ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
SCHOOL OF LIFE SCIENCES

MASTER PROJECT IN LIFE SCIENCES ENGINEERING

A DATA-DRIVEN DESCRIPTION OF SLEEP USING INTRACRANIAL EEG RECORDINGS

Carried out in the Kreiman Laboratory
At Harvard Medical School, Boston, USA
Under the supervision of Prof. Gabriel Kreiman

Done by

ALEXANDRE LUSTER

Under the direction of

Prof. Silvestro Micera
In the Translational Neural Engineering Laboratory
EPFL

30th June 2023

ABSTRACT

Sleep plays a crucial role in maintaining overall well-being, cognitive function and physical health. Polysomnography (PSG) has been used to establish the groundwork in sleep research, but its inability to fully characterize the intricate processes that occur in the brain has prompted the need for alternative approaches. In this work, we leverage the high spatiotemporal resolution of intracranial recordings to provide new insights into the organization of sleep. We employed a data-driven approach to analyze over 3,097 hours of data collected from 3,172 electrodes implanted in 34 patients with intractable epilepsy, thereby transcending the constraints and preconceptions associated with conventional PSG. We propose a simple and parameter-free approach to differentiate the sleep and wake states, which we validated against annotations made from video and labels provided by a commercial sleep tracker (in 2 patients). Our findings reveal that a location-dependent modulation of slow wave activity is a distinguishing feature of the quiescent state. Furthermore, we discovered oscillatory patterns of neural activity and their relationship to anatomical locations and demographics, as well as their evolution over the course of a night. We provide evidence for the existence of four clusters of cerebral spectral power in our recordings. These clusters indicate variations in the manifestation and timing of these states across brain areas, and exhibit a cyclical organization. In conclusion, we exploited intracranial recordings to provide a novel description of the composition and organization of sleep, thereby extending what is known about this fundamental physiological process. Improving our understanding of the local organization and regulation of sleep will be instrumental in attenuating the negative consequences of the innumerable diseases that manifest reciprocal interactions with sleep.

ACKNOWLEDGMENTS

Fist and foremost, I express my gratitude towards Prof. Gabriel Kreiman for welcoming me to his lab and mentoring me throughout the project. Thank you for taking the time to discuss on a weekly basis, being so encouraging and optimistic about my project all along, and fostering a stimulating and warm environment in the lab.

I would also like to thank my EPFL supervisor, Prof. Silvestro Micera for accepting to supervising my work from overseas, his advice on making the most out of this opportunity and the insights on the broader perspective of my project.

Finally, I would like to thank everyone in the Kreiman lab for the discussions and the good times. In particular, thank you to Marcelo Armendariz and Leonardo Pollina for their interactions with the patients and allowing me to collect data from the sleep tracker.

ABBREVIATIONS

AASM	American Academy of Sleep Medicine
ARI	adjusted rand index
BIRCH	Balanced Iterative Reducing and Clustering using Hierarchies
CAR	common average reference
ECoG	electrocorticography
EEG	electroencephalography
iEEG	intracranial EEG
MWU	Mann-Whitney U test
NREM	non-rapid eye movement
PIB	power in band
PSG	polysomnography
PSD	power spectral density
REM	rapid eye movement
SEEG	stereotatic EEG
SNR	signal-to-noise ratio
UMAP	uniform manifold approximation and projection

Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1 SLEEP

1.1.1 THE UNDENIABLE IMPORTANCE OF SLEEP

Humans spend around one third of their lives sleeping. Due to the opportunity cost of this behaviour, it is indisputable that it must be of the utmost importance. Indeed, it is necessary for all aspects of human well-being and optimal functioning: mental health[4], learning and memory[53], cognitive[19] as well as athletic performance[86], as well as short and long-term health[49]. Lack of consistent high quality sleep is a risk factor for all-cause mortality[13], having been linked to cardiovascular disease[26], type 2 diabetes[12], cancer[57] and Alzheimer's disease[70] among others.

1.1.2 POLYSOMNOGRAPHY

The gold standard technique to study sleep in the clinic or in research laboratories is a multi-modality recording procedure known as polysomnography (PSG). It monitors many physiological parameters during sleep: brain activity, eye movements, muscle tone (in the legs and chin) and heart rate with electroencephalography (EEG), electrooculography, electromyography and electrocardiography respectively, as well as air flow through the respiratory system.

After completion of the test, an expert scorer classifies every 30 second epoch into one of five sleep stages according to the rules of the American Academy of Sleep Medicine (AASM) scoring manual[42].

1.1.3 SLEEP STAGES

WAKE

The human waking state EEG is characterized by a dominant alpha (8-12 Hz) rhythm, or beta (12-30 Hz) rhythm if the eyes are open. In a fifth of the population, the alpha rhythm is not visible or weak and activity is similar during the eyes open and eyes closed conditions. In children, the dominant resting activity occurs with higher amplitude at a slower, age-dependent frequency. Rapid blinking at 0.5-2 Hz occurs during wakefulness and will slow down with drowsiness.

NREM

NREM is defined by exclusion as everything during sleep that isn't REM sleep and is further split into 3 stages: N1, N2, N3 in increasing order of sleep depth.

N1 is the lightest sleep of the three; it is a short transition period into sleep. During this stage, brain activity slows down and we observe mixed frequencies in the EEG over the majority of the epoch.

N2 is a deeper stage of sleep, and can be identified when K-complexes and sleep spindles are visible. The K-complex is a sharp negative deflection followed by a positive compensation, and is believed to

serve to suppress cortical arousal in response to external stimulus that is judged as undangerous in order to remain asleep[36]. Sleep spindles are bursts of activity in the 11-16 Hz range which play an important role in declarative memory consolidation and synaptic plasticity[31]. During N2, the heart rate and body temperature decrease, and waking up becomes more difficult. People spend close to half their sleeping time in N2 on average.

Finally, N3, also know as deep sleep or slow wave sleep (SWS), is the deepest state of sleep. During this stage, synchronization of broader pools of neurons results in large amplitude (at least 75 μV), slow frequency activity (0.5-2 Hz) over more than 20% of the epoch. This stage was formerly further divided into N3 and N4 when slow waves dominated more than half of the epoch. Deep sleep occurs mainly during the early phase of the night. Human growth hormone is released and contributes to maintenance and repair of tissues[14]. N3 and REM sleep are considered to be the most crucial stages of sleep because people will experience significantly more of their asleep time in these stages following sleep deprivation[66].

REM

REM sleep get its name from the characteristic Rapid Eye Movements that occur during this stage. It is characterized by wake-like brain activity with a sleeping body, earning it the description "paradoxical sleep". It is, however, absolutely essential for many cognitive functions, including consolidating procedural memory[2] and emotional processing[33]. GABA and glycine-mediated inhibition of motoneurons are responsible for REM skeletal muscle paralysis, a mechanism thought to be protective against vivid movements during intense dreams which mostly occur during REM.

SLEEP ARCHITECTURE

The sleep stages usually occur in sequential order, with the typical succession being N1, N2, and then N3 (progressively deeper NREM sleep), followed by an episode of REM sleep, which may or may not be concluded by a short waking period (which is generally unconscious). However, deviations from that typical sequence are quite common, with alternations between two clusters being completely normal (Figure 1.1). While the commonly-held belief that cycles last 90 minutes is true on average [15], there is a wide normal range around that average. Multiple studies indicate that they can range from as short as 30 minutes to 3 hours[7, 38], and this can even be true within the same person during a single night. Because of this, it seems unlikely that this sleep structure serves any timekeeping purpose. Interruption of sleep can reset the cycle, effectively lengthening the duration of a cycle.

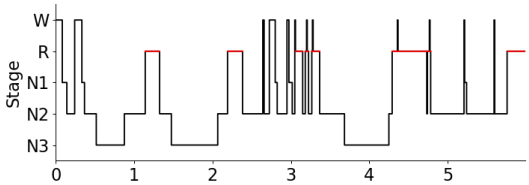

FIGURE 1.1: The hypnogram. A hypnogram shows the progression of sleep cycles as classified by a sleep expert from PSG over the course of a night. In this example, we can see that the different stages don't always take place in sequential order, and that the duration of individual cycles and stages can vary greatly throughout the night.

1.1.4 INTERPLAY WITH EPILEPSY

There are complex interactions between epilepsy and sleep. Epilepsy can alter sleep architecture[63] (e.g. increased wake after sleep onset[79]), continuity and quality. It often co-occurs with sleep disorders[17], in particular excessive daytime sleepiness and insomnia [61]. The interaction is bidirectional:

interictal epileptiform discharges (electrophysiological events that occur between seizures in epileptic patients) are correlated with synchronous slow wave oscillations observed in NREM[44] and inhibited during the asynchronous REM[60]. Furthermore, seizures can happen preferentially during wake or sleep depending on the location of the onset zone and type of epilepsy[44].

For these reasons, epileptic patients are not ideal subjects for sleep. In practice, sleep deprivation protocols are regularly used in epilepsy-monitoring units to increase the likelihood of the occurrence of seizures to facilitate the localization of seizure onset[87]. While antiepileptic drugs are known to negatively impact sleep quality and reduce REM sleep [16], they are usually tapered during monitoring. To limit the influence of these interactions on the generalizability of our findings to healthy populations, we excluded recordings taken during periods where patients exhibit particularly fragmented sleep due to antiepileptic drugs as recommended in the literature [29, 61].

1.1.5 LIMITATIONS OF POLYSOMNOGRAPHY

While PSG gives an overview of many bodily parameters during sleep, it does not place its focus on the brain. Indeed, the montage recommended by the AASM includes only 3 EEG derivations (F4-M1, C4-M1, O2-M1).

Better spatial coverage has been shown to be useful for differential diagnosis of sleep disorders[9], and higher spatial resolution is necessary to study local aspects of sleep, the importance of which be discussed in the next section.

The rules are subject to interpretation by expert scorers, and have limited applicability to patients with atypical sleep pattern. In particular, epilepsy is known to often present with disrupted sleep patterns due to seizures or medication use[22]. It has also been suggested that the current sleep stages are inhomogeneous should be further separated into a larger number of stages[59] or described as a continuum[39, 65]. Subjective sleep quality is poorly explained by variables of interest derived from PSG[45]. Data-driven clustering of sleep could enable new possible clinical interpretations by revealing patterns that vane-t been described yet[40].

Finally, manual scoring is time-consuming, costly and unreliable. Indeed, it has been shown that inter-rater agreement is very low even between highly trained scorers [89].

1.2 INTRACRANIAL EEG

EEG measures the electric potential generated by synchronized activity of ensembles of neurons with parallel geometric orientation in the superficial cortical layers[43]. Intracranial EEG is an invasive means of collecting neurophysiological data which provides much greater spatial accuracy than conventional EEG because of the small inter-electrode distance and absence of spatial blurring from the skull, scalp and dura mater[18], rendering it more useful for studying the organization of sleep in precise brain locations. Similarly to EEG, it measures the intracranial field potential originating from synchronized activity of neurons, but from populations of tens to hundreds of thousands of neurons, which is about 3 orders of magnitude less than what is recorded by EEG[64]. Besides the better spatial accuracy, iEEG also provides much greater signal fidelity with a signal-to-noise ratio (SNR) 20-100 times higher[68] due to close proximity to the signal and because the signal doesn't have to cross the dura mater and skull, plus the quality remains stable over long recording periods[68].

Of particular relevance for this work are electrocorticography (ECoG), which records from electrodes on the cortical surface and requires craniotomy, and stereotactic EEG (sEEG), a less invasive procedure in which electrodes on a needle record from the brain through small holes in the skull. While the first

provides denser coverage of a specific region of the brain, the latter can record bilaterally from deeper structures in the brain like the thalamus, basal ganglia, hippocampus or brainstem.

1.3 RELATED WORK

1.3.1 UNSUPERVISED LABELING OF SLEEP FROM IEEG

Few studies have leveraged unsupervised methods to label sleep from iEEG data. In a rat epileptic model, Runnova *et al* classified behavioral sleep compared to manual annotations from video using wavelet power in 5-10 Hz in the occipital and frontal cortical areas as a biomarker for sleep[73]. Their approach required tuning a threshold with some labeled data and achieved a median classification quality of 92%. They do not face the same difficulties as when working with human data since the electrode locations can be chosen for the scientific question of interest and be conserved across animals, the populations are more homogeneous and signal quality is higher.

In human epileptic patients, Sun *et al* [80] obtained a mean accuracy of 85.4% (using PSG as ground truth) in detecting sleep using a semi-hidden Markov model and spectral power features from a single ECoG electrode. The use of a semi-hidden model allowed them to prevent excessive state transitions which occurred with simpler models. Kremen *et al* used a rule-based approach to further perform labeling of all 5 behavioral states and achieved an overall weighted accuracy of 94% (against PSG) using spectral power from a single electrode with some prior knowledge from the literature to specify cutoffs[46], which might affect generalizability to heterogeneous groups of patients such as the one in our study.

1.3.2 DESCRIBING SLEEP BEYOND THE TRADITIONAL APPROACH

The main simplifications made by previously described model of sleep is the assumption that there are clearly distinct behavioral states, and that these always manifest synchronously across the entire brain. In reality, they exist on a continuum, and can often co-occur[84]. Indeed, sleep exists and is regulated locally[47]. Slow oscillations and sleep spindles can occur out-of-phase in different brain regions[62]. State-dissociation, the simultaneous occurrence of signs of different behavioral state across various parts of the brain[84], is involved in many sleep-related disorders such as sleepwalking, insomnia, sleep paralysis and narcolepsy[50]. There are also local disruptions to sleep organization in Alzheimer's disease, such as local deficits in slow wave activity in the frontal cortex and decreases in parietal sleep spindles[52]. Some of these changes occur even prior to the emergence of mild cognitive impairment, and have the potential to be useful for differential diagnosis of various forms of dementia, and can explain some of the mechanisms of cognitive decline[52].

Recent studies have shown that the stages of sleep defined by the AASM can be identified from iEEG only[24]. However, by aiming to replicate this traditional description of sleep, they fail to determine whether intracranial recordings can shine light patterns that polysomnography failed to or could not possibly reveal.

A complete description of the classical sleep stages as they relate to spectral power has been given by Prerau et. al.[71].

2. METHODS

2.1 DATASET

2.1.1 INTRACRANIAL RECORDINGS

Two different sets of intracranial field potentials were used in this project. The first consists of data recorded from 32 patients (17 female, 15 male) at Boston Children's Hospital and Brigham and Women's Hospital from 2007-2013. These patients had pharmacologically intractable focal epilepsy and were implanted with cortical electrodes to identify the epileptic focus for potential removal. Therefore, the electrode locations and numbers were dictated by clinical needs. Data was recorded at various sampling rates, from 250 to 2048 Hz. The platinum electrodes (Ad-Tech, Racine, WI, USA; 2.3 mm diameter exposed area, with inter-electrode distances of 10 mm) were organized either in grids, strips or a combination of both and implanted in the subdural space. Further details can be found in Wang et. al.[85]. Sixteen patients from the original study were excluded due to a lack of video monitoring.

The second dataset was recorded at Boston Children's Hopsital from 2 patients, an 11-year-old and a 19-year-old female, at a rate of 2048 Hz (referred to as patient m10000 and m10003). These patients were implanted with depth sEEG platinum-iridium electrodes (PMT Corporation; 0.8-mm diameter, 2.0-mm length cylinders; separated from adjacent contacts by 1.5 to 2.43 mm, 208 and 102 electrodes respectively). Simultaneously, they were wearing a commercial sleep-staging smartwatch (Fitbit Charge 5, San Francisco, CA) for automatic sleep detection and staging. While EEG devices are typically used for monitoring sleep stages and could provide higher accuracy, they were not suitable for these patients due to their invasive surgery. This device satisfied the necessary conditions, being small and convenient enough to obtain patient consent, having documented performance in sleep detection (estimated at around 85-90% [23, 35]), and providing a public API for extracting sleep classification and staging information.

Both datasets were collected using a Natus Quantum amplifier (Pleasonton, CA). Recordings where patients had highly fragmented sleep, indicated by the absence of visible oscillations in delta power over the recording, were excluded from further analysis. On average, the patients were 18.3 (2-42)[1] years old , had 89.4 (34-176) electrodes implanted, and the data was recorded continuously for an average of 3.8 (0.6-9.2) days.

2.1.2 SLEEP ANNOTATIONS

The Natus Neuroworks Wave software was used to read video recordings aligned to the neural data. Manual annotations of sleep and wake were performed from the videos once per hour for all 34 patients. For each time point, a binary label for sleep or wake was assigned based on the presence of movement or open eyes in the preceding 5 minutes.

[1]Throughout the text, this notation indicates the minimum and maximum values for fixed values (in this case, ranging from 2 to 42 years), and 95%, two-sided CI interval for the mean obtained by bootstrapping with 10,000 resamples in the case of statistics.

Furthermore, for the two patients wearing the smartwatch, Fitbit's Web API [2] was used to extract labels representing wake, light sleep, REM sleep and deep sleep at a frequency of 1 Hz.

2.1.3 PREPROCESSING AND FEATURE EXTRACTION

The DC bias was removed by the recording system. The neural signals were band-pass filtered in the range 0.5 to 100 Hz with a Hamming window. We applied a fifth-order Butterworth notch filter with a bandwidth of 3 Hz at 60, 120 and 180 Hz to eliminate power line interference and its harmonics. Following this, recordings were downsampled to 250 or 256 Hz. The common average reference (CAR) was subtracted from each signal. This step was crucial for enhancing the visibility of cyclical patterns in brain activity during sleep. These were less obvious with bipolar referencing (Figures A.3-A.7), whereby the signal from an electrode is referenced to one of its neighbours. Hence, this alternative re-referencing technique was not explored further. Additionally, CAR has been shown to be superior to local re-referencing schemes when studying spectral power from both ECOG and sEEG[90].

To represent each signal in low dimension, we divided the data into non-overlapping 10s epochs and computed power in band (PIB) features in 5 frequency bands commonly described in the sleep and neuroscience literature: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz) and gamma (30-100 Hz). Various studies have shown these to be effective in discriminating between the established sleep stages[46, 88]. We calculated the power spectral density (PSD) in each electrode using Welch's approximation with a Hamming window of 2.2 seconds and a 50% overlap, and Simpson's rule was used to integrate the power in the 5 frequency bands. An example of all computed PIB features can be found in Figures A.10-A.12.

Band	Frequency range	Interpretation
Delta	0.5-4 Hz	Deep sleep[48], cognitive processing[37]
Theta	4-8 Hz	Memory consolidation[28]
Alpha	8-12 Hz	Visual perception[75]
Beta	12-30 Hz	Conscious focus[1]
Gamma	30-100 Hz	High arousal, problem-solving[1]

TABLE 2.1: Frequency bands and related cognitive functions. Some examples of functions known to be related to high activity in each band are given.

To enhance the resolution of the PSD for specific analyses, we adopted a multitapering approach [82], combining multiple independent spectral estimates obtained through using various filters. Due to the much higher computational cost, we computed the spectrogram with a multitaper for a single electrode per patient. For most of the analyses conducted with the PIB, the multitaper yielded comparable features (Figure A.1 and A.2) as using Welch's approximation and produced similar results. A single taper was likely sufficient for most applications for two possible reasons. First, the downstream binning into broad frequency ranges reduces the need for a precise spectral estimate. Secondly, a single taper is sufficient for a decent estimation of the PSD due to the superior SNR of ECoG over EEG. Figures A.3-A.7 compares PIB features computed with the chosen methodology to features computed using a multitaper and to features computed from bipolar electrodes.

We defined any 10s signal as an artifact if any of the following conditions were met:

- there was a voltage range greater than 2000 μV over any 1s interval
- there was a supraphysiological (greater than 100 $\mu V/ms$) slope in voltage over time
- the voltage range was smaller than 10 μV

[2] https://dev.fitbit.com/build/reference/web-api/sleep/

Of all 10s signals, 1.01% were marked as artifacts based on these criteria. These were likely distorted by non-physiological events or by patient movement. Electrodes that had more than 50% of artifacts and electrodes presenting irregularities in the PIB features upon visual inspection were excluded from all downstream analyses. Out of the 3,172 electrodes we recorded data from, 108 electrodes (3.4%) were rejected. For the time windows marked as artifacts in the remaining electrodes, cubic interpolation was employed to impute missing values in the PIB.

2.1.4 ELECTRODE LOCATIONS

Electrode locations were dictated by clinical needs and patient-specific. With the exception of one subject, data was recorded unilaterally, with the highest number of electrodes being in the temporal lobe (Table 2.2). A post-operative computed tomography (CT) scan and a pre-operative T1-weighted magnetic resonance image (MRI) were acquired and co-registered using Freesurfer [27]. Bioimage Suite [67] was used to annotate the electrodes on the CT scan. Each electrode was mapped to a location in the Desikan-Killiany atlas using the iELVis Matlab toolbox [34]. These steps were performed by Jiarui Wang and Marcelo Armendariz for the first and second dataset respectively.

The remaining are spatially distributed as follows:

Lobe	Left	Right	Total
Temporal	857	371	1,228
Frontal	421	438	859
Parietal	282	336	618
Occipital	100	153	253
Other	67	39	106
Total	1,727	1,337	3,064

TABLE 2.2: Locations of the 3,064 electrodes kept for analysis.

The electrodes indicated as "other" are from the cingulate cortex (70), the insular cortex (24), the corpus callosum (3) or from a location that could not be determined (9). The number of electrodes in the finer scale Desikan-Killiany mapping are reported in Table 2.3

Middletemporal	332	**Lateraloccipital**	118	**Parstriangularis**	59	**Medialorbitofrontal**	17
Inferiortemporal	270	**Inferiorparietal**	113	**Caudalmiddlefrontal**	55	**Unknown**	9
Superiortemporal	266	**Parsopercularis**	98	**Parsorbitalis**	45	**Bankssts**	7
Rostralmiddlefrontal	178	**Precuneus**	96	**Parahippocampal**	45	**Pericalcarine**	7
Precentral	168	**Lingual**	95	**Cuneus**	33	**Frontalpole**	7
Fusiform	164	**Lateralorbitofrontal**	84	**Posteriorcingulate**	32	**Rostralanteriorcingulate**	6
Supramarginal	164	**Superiorparietal**	81	**Paracentral**	30	**Caudalanteriorcingulate**	6
Postcentral	164	**Entorhinal**	74	**Isthmuscingulate**	26	**Corpuscallosum**	3
Superiorfrontal	118	**Temporalpole**	67	**Insula**	24	**Transversetemporal**	3

TABLE 2.3: Number of electrodes in each of the 36 areas of the Desikan-Killiany mapping.

2.2 UNSUPERVISED SLEEP LABELING

2.2.1 SEPARATING SLEEP AND WAKE

We separated the recordings into two distinct clusters using the PIB from all electrodes. We used the relative bandpower (PIB over broadband power) as this proved more robust against drifts over long recording times. Recordings lasting longer than 48 hours were split into shorter segments of equal length

to reduce the impact of these long term drifts on clustering. Then, relative PIB features were normalized and projected to a 2D space using Uniform Manifold Approximation and Projection[55] (UMAP), a scalable non-linear dimensionality reduction technique based on neighbor embedding.

We employed OPTICS[5], a density-based clustering technique the to separate resulting embeddings. The minimum number of samples per cluster was set to one fifth of the recording. When this method produced more than 2 clusters, points in the smallest cluster were recursively reassigned to their other nearest cluster in Euclidean distance until 2 clusters remained. Lastly, in cases where the distribution between the 2 obtained clusters was extremely imbalanced (a cluster contained more than 80% of the points), KMeans++[6] was used to produce the clusters instead.

To assign the resulting cluster to behavioral state, we used the ratio of delta to beta power (delta/beta power), which has been previously shown to be the single best feature derived from EEG to discriminate sleep and wake[81]. We interpret the cluster with the highest average delta/beta power as sleep. In one of of the 34 patients, this ratio was lower during sleep (oscillations were present during the state with lower delta/beta power), so we exchanged the labels of the two clusters.

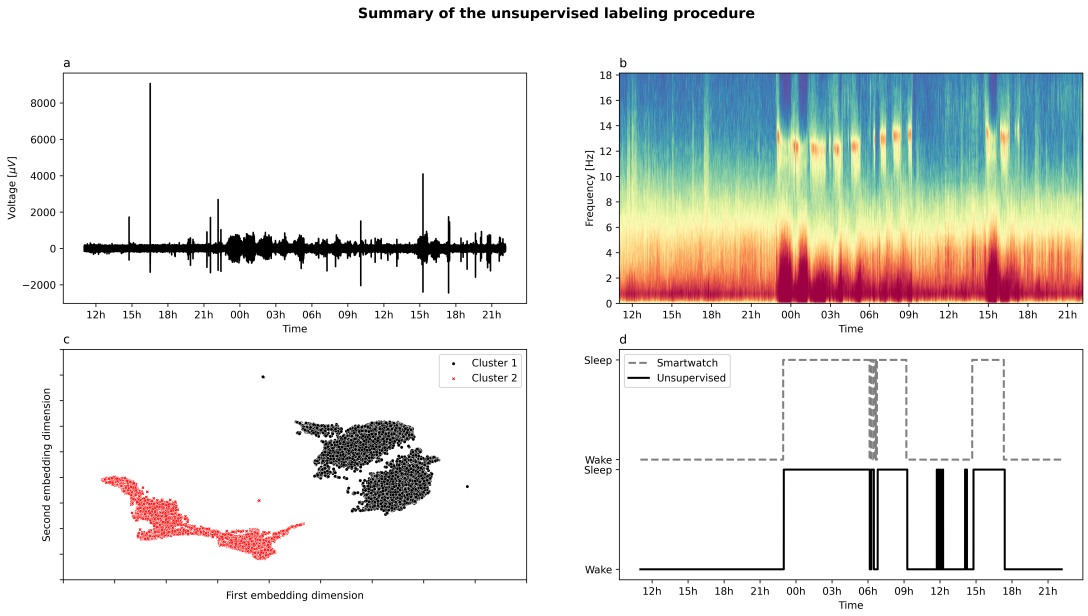

FIGURE 2.2: Summary of the unsupervised labeling method. (a) Raw voltage in a single electrode after bandpass filtering and CAR. (b) The spectrogram, computing for non-overlapping windows of 10 seconds, is integrated over 5 distinct frequency bands to compute band power, which is used as features for clustering. (c) Relative PIB is normalized and projected to a 2-dimensional space using UMAP and clustered using DBSCAN. (d) The resulting labels are evaluated against the predictions made using a commercial sleep tracker when available, and to manual annotations from video.

2.2.2 DESCRIBING SLEEP AND WAKE

To study the differences in brain activity during the wake and sleep state, we compared the mean PIB in both states and tested if the modulation of power by behavioral state is location-dependent. We tested if the activity in each band was different during sleep and wake using a two-sided, two-sample t-test. Then, we compared the distribution of the ratio of power during sleep and wake across the frontal, temporal, parietal and occipital lobe, as well as between the left and right hemisphere for each lobe using a Mann-Whitney U (MWU) test[56]. We applied Bonferroni correction to counteract multiple testing for every group of tests (across bands, across lobes for every band and between left and right hemisphere for every lobe).

2.2.3 DEFINING SLEEP EPOCHS

We defined sleep epochs as semi-continuous bouts of sleep lasting at least 3 hours. We implemented this smoothing the unsupervised labels by taking the majority label over a sliding window of 40 minutes, to tolerate awakenings shorter than 20 minutes, and then selecting strictly continuous bouts of sleep according to the smoothed labels.

To avoid including particularly fragmented sleep, recordings continuous wake of at least 3 hours were not considered. We also did not extract sleep epochs from recordings where the agreement with manual annotations was less than 80%. To serve as a control for some of the analyses, we extract wake epochs with the exact same procedure with the roles of sleep and wake being inverted.

2.3 SLEEP CYCLES AND OSCILLATIONS IN BRAIN ACTIVITY

The unsupervised labeling procedure raised the following question: why do some electrodes have obvious oscillations in the PIB features during sleep, while others appear visually similar during sleep and wake ? To identify electrodes exhibiting oscillatory patterns of activity during sleep, as well as to study the duration of sleep cycles, we calculated the autocorrelation in delta/beta power. A mean filter with a window size of 5 minutes is applied to the result. Finally, we extracted the first peak in the autocorrelation that exceeded the fixed value of 0.05, ignoring the first 30 minutes (the minimum duration of sleep cycles). We interpreted the presence of such peaks as indicative of cyclic activity in the electrode, and the time of the first peak as the average sleep cycle duration. All of this was performed both electrode-wise to study the location-dependence of these patterns. To study the correlation of cycle length with demographics, we tried two approaches. The first is to repeat the same procedure using the ratio of the average delta power across all electrodes to the average beta power across all electrodes. The second, more robust, is to simply use the median estimated cycle duration from all individual electrodes.

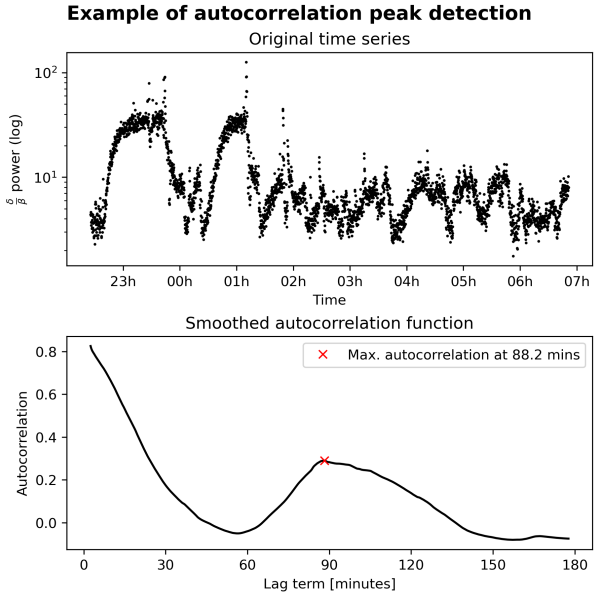

FIGURE 2.3: Autocorrelation method to detect cyclic patterns in delta/beta power. The first method tests for cyclic behavior and determines the typical cycle duration in a sleep epoch by detecting the first peak in the autocorrelation function.

This methodology has demonstrated robustness and has been utilized previously in the sleep literature[25]. However, while it provides a general understanding of the dominating periodicity in a sleep

epoch, does not allow for the characterization of the length of subsequent cycles. For this purpose, we employed a different approach. We extracted peaks in delta/beta power, interpreting them as the time of deepest sleep within a cycle. We first median-filter the input signal (delta/beta power) with a window of 5 minutes. Then, a peak detection algorithm is run such that peaks are separated by at least 30 minutes (the minimum duration of a sleep cycle), and such that they surpass the 80th percentile in delta/beta power over the entire sleep. Moreover, the prominence of each peak was required to be at least half of the 5 to 95% interpercentile range. This procedure was performed in a single electrode for each electrode, which we selected by visual inspection for displaying the most obvious oscillations of delta/beta power during sleep.

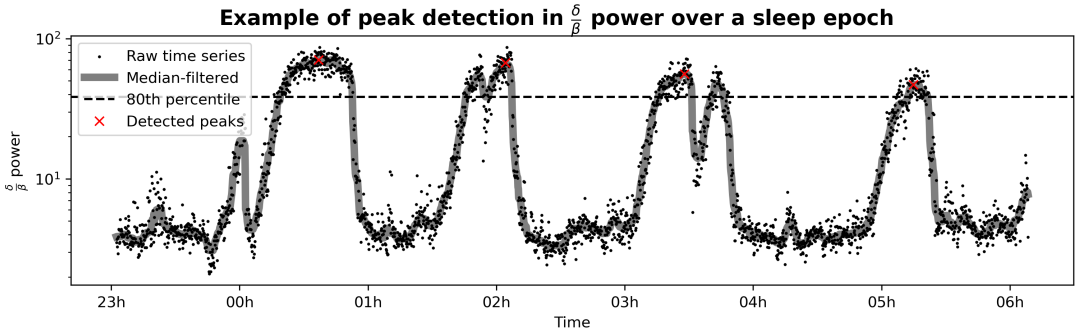

FIGURE 2.4: Peak detection method to determine the duration of individual cycles. The second method estimates the length of individual sleep cycles computes the distances between the detected peaks in delta/beta power (the 4 red crosses).

2.4 CLUSTERS OF ACTIVITY DURING SLEEP

2.4.1 NUMBER OF CLUSTERS IN BRAIN ACTIVITY DURING SLEEP

Previous studies on clustering sleep data have typically imposed a predefined number of clusters based on the existing literature[21, 78]. However, the objective of this experiment was to allow the data to guide the determination of the optimal number of clusters to capture the activity patterns in the recordings. To achieve this goal, we compared various clustering methods along with different clustering metrics. We also conducted analyses to confirm that the choice of preprocessing steps employed did not affect the outcome. We compared different ways of normalizing the PIB features, as well as two techniques (PCA and UMAP) to embed the resulting features in 2, 10 and 100-dimensional spaces. By adopting this data-driven approach, we aimed to uncover the inherent structure within the data and identify the most suitable number of clusters without relying on preconceived assumptions from prior research.

We compared k-means, agglomerative clustering, spectral clustering and BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) as clustering methods. The k-means algorithm aims to partition a set of data points into k distinct groups by minimizing the within-cluster sum of squares, where each data point is assigned to the cluster with the nearest mean, and the process is repeated until convergence. We also tested agglomerative clustering with Ward's linkage[3]. This method starts with every point as a cluster and iteratively merges the two clusters that will produce the smallest increase in the total within-cluster variance until a predetermined number of clusters is reached. Spectral clustering starts by constructing a similarity matrix based on pairwise distances between data points, which is projected to a lower-dimensional representation using singular value decomposition[77]. Finally, a simple clustering algorithm is applied to this representation to obtain the final clusters. Spectral clustering is particularly effective in capturing complex geometric structures and can handle non-convex clusters that traditional methods struggle with. We used a radial basis function kernel to define pairwise distances and k-means to

cluster the embedding space. Lastly, BIRCH is a hierarchical clustering algorithm that constructs a tree structure from the data and recursively partitions it into smaller clusters. At the end, the cluster centroids can be read off the leafs and provided as input to another clustering algorithm. We used them as input to the agglomerative clustering algorithm described above. For more details, see the original description in Zhang et. al. 1996[91].

The goodness of clustering metrics tested were the distortion score, the silhouette score, the Calinski-Harabasz score, the Davies-Bouldin index and the Gap statistic.

The distortion score is simply the sum of square distances to the assigned cluster center. The optimal number of clusters corresponds to an elbow in the plot of distortion versus the number of clusters. The Silhouette score[72] is the average of the silhouette values s_i of each sample i,

$$s_i = \frac{b_i - a_i}{max(a_i, b_i)}$$

where a_i is the average distance of point i to the other points in the same cluster, and b_i is the smallest average distance of i to all points in another cluster. We used the Euclidean distance. The score ranges from -1 to 1, with 1 indicating the highest clustering quality. The Calinski-Harabasz score[11] (also known as the variance ratio criterion) is defined as

$$CH = \frac{SS_{between}/(k-1)}{SS_{within}/(n-k)}$$

where SS denotes the sum of squared (Euclidean) distances, k is the number of clusters and n the number of samples. A higher score indicates better clustering. The Davies-Bouldin index is defined as

$$DB = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} \frac{s_i + s_J}{d_{ij}}$$

where s_i is the average distance between each point of cluster i its centroid, and d_{ij} is the distance between cluster centroids i and j. It is greater than 0, and smaller values indicate better clustering. Lastly, the gap statistic estimate the number of groups by comparing the within cluster dispersion with the expected dispersion under an appropriate null distribution. Mathematical details can be found in the original paper by Tibshirai et. al.[83].

2.4.2 CLUSTERING OF SLEEP EPOCHS

Relative PIB features from every electrode were used to separate all 113 sleep epochs into 4 clusters. These features were first normalized within each electrode and band. Then, UMAP was used to project the data to a 10-dimensional space. Finally, we used the KMeans++ algorithm with 4 clusters. The resulting clusters were labeled from 0 to 3 in order of increasing average delta power across all electrodes to reflect depth of sleep [51, 89]. As a final step, noise is removed from the labels by reassigning each time point to the most common label in the 5 minute window centered around that point.

The stability of the resulting clusters has been assessed in two ways. First, we measure the consistency of the procedure when clustering samples collected at even versus odd time points. Then, we compare the results when computing the clusters from one half of the electrodes to those obtained from the other half.

2.4.3 COMPARING CLUSTERS OF ACTIVITY ACROSS BRAIN AREAS

Next, we investigated regional differences in the clusters of activity. We performed this analysis by reapplying the clustering procedure using electrodes a fixed number of electrodes exclusively from a

specific brain lobe. For each lobe pair, we randomly selected two non-overlapping sets of 5 electrodes from the first lobe and two sets of non-overlapping sets of 5 electrodes from the second lobe. Next, we reassigned the cluster labels to a reference labeling in such a way that accuracy was maximized. The reference labeling for a given lobe was the clustering obtained by considering all electrodes from that lobe, and consistency between the references in both lobes was ensured by finding the best one-to-one mapping between these two.

This procedure was repeated 10 times for each sleep epoch in every patient who had sufficient coverage (i.e. at least 10 electrodes in both the first and second lobe). Only the lobe pairs with data available from at least 10 different sleep epochs were included in subsequent analyses. To assess the agreement between clusters obtained within the same lobe (estimated with 10 iterations) and across different lobes (estimated with 40 samples, considering the 4 possible pairings of splits between the lobes multiplied by 10 iterations), we computed the average adjusted rand index. The adjusted rand index (ARI) [41] is a clustering metric insensitive to permutation of the labels. The ARI between two sets of labels X and Y is defined as:

$$ARI = \frac{RI - \mathbb{E}(RI)}{max(RI) - \mathbb{E}(RI)}$$

where

$$RI = \frac{1}{|\{(i,j), j > i\}|} \sum_{(i,j),j>i} \mathbb{1}(\mathbb{1}(X_i = X_j) = \mathbb{1}(Y_i = Y_j))$$

.

Put in words, it is the average agreement across all pairs of points, where an agreement means that the two points are either attributed the same label in both sets or a different label in both sets, and disagreement means that the two points share the same label in a set but different labels in the other. The correction made to the adjusted rand index is such that random labeling is expected to give a value of 0, and is bounded by -0.5 and 1.

This analysis was repeated using 10 instead of 5 electrodes for clustering at each step. This allows to obtain more stable results, at the cost of having to exclude more patients with insufficient coverage in the relevant lobes.

2.4.4 TEMPORAL EVOLUTION OF CLUSTERS AND DYNAMICS

First, we sought to determine if the resulting clusters captured the cyclical patterns of activity that are present in the neural activity. To this end, re-used the autocorrelation method described in section 2.3.

Then, we explore the cyclical organization of the sleep clusters, by studying their occurrence over the course of sleep cycles. We implement this by unwrapping the time between the peaks in sleep depth, as defined in section 2.3. In other words, instead of thinking of the label as a point occurring at time T, we describe it as occurring at proportion P of the completion of the cycle it is in. With this new formulation, we can compare the cluster composition of subsequent cycles.

Next, we quantify the relative frequencies of the transitions and sequences of transitions between clusters as a way to understand the typical dynamical evolution of the clusters. For more robustness to noise in the clusters and to avoid frequent transitions between clusters, specifically for this analysis, we compute the most common label over a window of 10 minutes instead of 5 minutes.

3. RESULTS

3.1 UNSUPERVISED SLEEP LABELING

3.1.1 MANUAL ANNOTATIONS

The unsupervised method, when compared to the hourly manual annotations, achieved a mean balanced accuracy (average of the sensitivity and specificity) of 85.49% across the 34 patients. In certain cases, further improvements to the clustering procedure could be made to increase separability in the embeddings, improve the temporal consistency of the resulting clusters, agreement with manual labels and correlation with oscillatory behavior in delta over beta power. Several modifications were explored, including the use absolute instead of relative bandpower which in some cases resulted in more separable embeddings, splitting the recordings into chunks of various lengths and employing more elaborate clustering strategies. These were all performed prior to evaluation, i.e. without looking at the manual annotations (or labels from the sleep tracker). Implementing these adjustments led to a significant enhancement, with a mean labeling balanced accuracy of 90.74%. The overall contingency table summed over the 34 patients is shown in Table 3.4. To ensure greater reliability in subsequent findings, we employed these improved labels as the ground truth for distinguishing sleep and wake states throughout the remainder of this work.

Annotation	**Unsupervised label**	
	Sleep	Wake
Sleep	1,189	163
Wake	130	1,617

TABLE 3.4: Contingency table of the labels obtained with the unsupervised procedure (after tuning) versus manual annotations from video recordings. The same contingency table obtained without recording-specific adjustments to the algorithm described in section 2.2.1 are shown in Table B.1

We extracted 113 sleep epochs with an average duration of 6 hours and 22 minutes (3-14 hours) from 33 patients. Figure 3.5 shows an example comparing the unsupervised labels to manual annotations as well as extracted sleep epochs. Similarly, there are 172 wake epochs averaging 7 hours and 42 minutes (3-18 hours).

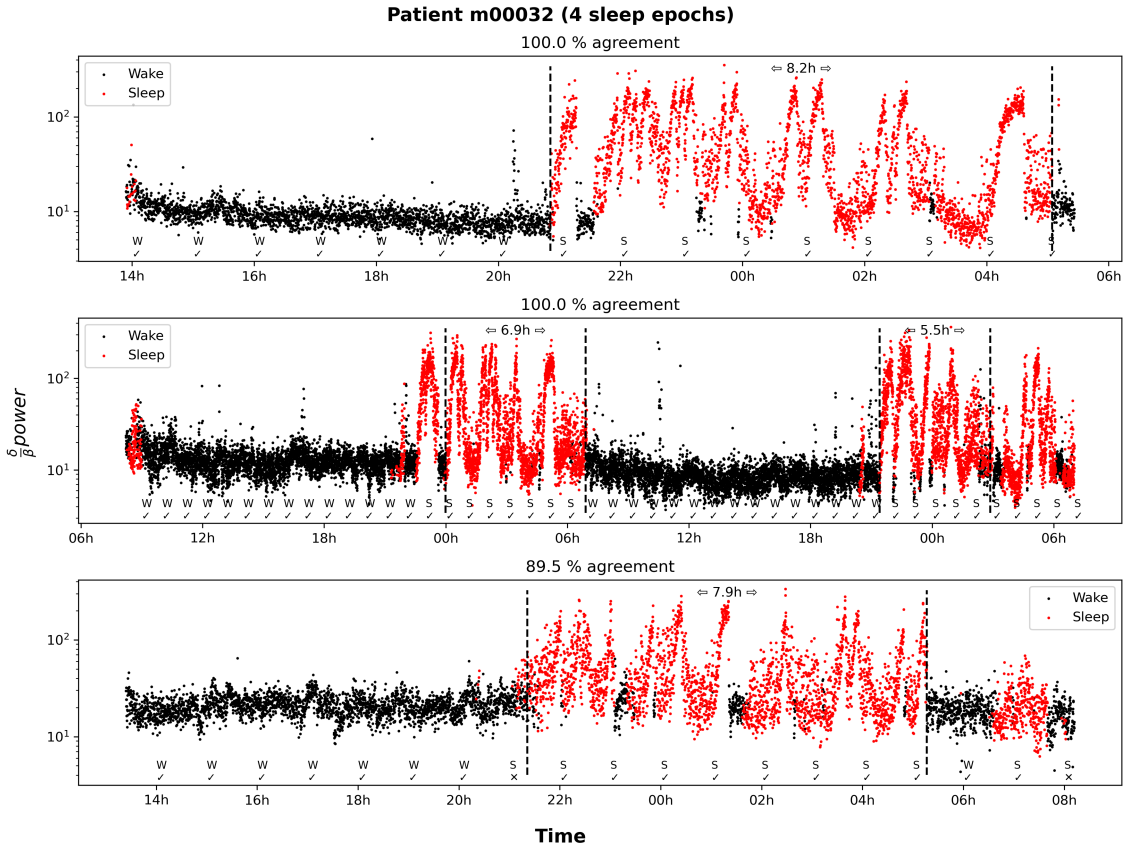

FIGURE 3.5: Unsupervised labeling of sleep and extraction of sleep epochs for patient m00032. Each time point is colored by the cluster it is assigned to. Agreement with the manual annotations (indicated by *S* for sleep and *W* for wake) is shown with ticks and mark on every hour. We also show the semi-continuous sleep epochs (dashed vertical lines) extracted for future analyses and their duration.

.

We compared the outcome of the same clustering procedure with different sets of features. First, we tried using the raw spectrogram instead of integrating power into finite bands. In addition to making clustering more computationally intensive, the high dimensionality and noise in this representation led to poor performance in separating sleep and wake.

We also evaluated a different set of features computed in the time domain (root mean square, mean absolute value, maximum and minimum) over 1s windows. These performed much worse in classifying the sleep versus wake condition than the PIB features (55.5 vs 75.5% decoding accuracy using a single electrode, averaged across all electrodes in one patient). Lastly, we obtained another set of features by training an autoencoder with a bottleneck layer of 10 dimensions, experimenting with convolutional and multi-layer perceptron architectures. These were trained to reconstruct either 1s signals directly or their Fast Fourier Transform[8]. While the features obtained this way had similar decoding power as the PIB features, they require dataset-specific training to be derived, are not interpretable and cannot be compared to the sleep literature. For this reason, we deemed that the PIB features would be the most useful for describing and clustering sleep.

3.1.2 SMARTWATCH

The wearable device provides the following labels: wake, wake_short (awakenings during sleep), rem, light and deep sleep. Additionally, during shorter bouts of sleep classified as naps, the labels are restless, asleep and awake. These were converted to a binary label indicating sleep and wake. Then, we compared the labels provided by the smartwatch to the manual annotations and found an accuracy of 93.3% (Table 3.5).

Smartwatch	**Annotation**	
	Sleep	Wake
Sleep	44	6
Wake	3	82

TABLE 3.5: Contingency table of the labels obtained from inspection of video recordings and those from the smartwatch.

The recordings from the second dataset were compared to labels obtained with the smartwatch. The agreement between these labels and those obtained by our unsupervised method was 91.4%. The accuracy with the fine-tuned version of the labels was unchanged for these two patients.

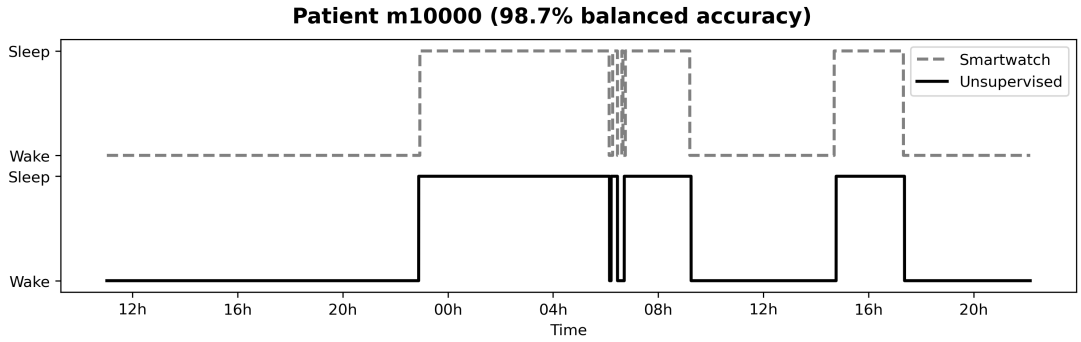

FIGURE 3.6: Labels obtained by our unsupervised method (solid line) versus those provided by the sleep tracker (dashed line) for patient m10000.

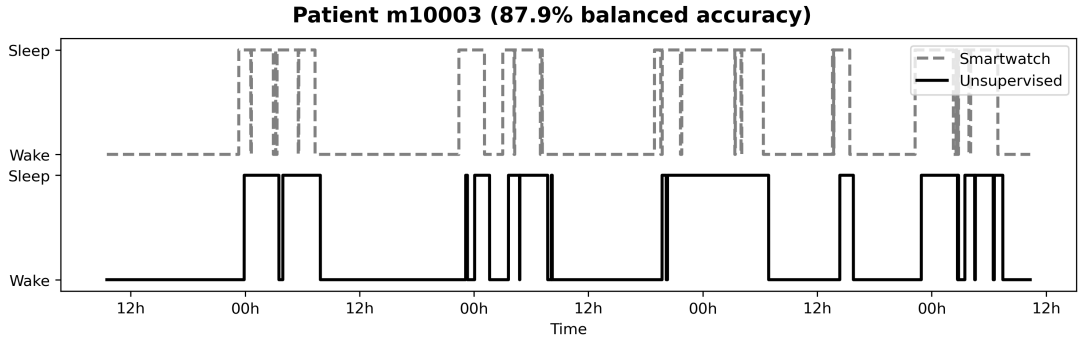

FIGURE 3.7: Labels obtained by our unsupervised method (solid line) versus those provided by the sleep tracker (dashed line) for patient m10003.

The structure of sleep stages was captured in the embeddings used to classify sleep versus wake (Fig. B.13).

3.1.3 DIFFERENCES BETWEEN THE SLEEP AND WAKE STATE

The power in delta and gamma were found to be modulated by behavioral state, with a median ratio of power during sleep versus wake of 2.2 and 0.9 respectively (Figure 3.8).

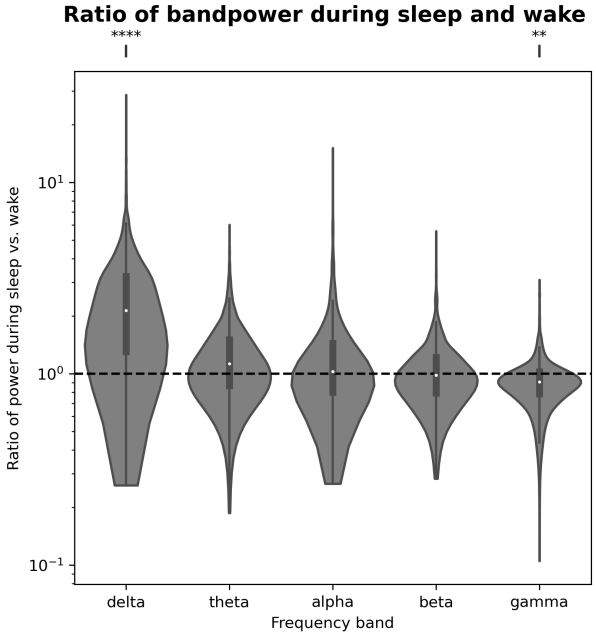

FIGURE 3.8: Distribution of the ratio of the mean band power measured during sleep versus during wake across all 3,064 electrodes.
The significance level of a two-sided, two-sample t-test comparing the observed values of the ratio to a null ratio of 1 are shown after Bonferroni correction. The boxplot inside each violin plot indicates the range of observed values (solid line), the first and third quartile (black box) and the median of the distribution (white dot).

A two-way analysis of variance shows that delta power is modulated differently by sleep in different lobes and in the two hemispheres, and that there is an interaction between the lobe and hemisphere on the modulation of delta power (Table 3.6). However, given the large F-value for the coefficient of the lobe (F(3, 2950)=148), it is clear that it is the most important variable in explaining the modulation of delta power.

Source of variation	SS	df	F	p-value
Lobe	1,362.671	3	148.0	2.418e-89
Hemisphere	27.168	1	8.852	2.951e-1
Interaction	37.581	3	4.082	6.66e-3
Residual	9,053.753	2,950		
Total	10,481.173	2,957		

TABLE 3.6: Two-way analysis of variance of the delta power ratio during sleep versus wake.
Electrodes located outside the temporal, frontal, parietal and occipital lobe were excluded from the analysis.

The distribution of the delta power ratio was found to be lobe-dependent (Figure 3.9) with the ratio being significantly different across all pairs of lobes. We observe that slow wave activity is most strongly modulated in the frontal lobe, consistent with what is known from the EEG literature [10, 54].

Remarkably, the distributions exhibit striking similarity between the left and right hemisphere even though the data for each hemisphere was obtained from separate sets of patients without any overlap. This observation rules out the possibility that the results obtained could be attributed to the fact that data from different lobes originates from different individuals. With the exception of the parietal lobe, we found that the delta power in the left and right hemisphere were not differentially modulated by behavioral state.

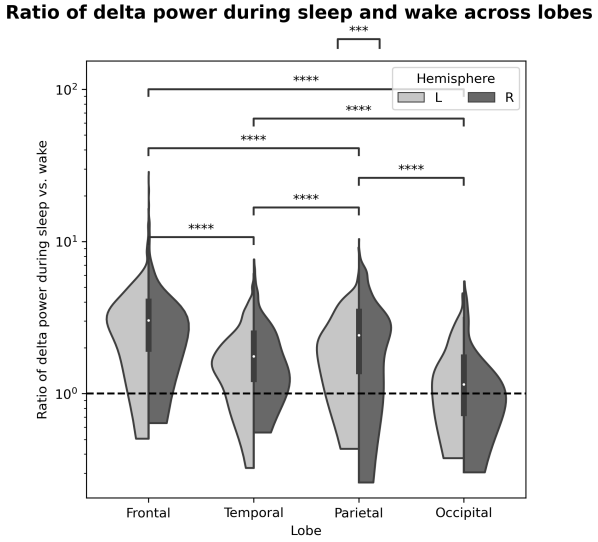

FIGURE 3.9: Distribution of the ratio of delta power during sleep versus wake. The significance level of a MWU test applied to all pairs of lobes and between the left and right hemisphere for every lobe is indicated, after Bonferroni correction.

Power in the gamma band is also modulated by the lobe, the hemisphere and there is an interaction between the two factors (Table D.2)

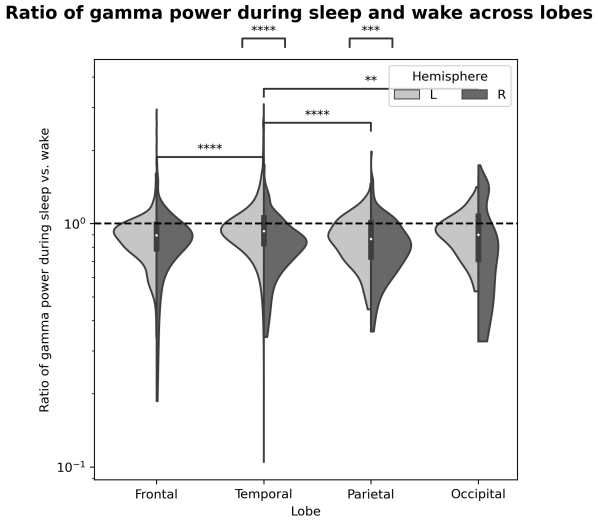

FIGURE 3.10: Distribution of the ratio of gamma power during sleep versus wake. The significance level of a MWU test applied to all pairs of lobes and between the left and right hemisphere for every lobe is indicated, after Bonferroni correction.

The result of the same analysis for the other frequency bands can be found in Appendix D.

3.2 SLEEP CYCLES AND OSCILLATIONS IN BRAIN ACTIVITY

3.2.1 SLEEP CYCLE AND DURATION

Contrary to popular belief, sleep cycles do not always last exactly 90 minutes. In my dataset, the average cycle duration in a cluster varied from 44 to 133 minutes, with the range in duration of individual cycles being even broader. The median cycle duration was found to be 95 minutes, and the most typical cycle length, defined as the mode of the kernel density estimate, was 90 minutes (Figure 3.11). We observed that the second and third cycles were on average longer than the first, while later cycles were again shorter (3.12), as had been shown previously by Brezinova et. al[7]. The differences in duration between subsequent cycles were not statistically significant.

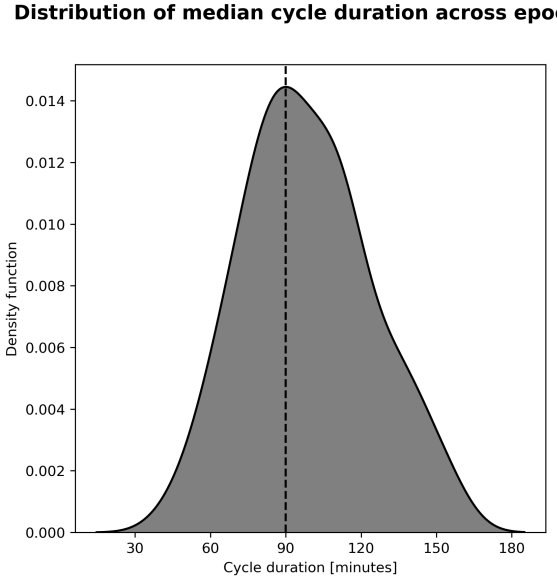

FIGURE 3.11: Kernel density estimate of the typical cycle duration in the 113 sleep epochs. For each epoch, the cycle duration was defined as the median cycle duration estimated by using the first peak in the autocorrelation function across all electrodes. The dashed vertical line indicates the mode of the distribution.

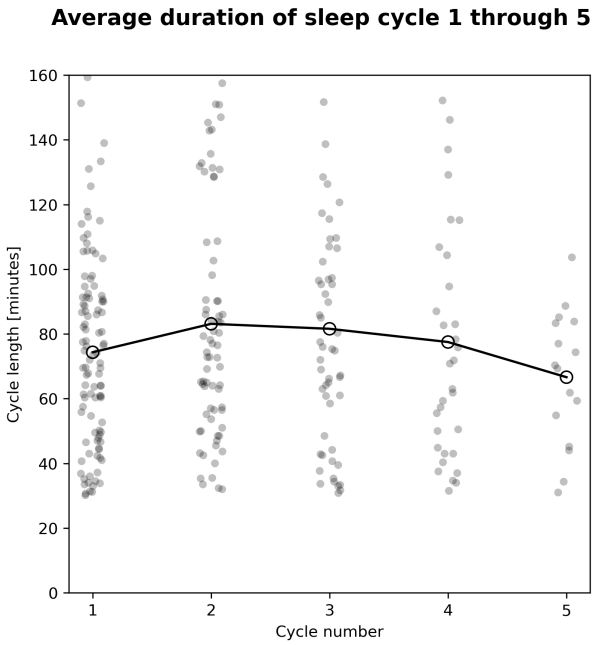

FIGURE 3.12: Duration of subsequent sleep cycles. The durations were estimated from the peaks in delta/beta power as described in 2.3. Data from all sleep epochs is shown. The solid line and circles indicate the mean duration of subsequent cycles.

No correlation was found between age ($r^2 = 0.048, p = 6.23e - 1$) or gender ($t = -0.678, p = 4.99e - 1$) with the median duration of sleep cycles (Figure E.18).

3.2.2 OSCILLATORY PATTERNS OF NEURAL ACTIVITY

In contrast, a weak but significant correlation was found between age and the proportion of electrodes displaying oscillatory patterns of delta/beta power ($r^2 = 0.134, p = 6.55e - 5$, Figure 3.13). No correlation was found with gender ($t = 0.509, p = 6.12e - 1$).

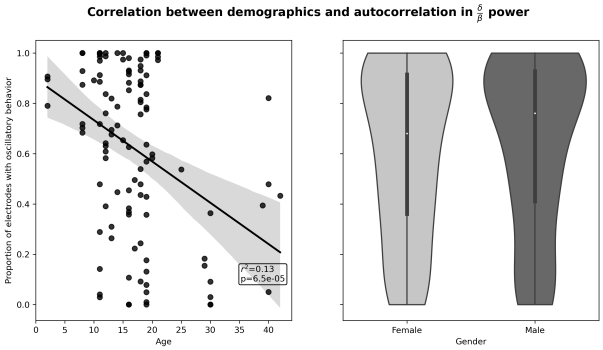

FIGURE 3.13: Correlation between demographics and the occurrence of autocorrelated delta/beta activity. Linear regression between patient age and the proportion of electrodes displaying oscillatory behavior as previously defined (left), and distribution of the proportion in patients of male and female gender (right). The solid line indicates the best linear fit to the data with the shaded area indicating a bootstrapped 95% confidence interval.

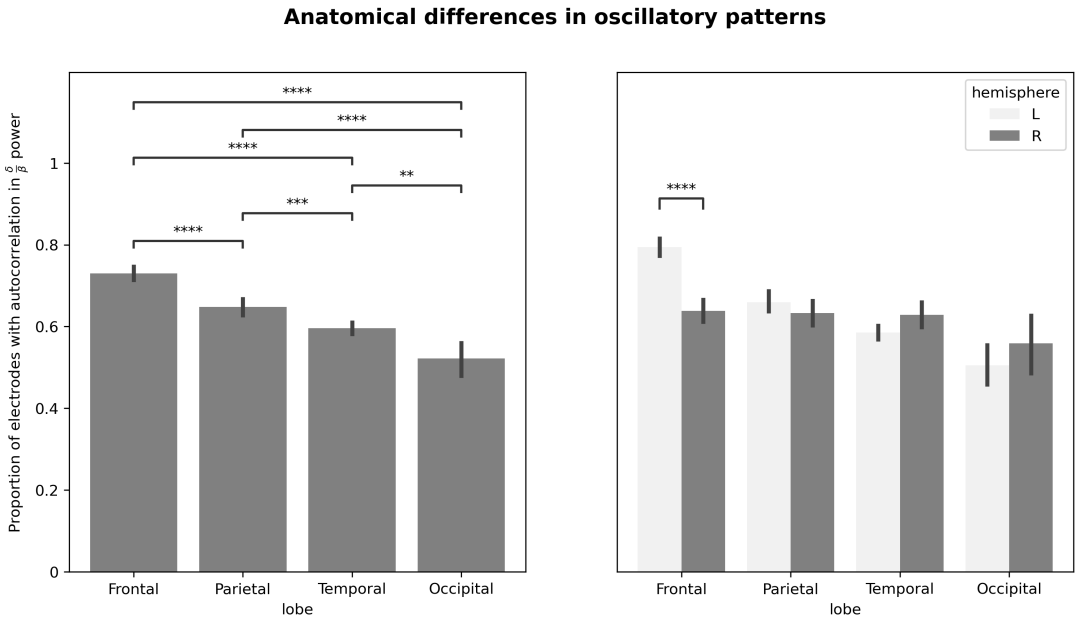

FIGURE 3.14: Proportion of electrodes displaying oscillatory behavior. The proportions are compared between lobes (left) and between hemispheres within a lobe (right). We used the MWU test with Bonferroni correction to test for differences in the proportion of electrodes displaying oscillatory behavior across the different lobes (left) and between the left and right hemisphere within each lobe (right).

3.3 CLUSTERS OF ACTIVITY DURING SLEEP

3.3.1 NUMBER OF CLUSTERS

We used the elbow method applied to the distortion score to define the optimal number of clusters resulting from KMeans clustering. With these parameters, we found the optimal number of clusters in most recordings to be 4 (Figure 3.15).

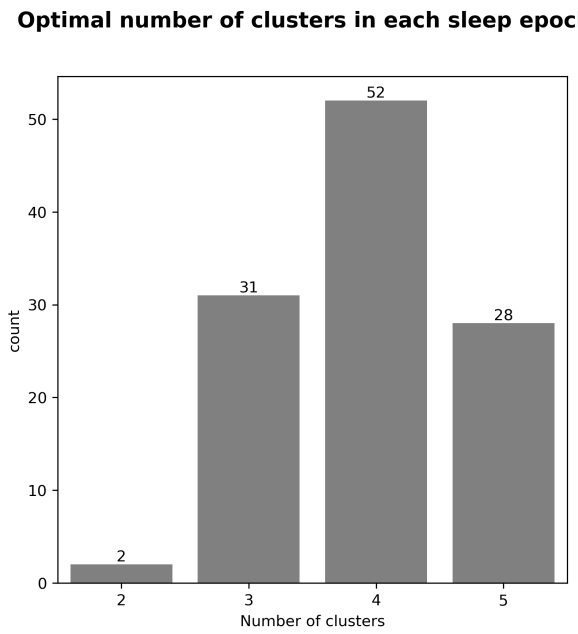

FIGURE 3.15: Distribution of the optimal number of clusters in all the sleep epochs using the elbow method in the distortion score.

The four other cluster scoring metrics could not be used as they suffered from ceiling effects (the gap statistic always returned 19 clusters, the maximum evaluated, as optimal) or floor effects (using the silhouette score, the Davies-Bouldin index or the Calinski-Harabasz index indicated that 2 clusters were optimal, which couldn't be trusted because it was the lowest number they could output) (Figure F.21). Because of this, we used the distortion score to define the optimal number of clusters for the dataset. Combined with the distortion score, 4 clusters were optimal regardless of the clustering method used (Figure 3.16), with the exception of Spectral Clustering which was rejected by majority vote and for being more computationally intensive.

The normalization method played no role in the optimal number of clusters (Figure 3.16), nor did the choice of the UMAP embedding dimension. When using PCA instead, more clusters were suggested unless dimensionality was reduced to 2 dimensions. We discarded PCA as it cannot represent nonlinear relationships in the data and proved to be less performant in discriminating sleep and wake (71% with PCA vs. 85% with UMAP).

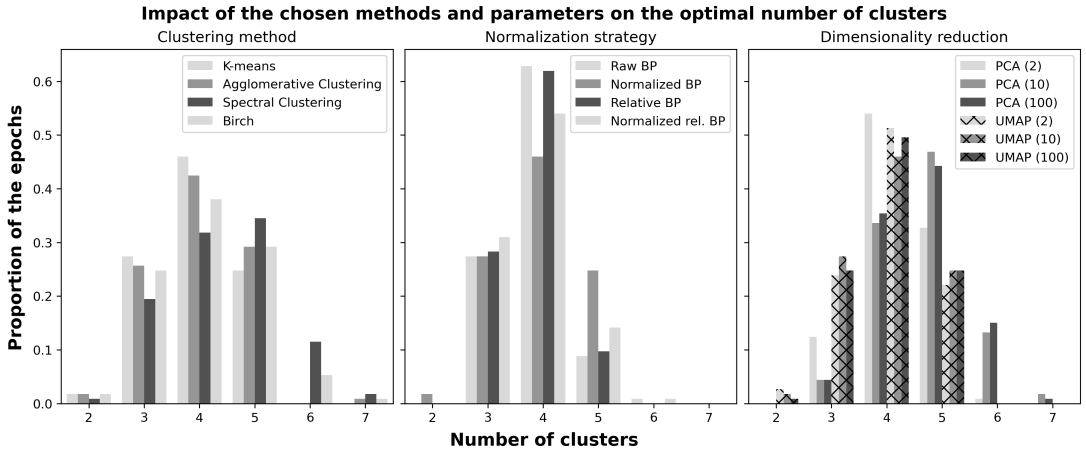

FIGURE 3.16: Distribution of the optimal number of clusters in all 113 sleep epochs. The results were obtained using the elbow method in the distortion score with different clustering methods (left), normalization strategies (middle) as well as dimensionality reduction techniques and parameters.

3.3.2 CHARACTERIZATION OF THE CLUSTERS

TEMPORAL CONSISTENCY

The clusters obtained are temporally consistent, meaning that time points close in time are much more likely to be assigned to the same cluster than predicted by chance, a necessary condition for them to represent sleep stages, which have been described to last from minutes to up to an hour in the literature (with occasional, short transitions). An example showing the temporal consistency of the clusters as well as their correlation with the visible oscillations in delta/beta power is shown in Figure 3.17.

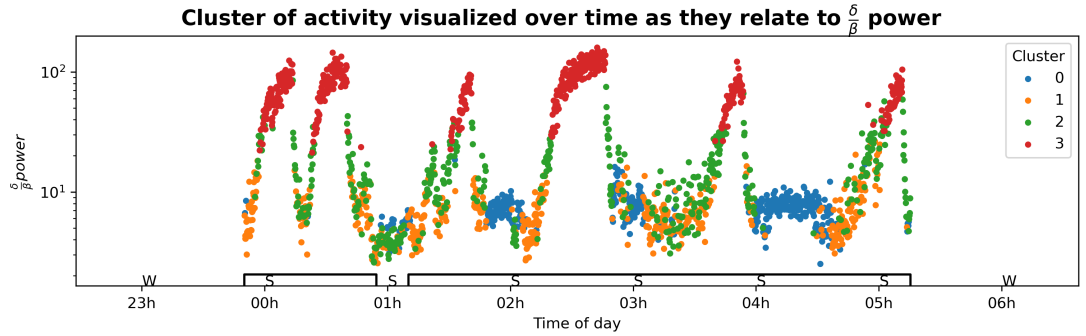

FIGURE 3.17: Temporal evolution of the four clusters as they relate to delta/beta power over a sleep epoch. The time points classifed as sleep by the unsupervised labeling procedure are shown with the solid black line at the bottom, and manual annotations from video are indicated on the hour with W for wake and S for sleep. We observe high temporal consistency, the tendency of neighbouring time points to be assigned to the same cluster.

Another example showing over every electrode in a patient with the obtained clusters can be found in Figure C.14. To quantify the temporal consistency, we compute the accuracy between the cluster assignment of even and odd time points. Before smoothing, the average accuracy is 0.79 (0.77-0.81), meaning that 79% of time points have been assigned the same label as the following time point, which is much higher than the chance level pf 0.28 (0.28-0.29) obtained by permuting these labels. After smoothing, it is increased to 0.91 (0.9-0.92), however, the consistency would be almost as high for smoothed, permuted labels (0.9, 0.89-0.9). This shows that the results that our procedure produces

labels are highly consistent in time, even prior to smoothing, although this step artificially increases their temporal consistency.

STABILITY

We also provide evidence that the clusters obtained are stable, since clustering using half of the electrodes or half of the time points leads to similar results. We obtain an average ARI of 0.74 (0.72-0.76) when using two different halves of the electrodes. This is far from chance level (0). The second test comparing the clustering outcomes of clustering from alternating time points gives a slightly higher ARI of 0.76 (0.72-0.79).

For comparison, the average ARI for the same tests conducted in wake epochs is 0.6 (0.58-0.62) and 0.73 (0.7-0.76) respectively. The first control indicates that the activity is more different throughout the brain during wake, and that we successfully clusters sleep into states that represent most of the brain pretty well. The second control is not very different from wake, indicating that this is probably a poor control. The activity between consecutive time points is highly correlated, resulting in quite similar clusters.

After having validated the robustness of the clusters we obtained, we were interested in providing a description of each individual cluster. We first consider the differences in the input PIB features, shown in Figure 3.18. To put the different frequency bands on a similar scale, we first apply min-max scaling to each feature. We observe many significant differences in the power of each frequency band within the different clusters. Delta power clearly increases through clusters 0 to 3, as expected by definition. This is true for theta and alpha power too. On the other hand, gamma power is comparatively lower in clusters representing deeper sleep. Another way to see it is that cluster 0 is characterized by high levels of high frequency activity (in particular, gamma), and lower levels of slow wave activity, while cluster 3 has more slow wave activity than the overnight average and less gamma activity. A more detailed comparison of these results separated by lobe is presented in appendix G.

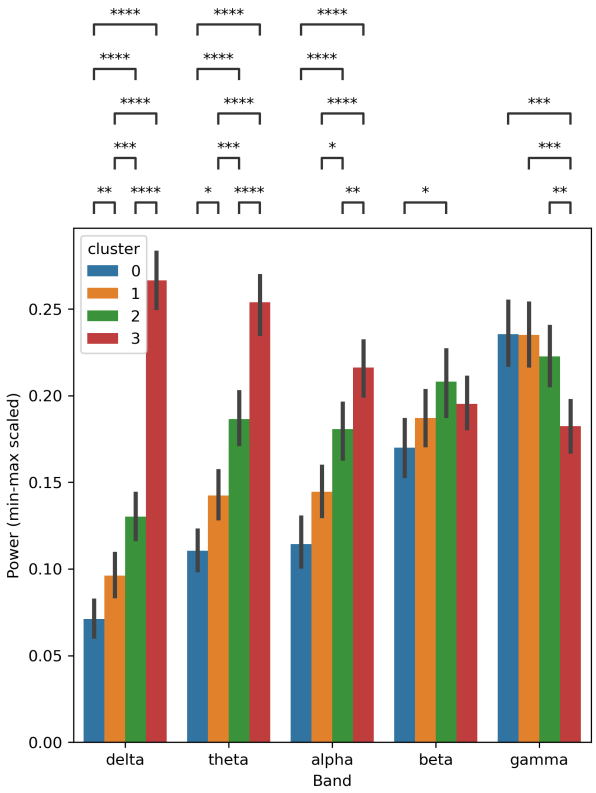

FIGURE 3.18: Comparison of power (min-max scaled) across each cluster for every band. The results of a MWU test with Bonferroni correction between all possible pairs of clusters within each band are presented.

TIME SPENT IN EACH CLUSTER

Next, we investigate the time spent in each cluster over the course of sleep epochs. Patients spent significantly more time in cluster 0 than 3 or 2, and cluster 1 than 3 (Figure 3.19). Overall, we can say that they spend more time in states of shallow sleep (characterized by low delta power).

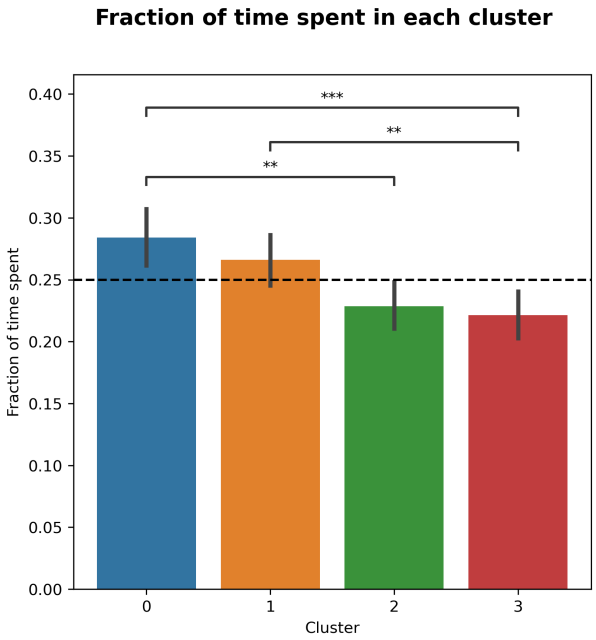

FIGURE 3.19: Average proportion of time spent in each cluster over all sleep epochs. The dashed line indicates chance level.

This pattern is conserved throughout cycles 1, 2, 3 and 4 (Figure 3.20). Later cycles were not examined due because we had fewer than 25 epochs with 5 cycles or more to extract statistics from.

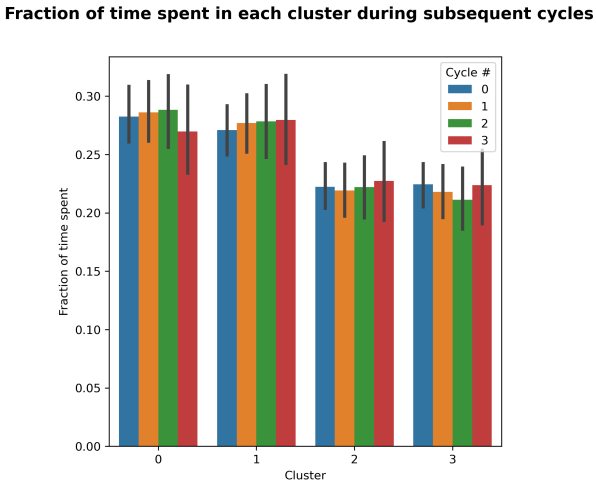

FIGURE 3.20: Evolution of the composition of cycles during subsequent sleep cycles.

AVERAGE PSD IN EACH CLUSTER

Furthermore, we compare the frequency content in the different clusters at a higher resolution using a multitaper spectrogram computed in a single electrode. For every epoch, we average the PSD across all time points in each cluster separately at frequencies below 17 Hz. We excluded higher frequencies which showed less variation with the different clusters, and in order to avoid visualizing the notch-filtered frequencies. These exhibit clear differences between the different clusters of sleep (Figure 3.21). In this example, cluster 3 shows the highest power in slow wave activity with a peak around 2 Hz (remember

that the data is bandpass-filtered between 0.5-100 Hz as a very first preprocessing step). Cluster 1 and 2 exhibit less power in low frequencies but an increase around 12 Hz. This could be due to slow sleep spindles, a hallmark sign of light NREM sleep[20]. However, these PSD signatures were not similar across patients. This could possibly be explained by the fact that they were generated from electrodes in different brain locations.

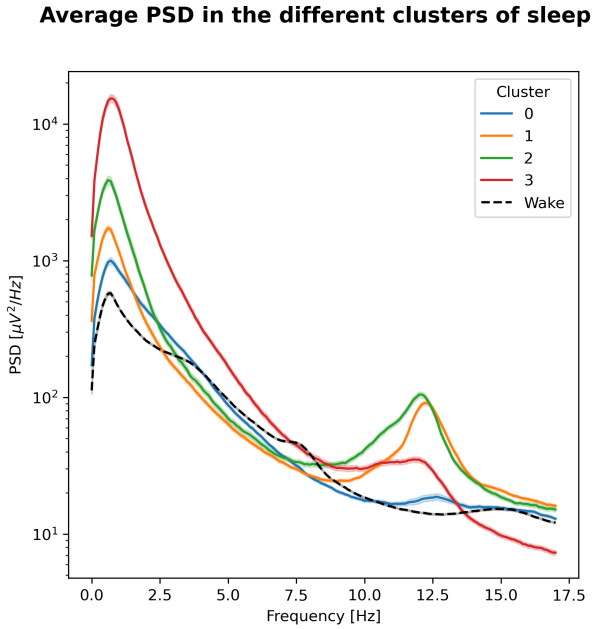

FIGURE 3.21: PSD signature of each cluster in a selected sleep epoch. The solid lines indicate the mean spectral density at each frequency, with the shaded area indicating a bootstrapped 95% CI for the mean. The average PSD during wake is shown for reference.

We attempted to summarize the spectral information in each clusters by aggregating these results across all sleep epochs (Figure 3.22). We observe that cluster 3 has higher slow wave activity, and, together with cluster 2, elevated power around 12 Hz (which could be due to sleep spindles). For the wake state, there is a slight increase in the alpha range.

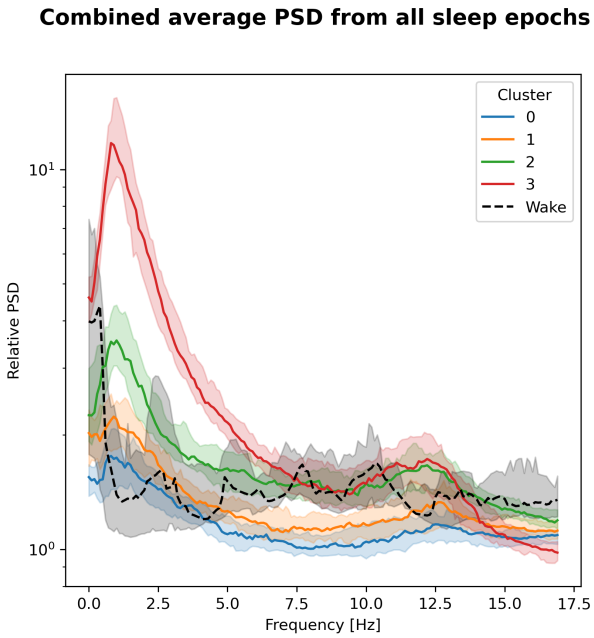

FIGURE 3.22: Median PSD signature of each cluster in a selected sleep epoch. This plot shows the median of each cluster signature, scaled by the average PSD in the corresponding epoch. The solid lines indicate the PSD signatures as in presented Figure 3.21 averaged across all epochs, with the shaded area indicating a bootstrapped 95% CI for the mean.

COMPARISON TO STAGES FROM THE SLEEP TRACKER

Lastly, as a means of getting an idea of how our clusters relate to the sleep stages defined by the AASM, we compare them to the labels provided by the sleep tracker (Figure 3.23). The sleep stages are to be taken with a grain of salt, as they are predicted by the sleep tracker unbeknownst to the brain activity. Moreover, the algorithm used by the sleep tracker relies on the typical sleep patterns to make predictions, when patients in monitoring are highly likely to exhibit disrupted sleep patterns. In the example shown, there is some degree of overlap between both sets of labels (ARI of 0.41 and 0.57 respectively). However, because we force the number of clusters to be 4, we could not possibly replicate the 3 stages in the second epoch (second row), which is lacking deep sleep according to the sleep tracker. In the other patient, the ARI between the stages produced by the sleep tracker and the clusters was much lower, with a mean of 0.09 (0.02-0.16). However, unlike for the patient shown, the labels from the sleep tracker did not correlate with the oscillatory delta/beta patterns, so we conclude that they were likely not accurate.

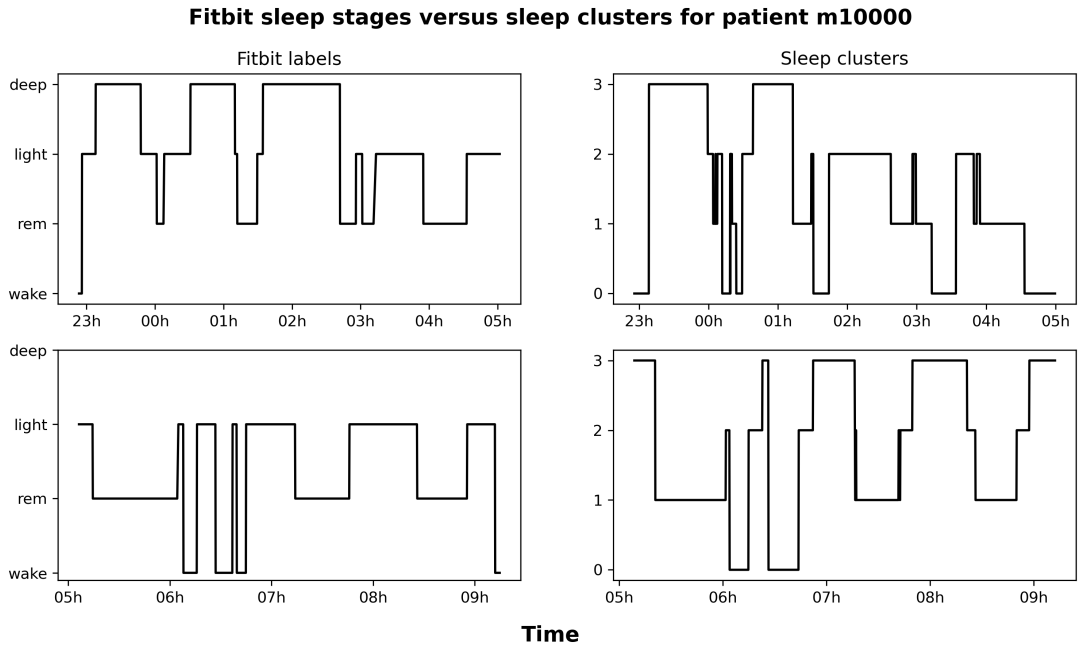

FIGURE 3.23: Sleep stages versus clusters. The ARI between the two sets of labels is 0.41 for the first sleep epoch (first row) and 0.57 for the second epoch (second row).

3.3.3 CLUSTERS OF ACTIVITY IN DIFFERENT LOBES

Clustering sleep using only electrodes from different lobes generally led to similar clusters (Figure 3.24).

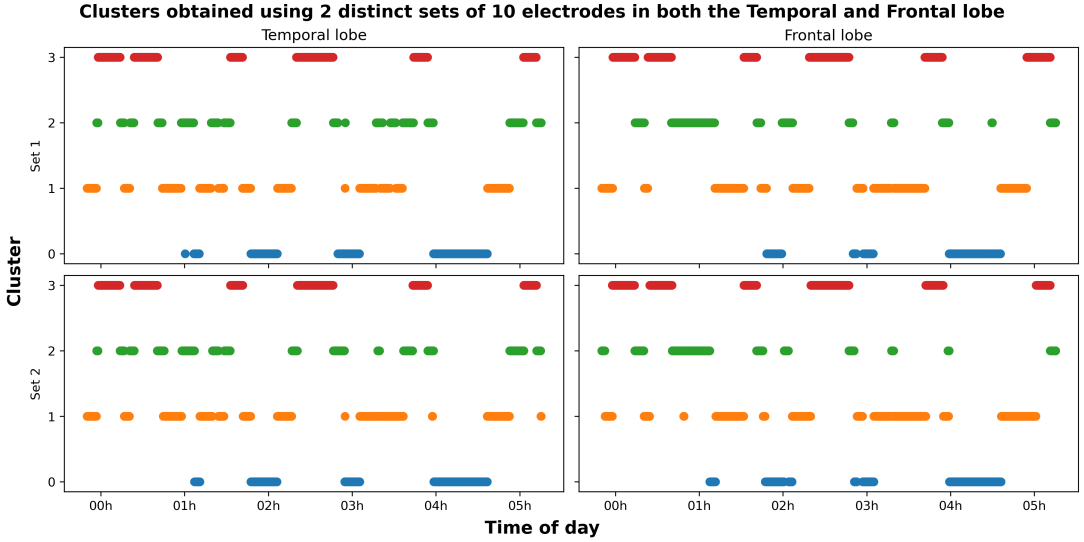

FIGURE 3.24: Comparison of the clusters obtained using 2 non-overlapping sets of 10 electrodes in the temporal and frontal lobe. The first row indicates the clusters derived from one set of electrodes in each lobe, and the second row shows those derived from 10 different electrodes from the same lobe. The ARI in this example is 0.74 and 0.7 within the temporal and frontal lobe respectively (comparing rows), and 0.47, 0.48, 0.5 and 0.51 for all possible combinations across the two lobes (comparing every pair between the left and right quadrants). This equates to an accuracy of 0.89 and 0.87 within, and 0.7, 0.72, 0.72 and 0.74 between the two lobes.

		5		10	
		Within	Across	Within	Across
Accuracy	Mean	0.71	0.62	0.78	0.57
	CI	0.69-0.74	0.56-0.68	0.75-0.8	0.52-0.59
ARI	Mean	0.44	0.32	0.69	0.42
	CI	0.4-0.5	0.23-0.42	0.67-0.72	0.38-0.49

TABLE 3.7: Average agreement scores. For each measure, a bootstrapped 95% CI is given. "Within" and "Across" indicates agreement between two sets of electrodes in the same lobe and across different lobes respectively.

Among all lobe pairs tested (frontal-parietal, frontal-temporal, parietal-occipital, temporal-occipital, temporal-parietal), we found that clusters obtained from two non-overlapping sets of electrodes within a lobe were significantly more similar to each other than those derived from electrodes in a different lobe (Figure 3.25 and Table 3.7). This supports the notion that different areas of the brain can be experiencing different types of sleep at the same time, or that the sleep stages are expressed differently depending on the brain region. The average ARI within a lobe was 0.44 and 0.32 across lobes with 5 electrodes, and these numbers were increased to 0.69 and 0.42 respectively when using 10 electrodes. Note that the ARI is naturally lower than the accuracy in all cases. The ARI is more reliable in case of incorrect mapping between the clusters, but the accuracy is provided for reference since it is likely to be more familiar to the reader. The frontal-occipital pair was excluded from the analysis due to an insufficient number of patients with the required coverage (10 electrodes in both regions). It is unsuprising that there aren't enough patients with electrodes in both the frontal and occipital lobe, given that electrodes are localized in a restricted area surrounding the suspected epileptogenic focus. The results of the same analysis performed with 10 electrodes per clustering iteration can be seen in Figure F.22.

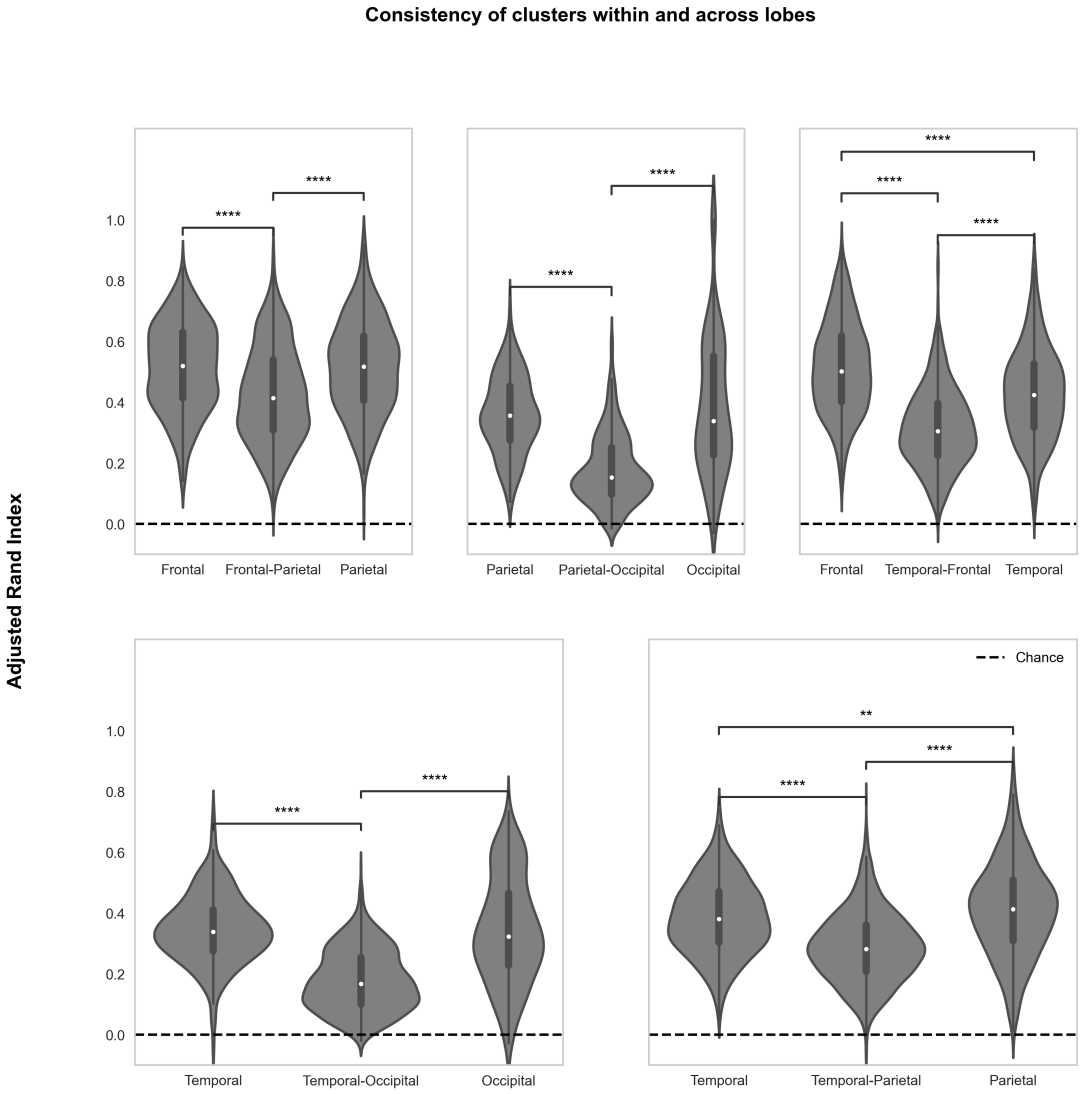

FIGURE 3.25: Distribution of the ARI of sets of cluster obtained within and between lobes using 10 electrodes to produce each clustering. The results of a MWU test with Bonferroni correction between the average ARI within the first lobe, within the second lobe, and between the first lobe and second lobe are shown.

The similarity of clusters derived using the exact same procedure within and across lobes during wake epochs is consistently much lower (average ARI of 0.32 and 0.17 with and across lobes respectively with 5 electrodes, 0.39 and 0.24 with 10 electrodes). This indicates that the brain activity is much more similar across areas during sleep, and that the clusters we obtain during sleep capture periods where the brain exhibits certain widespread patterns. It highlights that the signals we record are local in nature, which is less obvious during sleep due to synchronized activity. Put simply, brain areas are less different during sleep than they are during wake.

3.3.4 TEMPORAL EVOLUTION OF CLUSTERS AND DYNAMICS

In 88 of the 113 sleep epochs, we found that the cluster labels were periodic according to the autocorrelation method.

With the peak detection approach described in section 2.3, we could assess and visualize the consistency

in time of the occurrence of clusters throughout subsequent cycles. An example is shown in Figure 3.26.

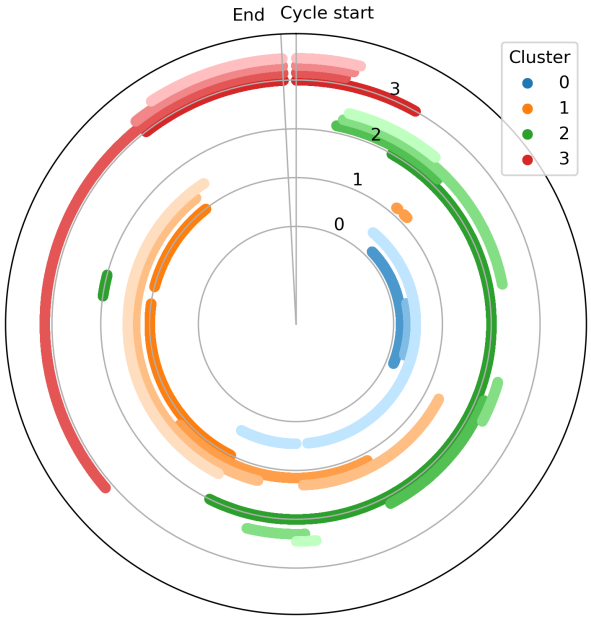

FIGURE 3.26: Composition of subsequent cycles in terms of the defined clusters. This example shows the temporal occurrence of the 4 clusters in 4 subsequent cycles (the first is the innermost and darkest line for each cluster and level) for a given sleep epoch. All 4 cycles start with cluster 3, and then transition to cluster 2. In the first cycle, this transition took place later (the first red line is longer than the others).

In most cases like the one presented here, the different clusters manifest roughly during the same phase of each cycle, consistently with the observation that the cluster labels are autocorrelated, although this sequence was not always conserved across sleep epochs and patients. By computing the most common label at each point along the evolution of a cycle (over a window of 5 minutes per 90 minute cycle), we produced the following plot indicating the typical evolution of clusters during sleep epochs:

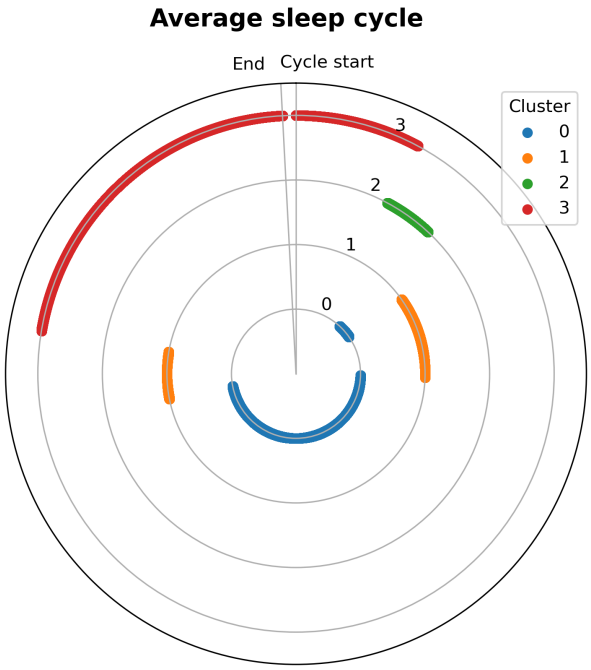

FIGURE 3.27: Typical temporal organization of the clusters within a cycle. The most common cluster label at each proportion of the evolution of a cycle across all sleep epochs is shown. The typical succession is cluster 3 at the start of a a new cycle, followed by 2, etc.

Lastly, we studied the frequency of different cluster transitions. It appears that all transitions are more or less equally frequent (Figure 3.28 and Table 3.8), with the exception of the transitions between 0 and 3. The same analysis separated by cycle number is presented in Figure H.24.

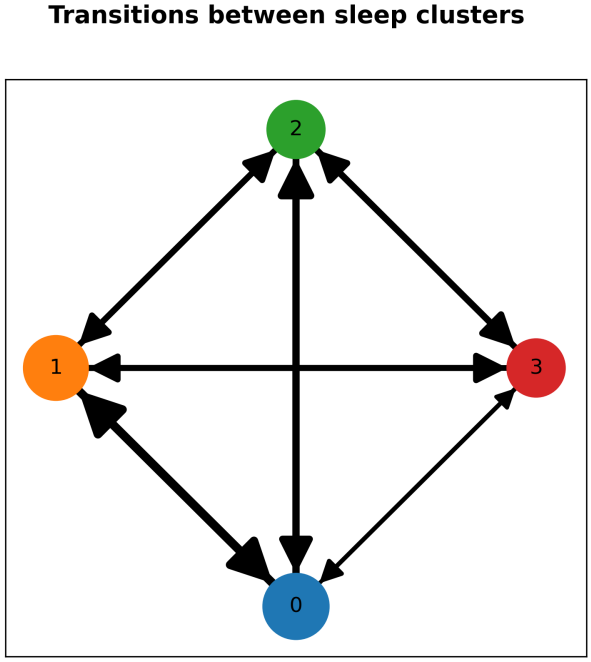

FIGURE 3.28: Transitions between the clusters. The edges are sized proportionally to the number of transitions (counts are reported in Table 3.8 between the clusters, represented by the nodes, sized proportionally to their duration (presented in Figure 3.19).

	To			
From	0	1	2	3
0	0	332	280	148
1	324	0	231	233
2	259	228	0	252
3	170	222	234	0

TABLE 3.8: Number of transition between the clusters.

By studying the frequency of occurrence of longer sequences transitions, it is clear that the most common transitions are oscillations between two states (e.g. the most frequent transition, 1-0-1, Figure 3.29) and not progressions into deeper or shallower sleep (e.g. the infrequent 3-2-1).

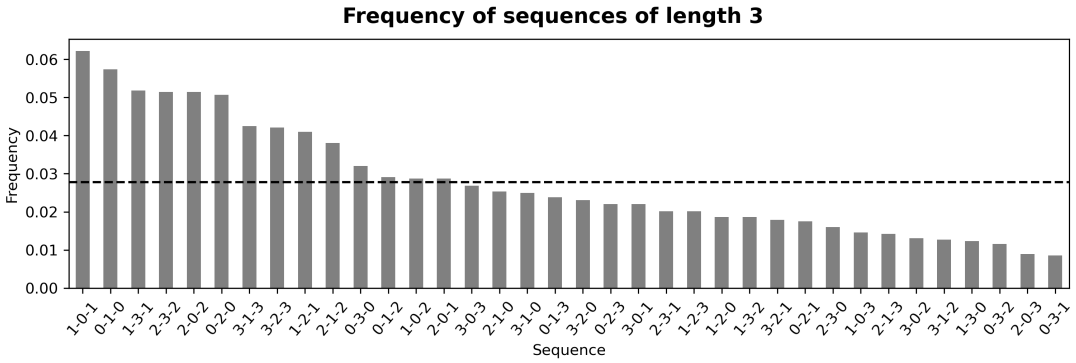

FIGURE 3.29: Frequency of all possible sequences of 3 clusters. The dashed line indicates chance level, and the solid line represent a bootstrapped 95% CI for the mean. Results of a MWU test with Bonferroni correction between all pairs of clusters are highlighted.

4. DISCUSSION

We provided a fully data-driven description of sleep. The only assumptions we made from the sleep literature were that cycles last between 30 and 180 minutes, which we could have deduced from our data, and more importantly that the delta, theta, alpha, beta and gamma band can be useful in describing sleep. A richer set of features could potentially have lead to different conclusions and reveal other properties of the organization of sleep. For example, it is known that brain activity during REM sleep and wake is similar at first glance, which justifies the use of the EOG in sleep studies. However, given how different these behavioral states are, it seems evident that it should be possible to clearly discriminate them with data such as the one used in this work, provided an appropriate description of the data is used.

UNSUPERVISED LABELING OF SLEEP AND WAKE STATES

We showed that a simple procedure could cluster iEEG recordings into sleep and wake with satisfying agreement with manual annotations or labels for an automatic sleep staging device. Given that neither of these two can be held as ground truth, it is not possible to determine what the actual accuracy in detecting sleep is. Nevertheless, it appears that this method can provide a simple way to increase the reliability of classification of behavioral state if used in combination with manual annotations or a sleep-staging device. If used in isolation, it is still possible to gain trust in the outcome by considering intrinsic properties of the resulting clusters. The first is the high temporal consistency of the obtained clusters. In other words, the two clusters are organized into alternating, long periods of one cluster followed by the other, and not always transitioning between the two clusters, which would be the expected result if we were not able to cluster sleep and wake. In some cases, we did notice that the clusters seemed to capture the first and second half of the recording because of long term drifts in the signals, which could be addressed by using the relative PIB and clustering shorter recordings separately. Another solution could have been to correct for these long term trends. The second is the high correlation between the clusters and the occurrence of oscillations in delta over beta power. And finally, two clusters were clearly separable in the embedding space in many cases, which is also strong evidence that the recordings could successfully be split into 2 very distinct behavioral states.

As expected, we found that delta power was increased during sleep. Interestingly, this effect was not lateralized in the frontal lobe, while the presence of oscillations in slow wave activity was. In the other bands, the modulation seemed to be different in the left versus right hemisphere.

SLEEP CYCLE PARAMETERS AND PERIODIC ACTIVITY

We were able to confirm the general parameters of sleep cycle duration and evolution in epileptic patients in hospital monitoring to be similar to what is known about healthy patients from PSG, despite using vastly different physiological recordings and methods.Fluctuating patterns of slow wave activity over the course of sleep cycles, are more likely to arise in some anatomical locations are more likely than others. As noted by Sekimoto et. al.[76] we saw that there is an asymmetry in slow wave activity during sleep between the left and right hemisphere but not in the other lobes. The presence of these patterns is

driven not only by recording location, but also strongly by individual variability. Previous iEEG studies had also shown that cyclical patterns of slow wave activity are not tightly related to recording location, and can even be detected locally during wakefulness in pathological cases[74]. The second method to estimate the duration of individual cycles based on peak detection in slow wave activity resulted in shorter cycles than the autocorrelation method (median of 75 minutes). This is because it is less reliable and is susceptible to detecting interrupted cycles.

CLUSTERS OF ACTIVITY DURING SLEEP

We provide solid evidence that our data is best represented with 4 clusters, but these results could be different with different physiological signals being recorded (i.e. rem and wake similar patterns potentially, use EOG), computation of a different subset of features, and maybe different methods. I suspect that information from the brain should be enough though, but need to describe it well. A recent study by also reached the conclusion that EEG sleep data could was best described using 4 data-driven stages[30].

Our approach provides further evidence that more time is spent in states of shallow sleep than deep sleep[69]. This is a powerful result given that KMeans is biased to find clusters of even size in the absence of clearly distinct groups [58]. However, we did not observe any differences in the composition of subsequent sleep cycles, contrarily to what as been described in the literature.

While we saw clear differences in the average PSD between the clusters in individual patients, it was harder to generalize those findings across patients. There are a few possible explanations for this. First, the PSD signatures could be different across patients because the chosen electrodes are in different brain locations and the clusters express different patterns of PSD across the brain. This is hard to correct for given that no area is covered in all patients. Second, there is inter- and intrapatient variability in spectral activity during sleep[20], which we cannot control for. Last but not least, it is likely that the 4 clusters don't always represent 4 distinct sleep stages (as could be the case when a sleep stage is not experience over an entire sleep epoch) or they don't perfectly map across patients (cluster 1 might not be the same sleep stage in patient as in patient 2, being the stage with the second lowest average delta power in both). Despite these challenges, we could still observe some differences in the typical PSD of each cluster. These findings suggest presence of slow waves (0.5-4 Hz) in cluster 3, sleep spindles in cluster 2 and 3 (around 12 Hz), and the elevated relative activity in the alpha range in wake could be attributed to the patients having their eyes open. Further tests would have to be conducted to confirm these hypotheses.

We show modest evidence of differences in the PIB within clusters and the timing of these clusters across the different lobes, although the second result would require stricter controls to exclude the possibility of the finding being solely attributable to volume conduction.

While we were able to reveal a cyclical structure to the obtained clusters in most cases, difficulties in mapping clusters across patients made it challenging to formulate conclusions about the structure of sleep within sleep cycles. The interpretation of data-driven clusters can be challenging, and could be aided in future studies by comparing to AASM stages obtained with PSG[40]. To further complicate matter, learned sleep representations are typically unstable between nights and show variability with demographics[40] which could be even greater with subject health status. Because of these difficulties, we could not find strong evidence for a sequential order between the clusters, despite what the sleep literature suggests. Our analysis suggests that oscillations between two different states are more common than a smooth, sequential evolution. Whether this is a correct finding enabled by a finer temporal resolution (10s epochs versus 30s in poysomnography, but more importantly because expert scorers tend to assign the same labels to sequential epochs) or an artifact caused by an imperfect mapping of the clusters across patients and noise in the clusters remains to be proven.

LIMITATIONS

While iEEG presents a special opportunity to study high quality brain signals recorded close to their source, there are several noteworthy limitations to this modality and to our approach in general. Because invasive surgery is only warranted in the case of a clinical need, we are constrained to study non-healthy subjects. Epileptic patients in hospital monitoring generally present with interrupted and poor quality sleep. Electrode locations are dictated by clinical needs and vary from patient to patient. This complicates generalization of findings about specific locations and implies that some structures of the brain are rarely recorded from. Because of the implanted electrodes, wearing an EEG headset for PSG is often not achievable or desirable for the patients, making comparisons to the established sleep literature challenging.

Our approach could have benefited from using a richer representation of the neural signals. We mostly used PIB features computed over 10s, but there are many other features that can be derived from neural signals and could be useful in descriing sleep. In particular, we did not fully exploit the high frequency information contained in the signals (for example, sharp wave ripples), which is generally less accessible to non-invasive recording modalities. We cannot guarantee that we would still find 4 clusters of sleep if clustering a more holistic representation of sleep, just like recording eye movements was originally necessary in PSG to distinguish wake and REM. One shortcoming of using data-driven clusters to partition sleep is the difficulty in mapping these clusters across recordings and especially across patients. Future studies should propose new strategies to address this problem so that conclusions can be extended to the population level.

APPLICATIONS

Extending our understanding of the local organization and regulation of sleep using iEEG holds promising translational applications. By identifying and characterizing abnormal neural activity associated with sleep disturbances, iEEG research could pave the way for the development of more effective therapeutic approaches, such as targeted neuromodulation or closed-loop stimulation techniques. These interventions have the potential to restore sleep functions in diseases that lead to or worsen sleep disorders and may even enhance the beneficial effects of sleep. Indeed, a very recent study demonstrated that real-time closed-loop deep brain stimulation delivering prefrontal pulses could enhance sleep spindles and their coordination with other aspects of sleep physiology, resulting in improved memory consolidation[32]. Furthermore, there is potential for commercial applications of brain-computer interfaces to enhance specific cognitive functions, combat daytime sleepiness, or mitigate sleep inertia.

BIBLIOGRAPHY

[1] Priyanka A Abhang, Bharti W Gawali and Suresh C Mehrotra. 'Technical aspects of brain rhythms and speech parameters'. In: *Introduction to EEG-and speech-based emotion recognition* (2016), pp. 51–79.

[2] Sandra Ackermann and Björn Rasch. 'Differential effects of non-REM and REM sleep on memory consolidation?' In: *Current neurology and neuroscience reports* 14.2 (2014), pp. 1–10.

[3] Michael R Anderberg. 'The broad view of cluster analysis'. In: *Cluster analysis for applications* 1.1 (1973), pp. 1–9.

[4] Kirstie N Anderson and Andrew J Bradley. 'Sleep disturbance in mental health problems and neurodegenerative disease'. In: *Nature and science of sleep* 5 (2013), p. 61.

[5] Mihael Ankerst et al. 'OPTICS: Ordering points to identify the clustering structure'. In: *ACM Sigmod record* 28.2 (1999), pp. 49–60.

[6] David Arthur and Sergei Vassilvitskii. 'K-means++ the advantages of careful seeding'. In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. 2007, pp. 1027–1035.

[7] Vlasta Březinová. 'Sleep cycle content and sleep cycle duration'. In: *Electroencephalography and clinical neurophysiology* 36 (1974), pp. 275–282.

[8] E Oran Brigham. *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.

[9] Ellen J Bubrick, Saad Yazdani and Milena K Pavlova. 'Beyond standard polysomnography: advantages and indications for use of extended 10–20 EEG montage during laboratory sleep study evaluations'. In: *Seizure* 23.9 (2014), pp. 699–702.

[10] Christian Cajochen, Rebecca Foy, Derk-Jan Dijk et al. 'Frontal predominance of a relative increase in sleep delta and theta EEG activity after sleep loss in humans'. In: *Sleep Res Online* 2.3 (1999), pp. 65–69.

[11] Tadeusz Caliński and Jerzy Harabasz. 'A dendrite method for cluster analysis'. In: *Communications in Statistics-theory and Methods* 3.1 (1974), pp. 1–27.

[12] Francesco P Cappuccio et al. 'Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis'. In: *Diabetes care* 33.2 (2010), pp. 414–420.

[13] Francesco P Cappuccio et al. 'Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies'. In: *Sleep* 33.5 (2010), pp. 585–592.

[14] David W Carley and Sarah S Farabi. 'Physiology of sleep'. In: *Diabetes Spectrum* 29.1 (2016), pp. 5–9.

[15] Mary A Carskadon, William C Dement et al. 'Normal human sleep: an overview'. In: *Principles and practice of sleep medicine* 4.1 (2005), pp. 13–23.

[16] Bruno Miguel Santos Carvalho, João Chaves and António Martins da Silva. 'Effects of antiepileptic drugs on sleep architecture parameters in adults'. In: *Sleep Science* 15.2 (2022), p. 224.

[17] Aslı Ece Çilliler and Bülent Güven. 'Sleep quality and related clinical features in patients with epilepsy: A preliminary report'. In: *Epilepsy & Behavior* 102 (2020), p. 106661.

[18] Nathan E Crone, Alon Sinai and Anna Korzeniewska. 'High-frequency gamma oscillations and human brain mapping with electrocorticography'. In: *Progress in brain research* 159 (2006), pp. 275–295.
[19] Giuseppe Curcio, Michele Ferrara and Luigi De Gennaro. 'Sleep loss, learning capacity and academic performance'. In: *Sleep medicine reviews* 10.5 (2006), pp. 323–337.
[20] Luigi De Gennaro and Michele Ferrara. 'Sleep spindles: an overview'. In: *Sleep medicine reviews* 7.5 (2003), pp. 423–440.
[21] Nicolas Decat et al. 'Beyond traditional sleep scoring: Massive feature extraction and data-driven clustering of sleep time series'. In: *Sleep Medicine* 98 (2022), pp. 39–52.
[22] Christopher P Derry and Susan Duncan. 'Sleep and epilepsy'. In: *Epilepsy & Behavior* 26.3 (2013), pp. 394–404.
[23] Xiaofang Dong et al. 'Validation of Fitbit Charge 4 for assessing sleep in Chinese patients with chronic insomnia: A comparison against polysomnography and actigraphy'. In: *Plos one* 17.10 (2022), e0275287.
[24] Nicolás von Ellenrieder et al. 'SleepSEEG: automatic sleep scoring using intracranial EEG recordings only'. In: *Journal of Neural Engineering* 19.2 (2022), p. 026057.
[25] Lorenz A Fenk, Juan Luis Riquelme and Gilles Laurent. 'Interhemispheric competition during sleep'. In: *Nature* 616.7956 (2023), pp. 312–318.
[26] Jane E Ferrie et al. 'A prospective study of change in sleep duration: associations with mortality in the Whitehall II cohort'. In: *Sleep* 30.12 (2007), pp. 1659–1666.
[27] Bruce Fischl. 'FreeSurfer'. In: *Neuroimage* 62.2 (2012), pp. 774–781.
[28] Michelle A Frazer et al. 'Shining a light on the mechanisms of sleep for memory consolidation'. In: *Current Sleep Medicine Reports* 7 (2021), pp. 221–231.
[29] Itzhak Fried et al. *Single neuron studies of the human brain: probing cognition*. MIT Press, 2014.
[30] Clifton Frilot II, David E McCarty and Andrew A Marino. 'An original method for staging sleep based on dynamical analysis of a single EEG signal'. In: *Journal of Neuroscience Methods* 308 (2018), pp. 135–141.
[31] Steffen Gais and Jan Born. 'Declarative memory consolidation: mechanisms acting during human sleep'. In: *Learning & Memory* 11.6 (2004), pp. 679–685.
[32] Maya Geva-Sagiv et al. 'Augmenting hippocampal–prefrontal neuronal synchrony during sleep enhances memory consolidation in humans'. In: *Nature Neuroscience* (2023), pp. 1–11.
[33] S Groch et al. 'The role of REM sleep in the processing of emotional memories: evidence from behavior and event-related potentials'. In: *Neurobiology of learning and memory* 99 (2013), pp. 1–9.
[34] David M Groppe et al. 'iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data'. In: *Journal of neuroscience methods* 281 (2017), pp. 40–48.
[35] Shahab Haghayegh et al. 'Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis'. In: *Journal of medical Internet research* 21.11 (2019), e16273.
[36] Péter Halász. 'The K-complex as a special reactive sleep slow wave–a theoretical update'. In: *Sleep medicine reviews* 29 (2016), pp. 34–40.
[37] Thalia Harmony. 'The functional significance of delta oscillations in cognitive processing'. In: *Frontiers in integrative neuroscience* 7 (2013), p. 83.
[38] Ernest Hartmann. 'The 90-minute sleep-dream cycle'. In: *Archives of General Psychiatry* 18.3 (1968), pp. 280–286.
[39] Werner Haustein et al. 'Automatic analysis overcomes limitations of sleep stage scoring'. In: *Electroencephalography and clinical neurophysiology* 64.4 (1986), pp. 364–374.
[40] Lieke WA Hermans et al. 'Representations of temporal sleep dynamics: Review and synthesis of the literature'. In: *Sleep Medicine Reviews* (2022), p. 101611.
[41] Lawrence Hubert and Phipps Arabie. 'Comparing partitions'. In: *Journal of classification* 2 (1985), pp. 193–218.

[42] Conrad Iber. 'The AASM manual for the scoring of sleep and associated events: Rules'. In: *Terminology and Technical Specification* (2007).
[43] Alice F Jackson and Donald J Bolger. 'The neurophysiological bases of EEG and EEG measurement: A review for the rest of us'. In: *Psychophysiology* 51.11 (2014), pp. 1061–1071.
[44] Bo Jin et al. 'Epilepsy and its interaction with sleep and circadian rhythm'. In: *Frontiers in Neurology* 11 (2020), p. 327.
[45] Katherine A Kaplan et al. 'When a gold standard isn't so golden: Lack of prediction of subjective sleep quality from sleep polysomnography'. In: *Biological psychology* 123 (2017), pp. 37–46.
[46] Vaclav Kremen et al. 'Automated unsupervised behavioral state classification using intracranial electrophysiology'. In: *Journal of neural engineering* 16.2 (2019), p. 026004.
[47] James M Krueger et al. 'Sleep as a fundamental property of neuronal assemblies'. In: *Nature Reviews Neuroscience* 9.12 (2008), pp. 910–919.
[48] Siyu Long et al. 'Sleep quality and electroencephalogram delta power'. In: *Frontiers in Neuroscience* 15 (2021), p. 803507.
[49] Faith S Luyster et al. 'Sleep: a health imperative'. In: *Sleep* 35.6 (2012), pp. 727–734.
[50] Mark W Mahowald and Carlos H Schenck. 'Insights from studying human sleep disorders'. In: *Nature* 437.7063 (2005), pp. 1279–1285.
[51] Beth A Malow et al. 'Relationship of interictal epileptiform discharges to sleep depth in partial epilepsy'. In: *Electroencephalography and clinical neurophysiology* 102.1 (1997), pp. 20–26.
[52] Bryce A Mander. 'Local sleep and Alzheimer's disease pathophysiology'. In: *Frontiers in Neuroscience* 14 (2020), p. 525970.
[53] Pierre Maquet. 'The role of sleep in learning and memory'. In: *science* 294.5544 (2001), pp. 1048–1052.
[54] Cristina Marzano et al. 'How we fall asleep: regional and temporal differences in electroencephalographic synchronization at sleep onset'. In: *Sleep medicine* 14.11 (2013), pp. 1112–1122.
[55] Leland McInnes, John Healy and James Melville. 'Umap: Uniform manifold approximation and projection for dimension reduction'. In: *arXiv preprint arXiv:1802.03426* (2018).
[56] Patrick E McKnight and Julius Najab. 'Mann-Whitney U Test'. In: *The Corsini encyclopedia of psychology* (2010), pp. 1–1.
[57] Maria Paola Mogavero et al. 'Sleep disorders and cancer: State of the art and future perspectives'. In: *Sleep Medicine Reviews* 56 (2021), p. 101409.
[58] Laurence Morissette and Sylvain Chartier. 'The k-means clustering technique: General considerations and implementation in Mathematica'. In: *Tutorials in Quantitative Methods for Psychology* 9.1 (2013), pp. 15–24.
[59] Bettina Müller, Wolf Dietrich Gäbelein and Hartmut Schulz. 'A taxonomic analysis of sleep stages'. In: *Sleep* 29.7 (2006), pp. 967–974.
[60] Marcus Ng and Milena Pavlova. 'Why are seizures rare in rapid eye movement sleep? Review of the frequency of seizures in different sleep stages'. In: *Epilepsy research and treatment* 2013 (2013).
[61] Yuval Nir et al. 'Microelectrode studies of human sleep'. In: *Single Neuron Studies of the Human Brain: Probing Cognition* (2014), pp. 165–188.
[62] Yuval Nir et al. 'Regional slow waves and spindles in human sleep'. In: *Neuron* 70.1 (2011), pp. 153–169.
[63] Lino Nobili et al. 'Sleep and epilepsy: A snapshot of knowledge and future research lines'. In: *Journal of Sleep Research* 31.4 (2022), e13622.
[64] PL Nunez and EEG In VS Ramachandran. 'Encyclopedia of the human brain'. In: *Sci. Rep* 20 (2002), p. 20.
[65] Robert D Ogilvie. 'The process of falling asleep'. In: *Sleep medicine reviews* 5.3 (2001), pp. 247–270.

[66] Levent Oztürk et al. ‘Effects of 48 hours sleep deprivation on human immune profile.’ In: *Sleep research online: SRO* 2.4 (1999), pp. 107–111.
[67] Xenophon Papademetris et al. ‘BioImage Suite: An integrated medical image analysis suite: An update’. In: *The insight journal* 2006 (2006), p. 209.
[68] Josef Parvizi and Sabine Kastner. ‘Human intracranial EEG: promises and limitations’. In: *Nature neuroscience* 21.4 (2018), p. 474.
[69] Aakash K Patel, Vamsi Reddy and John F Araujo. ‘Physiology, sleep stages’. In: *StatPearls [Internet]*. StatPearls Publishing, 2022.
[70] Laure Peter-Derex et al. ‘Sleep and Alzheimer’s disease’. In: *Sleep medicine reviews* 19 (2015), pp. 29–38.
[71] Michael J Prerau et al. ‘Sleep neurophysiological dynamics through the lens of multitaper spectral analysis’. In: *Physiology* 32.1 (2017), pp. 60–92.
[72] Peter J Rousseeuw. ‘Silhouettes: a graphical aid to the interpretation and validation of cluster analysis’. In: *Journal of computational and applied mathematics* 20 (1987), pp. 53–65.
[73] Anastasiya Runnova et al. ‘Automatic wavelet-based assessment of behavioral sleep using multichannel electrocorticography in rats’. In: *Sleep and Breathing* 25.4 (2021), pp. 2251–2258.
[74] Robert NS Sachdev et al. ‘Delta rhythm in wakefulness: evidence from intracranial recordings in human beings’. In: *Journal of neurophysiology* 114.2 (2015), pp. 1248–1254.
[75] Jason Samaha and Bradley R Postle. ‘The speed of alpha-band oscillations predicts the temporal resolution of visual perception’. In: *Current Biology* 25.22 (2015), pp. 2985–2990.
[76] Masanori Sekimoto et al. ‘Asymmetric interhemispheric delta waves during all-night sleep in humans’. In: *Clinical neurophysiology* 111.5 (2000), pp. 924–928.
[77] Jianbo Shi and Jitendra Malik. ‘Normalized cuts and image segmentation’. In: *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000), pp. 888–905.
[78] Xiao Shuyuan et al. ‘Notice of Removal: An improved K-means clustering algorithm for sleep stages classification’. In: *2015 54th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*. IEEE. 2015, pp. 1222–1227.
[79] Pedro Sudbrack-Oliveira et al. ‘Sleep architecture in adults with epilepsy: a systematic review’. In: *Sleep medicine* 53 (2019), pp. 22–27.
[80] Samantha Sun et al. ‘Unsupervised Sleep and Wake State Identification in Long-Term Electrocorticography Recordings’. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 629–632.
[81] Kristina Šušmakova and Anna Krakovska. ‘Discrimination ability of individual measures used in sleep stages classification’. In: *Artificial intelligence in medicine* 44.3 (2008), pp. 261–277.
[82] David J Thomson. ‘Spectrum estimation and harmonic analysis’. In: *Proceedings of the IEEE* 70.9 (1982), pp. 1055–1096.
[83] Robert Tibshirani, Guenther Walther and Trevor Hastie. ‘Estimating the number of clusters in a data set via the gap statistic’. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2 (2001), pp. 411–423.
[84] Mark W Mahowald, Michel A Cramer Bornemann and Carlos H Schenck. ‘State dissociation, human behavior, and consciousness’. In: *Current topics in medicinal chemistry* 11.19 (2011), pp. 2392–2402.
[85] Jiarui Wang et al. ‘Mesoscopic physiological interactions in the human brain reveal small-world properties’. In: *Cell reports* 36.8 (2021), p. 109585.
[86] Andrew M Watson. ‘Sleep and athletic performance’. In: *Current sports medicine reports* 16.6 (2017), pp. 413–418.
[87] Audry Devisanty Wuysang et al. ‘The relationship between sleep quality and the severity of insomnia with the frequency of seizures in epileptic patients’. In: *Medicina Clinica Practica* 4 (2021), p. 100215.

[88] Magdy Younes et al. ‘Odds ratio product of sleep EEG as a continuous measure of sleep state’. In: *Sleep* 38.4 (2015), pp. 641–654.
[89] Magdy Younes et al. ‘Reliability of the American Academy of Sleep Medicine rules for assessing sleep depth in clinical practice’. In: *Journal of Clinical Sleep Medicine* 14.2 (2018), pp. 205–213.
[90] James J Young et al. ‘Quantitative signal characteristics of electrocorticography and stereoelectroencephalography: the effect of contact depth’. In: *Journal of Clinical Neurophysiology* 36.3 (2019), p. 195.
[91] Tian Zhang, Raghu Ramakrishnan and Miron Livny. ‘BIRCH: an efficient data clustering method for very large databases’. In: *ACM sigmod record* 25.2 (1996), pp. 103–114.

Appendices

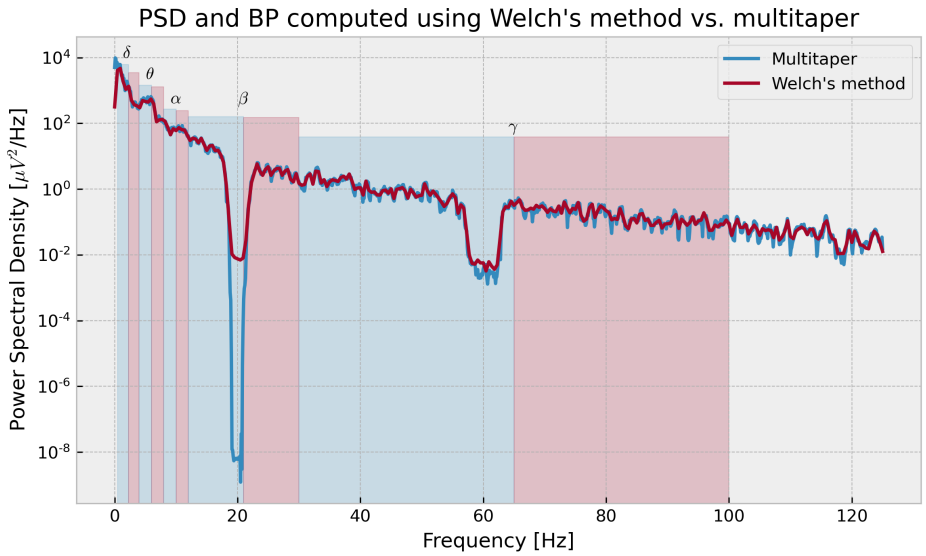

FIGURE A.1: PSD computed using Welch's periodogram versus the multitapering technique. The PSD computed using both approximations from a 10s signal in a single electrode are compared (solid lines). The resulting features (indicated by the height of the translucent rectangles) are indicated in units of μV^2 and are very similar.

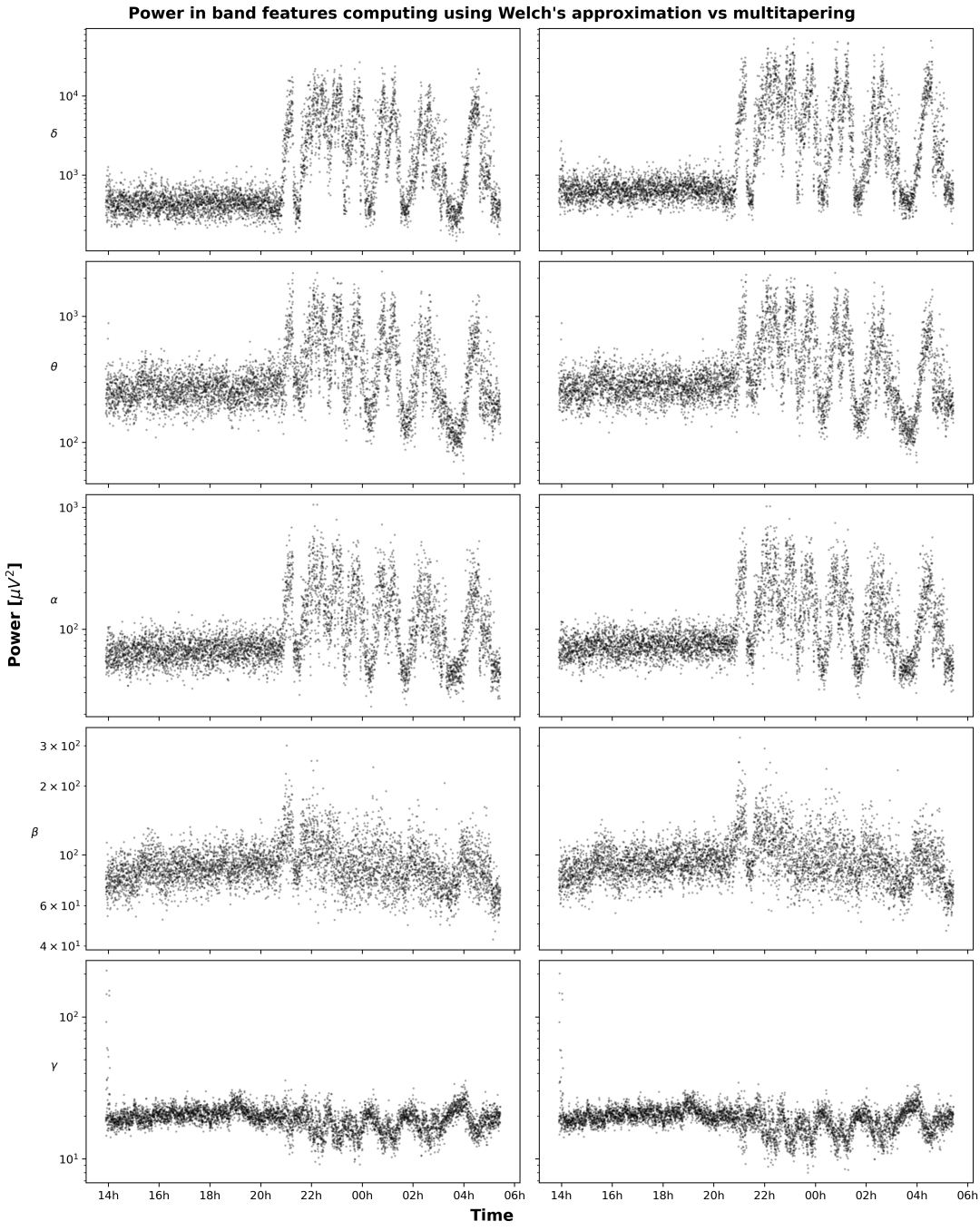

FIGURE A.2: Comparison of features obtained with Welch's approximation and the multitapering technique. The features computed using Welch's periodogram (left) are very similar to those obtained using the more computationally intensive multitapering technique and display identical patterns over time.

A FEATURES

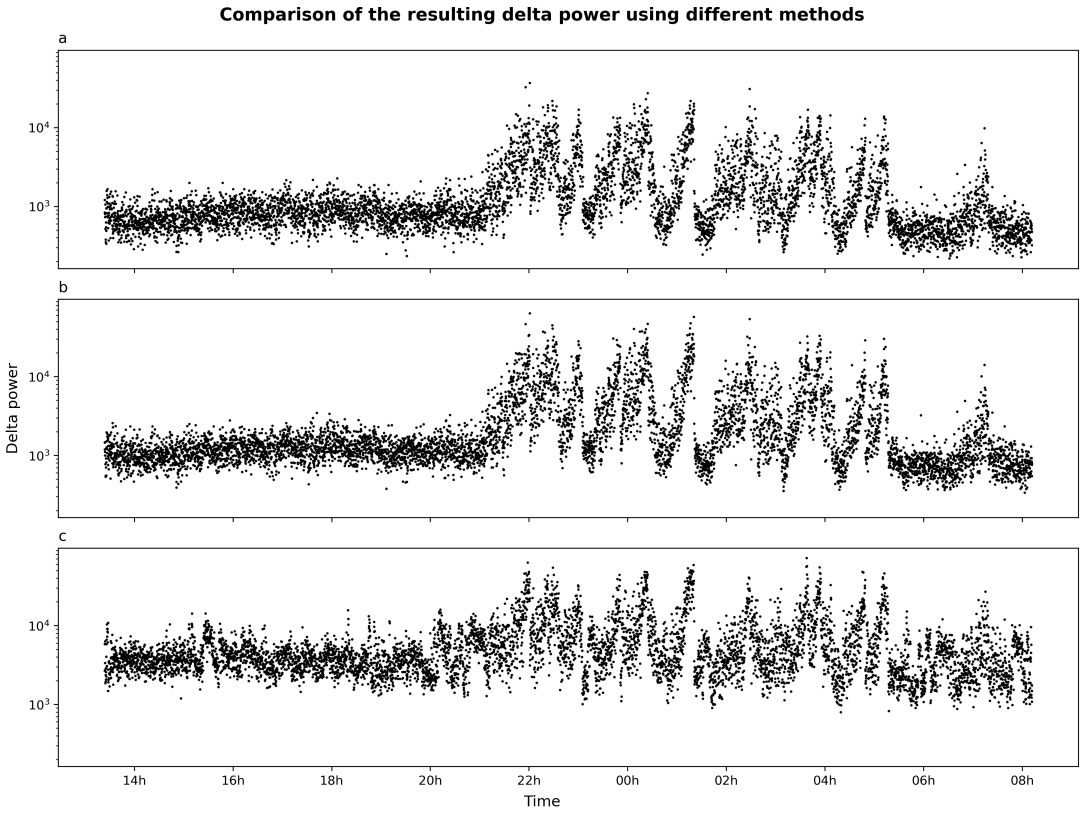

FIGURE A.3: Delta power computed with different re-referencing methods and spectral approximations. (a) Delta power in a single electrode computed with the chosen methods (common average referencing and Welch's periodogram to estimate the PSD) (b) Delta power in a single electrode computed using a multitaper spectrogram using the CAR signals (c) Delta power in a bipolar electrodes computed using a multitaper spectrogram after bipolar referencing

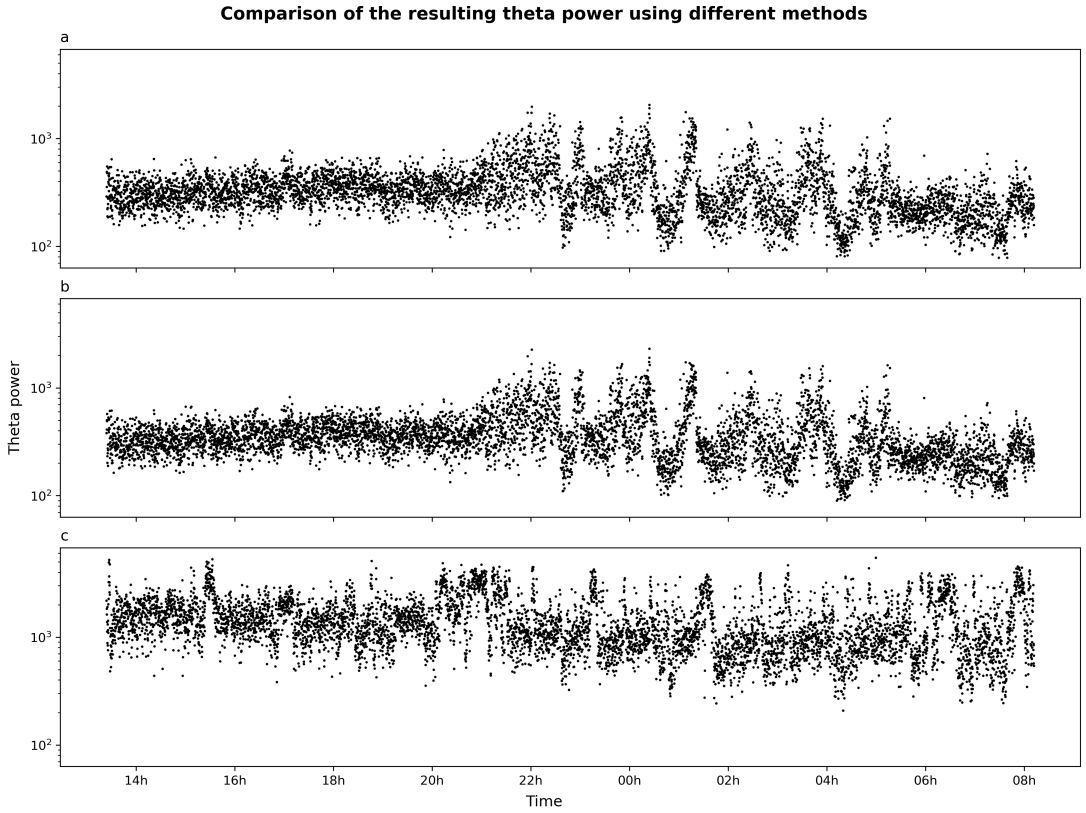

FIGURE A.4: Theta power computed with different re-referencing methods and spectral approximations. (a) Theta power in a single electrode computed with the chosen methods (common average referencing and Welch's periodogram to estimate the PSD) (b) Theta power in a single electrode computed using a multitaper spectrogram using the CAR signals (c) Theta power in a bipolar electrodes computed using a multitaper spectrogram after bipolar referencing

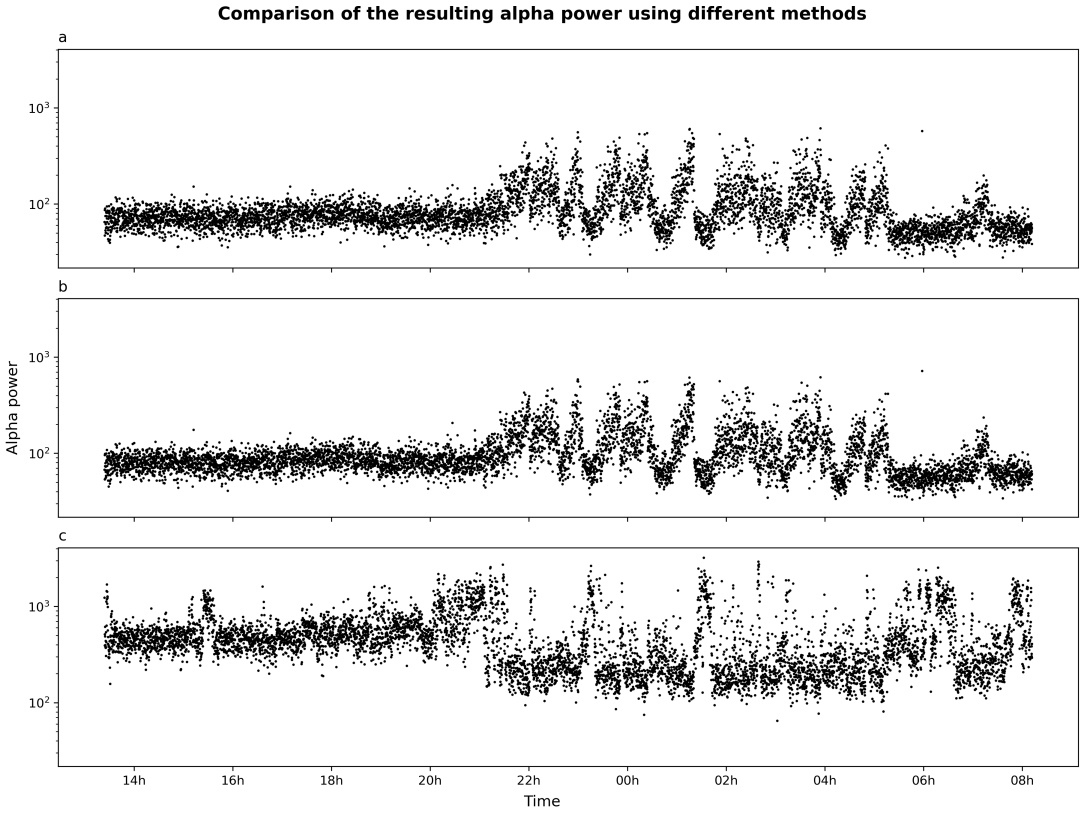

FIGURE A.5: Alpha power computed with different re-referencing methods and spectral approximations. (a) Alpha power in a single electrode computed with the chosen methods (common average referencing and Welch's periodogram to estimate the PSD) (b) Alpha power in a single electrode computed using a multitaper spectrogram using the CAR signals (c) Alpha power in a bipolar electrodes computed using a multitaper spectrogram after bipolar referencing

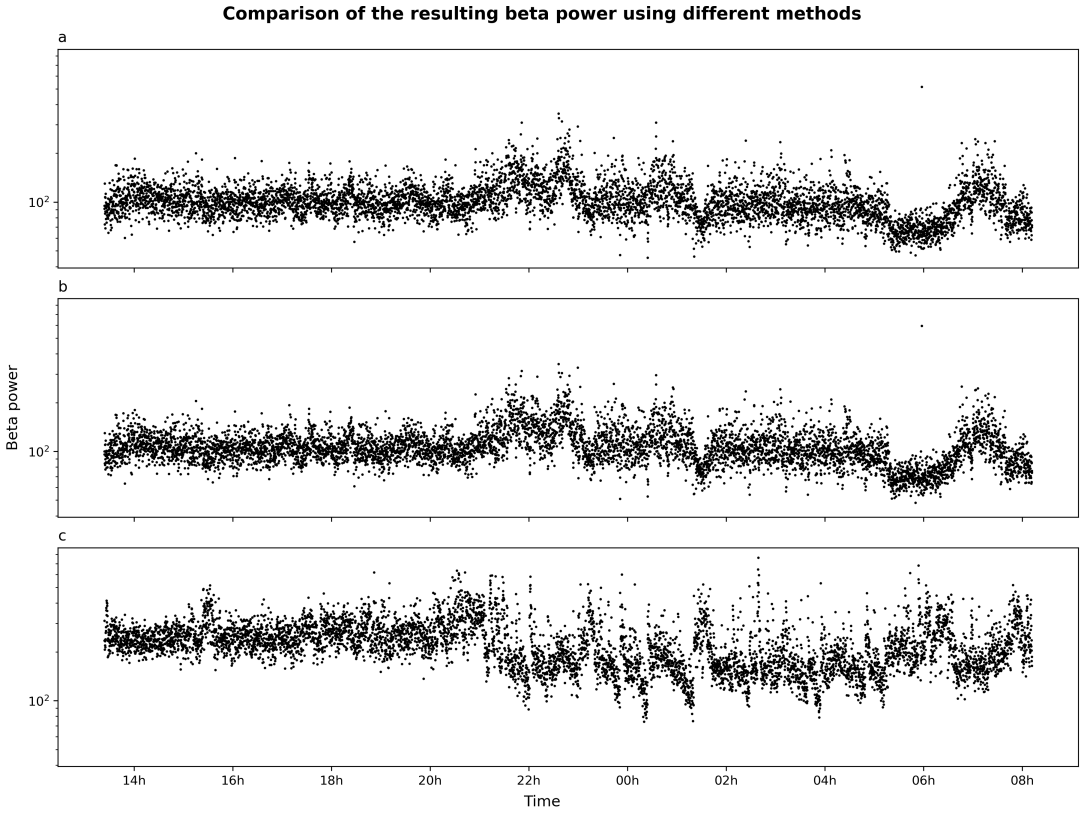

FIGURE A.6: Beta power computed with different re-referencing methods and spectral approximations. (a) Beta power in a single electrode computed with the chosen methods (common average referencing and Welch's periodogram to estimate the PSD) (b) Beta power in a single electrode computed using a multitaper spectrogram using the CAR signals (c) Beta power in a bipolar electrodes computed using a multitaper spectrogram after bipolar referencing

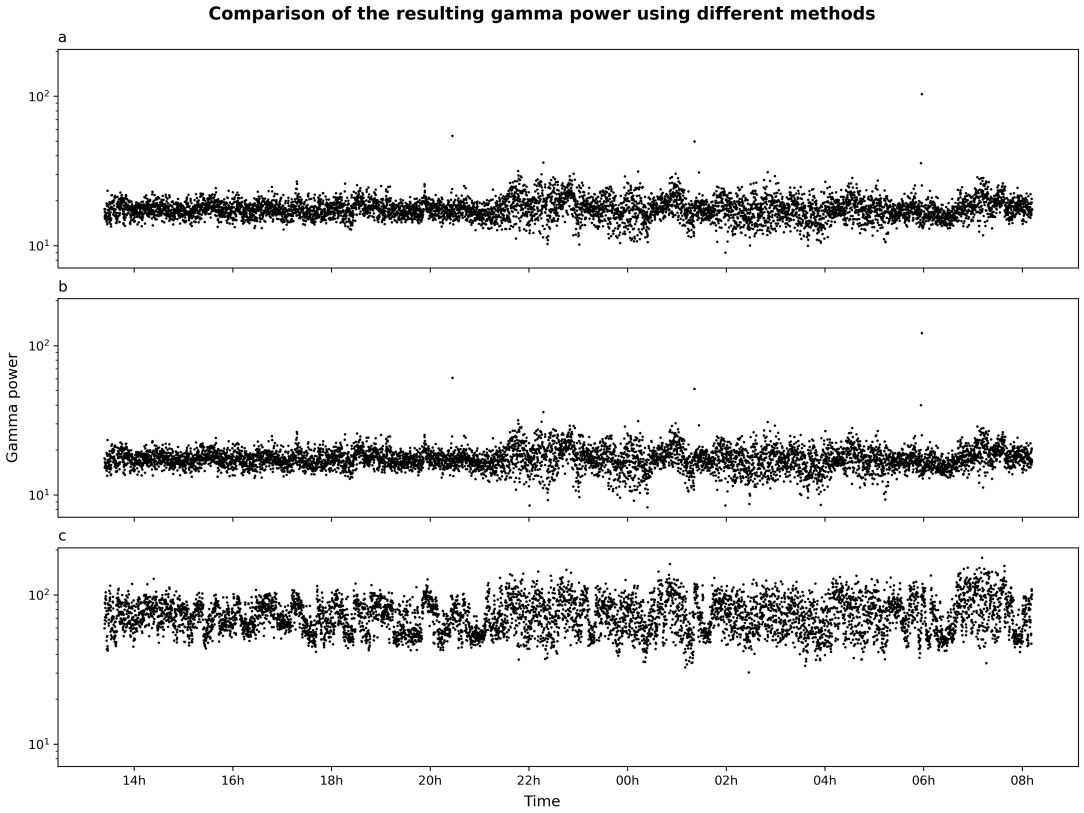

FIGURE A.7: Gamma power computed with different re-referencing methods and spectral approximations. (a) Gamma power in a single electrode computed with the chosen methods (common average referencing and Welch's periodogram to estimate the PSD) (b) Gamma power in a single electrode computed using a multitaper spectrogram using the CAR signals (c) Gamma power in a bipolar electrodes computed using a multitaper spectrogram after bipolar referencing

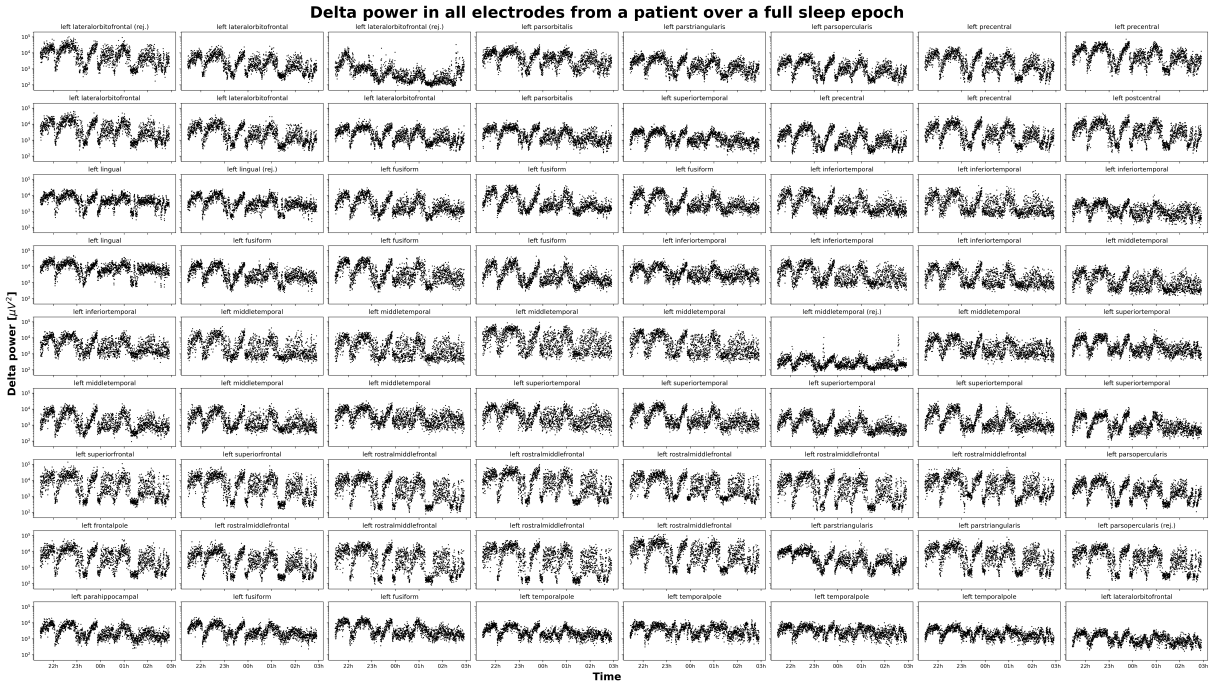

FIGURE A.8: Delta power in all electrodes over a sleep epoch. The anatomical locations of the electrodes in Desikan-Killiany parcellation are indicated, and those rejected upon visual inspection or because they contained artifacts (as defined in section 2.1.3) in more than 50% of epochs are marked *rej.*.

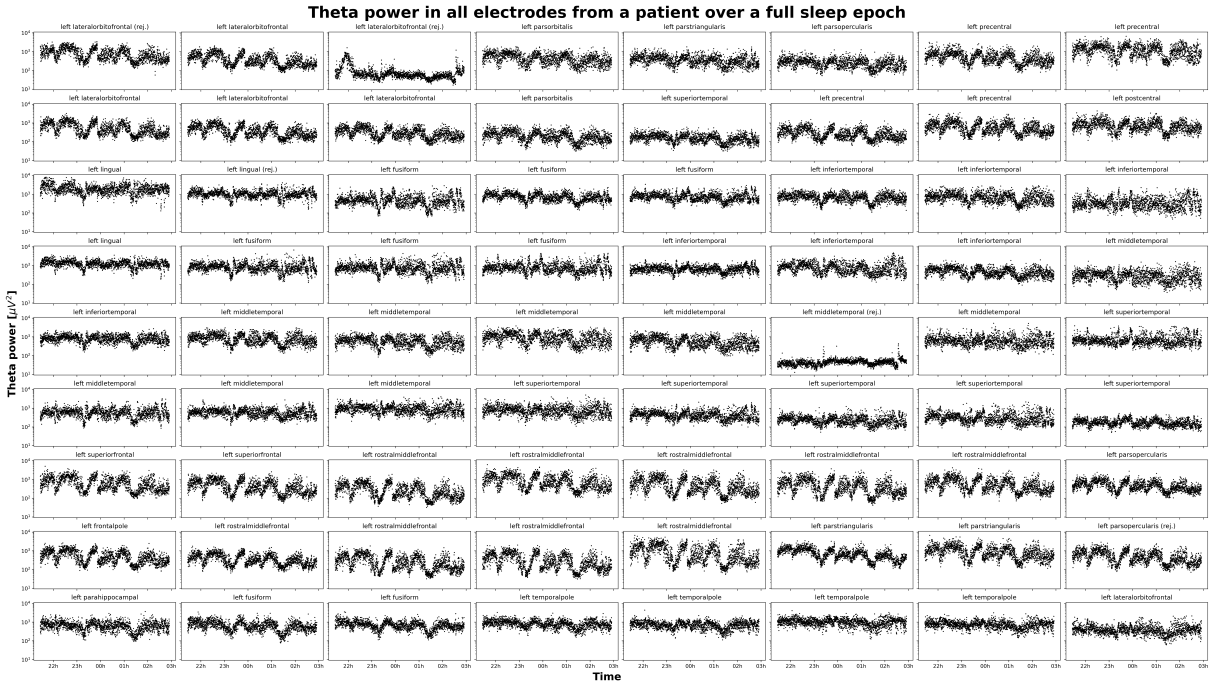

FIGURE A.9: Theta power in all electrodes over a sleep epoch. The anatomical locations of the electrodes in Desikan-Killiany parcellation are indicated, and those rejected upon visual inspection or because they contained artifacts (as defined in section 2.1.3) in more than 50% of epochs are marked *rej.*.

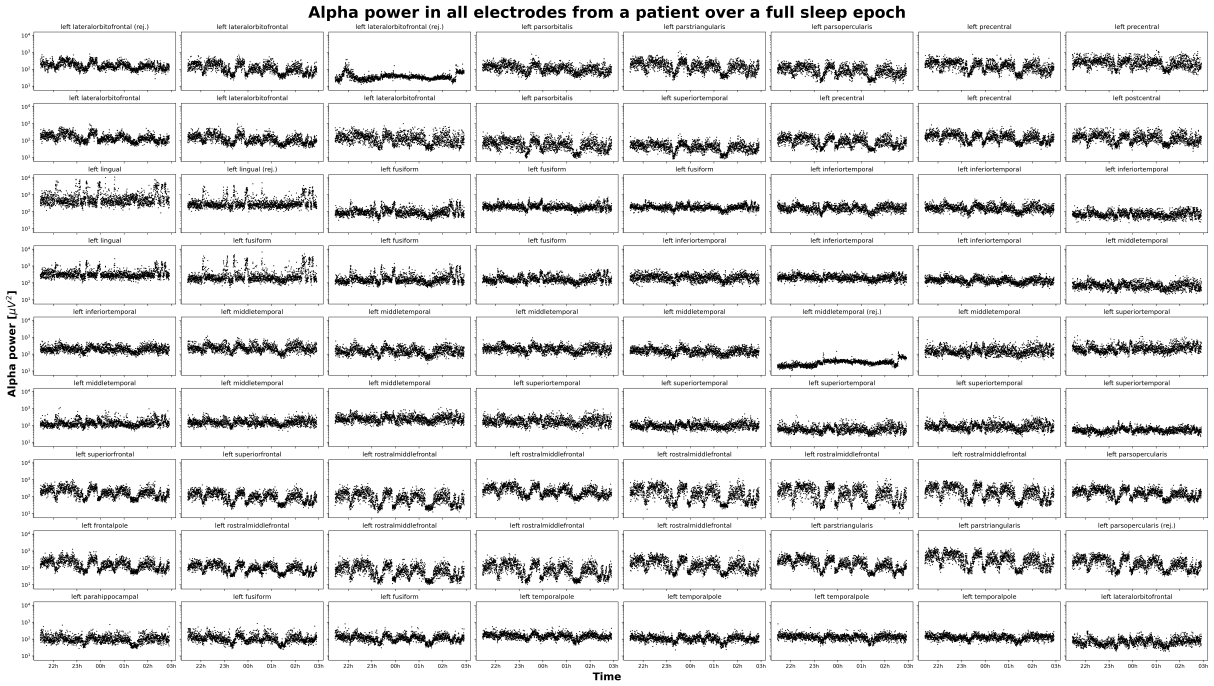

FIGURE A.10: Alpha power in all electrodes over a sleep epoch. The anatomical locations of the electrodes in Desikan-Killiany parcellation are indicated, and those rejected upon visual inspection or because they contained artifacts (as defined in section 2.1.3) in more than 50% of epochs are marked *rej.*.

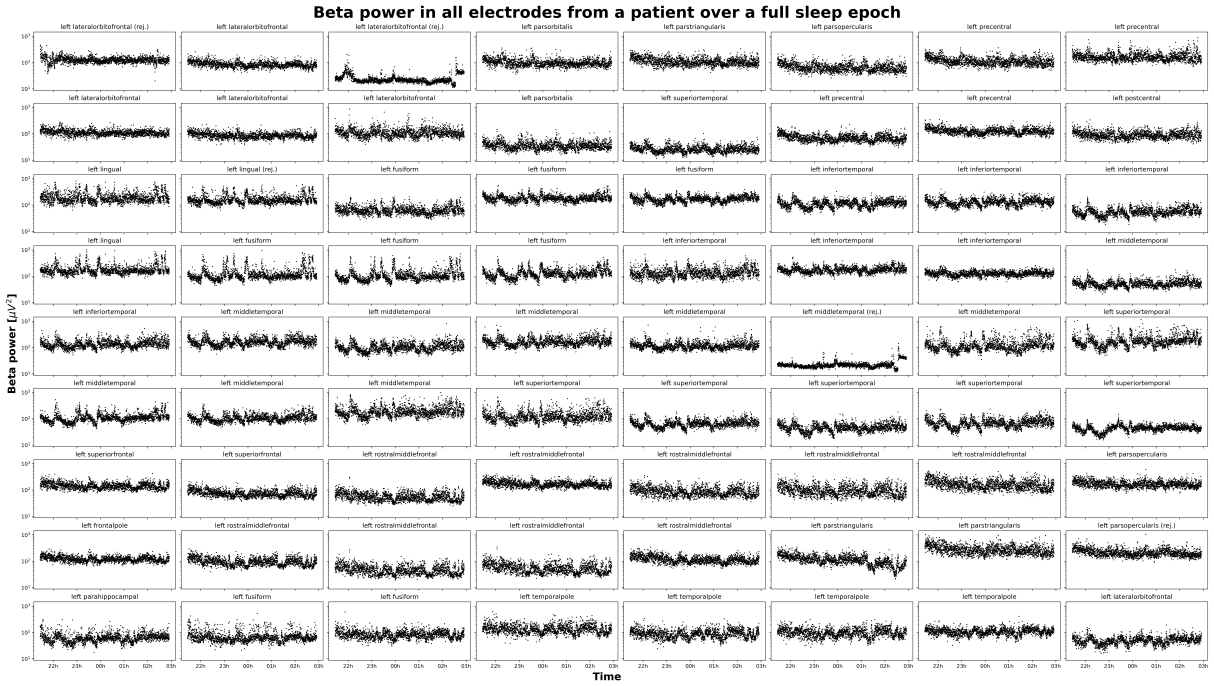

FIGURE A.11: Beta power in all electrodes over a sleep epoch. The anatomical locations of the electrodes in Desikan-Killiany parcellation are indicated, and those rejected upon visual inspection or because they contained artifacts (as defined in section 2.1.3) in more than 50% of epochs are marked *rej.*.

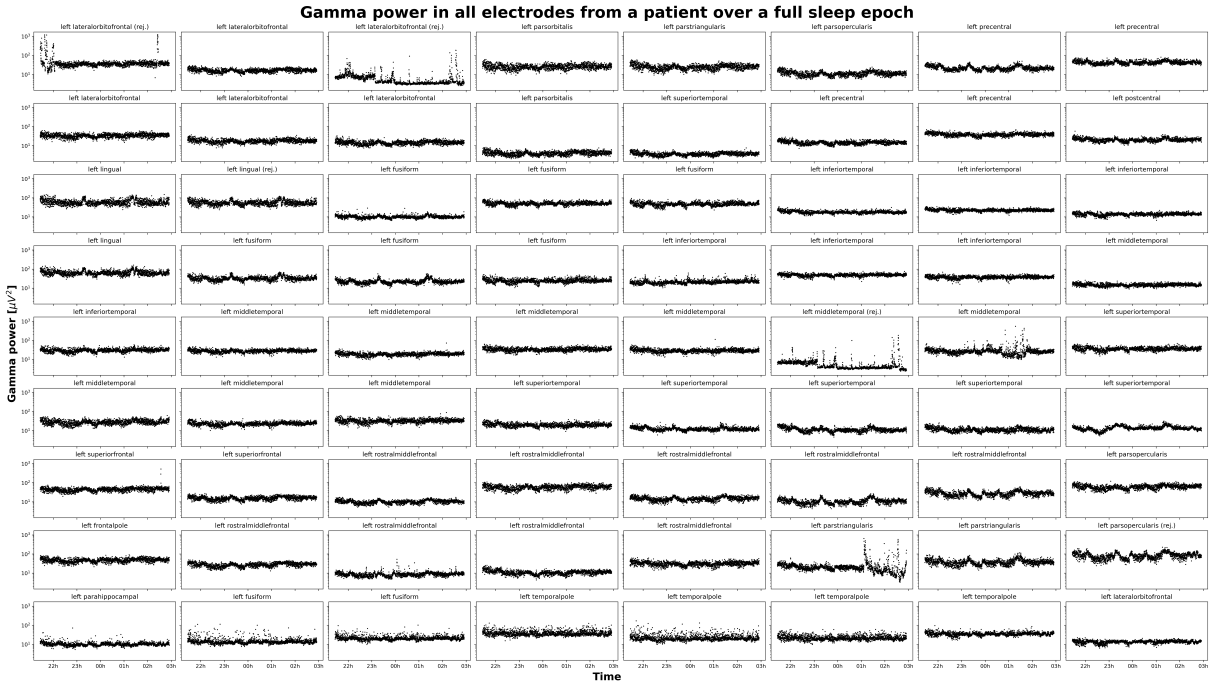

FIGURE A.12: Gamma power in all electrodes over a sleep epoch. The anatomical locations of the electrodes in Desikan-Killiany parcellation are indicated, and those rejected upon visual inspection or because they contained artifacts (as defined in section 2.1.3) in more than 50% of epochs are marked *rej.*.

B UNSUPERVISED LABELING

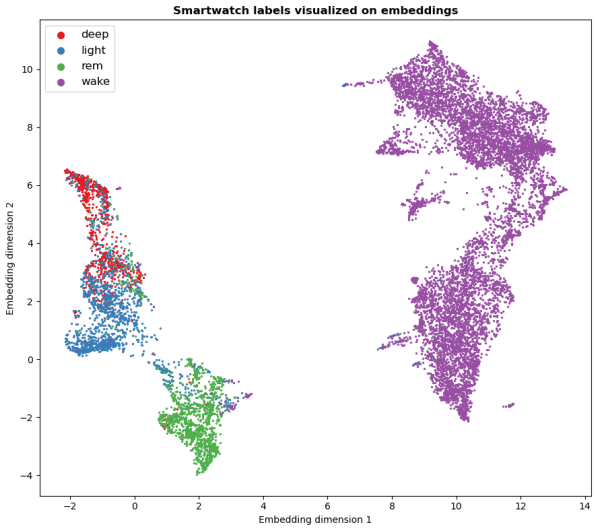

FIGURE B.13: Structure of the sleep stages in the embedding space used for clustering. The features projected on a 2-dimensional space using UMAP used for clustering are colored by the label provided by the automatic sleep-staging device. Sleep and wake are disjoint, and the same is true for REM and NREM. Sleep depth appears to be encoded by the second embedding dimension.

Annotation	Unsupervised label	
	Sleep	Wake
Sleep	1,119	286
Wake	200	1,494

TABLE B.1: Contingency table of the labels obtained with the unsupervised procedure with procedure described in section 2.2.1 versus manual annotations from video recordings.

Source of variation	SS	df	F	p-value
C(lobe)	3.621	3	17.486	2.985e-11
C(hemisphere)	4.16	1	60.262	1.137e-14
C(lobe):C(hemisphere)	4.045	3	19.532	1.548e-12
Residual	203.646	2950		
Total	215.473	2957		

TABLE D.2: Two-way analysis of variance of the ratio of gamma power during sleep versus wake. Electrodes that were not in the temporal, frontal, parietal or occipital lobe were excluded.

C CLUSTERS

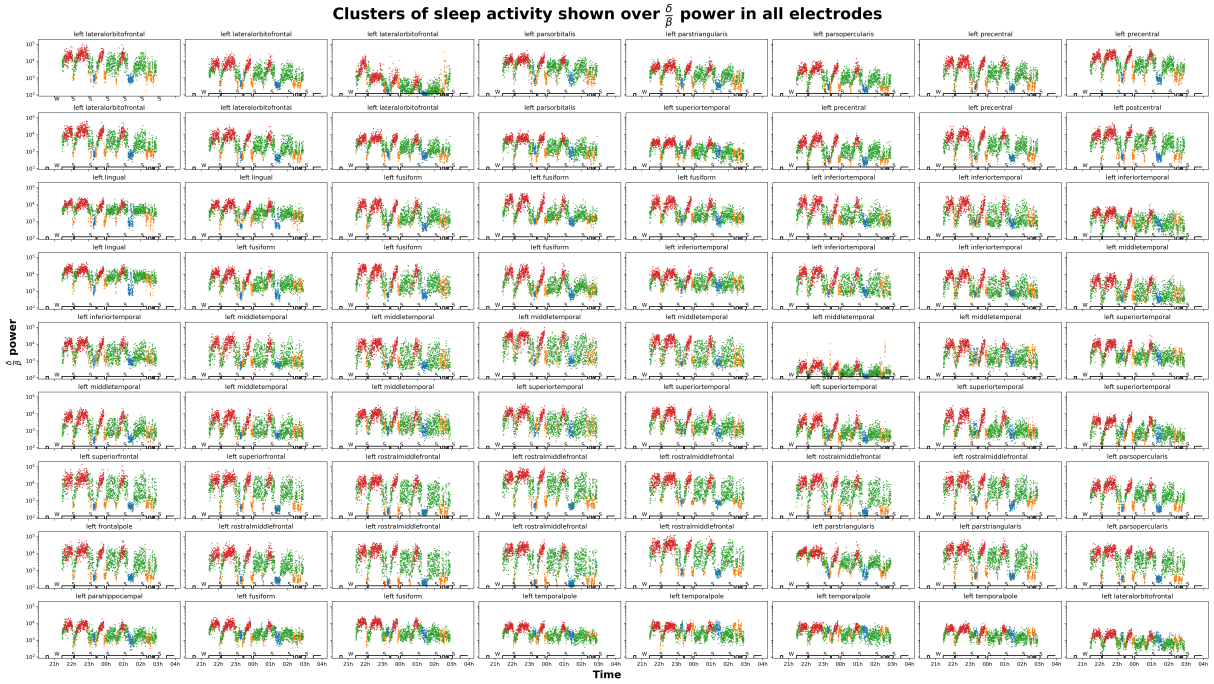

FIGURE C.14: Cluster of sleep activity visualized over delta/beta power in all electrodes over a full sleep epoch. The anatomical locations in the Desikan-Killiany parcellation as well as the lobe are indicated. The unsupervised labeling of sleep is shown for reference with the black line at the bottom indicating sleep, while manual annotations are shown with the symbols *S* and *W*.

D COMPARISON OF BANDPOWER DURING SLEEP AND WAKE

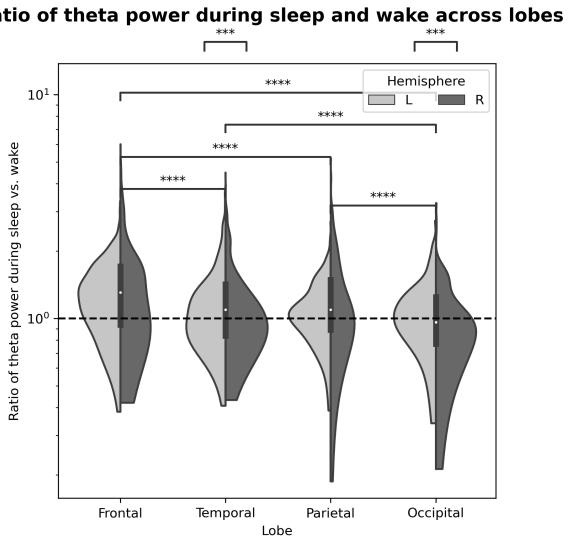

FIGURE D.15: Ratio of the mean theta power measured during sleep versus during wake across the four lobes. The significance level of a MWU test applied to all pairs of lobes and between the left and right hemisphere for every lobe is indicated, with Bonferroni correction.

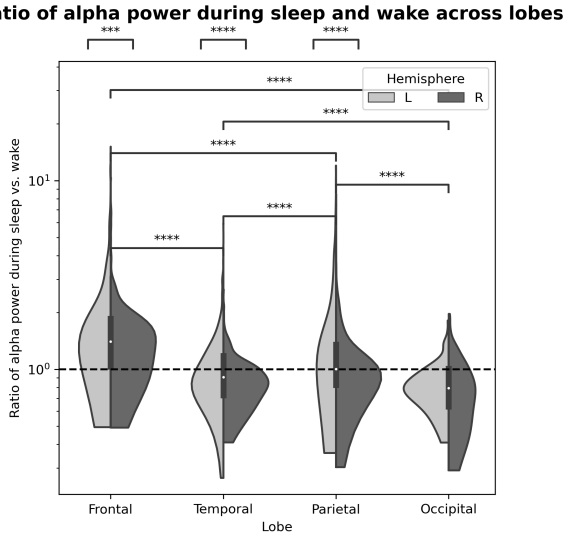

FIGURE D.16: Ratio of the mean alpha power measured during sleep versus during wake across the four lobes. The significance level of a MWU test applied to all pairs of lobes and between the left and right hemisphere for every lobe is indicated, with Bonferroni correction.

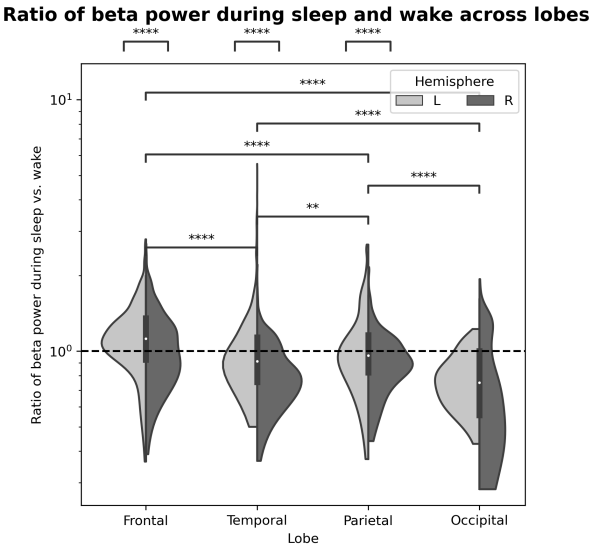

FIGURE D.17: Ratio of the mean beta power measured during sleep versus during wake across the four lobes. The significance level of a MWU test applied to all pairs of lobes and between the left and right hemisphere for every lobe is indicated, with Bonferroni correction.

E OSCILLATORY BEHAVIOR

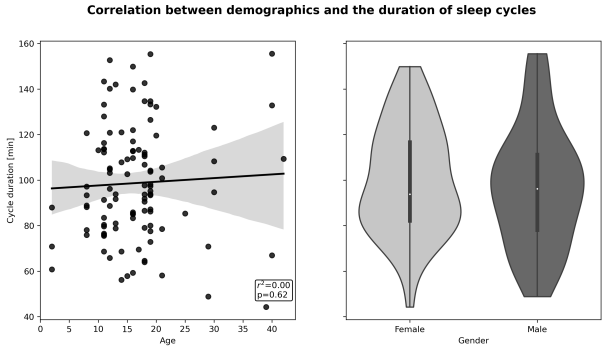

FIGURE E.18: Correlation between demographics and the duration of sleep cycles. Linear regression between patient age and sleep cycle duration as previously defined (left), and distribution of the proportion in patients of male and female gender (right). The solid line indicates the best linear fit to the data with the shaded area indicating a bootstrapped 95% confidence interval.

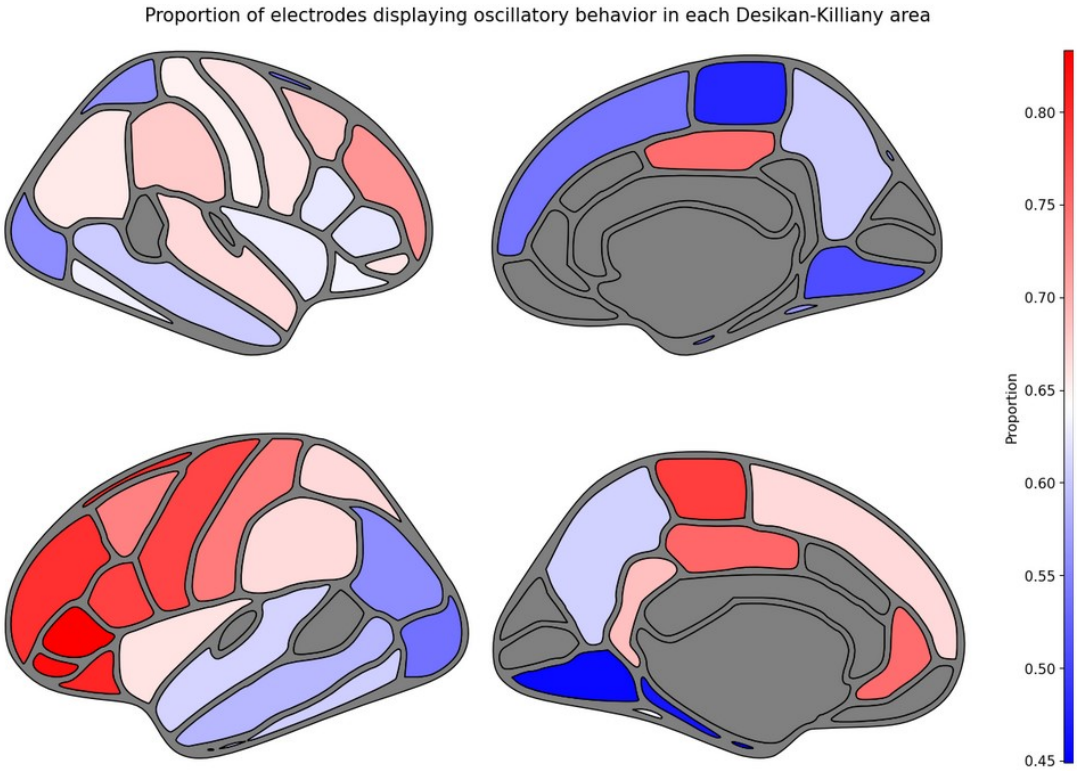

FIGURE E.19: Proportion of electrodes displaying oscillatory delta/beta power in all Desikan-Killiany areas visualized on a template brain. Areas that have data from fewer than 25 sleep epochs in a given lobe are shaded in grey.

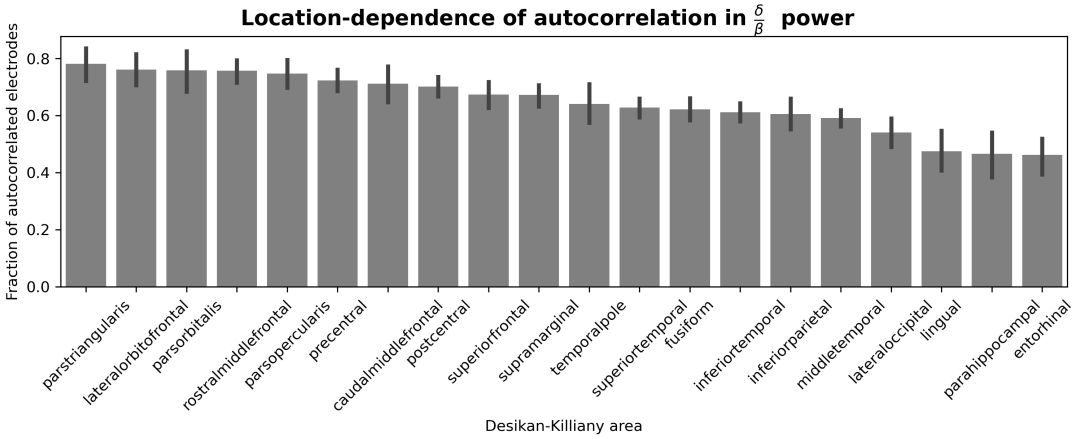

FIGURE E.20: Proportion of electrodes displaying oscillatory delta/beta power in all Desikan-Killiany. The bars show the overall proportion across all sleep epochs and patients, while the solid line represents a bootstrapped 95% CI interval for it. Only the areas that have data from at least 25 sleep epochs (in either lobe) are shown. The 7 areas with the highest average proportion of electrodes displaying oscillations in delta/beta power are all located in the frontal lobe.

F CLUSTERS OF ACTIVITY DURING SLEEP

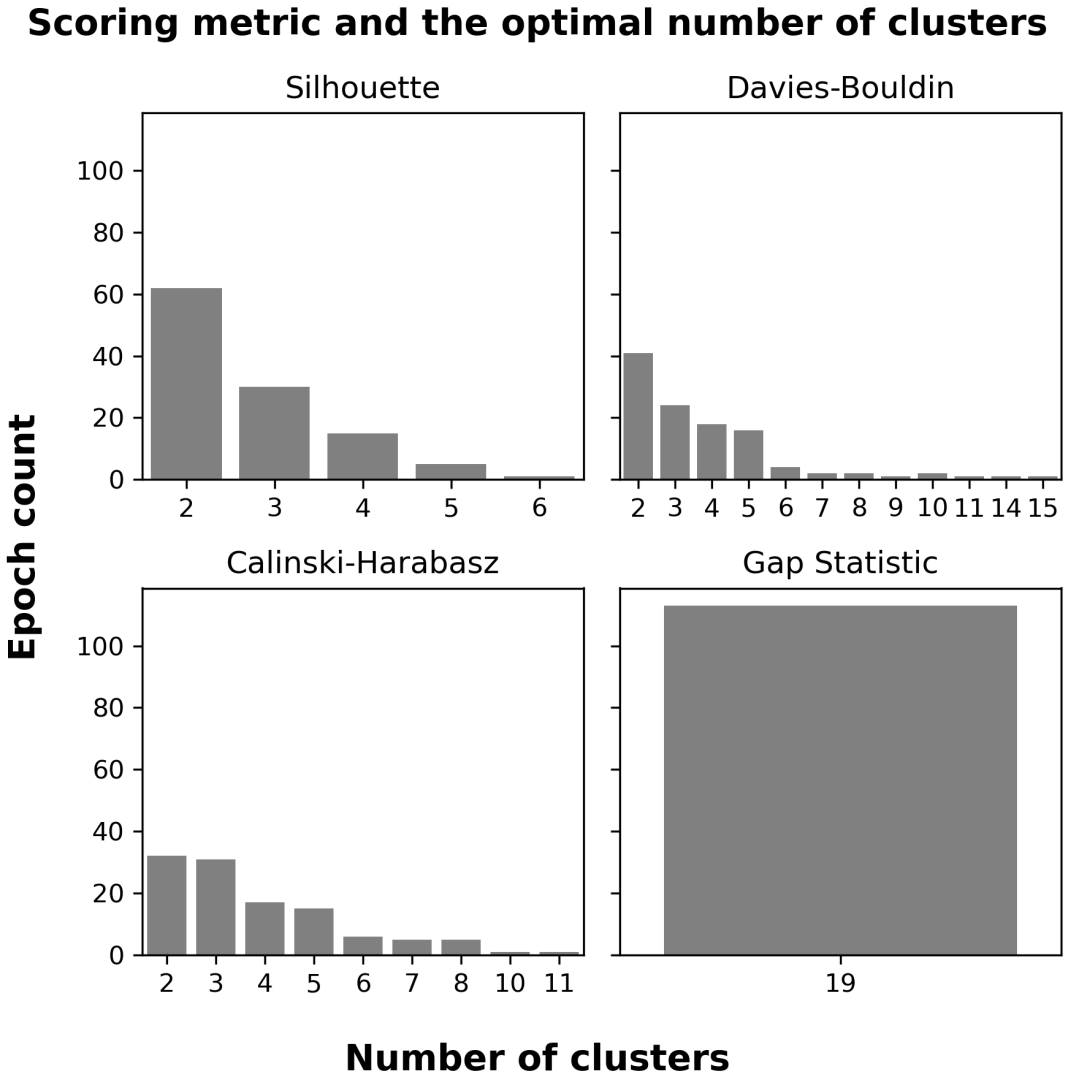

FIGURE F.21: Distribution of the optimal number of clusters in all the sleep epochs using different scoring metrics. The gap statistic method always returned the highest number of clusters tested (19) while the other methods were biased towards suggesting the lowest number of clusters evaluated (2).

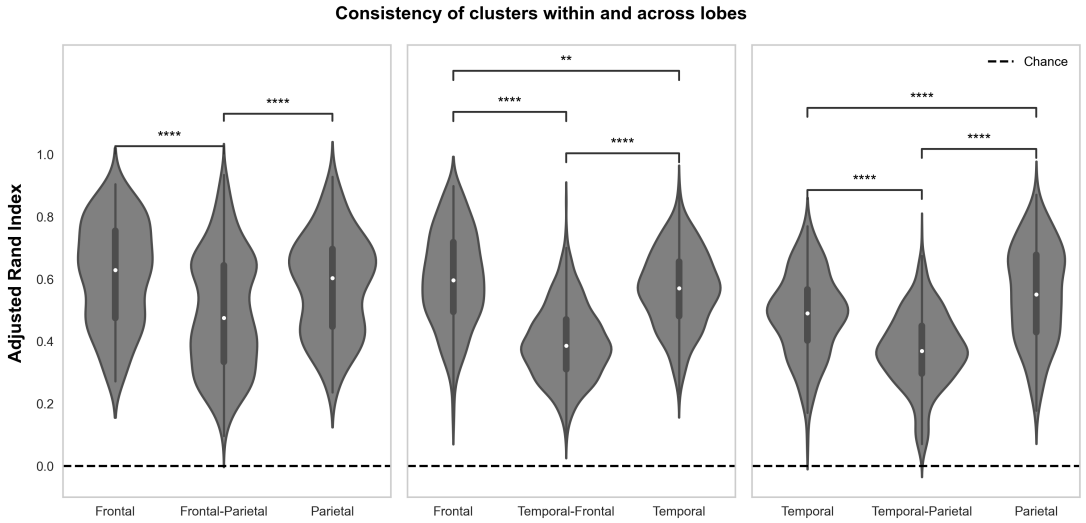

FIGURE F.22: Distribution of the ARI of sets of cluster obtained within and between lobes using 10 electrodes to produce each clustering. The parietal-occipital and temporal-occipital pairs are absent due to an insufficient number of patients with the required coverage (20 electrodes in both regions). The distribution was generated using data from 5, 12, 5 patients and 18, 24, 20 epochs for the frontal-parietal, temporal-frontal and temporal-parietal pairs respectively.

G CLUSTER COMPOSITION

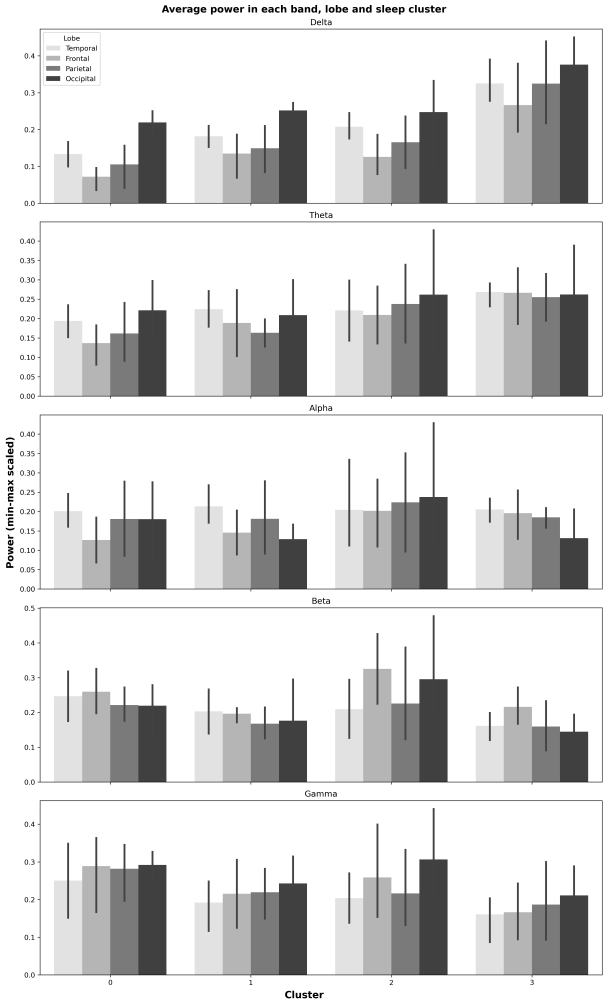

FIGURE G.23: Average power in every frequency band compared across each lobe and cluster. The power in each band was min-max scaled over the sleep epoch for every electrode to put the different bands on a comparable scale. The vertical bars indicate a bootstrapped 95% CI for the mean.

H CLUSTER DYNAMICS

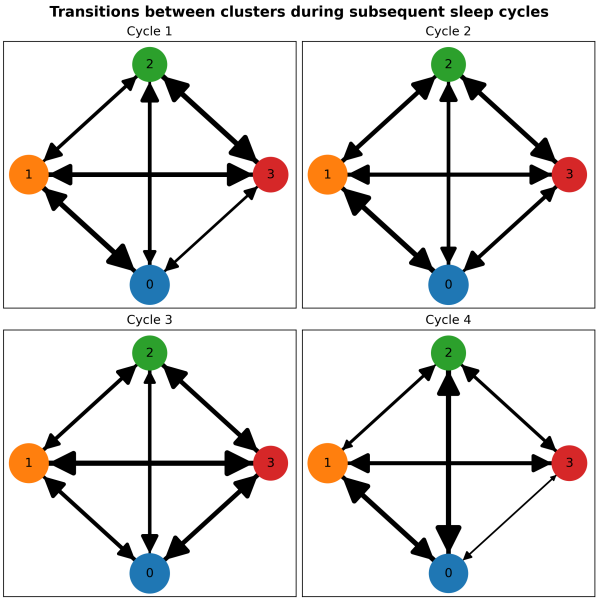

FIGURE H.24: Transitions between the clusters, separated by cycle. The edges are sized proportionally to the number of transitions between the clusters, represented by the nodes. The aforementioned are sized proportionally to their duration (presented in Figure 3.20).