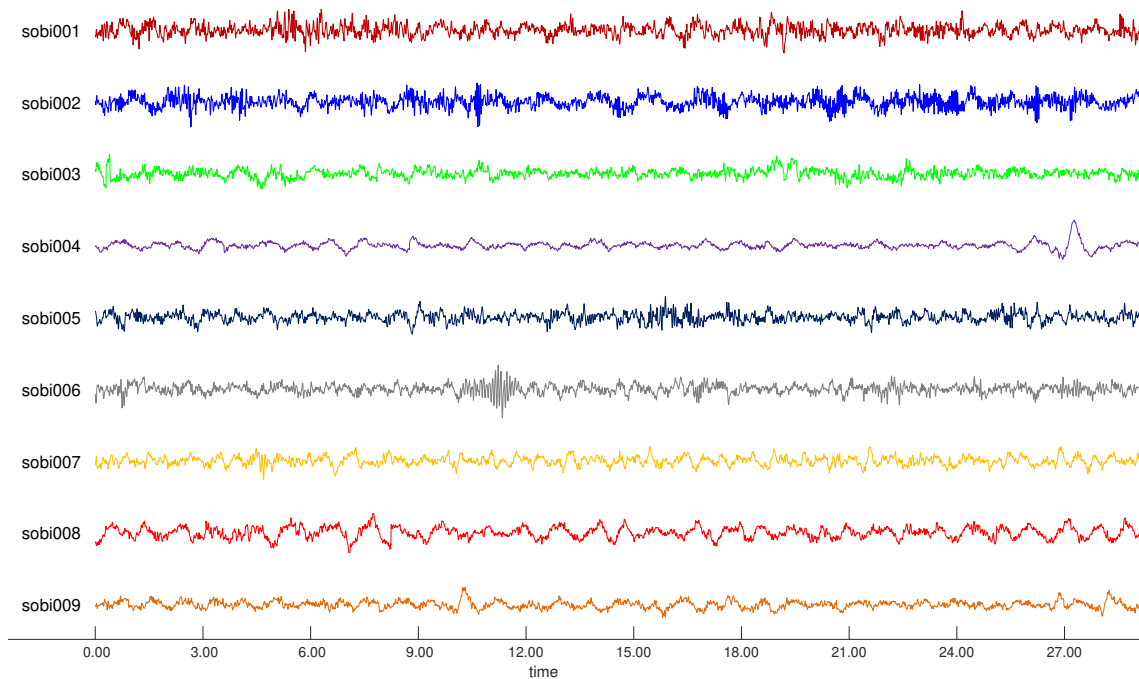
Obrázek 5.34: Průměrné topografické mapy nezávislých komponent vypočítaných metodou SOBI shluku číslo 1–12. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.

Průměrné topografické mapy shluků metody SOBI

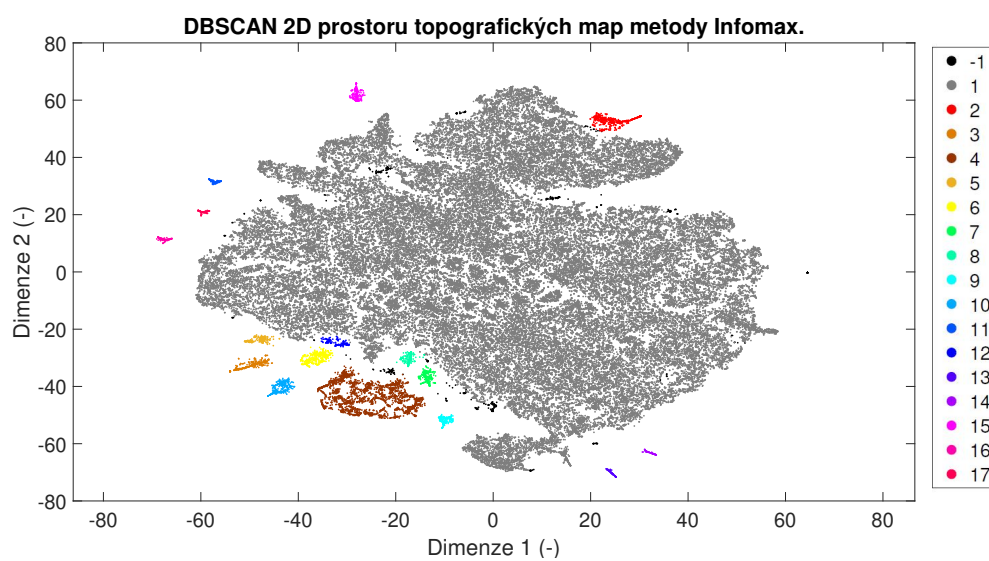


Obrázek 5.35: Průměrné topografické mapy nezávislých komponent vypočítaných metodou SOBI shluku číslo 13–19 a -1. Shluk s číslem -1 je skupina bodů, které DBSCAN definuje jako odlehlé. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.

Časové řady nezávislých komponent shluku číslo 12 metody SOBI

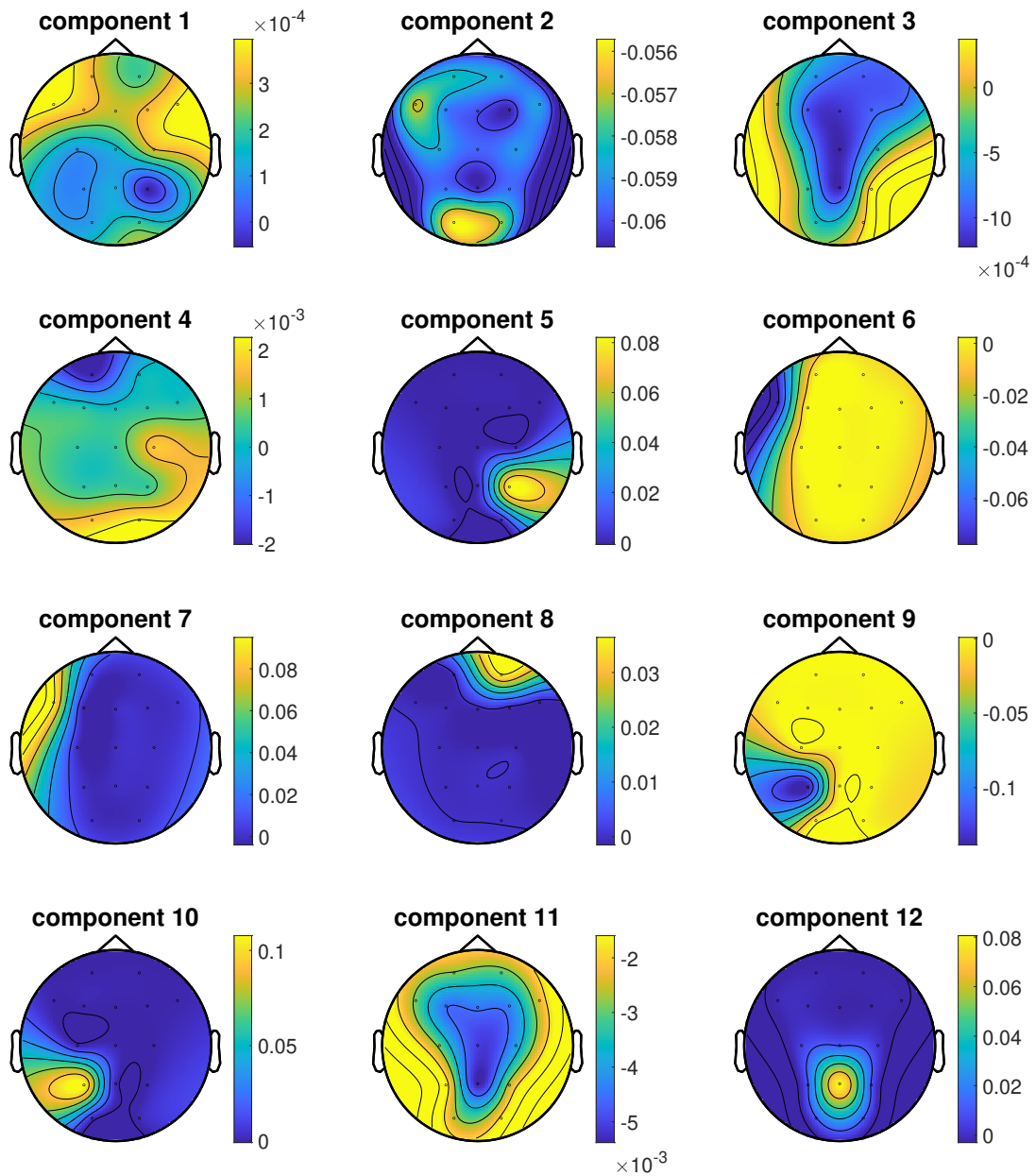


Obrázek 5.36: Ukázka časových řad (nahore) a normovaných výkonových spekter (dole) IC metody SOBI shluku číslo 12 získaného klasifikací DBSCAN.

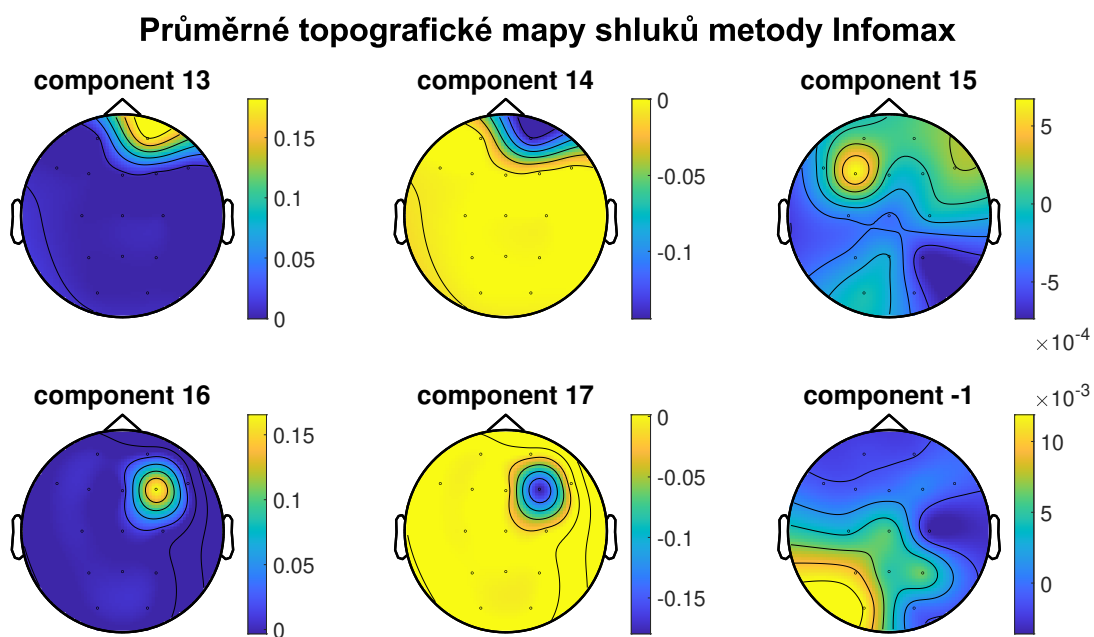


Obrázek 5.37: 2D prostor topografických map vypočítaných metodou Infomax redukováných metodou t-SNE a klasifikovaných metodou DBSCAN. Nastavení DBSCANu: $Eps(\varepsilon) = 2$ a $MinPts = 80$.

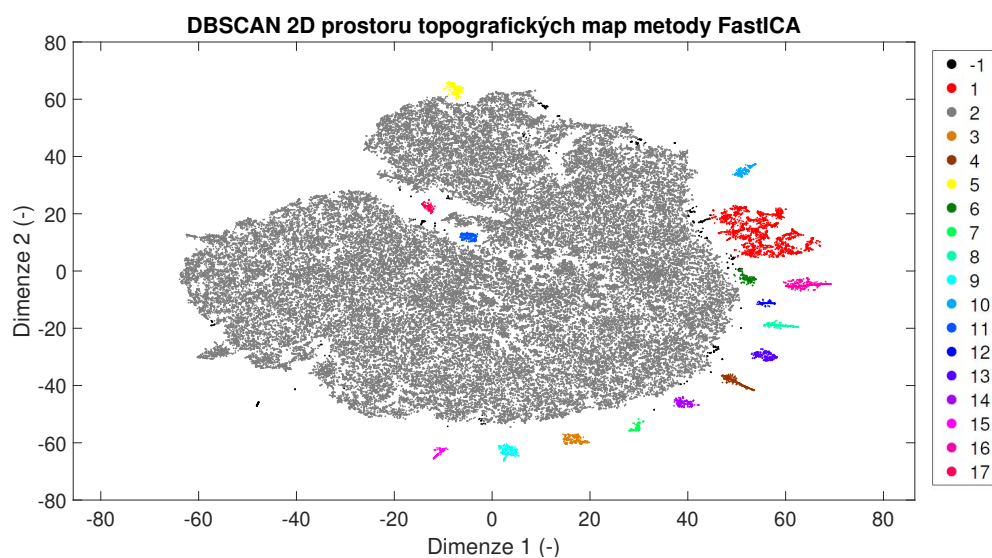
Průměrné topografické mapy shluků metody Infomax



Obrázek 5.38: Průměrné topografické mapy nezávislých komponent vypočítaných metodou Infomax shluku číslo 1–12. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.

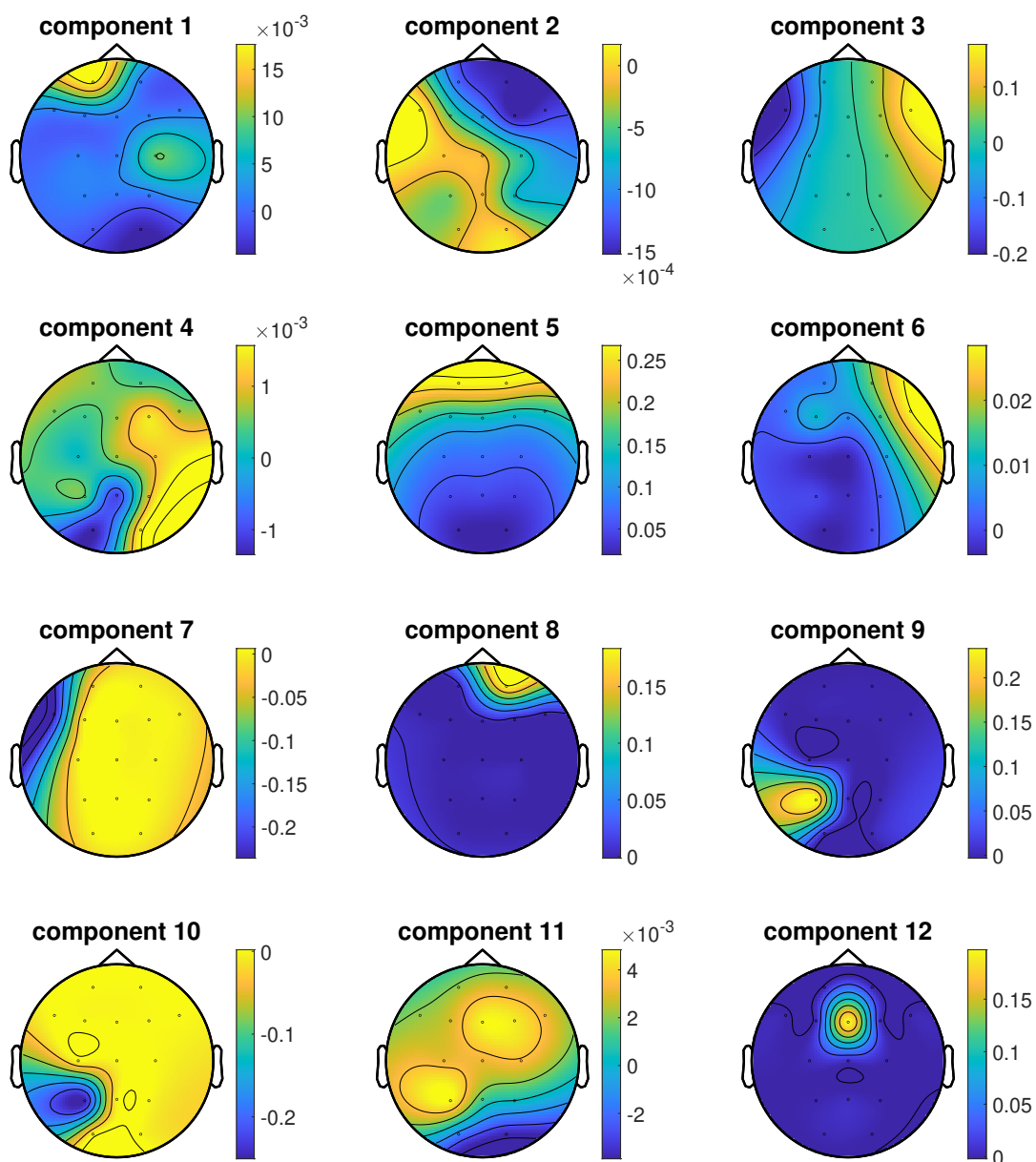


Obrázek 5.39: Průměrné topografické mapy nezávislých komponent vypočítaných metodou Infomax shluku číslo 13–17 a -1. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.



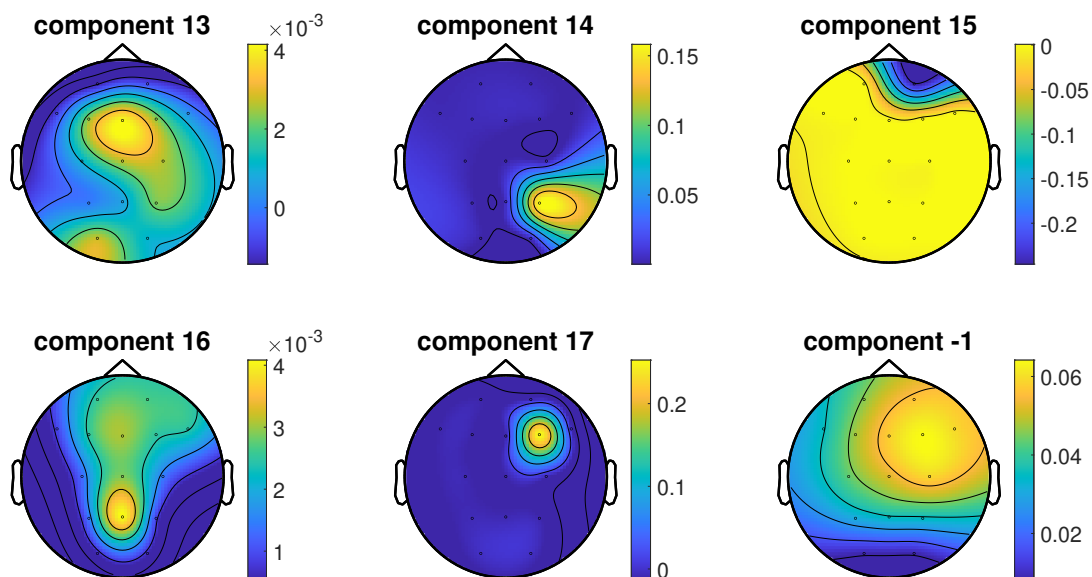
Obrázek 5.40: 2D prostor topografických map vypočítaných metodou FastICA redukováných metodou t-SNE a klasifikovaných metodou DBSCAN. Nastavení DBSCANu: $Eps(\epsilon) = 2$ a $MinPts = 80$.

Průměrné topografické mapy shluků metody FastICA

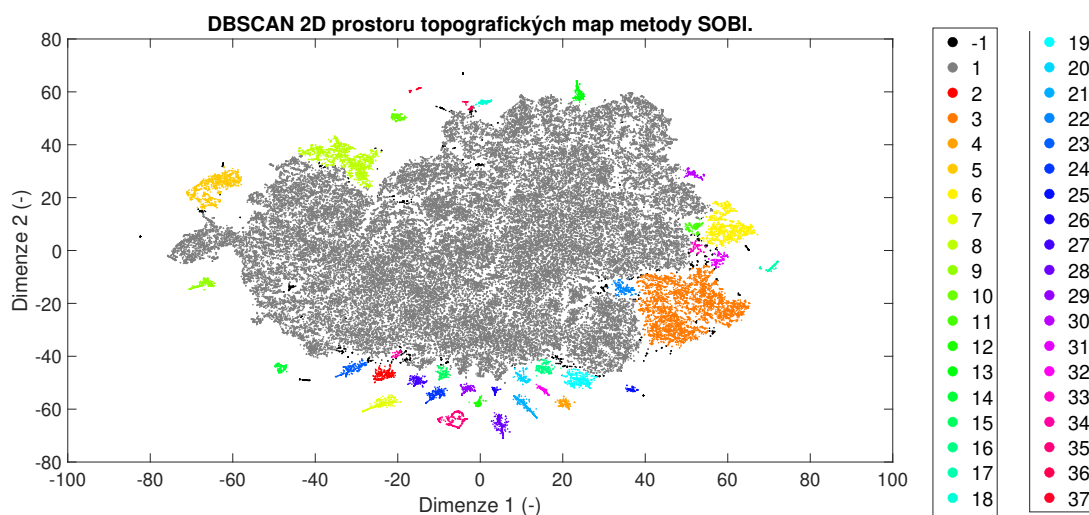


Obrázek 5.41: Průměrné topografické mapy nezávislých komponent vypočítaných metodou FastICA shluku číslo 1–12. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.

Průměrné topografické mapy shluků metody FastICA

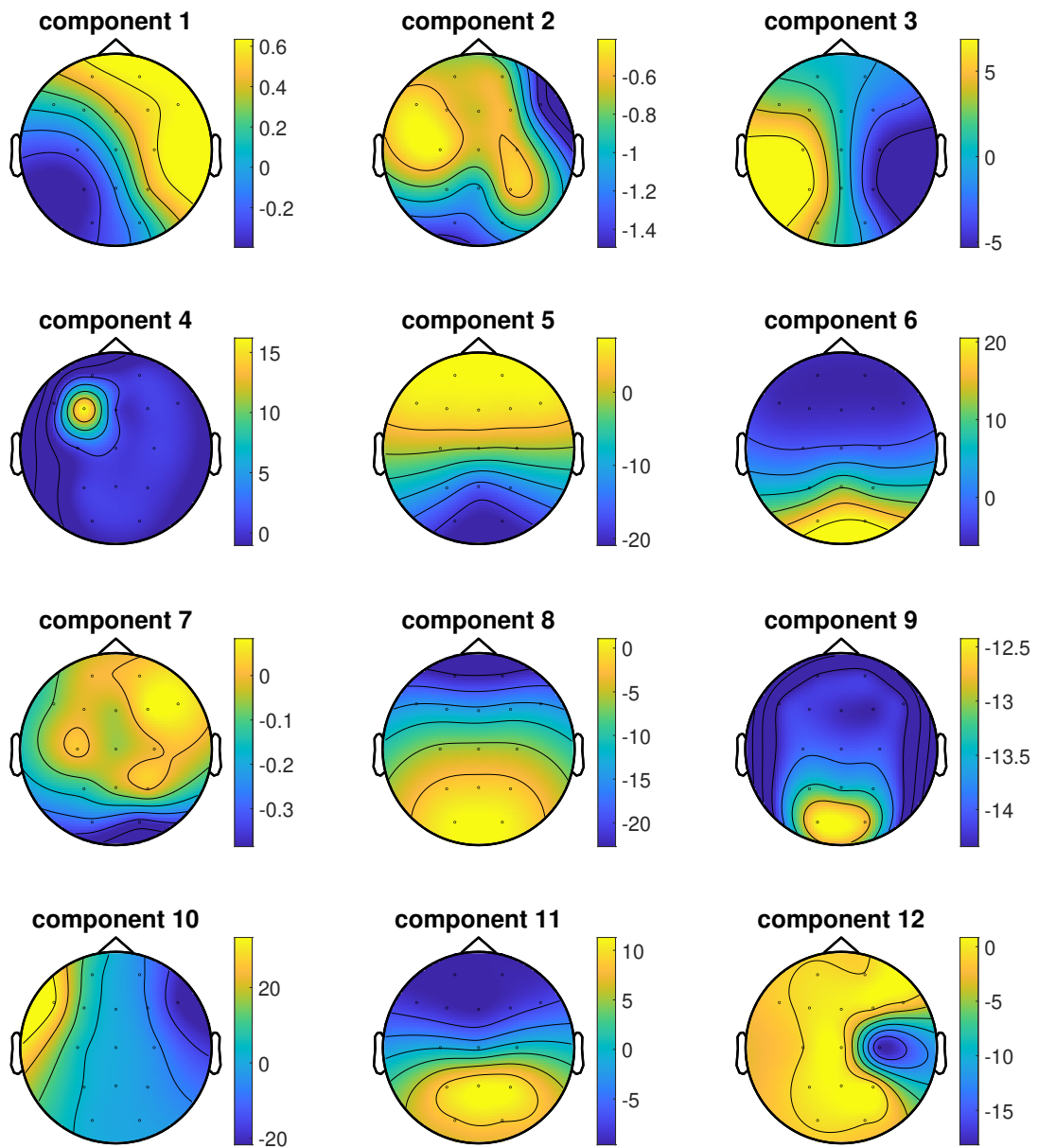


Obrázek 5.42: Průměrné topografické mapy nezávislých komponent vypočítaných metodou FastICA shluku číslo 13–17 a -1. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.



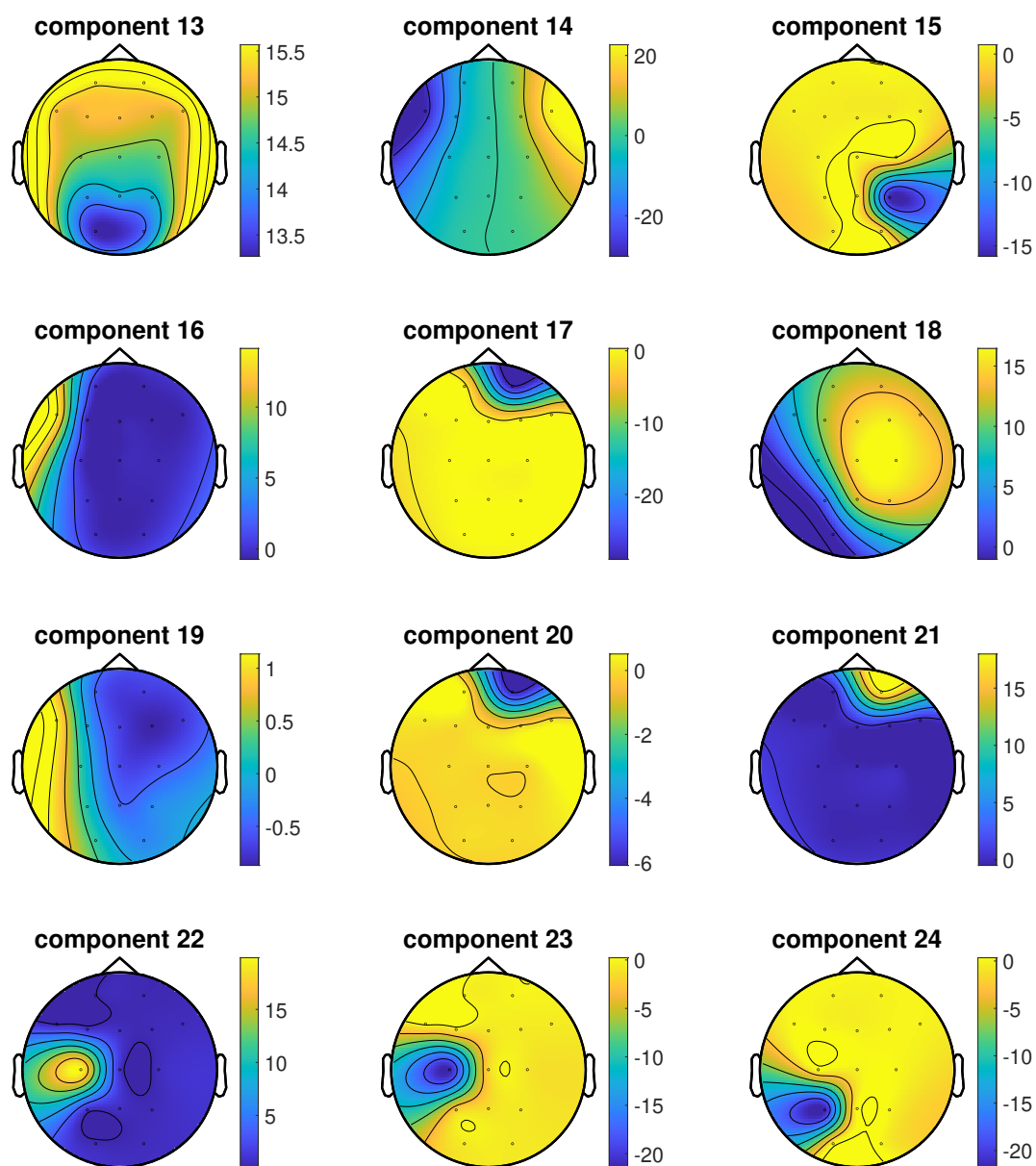
Obrázek 5.43: 2D prostor topografických map vypočítaných metodou SOBI redukováných metodou t-SNE a klasifikovaných metodou DBSCAN. Nastavení DBSCANu: $Eps(\epsilon) = 2$ a $MinPts = 80$.

Průměrné topografické mapy shluků metody SOBI



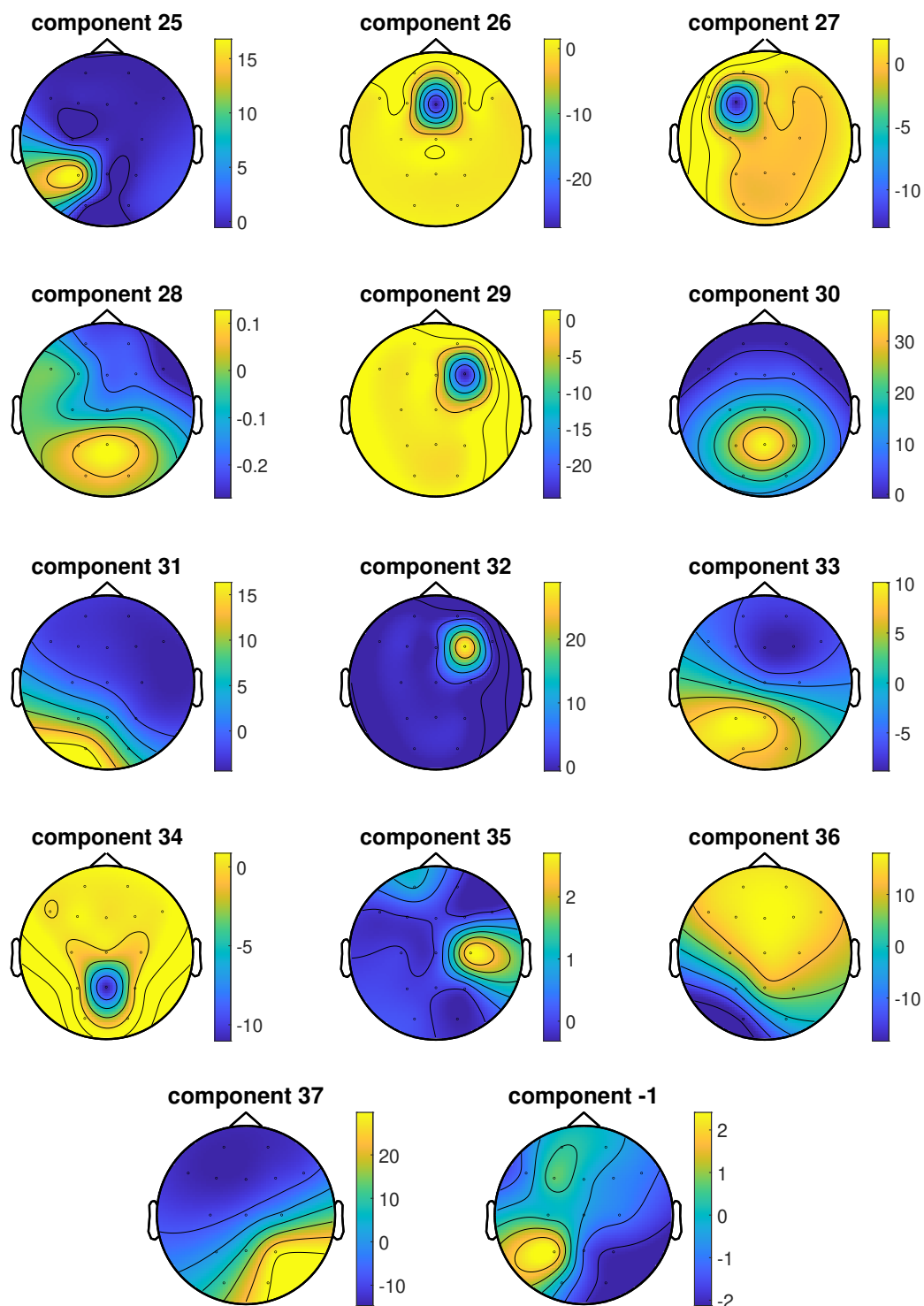
Obrázek 5.44: Průměrné topografické mapy nezávislých komponent vypočítaných metodou SOBI shluku číslo 1–12. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.

Průměrné topografické mapy shluků metody SOBI

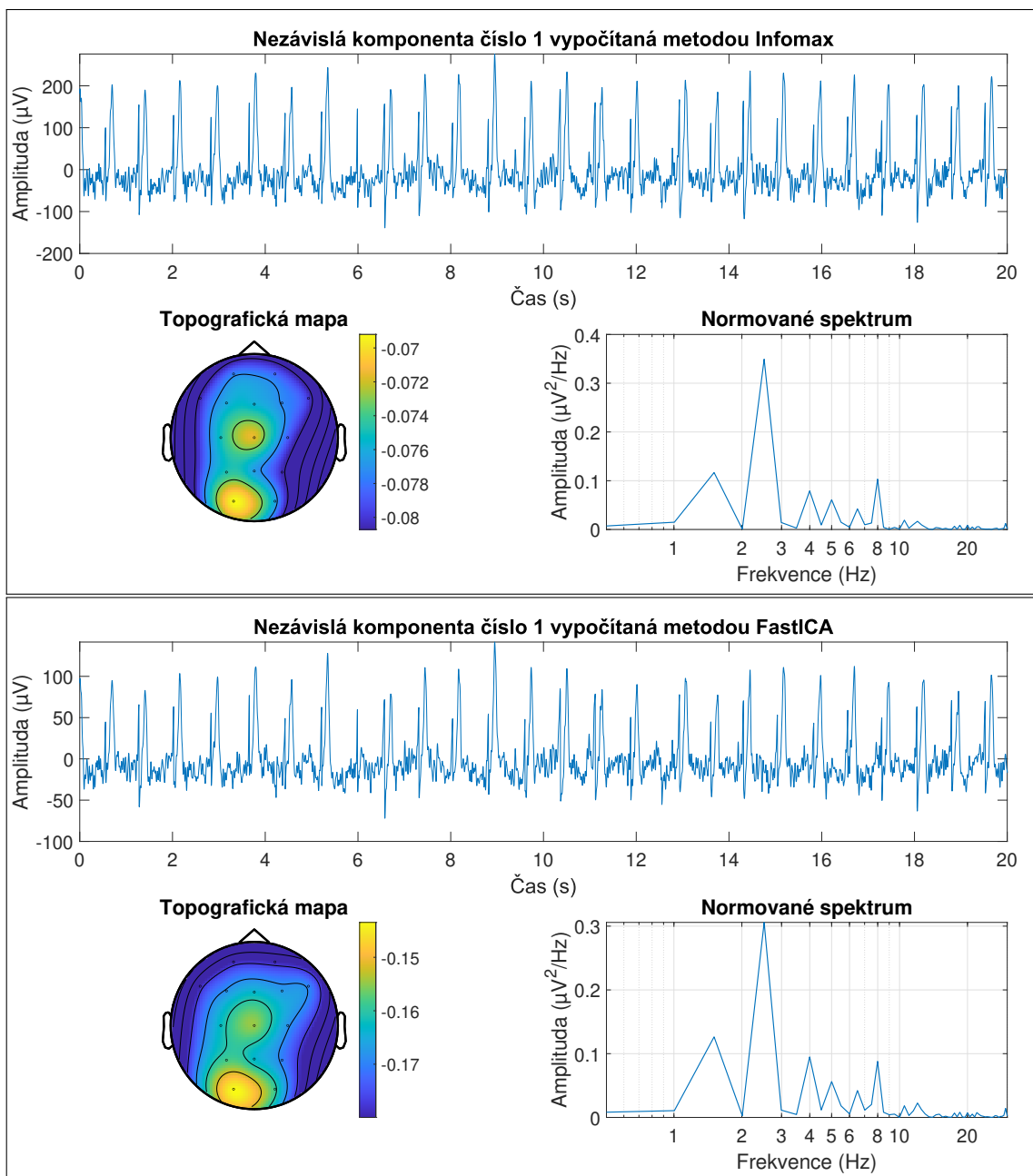


Obrázek 5.45: Průměrné topografické mapy nezávislých komponent vypočítaných metodou SOBI shluku číslo 13–24. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.

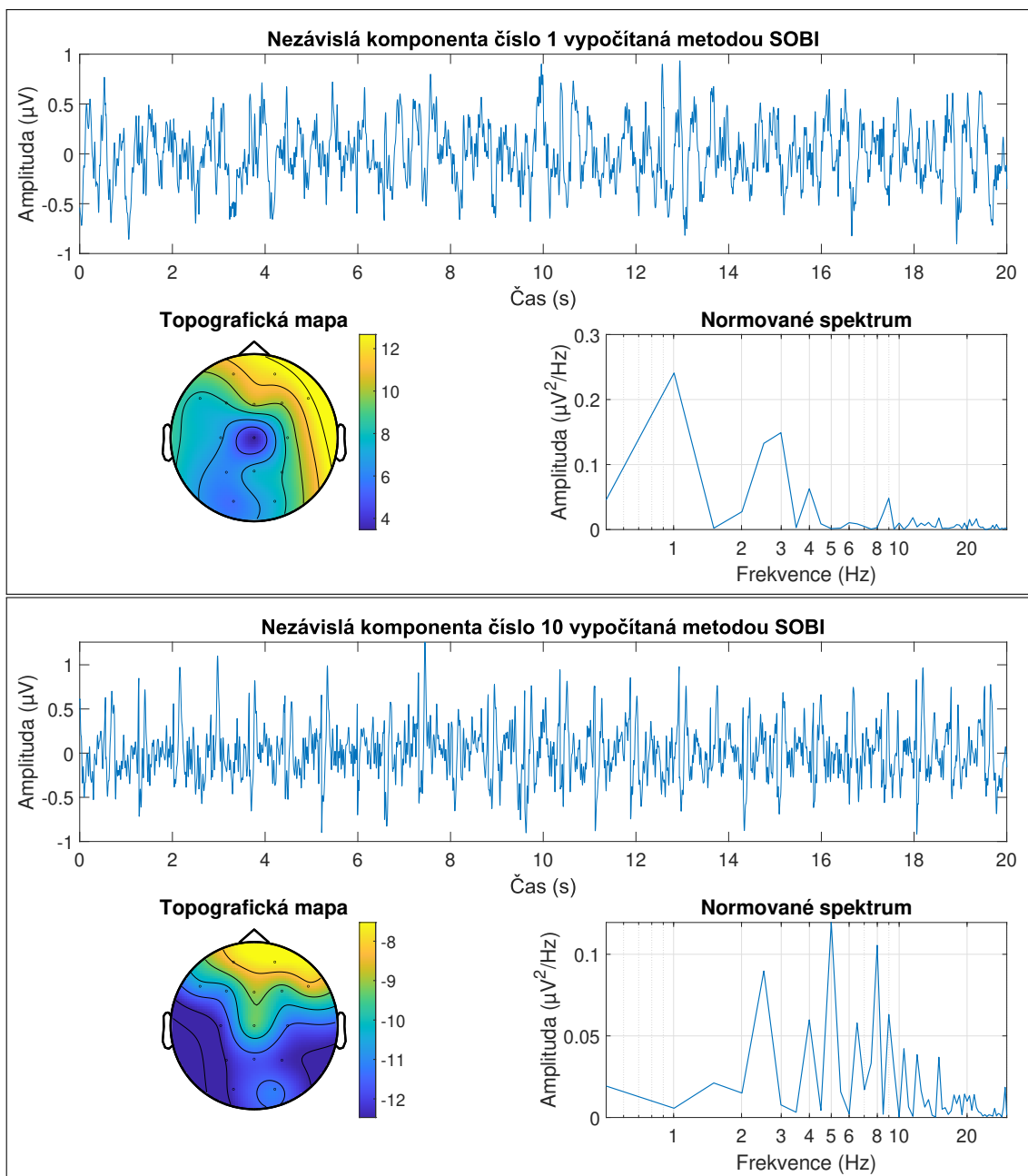
Průměrné topografické mapy shluků metody SOBI



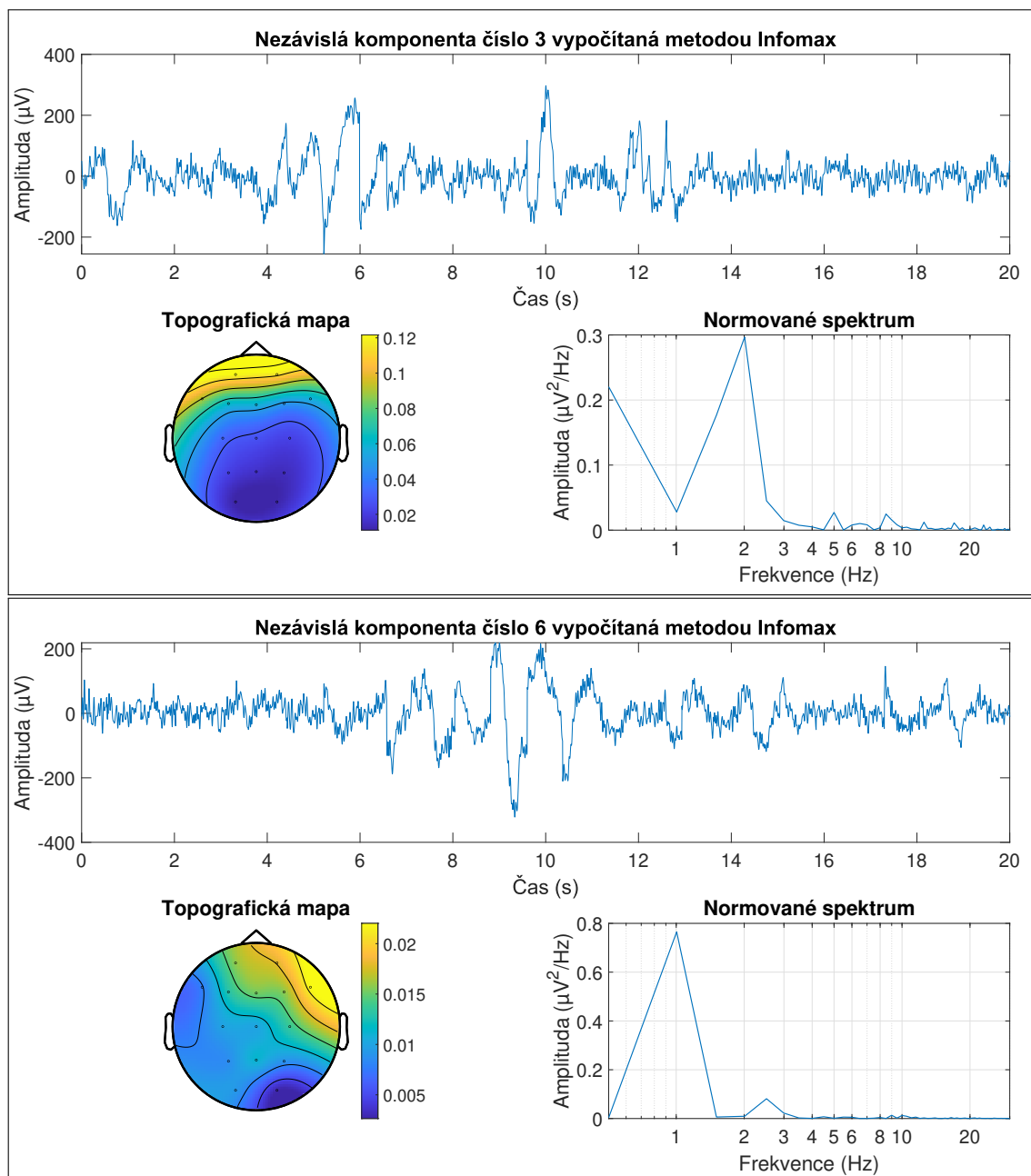
Obrázek 5.46: Průměrné topografické mapy nezávislých komponent vypočítaných metodou SOBI shluku číslo 25–37 a -1. Barevná škála představuje napětí v μV , component = komponenta, číslo komponenty odpovídá číslu shluku.



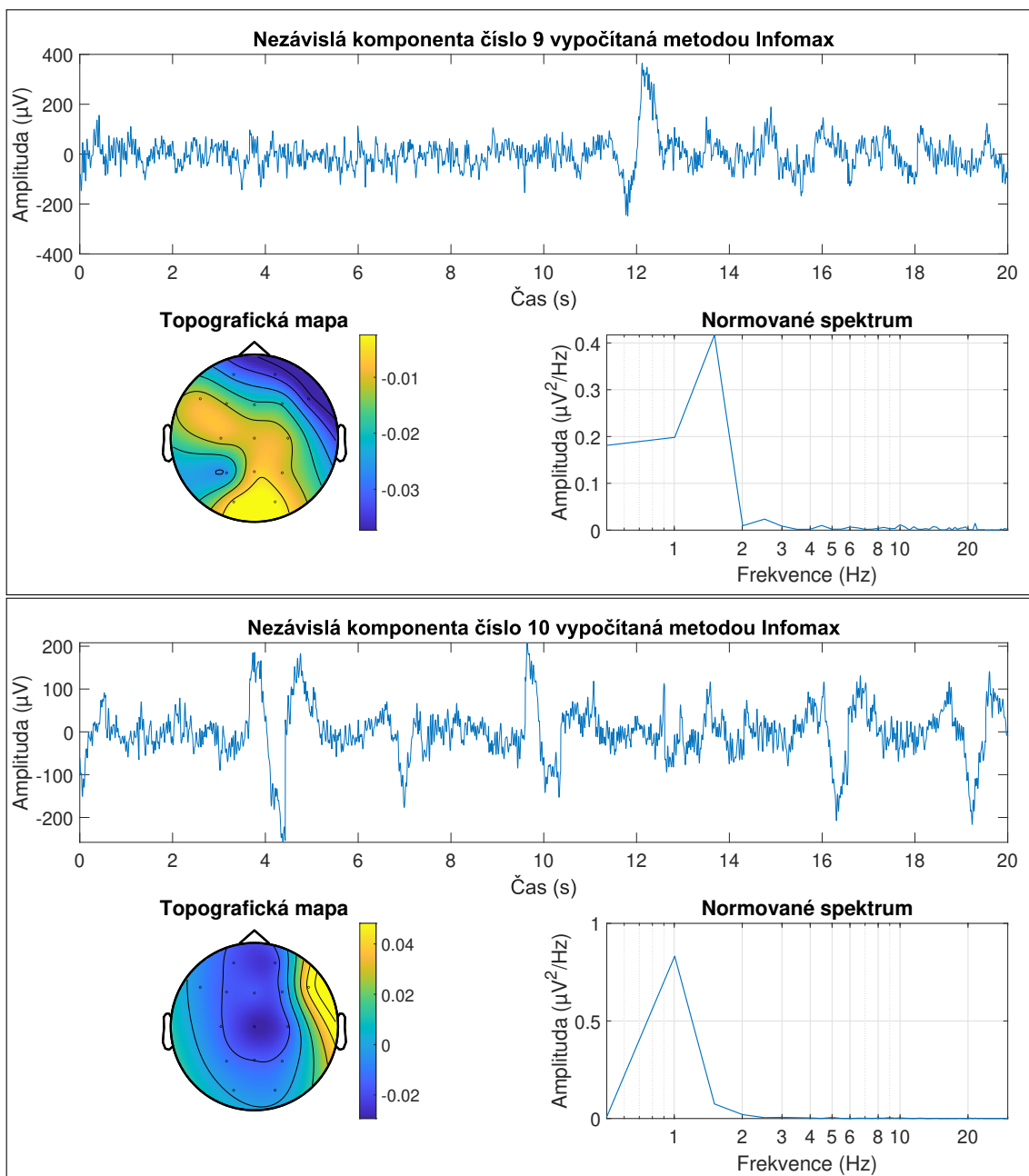
Obrázek 5.47: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 1 vypočítané metodou Infomax (nahore) a nezávislé komponenty číslo 1 vypočítané metodou FastICA (dole). Barevná škála představuje napětí v μV .



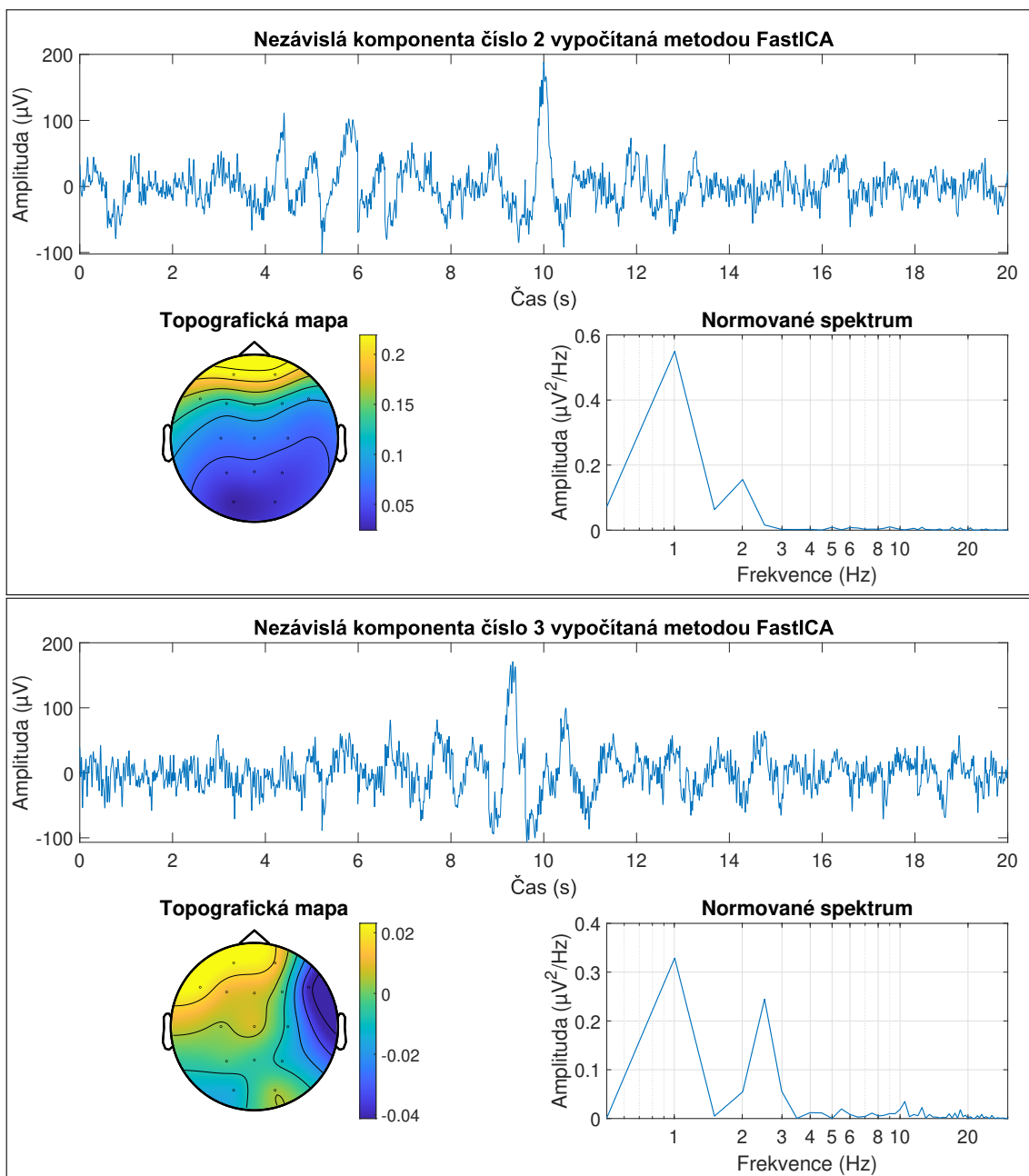
Obrázek 5.48: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 1 vypočítané metodou SOBI (nahore) a nezávislé komponenty číslo 10 vypočítané metodou SOBI (dole). Barevná škála představuje napětí v μV .



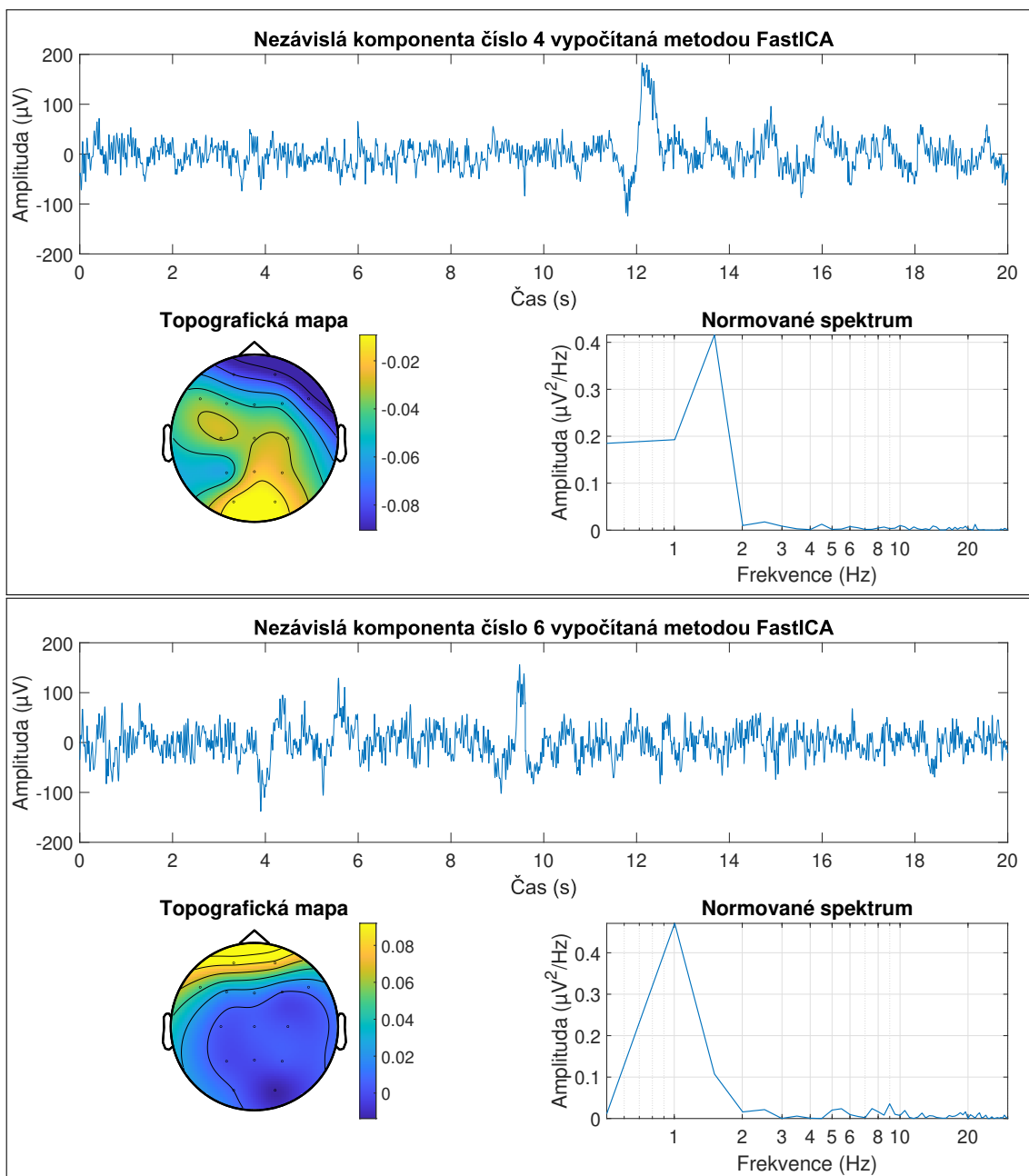
Obrázek 5.49: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 3 (nahore) a číslo 6 (dole) vypočítané metodou Infomax. Barevná škála představuje napětí v μV .



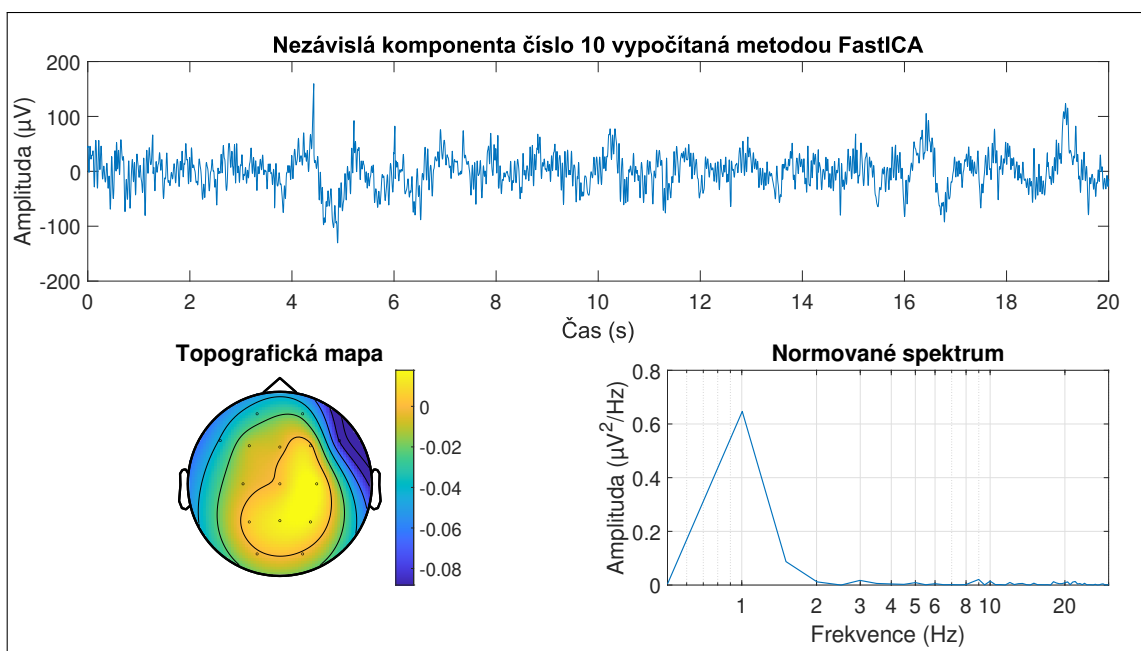
Obrázek 5.50: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 9 (nahore) a číslo 10 (dole) vypočítané metodou Infomax. Barevná škála představuje napětí v μV .



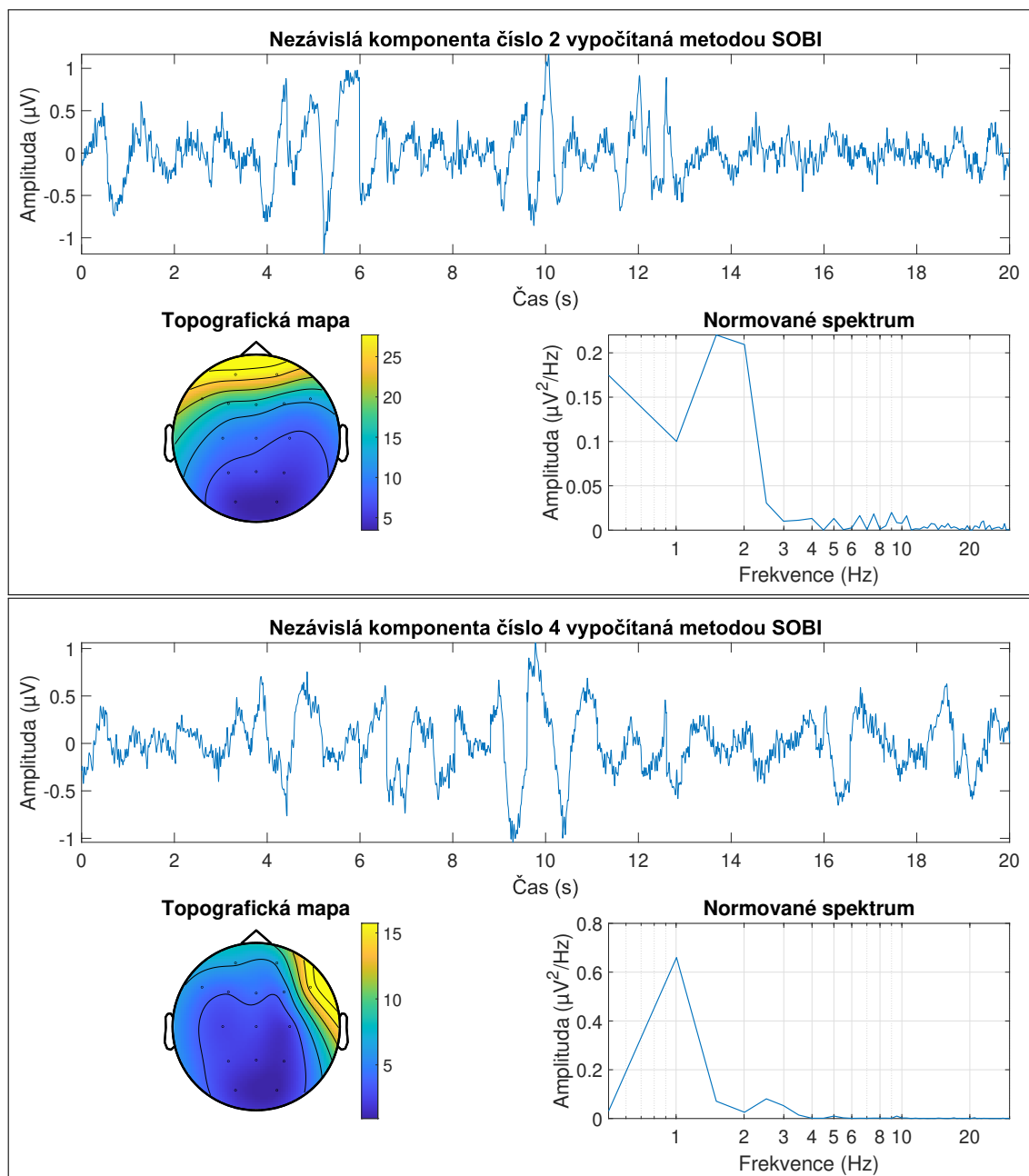
Obrázek 5.51: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 2 (nahore) a číslo 3 (dole) vypočítané metodou FastICA. Barevná škála představuje napětí v μV .



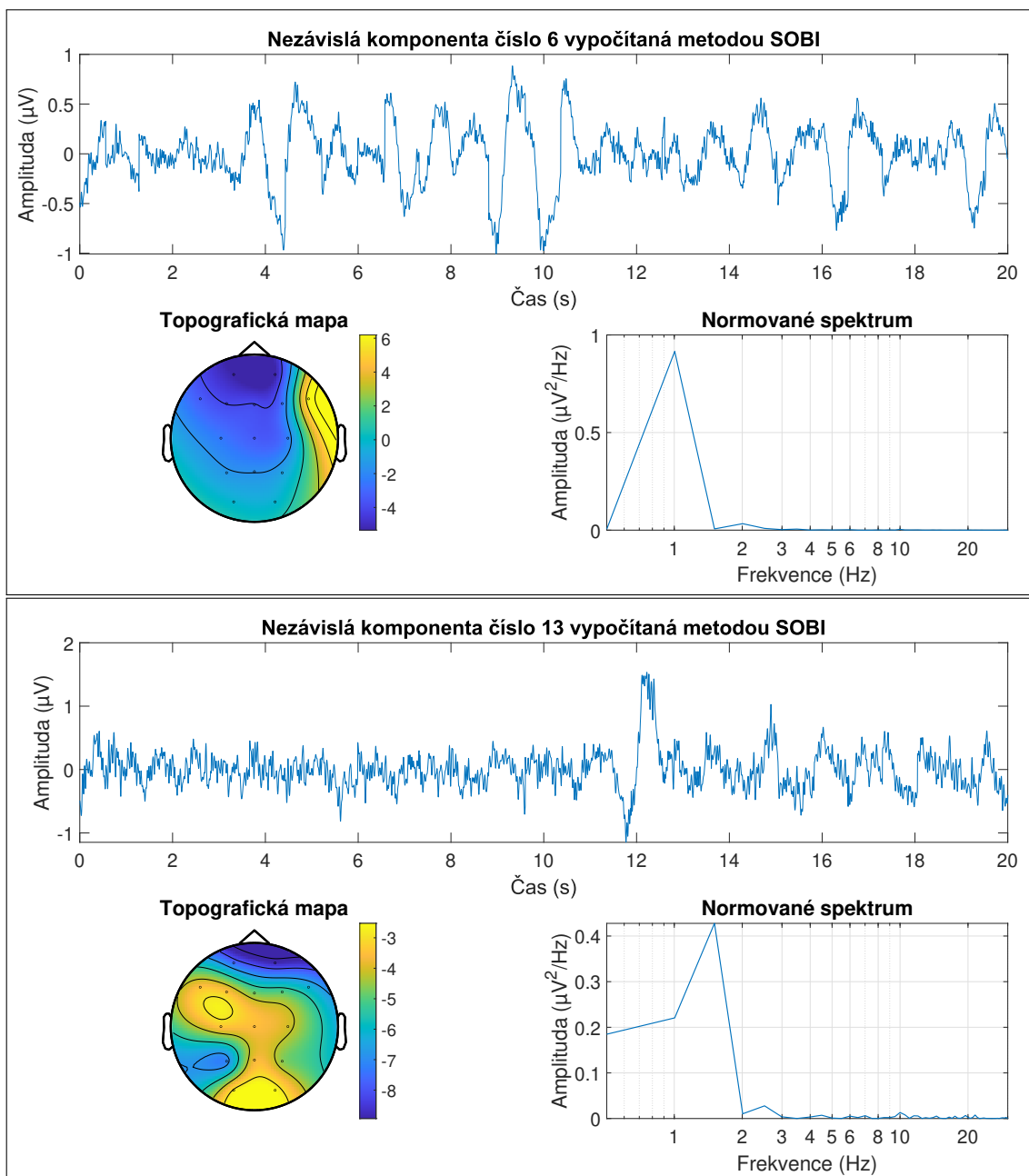
Obrázek 5.52: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 4 (nahore) a číslo 6 (dole) vypočítané metodou FastICA. Barevná škála představuje napětí v μV .



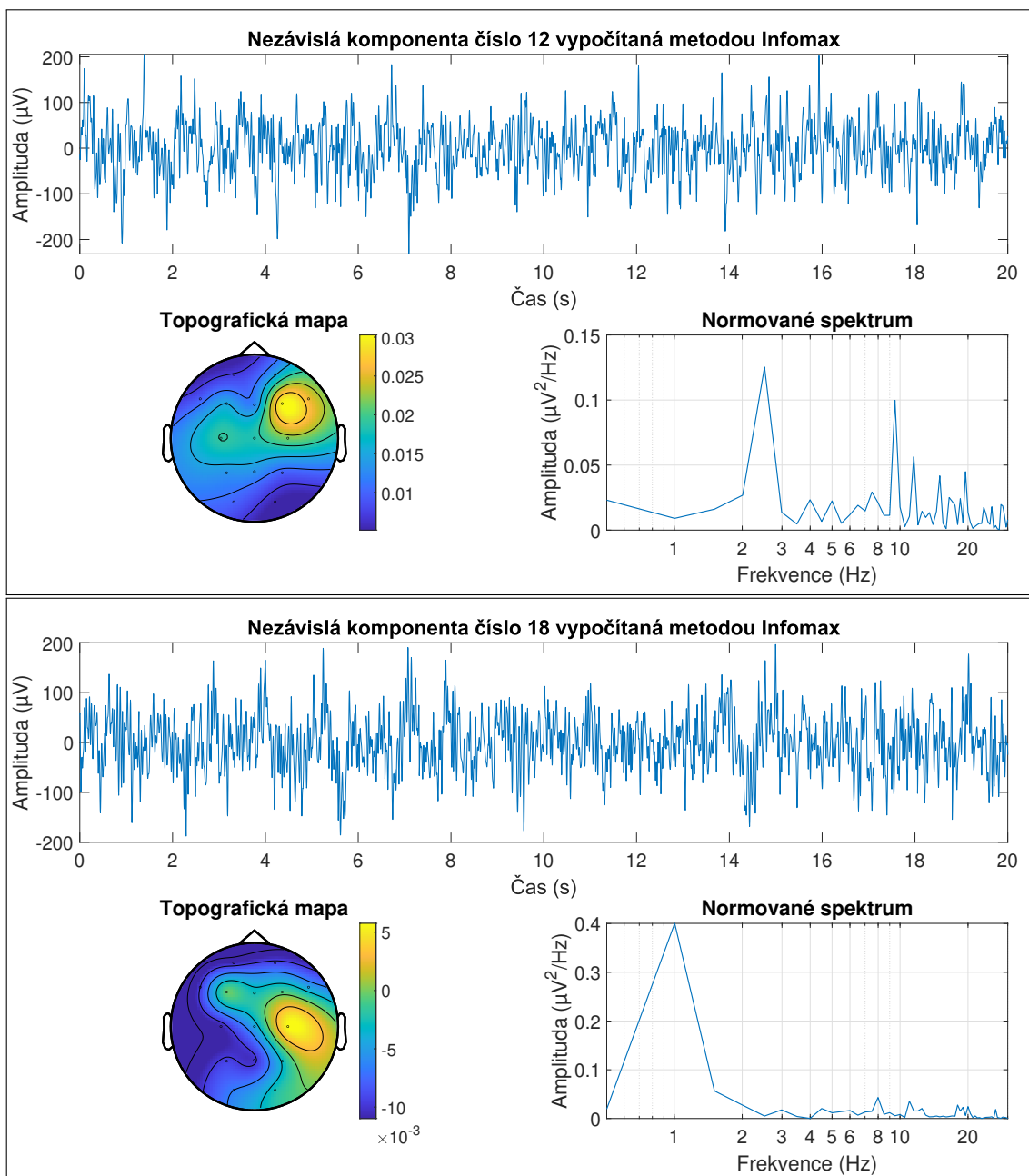
Obrázek 5.53: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 10 vypočítané metodou FastICA. Barevná škála představuje napětí v μV .



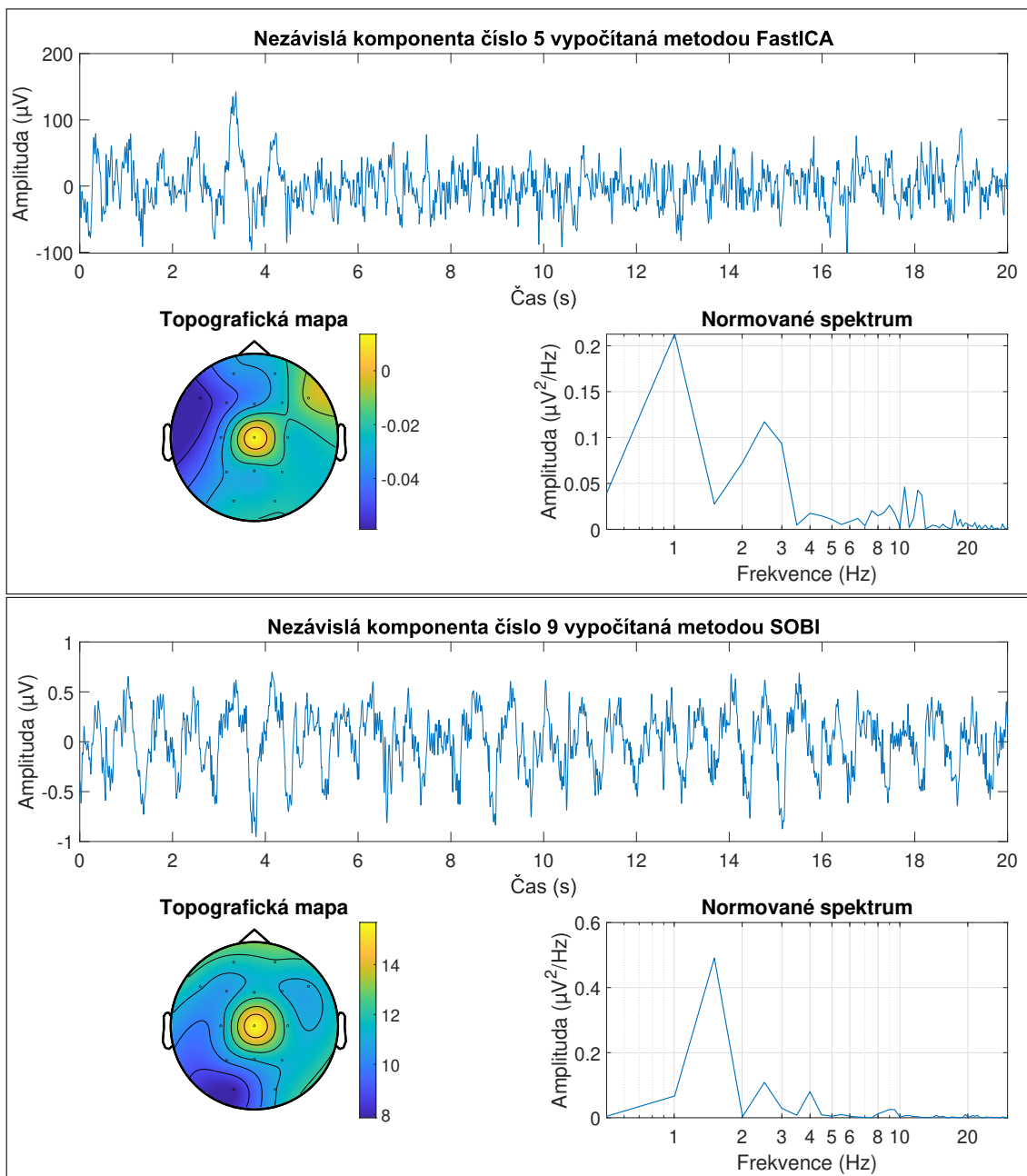
Obrázek 5.54: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 2 (nahore) a číslo 4 (dole) vypočítané metodou SOBI. Barevná škála představuje napětí v μV .



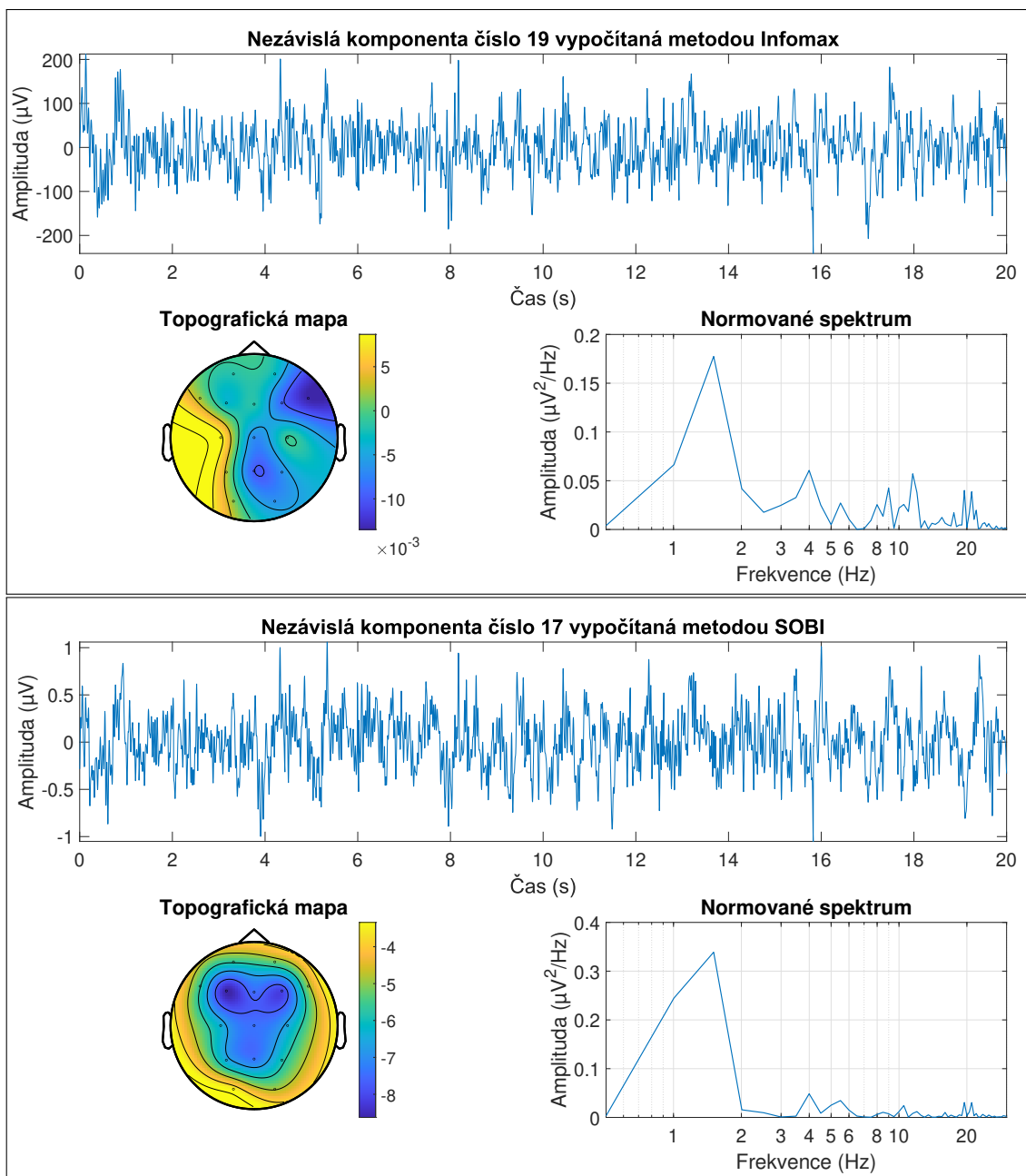
Obrázek 5.55: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 6 (nahore) a číslo 13 (dole) vypočítané metodou SOBI. Barevná škála představuje napětí v μV .



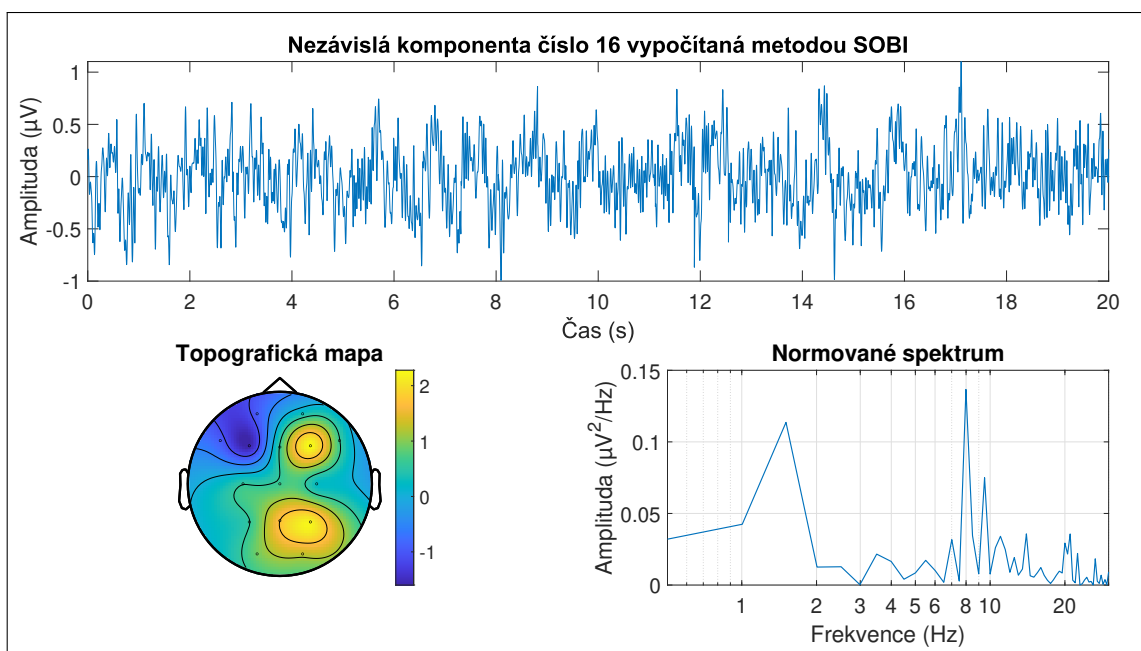
Obrázek 5.56: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 12 (nahore) a číslo 18 (dole) vypočítané metodou Infomax. Barevná škála představuje napětí v μV .



Obrázek 5.57: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 5 vypočítané metodou FastICA (nahore) a číslo 9 vypočítané metodou Infomax (dole). Barevná škála představuje napětí v μV .



Obrázek 5.58: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 19 vypočítané metodou Infomax (nahore) a nezávislé komponenty číslo 17 vypočítané metodou SOBI (dole). Barevná škála představuje napětí v μV .



Obrázek 5.59: Časová řada, topografická mapa a normované výkonové spektrum nezávislé komponenty číslo 16 vypočítané metodou SOBI. Barevná škála představuje napětí v μV .

6 Diskuse

Diskusi jsem kvůli obsáhlosti diplomové práce rozdělila na podkapitoly odpovídající podkapitolám ve výsledcích. Celkové shrnutí bude uvedeno na konci diskuse.

6.1 Předzpracování dat

Předzpracování dat, zejména filtrace a odstranění artefaktů velice ovlivňuje výsledky dalších použitých metod. K filtraci jsem použila dvoucestný filtr s konečnou impulzní odezvou řádu 1234. Řád filtru byl určen pomocí nástroje *Filter Designer* v programovacím prostředí MATLAB. Správnou filtraci dat jsem ověřila frekvenční analýzou provedenou před filtrací a po filtraci. Následně jsem záznamy rozdělila na jednotlivé spánkové fáze a fázi bdění a následně segmentovala na úseky dlouhé 1 minutu.

K odstraňování artefaktů z EEG záznamu se velice často používá analýza hlavních komponent. Nicméně použít ICA, odstranit artefakty a znovu použít ICA k jeho ověření by nebylo vhodné, obzvláště v případě kombinace různých druhů ICA. Proto jsem pro potlačení výrazných artefaktů navrhla jednoduchý detektor založený na principu interkvartilního rozsahu. Pokud se v segmentu objeví vzorek s amplitudou vyšší než nastavený práh, detektor ho označí jako dalekou hodnotu. U každé takto označené daleké hodnoty jsem odstranila část signálu o délce 25 vzorků před označenou hodnotou a 25 vzorků po ní. Celkovou délku odstraněného záznamu 51 vzorků jsem zvolila z toho důvodu, kdyby se v blízkosti odlehlé hodnoty nacházela další. Zůstatkový signál díky tomu nebyl nepřiměřeně dělen. Pokud segment obsahoval artefakty z více než 40 %, byl z dalšího zpracování vyřazen.

Takové odstraňování artefaktů může vést ke špatnému navázání zbývajících částí signálu a k nespojitostem. Použití tohoto detektoru by nebylo vhodné v případě, kdy by se zkoumaný signál dále zpracovával jako spojitý. Nicméně i tato situace by se dala vyřešit použitím spline funkce k plynulému navázání signálu. Další nevýhodou tohoto detektoru je možné detekování fyziologického signálu jako artefaktu. Tím lze pozbýt důležité informace ze signálu. Z hlediska dostupných možností bylo toto odstraňování artefaktů přiměřeným řešením.

6.2 Analýza nezávislých komponent paralelních pokusů

Analýza nezávislých komponent obecně při vícenásobném spuštění na stejných datech může vrátit rozdílné komponenty. Při testování této vlastnosti ICA jsem došla k odlišným závěrům. Metoda Infomax a metoda FastICA skutečně při každém spuštění vrátila trochu odlišné komponenty. Ale metoda SOBI vrátila pokaždé naprosto totožné IC. Tento rozdíl vychází z principů všech metod. Zatímco metody

Infomax a FastICA používají k výpočtu IC statistiky vyšších řádů (špičatost), metoda SOBI používá statistiku druhého řádu (korelační matice).

Korelační koeficienty metod Infomax byly ve většině segmentů vyšší než korelační koeficienty metody FastICA. A to v časových řadách, topografických mapách i výkonových spektrech. Metoda Infomax v každém cyklu extrahovala téměř shodné komponenty, u varianty FastICA se IC v každém pokusu lišily více.

Dalším rozdílem mezi variantami Infomax, SOBI a FastICA je různý počet extrahovaných nezávislých komponent. První dvě varianty ve všech svých pokusech vrátily 19 nezávislých komponent (počet původních smíšených zdrojů byl 19). Varianta FastICA extrahovala počet IC variabilně, ne vždy jich generovala 19. Počet generovaných IC se v některých segmentech napříč pokusy lišil. To vychází z předpokladu, že nesoulad v počtu generovaných IC a větší rozdíly v extrahovaných komponentách napříč pokusy vychází z principu metody FastICA. Při prvním spuštění spočítá jeden sloupec ortogonální nesměšovací matice, (tedy 1 nezávislou komponentu), při druhém spuštění druhý sloupec takový, aby komponenta měla maximální ne-Gaussovo rozložení (byla nezávislá na předchozí), atd. Neupravuje tedy již dříve nalezené sloupce nesměšovací matice. Metoda Infomax odhadne nesměšovací matici ze vstupních dat hned na začátku a poté ji iterativně optimalizuje kvůli maximalizaci entropie (zajištění nezávislosti). Metoda SOBI naproti tomu využívá časovou strukturu směsi zdrojových signálů a statistiku druhého řádu. Další informace bychom mohli získat, pokud bychom se na komponenty získané metodou SOBI podívali statistikou vyšších řádů, např. pomocí kurtózy.

Korelační koeficienty napříč metodami a napříč oblastmi zkoumání se lišily. Všechny hodnoty korelačních koeficientů byly vyšší než 0,99, kromě dvou. Metody Infomax–SOBI a FastICA–SOBI v oblasti časových řad měly nejvyšší koeficienty $> 0,98$. V případě korelace výkonových spekter byly nejvyšší hodnoty $= 1$. Z důvodu vyskytujících se hodnot korelačních koeficientů blížících se 1 a z důvodu zachování stejných podmínek pro všechny metody napříč oblastmi, jsem se rozhodla nastavit nejvyšší korelaci na hodnotu 1.

6.3 Analýza nezávislých komponent ukázkového záznamu

Z výsledků porovnání korelačních koeficientů napříč metodami u ukázkového záznamu vyplývá, že nejlépe spolu v časové oblasti korelují IC metod Infomax–FastICA. Metody měly největší počet korelovaných párů IC, a také největší počet těch silně korelovaných. Na počty korelovaných párů v časové oblasti dopadly metody Infomax–SOBI a SOBI–FastICA shodně. Nicméně absolutní hodnoty korelačních koeficientů měly varianty Infomax–SOBI vyšší než varianty SOBI–FastICA. S počty korelovaných IC souvisí i počty násobně korelovaných časových řad nezávislých kom-

ponent. Metody Infomax–FastICA a Infomax–SOBI měly shodný počet dvojnásobně korelovaných IC. Počet dvojnásobných korelací IC byl u metod SOBI–FastICA o polovinu menší. A od množství korelovaných IC vyplývá množství nespárovaných, nejméně jich bylo u variant Infomax–FastICA, shodně pak u variant Infomax–SOBI a SOBI–FastICA. Nejméně podobné by mohly být metody SOBI a FastICA, ale rozdíl oproti metodám Infomax a SOBI není tak výrazný.

V oblasti topografických map spolu nejlépe korelovaly IC metod Infomax–FastICA. Tyto metody měly největší počet párů korelovaných map i těch silně korelovaných. Nejmenší počet korelovaných párů tvořily metody Infomax–SOBI. Z toho vyplývá nejmenší počet nespárovaných topografických map ve prospěch metod Infomax–FastICA, v neprospěch metod Infomax–SOBI. Nejvíce vícenásobných korelovaných IC měly metody Infomax–FastICA, ale s nízkou násobností, v průměru jedna IC jedné metody korelovala se 2–3 IC metody druhé. U metod Infomax–SOBI násobně korelovalo nejméně párů, ale s vyšší násobností. Jedna IC jedné metody korelovala v průměru s více než 3 IC druhé metody. Z toho vyplývá, že v oblasti topografických map jsou metody Infomax a SOBI nejrozdílnější. Nejpodobnější si naopak jsou varianty Infomax a FastICA.

Nejvíce korelovaných párů výkonových spekter, kromě silně korelovaných, měly varianty SOBI–FastICA, nejméně korelovaných metody Infomax–FastICA. Ale počet silně korelovaných (tedy těch s absolutním korelačním koeficientem větším než 0,8) měly shodně metody Infomax–FastICA a Infomax–SOBI a větší než zbývající pár metod. Z výsledků korelace plynou počty nespárovaných výkonových spekter IC, kterých mají nejvíce metody Infomax–FastICA a nejméně SOBI–FastICA. V oblasti výkonových spekter jsou si nejpodobnější metody SOBI–FastICA, nejméně podobné komponenty mají metody Infomax–FastICA.

V oblasti časových řad spolu celkově korelovaly IC s výraznou specifickou časovou řadou, např. IC představující srdeční artefakty, či ty s výraznou delta aktivitou. V oblasti topografických map korelují IC často představující artefakty ze špatné elektrody, jedná se o výrazný prostorový projev IC. V oblasti výkonových spekter nejvíce korelovaly IC očních artefaktů a srdečního artefaktu. Tyto artefakty se ve spektrální oblasti projevují vysokými špičkami na nízkých frekvencích (většinou 1–2 Hz). Nezávislé páry komponent, které korelovaly ve všech třech oblastech, byly opět výrazné oční artefakty a fyziologická delta aktivita rušená artefaktem ze špatné elektrody.

Vztah mezi korelacemi časových řad, topografických map a výkonových spekter není lineární. V některých případech pár IC koreloval v časové řadě a nikoliv v topografické mapě či výkonovém spektru. V některých případech IC měla nejsilnější korelaci s IC druhé metody v časové oblasti, ale s jinou IC druhé metody v topografické mapě či výkonovém spektru. To může být způsobeno několika faktory. Prvním

z nich může být použití záznamů pouze z 19kanálového EEG. Je možné, že kdybych měla k dispozici data z více kanálů, mohla by se potvrdit linearita korelací, která se teoreticky předpokládá. Dalším důvodem vzniku této nelinearity může být fakt, že jsem brala v potaz pouze nejvyšší hodnoty korelačních koeficientů. Je tedy možné, že IC korelovala se stejnou IC v časové oblasti, topografické mapě i výkonovém spektru, pouze slaběji než s jinou komponentou. V oblasti výkonových spekter může být důvodem nelinearity výpočet výkonů v intervalu 0,5–30 Hz s krokem 0,5 Hz. Je možné, že při zjemnění kroku výpočtu výkonového spektra by se nelinearita mezi korelacemi mohla zmírnit.

Přestože některé topografické mapy matematicky korelovaly, opticky to nemusí být patrné. To může být způsobeno použitou interpolační technikou. Je možné, že při použití jiného způsobu výpočtu chybějících hodnot by topografické mapy vypadaly opticky lépe. Také musíme brát v potaz barevné škály u každé mapy. Pokud bych mapy v rámci jednoho obrázku zobrazila na stejné napěťové úrovni, znamenalo by to v některých případech ztrátu kontrastu a přechodů napětí. Barva mapy by se sjednotila a tím ztratila vypovídající hodnotu.

Otočení některých časových řad kolem osy x je způsobeno tím, že ICA nemůže určit rozptyly (energie) jednotlivých nezávislých komponent, nemůže tedy ani jednoznačně určit znaménko. Jakoukoliv nezávislou komponentu můžeme vynásobit -1 aniž bychom ovlivnili model ICA. Z toho důvodu je také podstatná informace o velikosti Pearsonova korelačního koeficientu, nikoliv o jeho znaménku.

6.4 Analýza nezávislých komponent všech záznamů

V prvním kroku jsem hodnotila, zda se liší množství vrácených nezávislých komponent metodou FastICA od množství, které extrahovaly metody Infomax a SOBI. Obě posledně jmenované metody extrahovaly v každém segmentu 19, metoda FastICA extrahovala množství komponent variabilně. Tento nesoulad souvisí s principem metody FastICA. Normalitu dat jsem testovala Shapiro-Wilkovým testem, který je celkem robustní. Oba výběry nepocházely z rozdělení s normální distribuční funkcí. Proto jsem pro test hypotéz použila neparametrický dvouvýběrový Wilcoxonův test. Na hladině významnosti 5 % jsem zamítla nulovou hypotézu, že oba výběry pochází z rozdělení se stejnou distribuční funkcí. Metoda FastICA se liší od metod Infomax a SOBI v množství extrahovaných IC. Je to jeden z matematických rozdílů mezi metodami. Nicméně rozdíl v počtu IC nemusí být z klinického hlediska v některých případech významné.

V dalších krocích jsem uvedla statistiky počtů korelovaných a nespárovaných IC ve všech třech zájmových oblastech, tedy v oblasti časových řad, topografických map

a výkonových spekter. Páry IC představovaly porovnání napříč všemi metodami, vznikly tak páry Infomax–FastICA, Infomax–SOBI a SOBI–FastICA.

Z porovnání statistiky počtů korelovaných časových řad IC uvedených kombinací metod vyplývá, že všechny varianty měli odlišný medián i modus. Normalitu výběrů jsem opět ověřila Shapiro–Wilkovým testem. Tento test jsem pro ověření normality kvůli jeho dobré robustnosti použila ve všech situacích. Také jsem ve všech situacích stanovila hladinu významnosti na 5 %. Ani jeden z výběrů nepocházel z rozdělení s normální distribuční funkcí. Pro testování nulové hypotézy, že výběry pochází z rozdělení se stejnou distribuční funkcí, jsem použila dvouvýběrový Wilcoxonův test. U všech kombinací korelovaných nezávislých komponent napříč metodami jsem nulovou hypotézu zamítla.

Nespárované časové řady IC napříč metodami měly také odlišný medián, ale menší než u korelovaných časových IC. Z průměrů korelovaných a nespárovaných časových řad vyplývá, že průměrně korelovalo více IC než zůstalo těch nespárovaných. Ani jeden z výběrů opět neměl normální rozdělení. Wilcoxonův test u všech zamítl nulovou hypotézu. Ani jeden výběr nepochází z rozdělení se stejnou distribuční funkcí. Je tedy pravděpodobné, že počty nespárovaných časových řad se liší napříč metodami ICA. Dále z výsledků můžeme předpokládat, že nejpodobnější časové řady nezávislých komponent mají metody Infomax a FastICA. To je stejný výsledek jako u ukázkového záznamu. Naopak nejslabší korelaci vykazovaly metody SOBI–FastICA, opět stejné jako v ukázkovém záznamu.

Korelace topografických map IC všech metod ukázala, že oproti časovým řadám korelovalo méně IC. Pokles byl nejvýraznější u metod Infomax–FastICA, kde se jednalo v průměru o 5 IC méně. Nejmenší pokles korelovaných IC zaznamenala metody SOBI–FastICA, v průměru to byly necelé 2 IC. Toto množství komponent se předvídatelně z korelovaných IC přesunulo do nespárovaných IC. Možné důvody této nelinearity jsem popsala v předchozí podkapitole 6.3. Ani jeden z výběrů korelovaných topografických map IC a nespárovaných topografických map IC nespĺňoval podmínky pro použití parametrických testů. Wilcoxonův test na hladině významnosti 5 % zamítl nulovou hypotézu o tom, že výběry pochází z rozdělení se stejnou distribuční funkcí, a to jak u skupiny korelovaných IC, tak i u nespárovaných. Z výsledků plyne, že nejpodobnější metody v oblasti topografických map jsou metody Infomax a FastICA, jak v případě korelace časových řad, tak v případě ukázkového segmentu. Nejslabší korelaci v případě topografických map vykazovaly metody Infomax a SOBI, což je rozdíl oproti korelaci časových řad, ale opět stejný výsledek jako v případě ukázkového záznamu. Je to zvláštní rozdíl, že v časových řadách korelují lépe/hůře jiné metody než v topografických mapách. Přitom tyto dvě oblasti spolu matematicky přímo souvisí.

Předpokládám, že by to mohlo souviset s násobnými korelacemi IC, jak bylo naznačeno u ukázkového segmentu. Metody SOBI–FastICA měly menší počet násobných korelací v časových řadách než metody Infomax–SOBI. Naopak metody Infomax–SOBI vykazovaly vyšší násobnost korelací topografických map než metody SOBI–FastICA. Dalším důvodem může být to, že komponenty, které spolu korelují, mohou sice v časové oblasti vypadat podobně, ale jejich zdroje se v mozku mohou nacházet na odlišných místech, potom i topografická mapa bude vypadat trochu odlišně. Např. pokud budeme mít v časové řadě 2 frekvenční zdroje, jeden bude pocházet z okcipitální oblasti mozku a druhý z frontální, bude topografická mapa vypadat jinak než u druhé korelované komponenty, která bude korelovat díky stejné první frekvenční složce v časové řadě, ale v topografické mapě bude zdroj pocházet jen z frontální oblasti.

Nejvyššího průměrného počtu korelovaných výkonových spekter IC dosáhly metody Infomax–FastICA, přestože tento počet nepatrně klesl oproti průměrnému počtu korelovaných map (o necelou 1 IC). U metod Infomax–SOBI vzrostl o více než 1 IC a průměrné množství korelovaných IC u metod SOBI–FastICA zůstalo téměř stejné. Modus a medián mají metody Infomax–SOBI a SOBI–FastICA stejné, nicméně ostatní popisné parametry se liší. Stejně je to pochopitelně i u nespárovaných výkonových spekter. Opět ani jeden z výběrů korelovaných a nespárovaných výkonových spekter nesplňoval předpoklad normality pro parametrický test. Výsledné p-hodnoty Wilcoxonova testu byly výrazně nižší než stanovená hladina významnosti. U všech výběrů korelovaných výkonových spekter všech metod jsem zamítla nulovou hypotézu, že by výběry pocházely z rozdělení se stejnou distribuční funkcí. Taktéž i u nespárovaných výkonových spekter IC. Je pravděpodobné, že se metody ICA liší i v oblasti výkonových spekter. Nejlépe korelovala výkonová spektra metod Infomax a FastICA, což je stejný výsledek jako v předchozích oblastech. Nicméně u ukázkového záznamu vyšly nejlépe varianty SOBI a FastICA. Jednoznačnou odpověď na to, které metody dopadly v oblasti výkonových spekter nejhůře, jsem nenalezla. Průměr korelovaných IC metod Infomax–SOBI byl větší než u metod SOBI–FastICA. Ale průměry nespárovaných IC vyšly lépe ve prospěch metod SOBI–FastICA. Tento zdánlivě protichůdný jev se dá vysvětlit odlišným rozložením množství IC u obou párů metod, které mohly vychýlit průměr. Mediány obou párů metod jsou totiž shodné u skupiny korelovaných i u skupiny nespárovaných výkonových spekter IC.

6.5 t-SNE a DBSCAN

Metoda t-SNE je nelineární metoda redukce dimenze. Při redukování prostoru k sobě přibližuje podobné objekty a husté shluky objektů rozšiřuje. Uživatel na vstupu nastavuje hodnoty perplexity, výsledek redukce je tedy částečně závislý na uživateli.

Ve své práci jsem redukovala 19D prostor topografických map. Nejlepší separace shluků jsem dosáhla s nastavenou perplexitou na hodnotu 40 a euklidovskou vzdálenostní metrikou. V redukovaném prostoru se sice nacházelo určité množství menších oddělených shluků, ale většina map zůstala v jednom velkém shluku, ve kterém se na určitých místech nacházely objekty s menší hustotou a na dalších s větší hustotou. Vnitřní struktura byla naznačena jen nepatrně. To mohlo být dáno tím, že je to z velké části aktivita mozku (artefakty by měly být minoritní), a že t-SNE není schopno oddělit jednotlivé aspekty fyziologických projevů mozku.

Na redukovaném 2D prostoru jsem provedla klasifikaci shlukovacím algoritmem DBSCAN. Výsledek klasifikace lze ovlivnit nastavením informací o velikosti požadovaných shluků a o minimálním počtu objektů ve shluku. Zpracovala jsem výsledky pro dvě různá nastavení DBSCANu. V prvním případě jsem se snažila najít nějakou vnitřní strukturu v redukovaném prostoru map, ve kterém byly velice mírně naznačeny hranice shluků již pomocí t-SNE. Na základě vlastností časových řad, topografických map a výkonových spekter náležejících k daným shlukům nešlo jednoznačně určit, že by shluk obsahoval pouze konkrétní aktivitu či artefakt. Většina shluků obsahovala směs různých aktivit a artefaktů.

Při druhém nastavení DBSCANu jsem se zaměřila naopak na analýzu okolních oddělených shluků. Okolní shluky byly většinou tvořeny artefakty ze špatné elektrody. Průměrné topografické mapy byly hodně výrazné, takže metoda t-SNE a následná klasifikace DBSCANem byla v tomto případě účinná. Úspěšné oddělení topografických map představujících artefakty ze špatné elektrody je způsobeno charakteristickým rozložením nezávislé komponenty v prostoru topomapy. Topografická mapa tohoto artefaktu se výrazně liší od topografických map jiných artefaktů či aktivity. Metoda t-SNE proto tyto výrazně jiné topografické mapy oddělila do samostatných shluků od ostatních map.

Každý shluk představoval jinou elektrodu případně jinou polaritu stejné elektrody. Důvodem, proč jeden shluk neobsahoval artefakt ze špatné elektrody v obou polaritách tkví v principu metody t-SNE. Ta totiž bere v úvahu i znaménko u vah topografické mapy, ne jen její absolutní hodnotu.

Při stejném nastavení klasifikačního algoritmu DBSCAN napříč metodami byl 2D prostor topografických map metody SOBI pokaždé klasifikován do více shluků než u metod Infomax a FastICA, aniž by odhalil nějaké specifické shluky pro tuto metodu. A dále pouze u varianty SOBI se vyskytly některé shluky artefaktů stejných elektrod se stejnou polaritou. Z těchto důvodů se zdá, že metoda SOBI se více liší od metod Infomax a FastICA, jejichž redukovaný a klasifikovaný prostor byl sobě více podobný.

6.6 Charakter komponent pro rozdílné metody potlačení artefaktů

Charakter komponent pro rozdílné metody potlačení artefaktů jsem prezentovala na ukázkovém záznamu. K prezentaci jsem vybrala segment bdění z toho důvodu, že se zde hojně vyskytovaly oční artefakty.

Srdeční artefakt dokázaly oddělit lépe metody Infomax a FastICA. I když ani zde nebyla extrakce ideální, protože v komponentě zůstaly artefakty ze špatné elektrody. Což bylo nejvíce vidět v topografické mapě. V časové řadě byl srdeční artefakt dobře viditelný. Metoda SOBI nedokázala extrahovat jednu komponentu se srdečním artefaktem, ale rozložila ho do minimálně dvou komponent. V časové řadě první komponenty nebyl artefakt dobře viditelný, lépe byl vidět v topografické mapě a výkonovém spektru. V první komponentě byl kromě srdečního artefaktu viditelný i artefakt ze špatné elektrody. V druhé komponentě metody SOBI je srdeční artefakt lépe viditelný v časové řadě a topografické mapě. Výkonové spektrum mu úplně neodpovídá, což je patrně způsobeno tím, že komponenta obsahuje jen část srdečního artefaktu.

Oční artefakty extrahované metodami Infomax byly dobře viditelné v časové oblasti, topografických mapách a zejména ve výkonovém spektru. Oční pohyby a mrkání zřejmě bylo natolik výrazné, že v některých topografických mapách je vidět aktivita táhnoucí se od fronto-polárních elektrod až do zadních částí mozku. Metoda Infomax a SOBI extrahovaly 4 IC s výraznými očními artefakty. Metoda FastICA extrahovala těchto komponent 5. Nicméně i u této metody jsou oční artefakty dobře rozpoznatelné v časové oblasti, topografické mapě i výkonovém spektru. Důvodů, proč metoda FastICA extrahovala více komponent představujících oční artefakty, může být více. Za prvé je možné, že metoda FastICA dokázala lépe extrahovat oční artefakty, zatímco u další metod mohly tyto artefakty zůstat smíchány s dalšími komponentami. Nebo za druhé FastICA mohla extrahovat oční artefakty a rozdělit je do 5 místo do 4 IC jako zbylé metody. V tom případě by musel být k očním artefaktům přimíchán i další signál, aby se zachovala nezávislost komponent. Metoda SOBI extrahovala 4 IC s očními artefakty, nicméně 2 IC si jsou velice podobné v časové řadě, topografické mapě i výkonovém spektru. Je možné, že kdybychom tyto 2 IC komponenty porovnali pomocí statistiky vyšších řádů, nebyla by jejich nezávislost tak silná, jak bychom předpokládali.

Zajímavé je, že metoda Infomax odhalila 2 špatné elektrody, ale jiné než jaké odhalily metody FastICA a SOBI. Artefakty ze špatné elektrody jsou dobře rozpoznatelné v záznamu EEG díky jejich vyšší amplitudě, ale nejsou dobře rozpoznatelné v časové řadě a ve výkonovém spektru nezávislých komponent. To z toho důvodu, že metoda ICA nezachová poměry amplitud a upravuje je na shodnou úroveň. Tyto

artefakty jsou jsou jednoznačně k poznání v topografické mapě. FastICA a SOBI extrahovaly komponentu s artefaktem ze špatné elektrody s označením Cz. Metoda Infomax extrahovala komponenty s artefakty ze špatné elektrody s označením F4 a C4. I když u metody Infomax nebyly topografické mapy tak výrazné, patrně k těmto artefaktům byl přimíchán i další signál či artefakt. Elektrody Cz a C4 spolu sousedí ve frontální rovině, elektrody Cz a F4 sousedí diagonálně. V časové oblasti ani ve frekvenčním spektru nejsou IC SOBI a FastICA příliš podobné, v obou případech se tak zřejmě kromě dominantního artefaktu bude vyskytovat i jiná aktivita či artefakt. Špatná elektroda Cz se částečně objevila i v komponentách představujících srdeční artefakty. Extrakce tohoto artefaktu tedy nebyla úplně ideální.

Ani jedna z metod nedokázala extrahovat komponenty představující pouze svalové artefakty. EMG artefakty tak byly rozprostřeny do více komponent, ve kterých se vyskytovala i jiná aktivita či jiné artefakty, např. fyziologická, či artefakt ze špatné elektrody. Z některých studií [45] zabývajících se separací EMG artefaktů z EEG signálu pomocí metod ICA vyplývá, že ICA nedokáže adekvátně oddělit myogenní aktivitu od neurogenních zdrojů. Jiné zdroje [46] naznačují, že některé metody ICA jsou lepší v extrakci komponent obsahujících svalové artefakty než jiné. Z mých výsledků nedokáži určit, zda je některá z ICA metod k odstraňování svalových artefaktů lepší.

Některé extrahované komponenty tvořily směsi EEG signálu a artefaktů. Konkrétně jsem prezentovala IC metody SOBI. Ta představovala směs fyziologického signálu, myogenní aktivity a artefaktů ze špatné elektrody. Konkrétně se jednalo o artefakty z elektrod F4 a C4, které dokázala metoda Infomax lépe oddělit.

6.7 Shrnutí

Nejsilnější korelace časových řád, topografických map a výkonových spekter dosahovaly nezávislé komponenty metod Infomax a FastICA. Tyto dvě metody jsou si po matematické stránce nejpodobnější. Metoda SOBI se více liší. Z pohledu izolování očních a srdečního artefaktu jsou si metoda Infomax a FastICA velmi podobné, nelze upřednostnit jednu z nich.

Redukce dimenze prostoru topografických map a následná klasifikace neodhalila žádnou vnitřní strukturu specifickou pro určitou variantu ICA. Ale odhalila, že metoda t-SNE dokáže dobře oddělit topografické mapy představující artefakty ze špatné elektrody. 2D prostor map byl u metody SOBI klasifikován do více shluků než u metod Infomax a FastICA, a také byl na první pohled více odlišný.

Omezením této studie může být realizace na 19 kanálových záznamech. Je možné, že počet zdrojů převyšoval počet kanálů. Nicméně zatím je stále 19 kanálové EEG nejběžněji využívané. Je pravděpodobné, že bych získala odlišné výsledky

při použití 256 kanálových EEG záznamů. V tomto případě by nejprve bylo nutné 256dimenzionální prostor nejprve redukovat, protože získáním 256 IC by byl původní signál příliš rozložen.

Velkým přínosem této studie je její robustnost. Analyzovala jsem 213 EEG záznamů od 60 pacientů. Po segmentaci obsahoval dataset téměř 4500 segmentů, které jsem následně analyzovala třemi metodami ICA.

Z analýzy vyplývá, že není zcela vhodné kombinovat postupnou analýzou různé druhy ICA. Například v současné době je často využíván algoritmus ADJUST (obsahující ICA Infomax) pro potlačení artefaktů a uživatelé dále využívají při zpracování svých datasetů následnou ICA analýzu. Konkrétní výběr algoritmu závisí na dané studii, nicméně je vhodné dodržet jednu variantu.

7 Závěr

Spánkové EEG záznamy jsem podrobila analýze nezávislých komponent pomocí tří nepoužívanějších variant ICA. Časové řady, topografické mapy a výkonová spektra jsem porovnála z hlediska jejich podobnosti či rozdílnosti pomocí Pearsonova korelačního koeficientu. Z tohoto matematického porovnání vyplývá, že metody Infomax a FastICA jsou sobě podobné. Metoda SOBI se více lišila od obou předchozích. Prostor topografických map jsem redukovala metodou t-SNE a klasifikovala shlukovacím algoritmem DBSCAN. Obojí při dvou různých nastavení klasifikátoru. První nastavení neodhalilo vnitřní strukturu specifickou pro některou z metod. Klasifikované shluky se nedaly přesně charakterizovat a přiřadit k některému artefaktu či fyziologické aktivitě. Nejpodobnější shluky a celkově nejpodobnější prostor vykazovaly metody Infomax a FastICA. 2D prostor topografických map metody SOBI a klasifikované shluky se lišily od předchozích metod více. Z porovnání charakteru IC představujících artefakty je jasné, že ani metoda ICA nedokáže plně oddělit fyziologickou aktivitu od artefaktů. Pro ukázkový záznam popsáný v textu práce se to nejlépe podařilo u variant Infomax a FastICA.

Závěr práce je takový, že varianty ICA nelze považovat za rovnocenné, a to jak z matematického hlediska, tak z klinického. Nejpodobnější jsou si metody Infomax a FastICA. Metoda SOBI se od obou předchozích více liší. Obecně nemohu říci, která z variant je nejlepší. Záleží vždy na dané studii, pro kterou z variant se rozhodnout. Z práce ale vyplývá, že je nutné dodržet jednotnou variantu ICA v rámci zpracování dat.

Seznam použité literatury

1. URIGÜEN, Jose Antonio; GARCIA-ZAPIRAIN, Begoña. EEG artifact removal—state-of-the-art and guidelines. *Journal of Neural Engineering*. 2015, roč. 12, č. 3, s. 031001. Dostupné z DOI: [⟨10.1088/1741-2560/12/3/031001⟩](https://doi.org/10.1088/1741-2560/12/3/031001).
2. SUBHA, D. Puthankattil; JOSEPH, Paul K.; U, Rajendra Acharya; LIM, Choo Min. EEG Signal Analysis: A Survey. *Journal of Medical Systems*. 2010, roč. 34, č. 2, s. 195–212. ISSN 0148-5598. Dostupné z DOI: [⟨10.1007/s10916-008-9231-z⟩](https://doi.org/10.1007/s10916-008-9231-z).
3. SANEI, Saeid.; CHAMBERS, Jonathon. *EEG signal processing*. Hoboken, NJ: John Wiley & Sons, 2007. ISBN 978-0-470-02581-9.
4. Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology* [online]. 1958, roč. 10, č. 2, s. 370–375 [cit. 2020-04-16]. ISSN 00134694. Dostupné z DOI: [⟨10.1016/0013-4694\(58\)90053-1⟩](https://doi.org/10.1016/0013-4694(58)90053-1).
5. MALMIVUO, Plonsey; MALMIVUO, Jaakko; PLONSEY, Robert. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.
6. SILBER, Michael H et al. The visual scoring of sleep in adults. *Journal of clinical sleep medicine*. 2007, roč. 3, č. 02, s. 22–22.
7. GORDON, S.M.; LAWHERN, V.; PASSARO, A.D.; MCDOWELL, K. Informed decomposition of electroencephalographic data. *Journal of Neuroscience Methods* [online]. 2015, roč. 256, s. 41–55 [cit. 2020-04-19]. ISSN 01650270. Dostupné z DOI: [⟨10.1016/j.jneumeth.2015.08.019⟩](https://doi.org/10.1016/j.jneumeth.2015.08.019).
8. ZAVALA-FERNÁNDEZ, Heriberto; SANDER, Tilmann H; BURGHOFF, Martin; ORGLMEISTER, Reinhold; TRAHMS, Lutz. Comparison of ICA algorithms for the isolation of biological artifacts in magnetoencephalography. In: *International Conference on Independent Component Analysis and Signal Separation*. 2006, s. 511–518.
9. ALBERA, Laurent; KACHENOURA, Amar; COMON, Pierre; KARFOUL, Ahmad; WENDLING, Fabrice; SENHADJI, Lotfi; MERLET, Isabelle. ICA-based EEG denoising: a comparative analysis of fifteen methods. *Bulletin of the Polish Academy of Sciences: Technical Sciences*. 2012, roč. 60, č. 3, s. 407–418.
10. DELORME, Arnaud; PALMER, Jason; ONTON, Julie; OOSTENVELD, Robert; MAKEIG, Scott; WARD, Lawrence M. Independent EEG Sources Are Dipolar. *PLoS ONE* [online]. 2012, roč. 7, č. 2 [cit. 2020-03-01]. ISSN 1932-6203. Dostupné z DOI: [⟨10.1371/journal.pone.0030135⟩](https://doi.org/10.1371/journal.pone.0030135).

11. DELORME, Arnaud; SEJNOWSKI, Terrence; MAKEIG, Scott. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*. 2007, roč. 34, č. 4, s. 1443–1449.
12. RAIESDANA, Somayeh. Automated sleep staging of OSAs based on ICA pre-processing and consolidation of temporal correlations. *Australasian Physical & Engineering Sciences in Medicine* [online]. 2018, roč. 41, č. 1, s. 161–176 [cit. 2020-04-19]. ISSN 0158-9938. Dostupné z DOI: <10.1007/s13246-018-0624-0>.
13. DHONGADE, Dayanand Vishwanath; RAO, TVKH. Classification of sleep disorders based on EEG signals by using feature extraction techniques with KNN classifier. In: *2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)*. 2017, s. 1–5.
14. JONMOHAMADI, Yaqub; POUDEL, Govinda R.; INNES, Carrie C. R. H.; JONES, Richard D. Microsleeps are Associated with Stage-2 Sleep Spindles from Hippocampal-Temporal Network. *International Journal of Neural Systems* [online]. 2016, roč. 26, č. 04 [cit. 2020-04-19]. ISSN 0129-0657. Dostupné z DOI: <10.1142/S0129065716500155>.
15. LEE, Te-Won; GIROLAMI, Mark; BELL, Anthony J; SEJNOWSKI, Terrence J. A unifying information-theoretic framework for independent component analysis. *Computers & Mathematics with Applications*. 2000, roč. 39, č. 11, s. 1–21.
16. LAUBACH, Mark; SHULER, Marshall; NICOLELIS, Miguel A.L. Independent component analyses for quantifying neuronal ensemble interactions. *Journal of Neuroscience Methods* [online]. 1999, roč. 94, č. 1, s. 141–154 [cit. 2020-05-13]. ISSN 01650270. Dostupné z DOI: <10.1016/S0165-0270(99)00131-4>.
17. ESPOSITO, Fabrizio; FORMISANO, Elia; SEIFRITZ, Erich; GOEBEL, Rainer; MORRONE, Renato; TEDESCHI, Gioacchino; SALLE, Francesco Di. Spatial independent component analysis of functional MRI time-series: To what extent do results depend on the algorithm used? *Human Brain Mapping* [online]. 2002, roč. 16, č. 3, s. 146–157 [cit. 2020-05-13]. ISSN 1065-9471. Dostupné z DOI: <10.1002/hbm.10034>.
18. PIORECKY, Marek; RUBESOVA, Pavlina; CERNA, Eva; MILETINOVA, Eva; KOPRIVOVA, Jana; BUSKOVA, Jitka. Computer Analysis of Hidden Layers of Sleep EEG: Dreams Correlates. In: *2019 E-Health and Bioengineering Conference (EHB)* [online]. IEEE, 2019, s. 1–4 [cit. 2020-03-06]. ISBN 978-1-7281-2603-6. Dostupné z DOI: <10.1109/EHB47216.2019.8969907>.

19. ABBAS, Nasir. A robust S 2 control chart with Tukey's and MAD outlier detectors. *Quality and Reliability Engineering International* [online]. 2020, roč. 36, č. 1, s. 403–413 [cit. 2020-03-13]. ISSN 0748-8017. Dostupné z DOI: [10.1002/qre.2588](https://doi.org/10.1002/qre.2588).
20. TUKEY, John Wilder. *Exploratory data analysis*. Reading, Mass.: Addison-Wesley Pub. Co., 1977. ISBN 978-0201076165.
21. HYVÄRINEN, Aapo; KARHUNEN, Juha; OJA, Erkki. *Independent component analysis*. 1st. New York: John Wiley & Sons, Inc., 2001. ISBN 9780471405405.
22. HYVÄRINEN, Aapo; OJA, Erkki. Independent component analysis: algorithms and applications. *Neural networks*. 2000, roč. 13, č. 4-5, s. 411–430.
23. STONE, James V. *Independent component analysis: a tutorial introduction*. MIT press, 2004.
24. PLANK, Markus. *Independent Component Analysis – demystified!: Support & Tips* [online]. Germany: © Brain Products GmbH, 2014 [cit. 2020-03-29]. Dostupné z: https://www.brainproducts.com/files/public/products/brochures_material/pr_articles/1402.ST_ICA.pdf.
25. HYVÄRINEN, Aapo; OJA, Erkki. A fast fixed-point algorithm for independent component analysis. *Neural computation*. 1997, roč. 9, č. 7, s. 1483–1492.
26. ULLSPERGER, Markus; DEBENER, Stefan. *Simultaneous EEG and fMRI: recording, analysis, and application*. Oxford University Press, 2010.
27. BELL, Anthony J; SEJNOWSKI, Terrence J. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*. 1995, roč. 7, č. 6, s. 1129–1159.
28. GARVEY, Jennie Hill. *Independent Component Analysis by Entropy Maximization (Infomax)*. Monterey, California, United States of America, 2007. Master's Thesis. Naval Postgraduate School.
29. LANGLOIS, Dominic; CHARTIER, Sylvain; GOSSELIN, Dominique. An introduction to independent component analysis: InfoMax and FastICA algorithms. *Tutorials in Quantitative Methods for Psychology*. 2010, roč. 6, č. 1, s. 31–38.
30. BELOUHRANI, Adel; ABED-MERAIM, Karim; CARDOSO, J-F; MOULINES, Eric. A blind source separation technique using second-order statistics. *IEEE Transactions on signal processing*. 1997, roč. 45, č. 2, s. 434–444.
31. SAHONERO, Guillermo; CALDERON, Humberto. A Comparison of SOBI, FastICA, JADE and Infomax Algorithms. In: 2017.

32. XU, Honghui; DENG, Yong. Dependent evidence combination based on shearman coefficient and pearson coefficient. *IEEE Access*. 2017, roč. 6, s. 11634–11640.
33. CHAN, YH. Biostatistics 104: correlational analysis. *Singapore Med J*. 2003, roč. 44, č. 12, s. 614–9.
34. LAMORTE, Wayne W. *Nonparametric Tests* [online]. Boston (USA), 2017 [cit. 2020-05-17]. Dostupné z: http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric4.html%5C#.
35. PRŮCHA, Ladislav. *6. Testování hypotéz* [online]. Praha (CZ) [cit. 2020-05-17]. Dostupné z: <https://math.feld.cvut.cz/prucha/mstp/6pu.pdf>.
36. BIRJANDTALAB, Javad; POUYAN, Maziyar Baran; COGAN, Diana; NOURANI, Mehrdad; HARVEY, Jay. Automated seizure detection using limited-channel EEG and non-linear dimension reduction. *Computers in Biology and Medicine*. 2017, roč. 82, s. 49–58. ISSN 00104825. Dostupné z DOI: [10.1016/j.combiomed.2017.01.011](https://doi.org/10.1016/j.combiomed.2017.01.011).
37. KIM, Donghyeon; KIM, Kiseon. Resting State EEG-Based Biometric System Using Concatenation of Quadrantal Functional Networks. *IEEE Access*. 2019, roč. 7, s. 65745–65756.
38. MAATEN, Laurens van der; HINTON, Geoffrey. Visualizing data using t-SNE. *Journal of Machine Learning Research*. Roč. 2008, č. 9.Nov, s. 2579–2605.
39. WATTENBERG, Martin; VIÉGAS, Fernanda; JOHNSON, Ian. How to Use t-SNE Effectively. *Distill*. 2016. Dostupné z DOI: [10.23915/distill.00002](https://doi.org/10.23915/distill.00002).
40. THE MATHWORKS, INC. *T-SNE*. San Francisco (CA), 1994-2018. Dostupné také z: <https://www.mathworks.com/help/stats/t-sne.html>. Online; cit. [2020-03-03].
41. BIRANT, Derya; KUT, Alp. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering* [online]. 2007, roč. 60, č. 1, s. 208–221 [cit. 2020-04-02]. ISSN 0169023X. Dostupné z DOI: [10.1016/j.datak.2006.01.013](https://doi.org/10.1016/j.datak.2006.01.013).
42. PIORECKÝ, Marek; ŠTROBL, Jan; KRAJČA, Vladimír. AUTOMATIC EEG CLASSIFICATION USING DENSITY BASED ALGORITHMS DBSCAN AND DENCLUE. *Acta Polytechnica* [online]. 2019, roč. 59, č. 5, s. 498–509 [cit. 2020-04-02]. ISSN 1805-2363. Dostupné z DOI: [10.14311/AP.2019.59.0498](https://doi.org/10.14311/AP.2019.59.0498).
43. ABDOLZADEGAN, Donya; MOATTAR, Mohammad Hossein; GHOSHUNI, Majid. A robust method for early diagnosis of autism spectrum disorder from EEG signals based on feature selection and DBSCAN method. *Biocybernetics and Biomedical Engineering* [online]. 2020, roč. 40, č. 1, s. 482–493 [cit. 2020-04-02]. ISSN 02085216. Dostupné z DOI: [10.1016/j.bbe.2020.01.008](https://doi.org/10.1016/j.bbe.2020.01.008).

44. KHAN, Kamran; REHMAN, Saif Ur; AZIZ, Kamran; FONG, Simon; SARASVADY, Sababady. DBSCAN: Past, present and future. In: *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*. 2014, s. 232–238.
45. MCMENAMIN, Brenton W.; SHACKMAN, Alexander J.; MAXWELL, Jeffrey S.; BACHHUBER, David R.W.; KOPPENHAVER, Adam M.; GREISCHAR, Lawrence L.; DAVIDSON, Richard J. Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *NeuroImage* [online]. 2010, roč. 49, č. 3, s. 2416–2432 [cit. 2020-05-17]. ISSN 10538119. Dostupné z DOI: [10.1016/j.neuroimage.2009.10.010](https://doi.org/10.1016/j.neuroimage.2009.10.010).
46. CRESPO-GARCIA, Maite; ATIENZA, Mercedes; CANTERO, Jose L. Muscle Artifact Removal from Human Sleep EEG by Using Independent Component Analysis. *Annals of Biomedical Engineering* [online]. 2008, roč. 36, č. 3, s. 467–475 [cit. 2020-05-17]. ISSN 0090-6964. Dostupné z DOI: [10.1007/s10439-008-9442-y](https://doi.org/10.1007/s10439-008-9442-y).