

Accepted Manuscript

Title: Diagnostic value of sleep stage dissociation as visualized on a 2-Dimensional sleep state space in human narcolepsy

Author: Anders Vinther Olsen Jens Stephansen Eileen Leary
Paul E Peppard Hong Sheungshul Poul Jennum Helge Sorensen
Emmanuel Mignot



PII: S0165-0270(17)30046-8
DOI: <http://dx.doi.org/doi:10.1016/j.jneumeth.2017.02.004>
Reference: NSM 7681

To appear in: *Journal of Neuroscience Methods*

Received date: 5-8-2016
Revised date: 11-2-2017
Accepted date: 13-2-2017

Please cite this article as: Anders Vinther Olsen, Jens Stephansen, Eileen Leary, Paul E Peppard, Hong Sheungshul, Poul Jennum, Helge Sorensen, Emmanuel Mignot, Diagnostic value of sleep stage dissociation as visualized on a 2-Dimensional sleep state space in human narcolepsy, *Journal of Neuroscience Methods* (2017), <http://dx.doi.org/10.1016/j.jneumeth.2017.02.004>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Diagnostic value of sleep stage dissociation as visualized on a 2-Dimensional sleep state space in human narcolepsy

Anders Vinther Olsen^{a,b}, Jens Stephansen^{a,b}, Eileen Leary^a, Paul E Peppard^d, Hong Sheungshul^e, Poul Jennum^c, Helge Sorensen^b, Emmanuel Mignot^a

^aCenter for Sleep Sciences and Medicine, Stanford School of Medicine, Palo Alto, California, USA
^bDepartment of Electrical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark
^cDanish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet, Denmark
^dDepartment of Preventive medicine, U Madison Wisconsin Madison, Wisconsin, USA
^eSleep Disorder Center, Catholic University, Seoul, Korea

Abstract

Background: Type 1 narcolepsy (NT1) is characterized by symptoms believed to represent Rapid Eye Movement (REM) sleep stage dissociations, occurrences where features of wake and REM sleep are intermingled, resulting in a mixed state. We hypothesized that sleep stage dissociations can be objectively detected through the analysis of nocturnal Polysomnography (PSG) data, and that those affecting REM sleep can be used as a diagnostic feature for narcolepsy.

New Method: A Linear Discriminant Analysis (LDA) model using 38 features extracted from EOG, EMG and EEG was used in control subjects to select features differentiating wake, stage N1, N2, N3 and REM sleep. Sleep stage differentiation was next represented in a 2D projection. Features characteristic of sleep stage differences were estimated from the residual sleep stage probability in the 2D space. Using this model we evaluated PSG data from NT1 and non-narcoleptic subjects. An LDA classifier was used to determine the best separation plane.

Comparison with existing methods: This method replicates the specificity/sensitivity from the training set to the validation set better than many other methods.

Results: Eight prominent features could differentiate narcolepsy and controls in the validation dataset. Using a composite measure and a specificity cut off 95% in the training dataset, sensitivity was 43%. Specificity/sensitivity was 94%/38% in the validation set. Using hypersomnia subjects, Specificity/sensitivity was 84%/15%. Analyzing treated narcoleptics the Specificity/sensitivity was 94%/10%.

Conclusion: Sleep stage dissociation can be used for the diagnosis of narcolepsy. However the use of some medications and presence of undiagnosed hypersomnolence patients impacts the result.

Keywords: Narcolepsy, Diagnostic, Sleep Stage Dissociation, Machine learning, LDA classifier

2016 MSC: 60G35 92C55 94A12

1. Introduction

Type 1 narcolepsy, a disorder caused by a loss of hypocretin neurons, is diagnosed based on the presence of daytime sleepiness, cataplexy, and a positive Multiple Sleep Latency Test (MSLT). The MSLT is a 4 to 5 daytime nap test opportunity where sleep latency and the occurrence of REM sleep within 15 minutes of onset is noted in each test, a feature called a Sleep Onset REM Period (SOREMP). The MSLT is performed following a night of nocturnal polysomnography (PSG), a procedure mostly performed to exclude confounding sleep disorders and note for the presence of a nocturnal SOREMP, a highly predictive diagnostic feature.

The cause of hypocretin cell loss in narcolepsy type 1 is likely of autoimmune origin. The disorder is strongly associated with HLA-DQB1*06:02, and has been shown to be triggered by upper airway infections such as streptococcus and influenza-A, notably 2009H1N1 [1] [2]. Documentation of low cerebrospinal fluid hypocretin 1 (csf-hcrt-1) below 110 pg/ml, is regarded as a definitive diagnostic test for diagnosing narcolepsy type 1 [3] [4].

The MSLT gained popularity due to the fact that it is quite specific ($\approx 97\%$) and sensitive ($\approx 93\%$) in diagnosing Type 1 narcolepsy [5] [1]. Until recently, a mean sleep latency (MSL) 8 min, and 2 SOREMPs during the daytime naps were considered as diagnostic for narcolepsy. Based on recent findings that a nSOREMP is highly specific ($\approx 99\%$) for type 1 narcolepsy, but has only moderate sensitivity ($\approx 45\%$), the most recent classification of sleep disorders (ICSD3) revised the definition of a positive MSLT to include the nSOREMP as part of the 2 SOREMP requirement. At the practical level, this change is insignificant, as in our study of over 800 type 1 narcoleptic patients all subjects with a nSOREMP also had 2 SOREMPs during naps [1]. For example, comparing 516 type 1 narcolepsy cases with a similar number of age matched controls, specificity for the ICSD3 defined PSG-MSLT was 98.6% and sensitivity 93.3% versus 98.6% and 92.9% for the older definition [1].

Because a nSOREMP is highly specific ($\approx 99\%$) for narcolepsy but has only modest sensitivity ($\approx 40 - 50\%$), it has been proposed that other diagnostic biomarkers should be sought within the PSG to add sensitivity, so that a nocturnal PSG alone could be used to diagnose narcolepsy without the need for a subsequent MSLT. Following on this hypothesis, sleep onset REM periods during nocturnal sleep, long periods of wakefulness within sleep, or the presence of specific EEG spectra abnormalities have been found to differentiate narcolepsy versus controls [6] [7]. Unfortunately, however, none of these new biomarkers were sufficiently predictive alone, although combining them may be helpful and has not yet been systematically tested. The advantage of using a PSG alone for diagnosis is that widely used to diagnose sleep disordered breathing, and other sleep disorders, and could thus be used as a screening tool for type 1 narcolepsy.

Using various components of the EEG, EOG and EMG that are known to differentiate sleep stages, it is possible to artificially project a multi component difference on a 2-D plane that display a clear clustering of the various sleep stage, a procedure called sleep state space analysis. In essence, a sleep state space analysis displays for any subject how distinct various sleep stages are from each other. Sleep state space abnormalities has been proposed as a new possible narcolepsy biomarker based on EEG studies in narcoleptic mice [8]. In this study, Diniz Behn et. al. [8] used a sleep stage space projection analysis to cluster sleep and wake and shows that narcoleptic mice have less distinct and more labile states of sleep and wakefulness. Similar work [9] has shown that it is possible to separate human sleep stages and look at their dynamics in a 2-D space, although these projections did not include stage 1 and did not attempt to study narcoleptics. Stage 1 may be particularly important to analyze as increased stage 1 is the most consistent difference in sleep architecture found in patients with narcolepsy [10]. Further, stage 1, REM sleep and wake are the most similar states at the electrophysiological level. In this study, we expanded on this idea, creating sleep state space projections in controls and narcolepsy, demonstrating that sleep stage dissociation in patients with narcolepsy (i.e. reduced distance between sleep stages) can be used to discriminate Type 1 patients from controls and other patients.

2. Population samples used in the study

This section applies the original number of patients in the datasets used in this study. It is important to note that these dataset have been analyzed by technicians and if the validity of the data were questioned by the technician, the subject has been removed from the dataset. This section reflects the original dataset numbers before removal.

2.1. Stanford sleep cohort (SSC)

This sample is a naturalistic sample of 862 successive patients that were seen at the Stanford Sleep Clinic between October 1999 and March 2007 at the Stanford Sleep Clinic [11] [12]. Data from the sample only includes 25 patients with type 1 narcolepsy (none taking sodium oxybate) and other patients with Delayed Sleep Phase (DSP) $n=14$, Insomnias $n=141$, REM Behavior disorders (RBD) $n=4$, Restless Leg Syndrome (RLS), $n=23$, Sleep Disordered Breathing (SDB) $n=607$ plus 39 subjects suffering from others disorders. Summary statistics for these subjects are provided in Table 1.

PSGs were all recorded using the Sandman Elite digital sleep software and Sandman SD32+ amplifiers. The recording was done using the Stanford Sleep Disorder clinic protocol, which is an extension of the AASMs clinical guidelines in the fact that it records additional respiratory signals with increased precision and processing. It best corresponds to updated AASM 2012 scoring rule criteria. In total 18 channels of information are recorded: EEG (using 10/20 electrode placement), EOG, EMG (measured on the submental muscle but also on the anterior tibialis muscles of each leg combined into a single EMG channel for both legs), ECG, vibration snoring, breathing effort, airflow, nasal pressure

and oxygen saturation (SpO₂). EEG is recorded using the international 10-20 system with a sampling frequency of 256 Hz. EMG, ECG and snoring signals are recorded using a sampling rate of 512 Hz, breathing effort and airflow recorded with a sampling rate of 64 Hz and SpO₂ recorded at 4 Hz. Filtering is done on electrophysiological signal channels by bandpass filtering in the range of 0.1 Hz and 0.45 times the sample rate. EMG channels are further high pass filtered at 10 Hz. Sleep stage scoring is performed by trained Registered Polysomnographic technicians under the supervision of board certified sleep disorder specialists at the Stanford Sleep Clinic.

2.2. Wisconsin sleep cohort (WSC)

This sample includes 1522 PSGs from the Wisconsin Sleep Cohort, a longitudinal study of sleep habits and disorders in the general population [13]. The cohort was established in 1988 from a sample of employees of 4 state agencies in south central Wisconsin, aged 30 to 60 years. Beginning in 2000, approximately 800 participants enrolled in the Wisconsin Sleep Cohort were recruited for an MSLT following a PSG study. Summary statistics for the sample are provided in Table 1. PSGs were all recorded using a Grass model 78 polysomnography machine and scoring performed using standard criteria at the time based on AASM guidelines. In total 18 channels were recorded including EEG (using 10/20 electrode placement), EOG and EMG. Breathing was measured using respiratory inductance plethysmography, nasal and oral airflow using thermocouples and SpO₂ using pulse oximetry. All signals had been resampled to 100 Hz from its original sampling rate before it was made accessible in this study without influence from the authors.

2.3. Jazz narcolepsy study sample (Jazz)

PSGs for this sample are all from patients with cataplexy and were obtained from baseline SXB-15 and SXB-22 studies for a sodium oxybate clinical trial conducted by Orphan Medical, now Jazz Pharmaceuticals [14] [15]. These were scored according to AASM 2004. The sample includes 270 adult patients with Type 1 narcolepsy. Recordings were all performed prior to sodium oxybate administration but many patients were on centrally acting stimulants or/and a few on antidepressants. Recordings were performed in 44 sites in the United States, Canada, Czech Republic, France, Germany, Netherlands, Switzerland and United Kingdom. Recordings featured 16 channels of PSG information with EEG (using 10/20 electrode placement), EOG, EMG, breathing, nasal and oral thermistors, nasal pressure and SpO₂. All signals were resampled to 100 or 128 Hz depending on patient before it was made accessible for this study. Scoring was done locally by experts and technicians using AASM recommendations.

2.4. Korean hypersomnia and narcolepsy sample (Korea)

This sample consists of 308 PSG recordings from Korea but information was missing for some patients therefore these were excluded so only 151 narcolepsy subjects remain.

Recordings were done on Embla system PSG recorders using S7000 and N7000 amplifiers and the Rem Logic software (v. 3.4 and 3.2). The recordings were set up using the AASM guidelines. In total 40 channels of data was recorded including 4 EEG channels (placement compared to the central 10/20 electrodes), 2 EOG, channels, chin EMG, 2 leg EMG, nasal and oral thermistors and SpO₂.

EEG, EMG and EOG were all sampled at 1000 Hz. All pressure signals were sampled at 50 Hz and saturation signals were sampled at either 5 or 10 Hz.

EEG and EOG filtering was done using a high pass filter with cutoff frequency at 0.5 Hz and a low pass filter with a cutoff frequency at 70 Hz. For chin EMG and leg EMG filtering was done using a high pass filter with 10 Hz as cutoff frequency and low pass filter with 70 Hz as cutoff frequency.

2.5. Construction of training and validation sets

The datasets used are described in Table 1. Five non-overlapping datasets were created, the first two to create and validate our 2-D sleep state space model, thereby identifying features differentiating sleep stages, and the next three to test the feature model in Type 1 narcolepsy versus other subjects. The first sleep state space training dataset included 200 random samples without replacement from the SSC database and 200 random samples without replacement drawn from the Wisconsin database. This was selected to ensure representation of a diverse set of PSGs, both drawn from the general population and from sleep clinics. The sleep stage space validation dataset consisted of all remaining SSC samples, including a few narcoleptic patients. The first narcolepsy-non narcolepsy comparison dataset included

200 subjects drawn from the WSC in comparison to 17 narcoleptic patients extracted from the Jazz Narcolepsy study sample (all taken from the SXB-15 database which were recorded at only two hospitals therefore minimizing the effect of inter-scorer variability). None of these subjects were receiving anti-depressants or centrally acting stimulants. It was used as a training dataset to construct a separation plane that would distinguish narcolepsy versus controls in the 2-D sleep stage plane. The second dataset, used to validate separation variables, was constructed using patients from the Korean database. In this sample, non Type 1 narcoleptic patients consist of subjects suffering from idiopathic hypersomnia or type 2 narcolepsy patients (HLA negative or/and hypocretin level < 200 mg/L) while Type 1 narcolepsy subjects were patients with low hypocretin level (≤ 110 mg/L) or HLA positivity plus cataplexy. A final dataset consisted of subjects treated with anti-depressants or centrally acting stimulants, whether population controls (WSC) or Type 1 narcoleptic (Jazz Narcolepsy study sample SXB-22 featuring the recordings from 42 sites). This dataset was also used to test the effects of medications on the classifier.

3. Homogenization of datasets

Because datasets were collected using differing filtering set ups and at different sites, a common preprocessing EEG algorithm was used to ensure that signal content was similar all across. First all signals was resampled to the lowest frequency of the electrophysiological channels which was 100 Hz then homogenized using an algorithm that included: (1) an initial 5th order high pass Butterworth filter with a cutoff frequency at 0.2 Hz; this was used to remove baseline drifting. (2) a 5th order low pass Butterworth filter with a cutoff frequency at 49 Hz; this was done to remove artifacts such as 50/60 Hz power line interferences and to homogenize data as some of the samples were not collected past 75 Hz. (3) a zero-phase digital filtering method filtering both in the forward and the reverse direction; this ensured that phase content was not distorted and no group delay after filtering. (4) Finally, the EEG signal was truncated so that values above $300 \mu V$ was set to a maximum value of $300 \mu V$ and values below $-300 \mu V$ was set to a minimum value of $-300 \mu V$; this was done to minimize outlier effects. It is known that this introduces some artifacts (especially in the form of DC frequency content) but values outside of this span is not biological signals so this measurement would be an artifact either way. Furthermore the DC frequency is not utilized in the feature extraction. For EMG no preprocessing was done as the features extracted from the EMG is purely time series features therefore filtering would disrupt the stability of the features across subjects. For EOG the features extracted are correlation outputs after filtering therefore the preprocessing is included in the feature extraction algorithm.

4. Pre processing of hypnograms

AASM guidelines state that each 30 sec epoch is labeled with its most prominent sleep stage. Any scored epoch can therefore be a mix of different stages, lowering the value of our sleep stage epoch gold standard. To minimize this effect, 30 sec epochs were split into 15 sec segments as shown in Appendix Fig. 9. Because the 15 sec epoch preceding or following a sleep transition was most likely to be composed of 2 sleep stages, it was then removed. This ensured that our gold standard sleep stages were more homogeneous. This was used for all analysis.

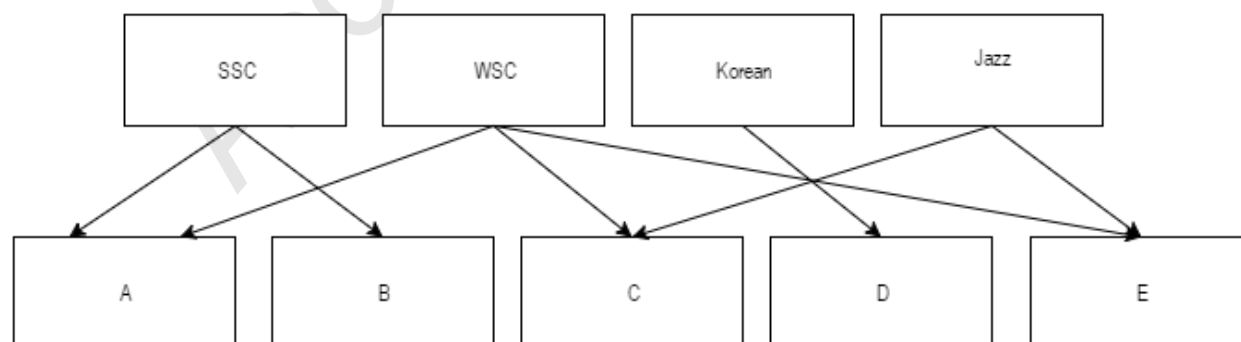


Figure 1: Flowchart on the development of datasets.

In this way the sleep stages that were investigated are pure and do not contain content from other sleep stages.

Table 1: Datasets and their use in the study. Data have been reported in means with their standard error mean unless otherwise noted.

Dataset:	Sleep Stage Space Model (Dataset A)	Training sample (Dataset B)	Validation in clinical sample (Dataset C)	Validation in hypersomnia sample (Dataset D)	Validation in treated sample (Dataset E)
Controls	SSC: 200 WSC: 200	SSC: 610	WSC: 200	Korean: 68	WSC: 200
Narcoleptics	None	SSC: 19	Jazz: 17	Korean: 47	Jazz: 113
Age	SSC:47.4± 1.3 WSC:58.8 ± 1.1	SSC: 45.4 ± 0.6	WSC:55.4 ± 1.5 Jazz:54 ± 3.5	28.3 ± 1.5	WSC:58.8 ± 1.1 Jazz: 56.8 ± 0.9
AHI(mean)	SSC: 2.75±2.6 WSC: 19.1±0.8	SSC: 0.2 ± 0.1	WSC: 1.7 ± 0.6	2.4 ± 0.6	WSC: 1.6±0.5
PLM	SSC: 8.6	SSC:7.33±0.66	N/A	N/A	N/A
TST(mean)	SSC: 325.8± 2.9 WSC: 357.5± 3.8	SSC:473.3 ± 2.0	WSC: 454.3±4.0 Jazz: 474±6.9	476.9±3.3	WSC:462.0±2.9 Jazz: 483±4.1
Rem Lat.	SSC: 120.6 ± 5.3 WSC: 125.2 ±6.2	SSC: 133.19±3.6	WSC:127.8±6.7 Jazz: 64±17.1	86.0±7.4	WSC:105.8±5.3 Jazz: 61±9.0
Sleep Eff.	SSC: 76,7 ± 1.3	SSC:77.4 ± 0.5	WSC: N/A	92.6 ± 1.0	N/A
W(%)	SSC: 23 ± 3.4 WSC:20.8 ± 5.1	SSC: 21.9± 2.4	WSC:20.3±4.6 Jazz: 17.1±9.4	9.1±4.4	WSC:24.2±4.8 Jazz:12.7±15.4
N1(%)	SSC:9.2 ± 1.6 WSC:9.0±2.7	SSC: 8.9±1.2	WSC:9.3±2.3 Jazz: 10.6±6.8	8.9±3.3	WSC:10.1±1.6 Jazz: 12.7±3.5
N2(%)	SSC:48.5 ±3.6 WSC:53.2±4.5	SSC: 48.8±2.3	WSC:51.8±5 Jazz: 43.29±12.5	51.7±4.9	WSC:50.7±3.7 Jazz: 45.6±7.1
N3(%)	SSC:4.3 ± 1.9 WSC:4.4±1.7	SSC: 4.5±0.8	WSC:5.9±2.5 Jazz: 8.3±8.7	12.1±3.6	WSC:5.0±2.0 Jazz: 5.0±2.5
REM(%)	SSC:12.8 ± 1.8 WSC:12.4±2.0	SSC: 13.1±1.1	WSC:12.4±2.1 Jazz: 14.7±10,1	18.7±3.2	WSC:13.4±1.9 Jazz: 15.7±3.9
Use:	Used to create a sleep state space and to model sleep stages differences in control patients	Used to establish classifier boundaries.	Used to verify the classifier in a replication sample.	Used to verify if the biomarker can be used to distinguish idiopathic hypersomnia and narcolepsy	Used to analyze whether the biomarker is still useable in subjects receiving treatment .

5. Methods

This section introduces the methods used during the study. All computations were done using MATLAB[®] for feature extraction (version 2010 was utilized) and for classification (version 2015a was utilized). All data files had been provided in the EDF format.

5.1. Overall strategy

To define boundaries of sleep stages, a state space model that could separate stages into reasonable clusters was first developed using 400 controls. The sleep stages were then projected on to this space and a Gaussian Mixture Model (GMM) was fitted to give it a normal distribution. These distributions were then used as input to an LDA classifier so that narcoleptic versus control subjects are classified (Fig. 2).

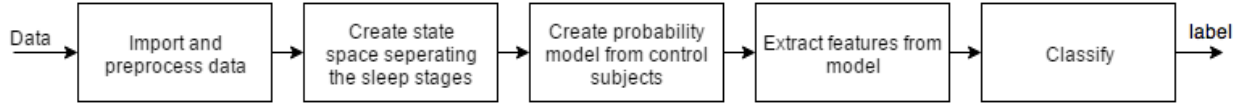


Figure 2: Flowchart explaining the strategy used, in simple steps. As seen, we first extracted features from the PSGs to create and validate a sleep-stage space model. Differences in sleep stage distributions were then used as features to classify narcoleptics versus control subjects.

5.2. Determining features predictive of sleep stages and creation of a sleep stage 2-D space

The approach used was similar to that used by Imbach et. al. [9] except that we also included sleep stage N1, which is essential to the study of narcolepsy considering reported abnormalities in this stage [10]. The LDA projection used 38 features extracted from the each of the 15 sec epoch segments of EEG, EOG and EMG. An LDA projection was used because it creates a space which has the highest variance between classes and the lowest variance inside classes.

5.2.1. EEG features

Twenty-eight features were extracted from the EEG (using signals from the central derivatives C3-A1 or C4-A2). Twenty features were simple power spectrum estimates (PSE) in equally spaced bands ranging from 0.2 Hz to 25 Hz. The bands are calculated using the Welch method with three second windows and a 50% overlap as this allows for quasi stable EEG signal. Two additional EEG features were derived using power band ratios as described in [9]. These power band ratios were computed as follows:

$$PB1 = \frac{P(8.6[Hz] - 19.3[Hz])}{P(1[Hz] - 10.9[Hz])} \quad (1)$$

$$PB2 = \frac{P(11.5[Hz] - 20.3[Hz])}{P(17.9[Hz] - 32.5[Hz])} \quad (2)$$

An additional two EEG features were derived based on the cumulative sum of the 90th and 95th percentile of the power spectrum. The three final EEG features were Hjorts parameters for activity, mobility and complexity:

$$Activity = var(y(n)) \quad (3)$$

$$Mobility = \sqrt{\frac{var\left(y(n) \frac{dy}{dn} \right)}{Activity}} \quad (4)$$

$$Complexity = \frac{Mobility\left(y(n) \frac{dy}{dn} \right)}{Mobility(y(n))} \quad (5)$$

5.2.2. EMG features

The three features selected were Hjorts parameters for activity, mobility and complexity.

5.2.3. EOG features

Eight features derived from the EOG modality were inspired by the work done in Kempfner et. al. [16]. By taking the cross correlation of filtered EOG signals, it was possible to discern between different types of eye movements such as Rapid Eye Movements (REMs) and Slow Eye Movements (SEMs). Table 2 shows the different filter cut off frequencies used in this modality.

After filtering the different features values were calculated as:

Table 2: Table showing the bandpass filtering cut off frequencies for the different filters used for the EOG feature extraction

Filter 1	Filter 2	Filter 3	Filter 4
0.25 - 5 [Hz]	0.5 - 5 [Hz]	0.75 - 5[Hz]	1 - 5[Hz]
Filter 5	Filter 6	Filter 7	Filter 8
1.25 - 5[Hz]	1.5 - 5[Hz]	1.75 - 5[Hz]	2 - 5[Hz]

$$r_k(n) = \frac{\sigma_{ab}(n)}{\sqrt{\sigma_{aa}\sigma_{bb}}} \quad (6)$$

By calculating EOG features as described above, genuine REMs occurring during REM and non REM sleep create a negative cross-correlation. These are then split into different variables (SEM vs REM) based on slopes. Background EOG signal is either not or positively correlated and provides zero values.

5.2.4. Feature normalization:

All 38 features were normalized for each patient so that the feature had zero mean and unity variance. To make sure features generalize forward feature selection was used.

5.3. Sleep stage dissociation modeling in narcoleptic versus controls

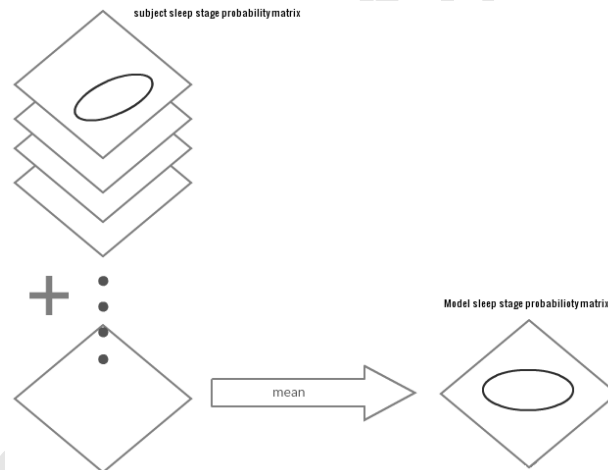


Figure 3: Demonstration on how the model sleep stages were created. First the sleep stages was projected onto the state space and then they were added together and then normalized to probabilities.

For each subject, epochs derived from each sleep stage were filtered with a 30-point hamming window (A 30 point filtering on 15 sec. epochs was used to extend on findings described previously [9] that found that 5 sec. epoch segments were best filtered using a 10 point Hamming window). This was done to account for within state variability during the recording. The filtered epochs of each subject were then projected onto the state space and a GMM with order 1 fitted to the data. This was done so that the distribution within the sleep stage was expressed in probabilities. Due to the accuracy of the state space noted in section 6.1, probabilities outside the 15th percentile were set to zero so that stage overlap was minimized in the model. This was done for each sleep stage and averaged across all epochs for the model dataset as seen in Fig. 3.

Probability features were next extracted from the remaining datasets similarly to the way the 2-D space model was created. Epochs are first projected onto the 2-D state space plane and GMM fitted to data points for each subjects sleep

stages. Features are then created by subtracting the sleep stage probability matrix of the model from the probability matrix of each subject as seen in Fig. 4.

$$P_{res} = P_{subject, sleepstage} - P_{sleepstageModel} \quad (7)$$

This is done using all model sleep stages giving in total 25 residual probability matrixes. To make it a single quantity the square sum of the residuals is computed using:

$$R_p(cs, ms) = \sum_m \sum_n p_{res(cs, ms)_{m,n}}^2 \quad (8)$$

where p_{res} is the residual probability matrix, cs is the current sleep stage and ms is the model sleep stage. When subtracting the sleep stage model probability matrix from the subjects sleep stage probability matrix a residual probability matrix is created. A contribution of 1 originates from the subjects sleep stage probability matrix and a contribution of -1 originates from the subtraction of the sleep stage model probability matrix. Therefore, due to the squaring operation, if the sleep stages were totally similar, the sum of square residuals would effectively give a zero and if the sleep stages were totally dissimilar, a maximum value of two will be obtained.

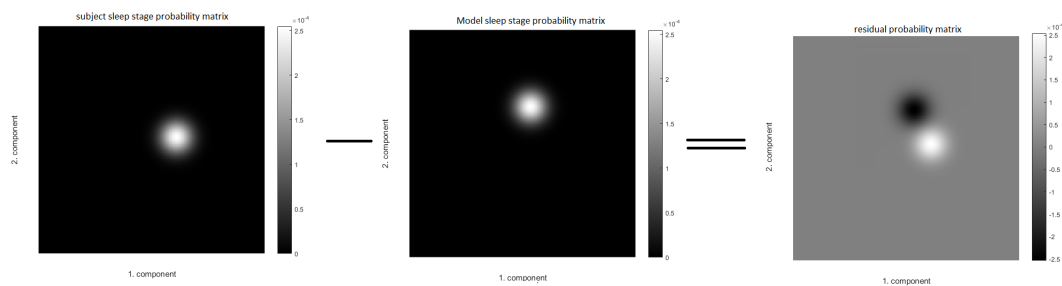


Figure 4: Demonstration on how the features were extracted. The General Mixture Model were fitted to the data points and then subtracted from the model sleep stage.

5.4. Classifier

A LDA classifier was used to facilitate interpretation as it is then possible to analyze the effect derived from each individual features. The LDA classifier also gives a linear decision boundary that can be moved by adjusting the cost of misclassification so that a classifier with any desired specificity can be selected (in this study, we are seeking high specificity classifiers that will be used in conjunction with previously described markers such as a nocturnal SOREMP). A forward feature selection algorithm was used to select features with the best effect. It should be noted that subjects not containing any REM sleep were considered non-narcoleptic subjects. In future work a non-linear classifier could be implemented but this would be on the cost of losing the biological knowledge obtained from the results.

6. Results

6.1. State space in narcolepsy versus control

Figure 5 shows 1000 random epochs from each sleep stage extracted from the model dataset (Dataset A) and projected onto a 2-D projection plane using LDA where inner class variance is minimized and between classes variance is maximized. As can be seen, the state space created gave excellent separation of each individual sleep stage. A simple LDA classifier achieved an accuracy of 82% for sleep stage classification, a result equivalent to that obtained in mice using a similar methodology [8]. In Appendix a comparison between the average of the NT1 vs. Controls for each sleep stage can be seen in Fig. 8

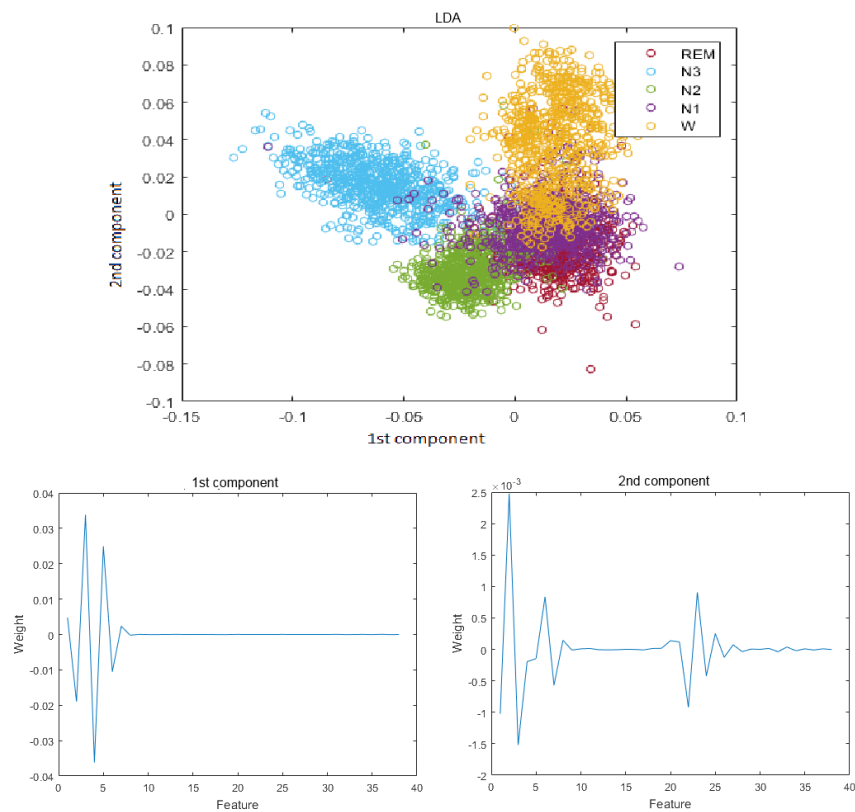


Figure 5: Top: The state space is shown with 1000 random sleep stage epochs from each subject projected onto the space. It is seen separation between classes is good. An LDA classifier yielded an accuracy of 82%. Bottom: The weights from the two components defining the sleep space. As it is seen the first component is mainly composed of EEG features while the second component also has influences from the EMG and EOG modalities. This property is valuable for analysis as it is possible to determine how a change in the signal will affect the position in sleep space. E.g. a change in EEG will mainly adjust the first component while changes in EOG or EMG will only make a positional change on the second component.

In Fig. 5 and Eqn.(9) the components are described. It is seen that the first component is mainly describing the EEG modality features, while the second component includes the EOG and EMG modalities. Therefore, all movement along the first component is describing changes to the EEG while movement along the second component shows changes in EEG but also changes in the other modalities.

The formula for the two component state space is shown in Eqn.(9) (insignificant values < 95th percentile omitted).

$$\begin{aligned}
 PC1 = & \begin{bmatrix} PSE1 \\ PSE2 \\ PSE3 \\ PSE4 \\ PSE5 \\ PSE6 \\ PSE7 \\ PSE8 \end{bmatrix} \cdot \begin{bmatrix} 4.78e-3 \\ -1.90e-2 \\ 3.38e-2 \\ -3.62e-2 \\ 2.49e-2 \\ -1.05e-2 \\ 2.34e-3 \\ -2.12e-4 \end{bmatrix} \\
 PC2 = & \begin{bmatrix} PSE1 \\ PSE2 \\ PSE3 \\ PSE4 \\ PSE5 \\ PSE6 \\ PSE7 \\ PSE8 \\ PSE9 \\ PSE11 \\ PSE18 \\ PSE19 \\ PSE20 \\ PB1 \\ PB2 \\ EEG_{90thPercentile} \\ EEG_{95thPercentile} \\ EEG_{Activity} \\ EEG_{Mobility} \\ EEG_{Complexity} \\ EMG_{Activity} \\ EOG_{Filter1} \\ EOG_{Filter2} \\ EOG_{Filter3} \\ EOG_{Filter4} \\ EOG_{Filter6} \end{bmatrix} \cdot \begin{bmatrix} -1.0e-3 \\ 2.48e-3 \\ -1.52e-3 \\ -1.95e-4 \\ -1.45e-4 \\ 8.36e-4 \\ -5.71e-4 \\ 1.44e-4 \\ -1.06e-5 \\ 1.48e-5 \\ 1.42e-5 \\ 1.68e-5 \\ 1.38e-4 \\ 1.18e-4 \\ -9.21e-4 \\ 9.06e-4 \\ -4.24e-4 \\ 2.51e-4 \\ -1.25e-4 \\ 7.27e-5 \\ -3.64e-5 \\ 1.53e-5 \\ -3.74e-5 \\ 3.71e-5 \\ -2.19e-5 \\ -1.03e-5 \end{bmatrix} \quad (9)
 \end{aligned}$$

Where PSE_x is the power spectrum estimated in the x 'th band as explained in the method section. The EOG_{Filter_y} is the $r_k(n)$ value extracted from the y 'th filter as explained in the method section.

6.2. Construction of an overall classifier

Training an LDA classifier on the training dataset (Dataset B) containing 728 clinical subjects and 19 untreated narcoleptics, it is possible to get a classifier that can separate clinical subjects and narcoleptics. To reduce the amount of features, a forward feature selection algorithm selected the 8 most prominent probability features; these are described in Table 3. By varying the cost of a misclassification, the classifier is trained so that a linear separation plane is described by an equation. If this equation yields $y \geq 0$, the subject is classified as narcoleptic. The ROC curve generated by varying the cost can be seen in Fig. 6. As noted in the introduction a high specificity (> 95%) is required, therefore an optimal separation plane is chosen and described by Eqn.(10). This is the separation plane used to create the confusion matrixes in Fig. 6 and the plane is also marked on the ROC curves.

$$y = \begin{bmatrix} R_p(W, N2) \\ R_p(W, N3) \\ R_p(N1, N1) \\ R_p(N1, N3) \\ R_p(N2, N2) \\ R_p(N3, N3) \\ R_p(REM, N1) \\ R_p(REM, N2) \end{bmatrix} \cdot \begin{bmatrix} -10.3 \\ 12.2 \\ -5.5 \\ -15.2 \\ -3.5 \\ 1.7 \\ 4.8 \\ 12.6 \end{bmatrix} \quad (10)$$

Table 3: Table summarizing the prominent probability features that can separate narcoleptics using sleep stage confusion. It is seen that some sleep stages show higher mismatch and some show higher match when compared.

Higher Mismatch than in controls with template	Wake compared to N3
	N3 compared to N3
	REM compared to N1
	REM compared to N2
Higher Match than in controls with template	Wake compared to N2
	N1 compared to N1
	N1 compared to N3
	N2 compared to N2

As seen in Eqn.(10), sleep stage N1 is involved in three out of eight features while the wake sleep stage is involved in two out of eight. Some features are based on a higher mismatch between sleep stages while others are based on a higher match. This is seen by analyzing the sign of the weight.

6.3. Testing the classifier in the various datasets

Using the classifier, that was trained on the Training dataset (Dataset B), on the Validation dataset (Dataset C) it is seen in Fig. 6B that a sensitivity of 38% and specificity of 94% is found. This is the same specificity as was obtained on the training set. This means that the classifier is valid for both clinical subjects and control subjects obtained from multiple recording sites. In Fig. 6C a sensitivity analysis is made on the high pretest probability sample of hypersomnia subjects vs. narcoleptic dataset (Dataset D). It is seen that there has been a drop in both the specificity, which is at 84%, and the sensitivity, which is at 15%, showing that the classifier is not performing well when the control group is consisting of hypersomnia patients.

6.3.1. Analysis of the effect of anti-depressant and central acting stimuli

Other work [6] suggests that antidepressants and other central acting stimuli can have an effect on the results of classification. Therefore, an analysis was done on Dataset E that contained controls versus the narcoleptics receiving antidepressants and other central acting stimuli. In Fig. 7 it is seen that the introduction of antidepressants and other central acting stimuli reduces the specificity to 94% and sensitivity to 10%. Therefore, a drop in results is seen when stimuli is introduced.

7. Discussion

In this study, we explored whether sleep-wake state dissociations could be used to diagnose narcolepsy using a night time PSG study alone, extending on studies of other PSG biomarkers such as SOREMS [1], sleep stage transitions [6] and power spectrum ratios [17] in the disorder. Using a combination of EEG, EOG and EMG features, we created a sleep state space that can be used to distinguish sleep stages in a 2 D-plan. The sleep state of each narcolepsy patient and each control was next match to this template, and similarities/dissimilarities based on 8 most distinguishing features were computed into a weighted score that varied from X to Z, with $y \geq 0$ being the threshold for defining narcolepsy. In essence, we are looking at variation to a control template of the various sleep stages in narcolepsy versus controls.

7.1. Feature analysis

Distinguishing narcolepsy features found following this analysis are summarized in Table 3. Four features were based on the observation of higher sleep stage mismatch with the template and four features were based on a higher match with the template. Starting with wake, this state shows mismatch with sleep stage N3, i.e. it has moved away from N3 of controls but increased match with stage N2, suggesting wake in narcolepsy has moved closer to stage N2. This result is in line with prior findings showing increased delta and theta activity and decreased alpha and sigma activity in the wakefulness of narcolepsy patients [18] pointing towards a general lowering of frequencies. This may

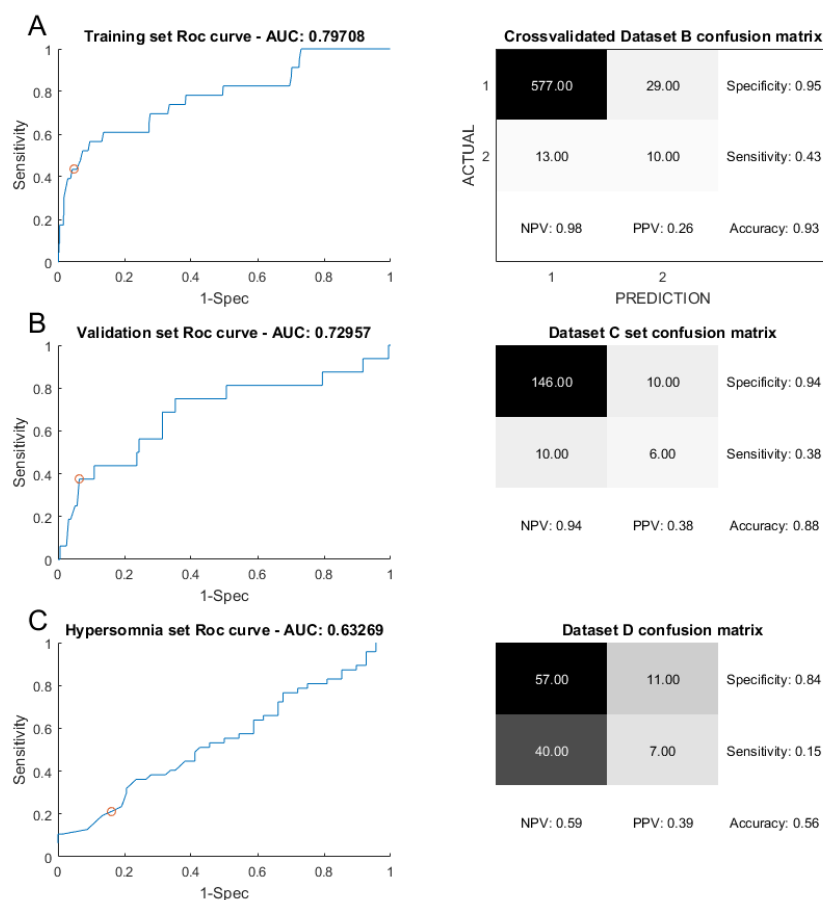


Figure 6: A: Shows ROC curve and confusion matrix from the training dataset (Dataset B) using 10-fold cross validation, B: Shows ROC curve and confusion matrix from the first validation dataset (Dataset C), C: Shows ROC curve and confusion matrix from the second validation dataset (Dataset D).

also reflect decreased muscle activity in wake. Regarding stage N1, a higher match was observed with template control N1 and N3 stages. This observation may be interpreted in several ways. First, the N1 sleep stage is known to be significantly increased in narcoleptic patients, representing about 21.5% of sleep versus 10.3% in controls [19]. The higher amount of N1 may lead to a more diffuse and variable N1 state than in controls that encompass more similarities with N3. Second, the higher match observed between the N1 of narcoleptic and template N3 sleep stage may reflect decreased slow wave activity during N1 in narcolepsy [20]. The narcoleptic N2 sleep stage is only matched with the model N2 sleep stage. This finding can also be explained by the change in amount of sleep as the percentage increase in N1 in narcoleptics comes from this stage meaning that there is less variability in N2. Finally, REM sleep stage is having a higher mismatch with N1 and N2 which could be due to multiple factors as reported by Mukai et. al. [21] who found power spectral density changes in the 0.3-0.9, 1.5-3.4 and 4.0-4.9 Hz frequency range of REM sleep in narcolepsy. Other factors, such as the lack of atonia in REM sleep [22], could also have an influence.

7.2. Strength and weaknesses of the study

The study has several strengths and a number of weaknesses. In term of strengths, exploration and validation were done in distinct datasets and yet gave similar performance. The sensitivity/specificity of this novel combined biomarker

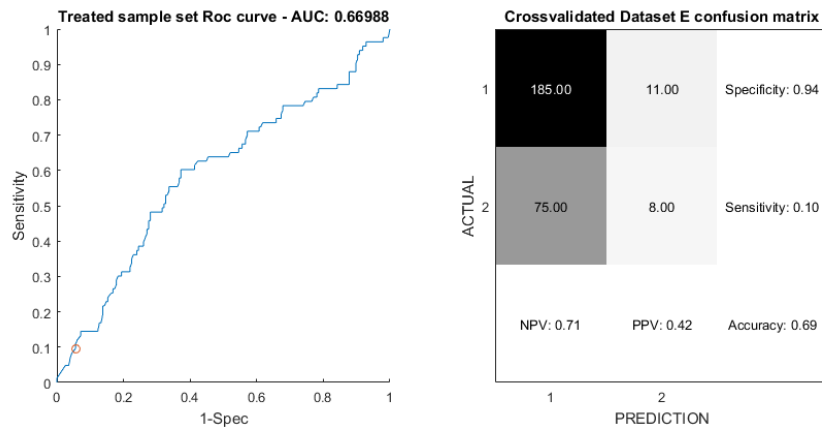


Figure 7: ROC and confusion matrix from the dataset containing the narcoleptics that are receiving anti depressants or other central acting stimuli.

is also quite good, approximately 40% and 95% in the exploration and validation dataset, although it behaved slightly worse in the type 1 narcolepsy versus other hypersomnia dataset (not surprisingly as this dataset may contain some mild spectrum narcolepsy cases). A weakness of the study pertains to the difficulties in interpreting the results, as similarities and dissimilarities are multifactorial in the model. For example, we have not discussed the potential role of REMs versus Slow Eye Movements (SEMs) in mediating the component diagnostic features discussed above, although, a comparison between lucid dreamers and narcoleptics found no difference in the EOG characteristics [23]. Further, because it is most discriminatory of sleep stages, the component examined are mostly influenced by EEG rather than by other modalities (EOG and EMG). It may be that examining other component projections would yield better results if abnormalities are more complex than simple sleep-stage mixing, which seems to be the case based on the complex results obtained. In summary, we found that sleep stage dissociation metrics as projected on a 2-D space can be used as a biomarker for the diagnosis of NT1 with a specificity of 95%, 94% and a sensitivity of 94% and 38% in the exploratory and confirmatory dataset. When tested on a dataset compiled of narcoleptics and other hypersomnias, specificity and sensitivity fell to 84% and 15%, but it is likely cases contain some undiagnosed Type 1 narcolepsy without cataplexy it is also worth knowing that only this dataset was diagnosed using lumbar punctures and this also introduces some variation compared to the MSLT diagnosis of the other datasets. Use of antidepressants or stimulants decreased performance.

7.3. Dataset manipulation

To conduct a study of this magnitude, including datasets recorded at multiple sites, on different equipment and scored by multiple people sometimes even using different criteria, required a huge normalization to make the datasets comparable without loss of data characteristics. This will always have an effect on the results. However as multiple scorings were not available for the data, an interscorer reliability analysis could not be performed to mitigate these issues.

7.4. Future work

This study expands on a collection of features that can be extracted from nocturnal PSGs to diagnose narcolepsy, namely nocturnal SOREMPs [1], sleep stage transition abnormalities [6] [24] [25], EEG spectral analysis differences [7] REM sleep atonia [3] and others. A major goal of the next study will be to combine these biomarkers together in a single detector, and to test this detector in a novel, independent sample of untreated subjects. Another goal would be to conduct a study using datasets with the same quantity of subjects where more of the relevant clinical information (HLA and hypocretin measurements) was available the solidify the findings according to the clinical information.

8. Conclusion

Sleep stage dissociation features of narcolepsy can be evaluated using sleep stage space projection. The biomarker have some diagnostic value. However, the value of the biomarker is decreased in treated patients.

9. Acknowledgment

Thanks to Oscar Carillo and Hyatt More, Stanford Sleep Center, and Sarah Alvarez-Horine, Jazz Pharmaceuticals, for assisting in procuring PSG data. The study was supported by grants from H. Lundbeck A/S, the Lundbeck Foundation, Jazz Pharmaceuticals, and the Technical University of Denmark. This project was also supported by R01HL62252 and 1UL1RR025011.

10. Appendix

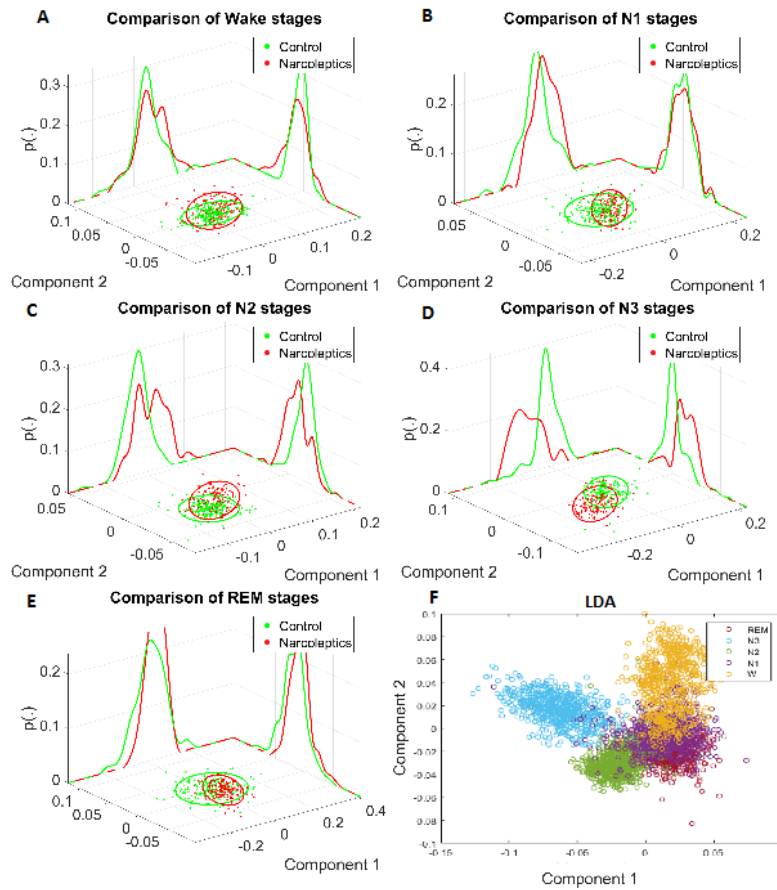


Figure 8: A-E: Figures show the average position in the LDA space for the different sleep stages from a subsample of 199 controls and 148 narcoleptics sampled from the clinical controls and the NT1 patients. The distributions on the different LDA components is shown on the axes and shown combined on the plane. F: shows the location in the state space to enhance the understanding for the reader.

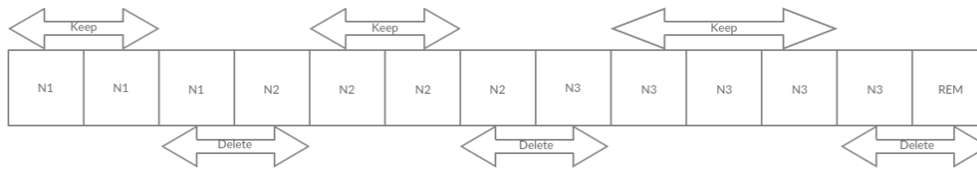


Figure 9: The figure depicts how the transition phases were discarded from the data to enhance purity in the sleep stages.

References

- [1] O. Andlauer, H. Moore, L. Jouhier, C. Drake, P. E. Peppard, F. Han, S.-C. Hong, F. Poli, G. Plazzi, R. O'Hara, E. Haffen, T. Roth, T. Young, E. Mignot, Nocturnal rapid eye movement sleep latency for identifying patients with narcolepsy/hypocretin deficiency, *Jama Neurology* 70 (7) (2013) 891–902. doi : 10.1001/jamaneuro.1.2013.1589.
- [2] G. Sorensen, S. Knudsen, P. Jennum, Sleep transitions in hypocretin-deficient narcolepsy, *Sleep (Online)* 36 (8) (2013) 1173–1177. doi : 10.5665/sleep.2880.
- [3] S. Knudsen, P. J. Jennum, J. Alving, S. P. Sheikh, S. Gammeltoft, Validation of the icdsd-2 criteria for csf hypocretin-1 measurements in the diagnosis of narcolepsy in the danish population, *Sleep* 33 (2) (2010) 169–76.
- [4] J. Williams, D. C. B. Lye, T. Umapathi, Diagnostic lumbar puncture: minimizing complications, *Internal Medicine Journal* 38 (7) (2008) 587–591. doi : 10.1111/j.1445-5994.2008.01631.x.
- [5] M. S. Aldrich, R. D. Chervin, B. A. Malow, Value of the multiple sleep latency test (mslt) for the diagnosis of narcolepsy, *Sleep (Rochester)* 20 (8) (1997) 620–629.
- [6] J. Christensen, O. Carrillo, E. Leary, P. Peppard, T. Young, H. Srensen, P. Jennum, E. Mignot, Sleep-stage transitions during polysomnographic recordings as diagnostic features of type 1 narcolepsy, *Sleep Medicine* 16 (2015) 15581566. doi : 10.1016/j.sleep.2015.06.007.
- [7] J. A. E. Christensen, E. G. S. Munk, P. E. Peppard, T. Young, E. Mignot, H. B. D. Srensen, P. Jennum, The diagnostic value of power spectra analysis of the sleep electroencephalography in narcoleptic patients, *Sleep Medicine* 16 (12) (2015) 1516–1527. doi : 10.1016/j.sleep.2015.09.005.
- [8] C. G. Diniz Behn, E. B. Klerman, T. Mochizuki, S. C. Lin, T. E. Scammell, Abnormal sleep/wake dynamics in orexin knockout mice, *Sleep, Sleep (rochester)* 33 (3) (2010) 297–306.
- [9] L. L. Imbach, E. Werth, U. Kallweit, Inter-hemispheric oscillations in human sleep., *PLoS ONE* 7 (11) (2012) 1.
- [10] T. Roth, Y. Dauvilliers, E. Mignot, J. Montplaisir, J. Paul, T. Swick, P. Zee, Disrupted nighttime sleep in narcolepsy, *Journal of Clinical Sleep Medicine* 9 (9) (2013) 955–965. doi : 10.5664/jcsm.3004.
- [11] H. Moore, E. Leary, S.-Y. Lee, O. Carrillo, R. Stubbs, P. Peppard, T. Young, B. Widrow, E. Mignot, Design and validation of a periodic leg movement detector, *Plos One* 9 (12) (2014) e114565. doi : 10.1371/journal.pone.0114565.
- [12] H. Moore, E. Leary, S.-Y. Lee, O. Carrillo, R. Stubbs, P. Peppard, T. Young, B. Widrow, E. Mignot, Design and validation of a periodic leg movement detector (vol 9, e114565, 2014), *Plos One* 10 (9) (2015) e0138205. doi : 10.1371/journal.pone.0138205.
- [13] A. Goldbart, P. Peppard, L. Finn, C. M. Ruoff, J. Barnet, T. Young, E. Mignot, Narcolepsy and predictors of positive mslts in the wisconsin sleep cohort, *Sleep (Rochester)* 37 (6) (2014) 1043–1051. doi : 10.5665/sleep.3758.
- [14] C. F. P. George, N. Feldman, Y. Zheng, T. L. Steininger, S. M. Grzeschik, C. Lai, N. Inhaber, A 2-week, polysomnographic, safety study of sodium oxybate in obstructive sleep apnea syndrome, *Sleep and Breathing* 15 (1) (2011) 13–20. doi : 10.1007/s11325-009-0320-0.
- [15] J. Black, C. Guilleminault, R. Bogan, H. Emsellem, N. Feldman, M. Hagaman, G. Hertz, V. Iyer, S. Kathawalla, D. Lankford, M. Mitler, R. Hayduk, R. Pascualy, P. Sahota, T. Roth, M. Scharf, L. Scrima, J. Schwartz, G. Zammit, U. X. M. S. Grp, A randomized, double blind, placebo-controlled multicenter trial comparing the effects of three doses of orally administered sodium oxybate with placebo for the treatment of narcolepsy, *Sleep* 25 (1) (2002) 42–49.
- [16] J. Kempfner, P. Jennum, H. B. D. Sorensen, J. A. E. Christensen, M. Nikolic, Automatic sleep staging: From young adults to elderly patients using multi-class support vector machine, 2013 35th Annual International Conference of the Ieee Engineering in Medicine and Biology Society 2013 (2013) 5777–5780. doi : 10.1109/embc.2013.6610864.
- [17] J. A. E. Christensen, E. G. S. Munk, P. E. Peppard, T. Young, E. Mignot, H. B. D. Srensen, P. Jennum, The diagnostic value of power spectra analysis of the sleep electroencephalography in narcoleptic patients, *Sleep Medicine* 16 (12) (2015) 1516–1527. doi : 10.1016/j.sleep.2015.09.005.
- [18] C. E. D. Alloway, R. D. Ogilvie, C. M. Shapiro, Eeg spectral analysis of the sleep-onset period in narcoleptics and normal sleepers, *Sleep (Rochester)* 22 (2) (1999) 191–203.
- [19] J. Li, X. S. Dong, X. Han, Z. M. He, Y. H. L. L. Wang, Q. Y. He, F. Han, Changes of sleep architecture in patients with narcolepsy, *National Medical Journal of China* 87 (9) (2007) 619–621.
- [20] R. Khatami, H.-P. Landolt, P. Achermann, J. V. Retey, E. Werth, J. Mathis, C. L. Bassetti, Insufficient non-rem sleep intensity in narcolepsy-cataplexy, *Sleep* 30 (8) (2007) 980–989.
- [21] J. Mukai, S. Uchida, S. Miyazaki, K. Nishihara, Y. Honda, Spectral analysis of all-night human sleep eeg in narcoleptic patients and normal subjects, *Journal of Sleep Research* 12 (1) (2003) 63–71. doi : 10.1046/j.1365-2869.2003.00331.x.
- [22] C. Schenck, M. Mahowald, Motor dyscontrol in narcolepsy - rapid-eye-movement (rem) sleep without atonia and rem-sleep behavior disorder, *Annals of Neurology* 32 (1) (1992) 3–10.
- [23] P. Dodet, M. Chavez, S. Leu-Semenescu, J.-L. Golmard, I. Arnulf, Lucid dreaming in narcolepsy, *Sleep* 38 (3) (2015) 487–497. doi : 10.5665/sleep.4516.
- [24] Y. Liu, J. Zhang, V. Lam, C. K. W. Ho, J. Zhou, S. X. Li, S. P. Lam, M. W. M. Yu, X. Tang, Y.-K. Wing, Altered sleep stage transitions of rem sleep: A novel and stable biomarker of narcolepsy, *Journal of Clinical Sleep Medicine* 11 (8) (2015) 885–894. doi : 10.5664/jcsm.4940.
- [25] F. Pizza, S. Vandi, M. Ilioti, C. Franceschini, R. Liguori, E. Mignot, G. Plazzi, V. Donadio, Nocturnal sleep stage transitions identify narcolepsy type-1 among central disorders of hypersomnolence, *European Journal of Neurology* 22 (2015) 218–218.