Estimation of Sleep Quality by Using Microstructure Profiles

Zuzana Rošťáková[1,2(✉)], Georg Dorffner[2], Önder Aydemir[2,3], and Roman Rosipal[1]

[1] Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia
zuzana.rostakova@savba.sk
[2] Section for Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria
[3] Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, Turkey

Abstract. Polysomnograhy is the standard method for objectively measuring sleep, both in patient diagnostics in the sleep laboratory and in clinical research. However, the correspondence between this objective measurement and a person's subjective assessment of the sleep quality is surprisingly small, if existent. Considering standard sleep characteristics based on the Rechtschaffen and Kales sleep models and the Self-rating Sleep and Awakening Quality scale (SSA), the observed correlations are at most 0.35. An alternative way of sleep modelling - the probabilistic sleep model (PSM) characterises sleep with probability values of standard sleep stages Wake, S1, S2, slow wave sleep (SWS) and REM operating on three second long time segments. We designed sleep features based on the PSM which correspond to the standard sleep characteristics or reflect the dynamical behaviour of probabilistic sleep curves. The main goal of this work is to show whether the continuous sleep representation includes more information about the subjectively experienced quality of sleep than the traditional hypnogram. Using a linear combination of sleep features an improvement in correlation with the subjective sleep quality scores was observed in comparison to the case when a single sleep feature was considered.

Keywords: Probabilistic sleep model · Hypnogram · Self–rating Sleep and Awakening Quality scale · Sleep features

1 Introduction

Polysomnography (PSG) is the standard method for objectively measuring sleep, both in patient diagnostics in the sleep laboratory and in clinical research. Besides revealing important events pointing towards sleep disorders, such an objective biomarker can also be expected to reflect the quality of sleep in

A. ten Teije et al. (Eds.): AIME 2017, LNAI 10259, pp. 105–115, 2017.
DOI: 10.1007/978-3-319-59758-4_12

terms of how rested the subject feels in the morning. Yet, if one looks at the correspondence between this objective measurement and a person's subjective assessment of one's sleep quality then it is found to be surprisingly small, if existent. In [1] the authors cluster patients by values of the Pittsburgh Sleep Quality Index (PSQI) and the Epworth Sleepiness Scale (ESS), both subjective measures of long-term (typically a month) sleep and daytime wakefulness qualities. They found no significant differences in any PSG sleep variable between those clusters, indicating that those subjective variables measure something distinct from the objective sleep recording.

In [2] the authors perform a multi–variable regression predicting different psychomotor performance results from subjective and PSG–based sleep variables. The highest correlation was between a set of variables containing total sleep time (TST), sleep efficiency (EFF), wake after sleep onset (WASO) and sleep onset latency (SLAT), with $R^2 = 0.21$ predicting the performance in a simple reaction time test, but only within a group of normal sleepers.

Our own previous results [3,4] show that correlations between any PSG sleep variable and subjective assessments on the same night (in particular Saletu's Subjective Sleep and Awakening Scale, or SSA, [5]) are poor, at best, with Pearson or Spearman correlation coefficients hardly above 0.4. The same work, however, showed that the novel probabilistic model of sleep, representing the microstructure of sleep as compared to standard hypnograms, can lead to variables with significantly higher correlation coefficients, pointing to the fact that standard sleep scoring does not extract the maximum information about sleep from the electrophysiological signals, in particular, electroencephalography (EEG).

In this paper we investigate whether linear combinations of several variables can achieve a higher correlation with subjective sleep quality than single variables. We do this for both probabilistic and traditional stage–based sleep profile to also investigate whether in a multi–variable setting the former can also outperform the latter.

2 Data Set Description

In this study, the electroencephalographic (EEG) data from 540 polysomnographic sleep recordings from the SIESTA database [6] were used. The 540 PSG nights were recorded from 270 subjects in two consecutive nights spent in the sleep laboratory.

The microstructure of each sleep recording was calculated by using the probabilistic sleep model (PSM) introduced in [4]. In the PSM method three seconds long time segments are used to calculate probability values to be in a certain stage (Wake, S1, S2, slow wave sleep (SWS) and REM). Figure 1 shows an example of the microstructure of sleep by depicting the probabilities for each sleep stage over time. The standard Rechtschaffen and Kales scores[1] obtained by the automatic scoring system Somnolyser 24×7 [9] are plotted as well.

[1] Nowadays, the American Academy of Sleep Medicine (AASM) sleep model is preferred in the clinical praxis, but we do not expect significant changes in results when using the AASM scores instead of the Rechtschaffen and Kales sleep model.

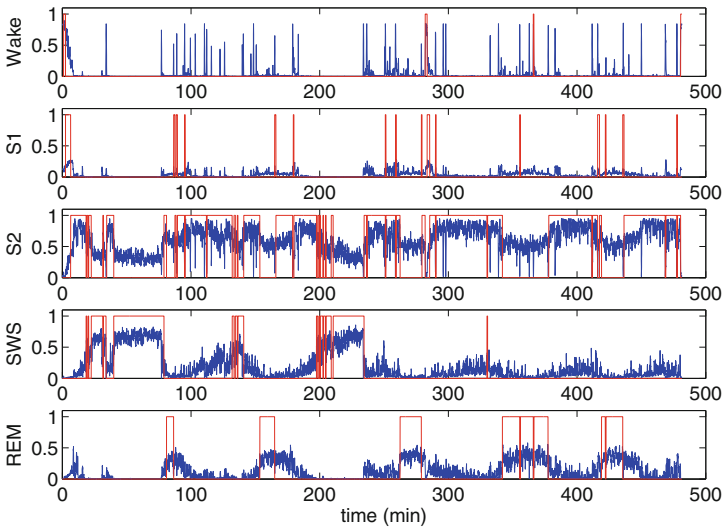

Fig. 1. An example of the microstructure of sleep for an all-night recording (blue) and corresponding Rechtschaffen and Kales scores (red). (Color figure online)

After the subjects woke up, they were asked to fill in the Self–rating Sleep And Awakening Quality (SSA) questionnaire [5]. The scale consist of 7, 8 and 5 questions on sleep quality, awakening quality and somatic complaints, leading to a total score with a value between 20 (best quality) and 80 (worst quality).

3 Sleep Features

3.1 Hypnogram Features

A set of 25 standard sleep variables derived from a hypnogram were calculated (Table 1). The descriptions and abbreviations of the features are given in the second and third column of Table 1, respectively.

3.2 PSM Based Sleep Features

For a continuous probabilistic sleep profile X observed over a time interval T we aimed to design variables that have a correspondence to the standard sleep measures and to include variables that optimally exploit potentially characteristics found in the continuous profiles.

Band Power. The band power (BP) was computed by the following formula

$$BP|_{f_1}^{f_2} = \sum_k \|F_x(k)|_{f_1}^{f_2}\|^2, \tag{1}$$

Table 1. The hypnogram features

Hypnogram feature	Description	Abbreviation
Time in bed (min)	Time from lights out to the end of the recording	TIB
Total sleep period (min)	Time from the first to the last epoch in any sleep stage	TSP
Total sleep time (min)	Sum of epochs in one of the sleep stages S1, S2, SWS, REM	TST
Wake within TSP (min)	Sum of wake-epochs within the total sleep period	WTSP
Wake after final awakening (min)	Time from final awakening to the end of the recording	WAFA
Lights out to S1	Time from lights out to the first occurrence of stage 1	LS1
Lights out to S2	Time from lights out to the first occurrence of stage 2	LS2
Sleep latency	Time from lights out to the first occurrence of three consecutive epochs in stage S1 or to the first occurrence of stage S2	SLAT
Sleep efficiency (%)	$\frac{TST}{TIB} \times 100$	EFF
Stage (min)	Time spent in a given sleep stage	S1, S2, S3, S4, SWS, REM
Stage (%)	$\frac{\text{Time spent in a given sleep stage}}{TST} \times 100$	S1p, S2p, S3p, S4p, SWSp, REMp
Frequency of awakening	Number of awakenings within the total sleep period	FW
Awakening-index	Number of awakenings within the total sleep period per hour sleep	FWTST
Frequency of stage shifts	Number of stage shifts within the total sleep period	FS
Stage shift-index	Number of stage shifts within the total sleep period per hour sleep	FSTST

where $F_x(k)|_{f_1}^{f_2}$ denotes the coefficients of the Fast Fourier Transform of x between a lower cut-off frequency $f_1 = 0$ and an upper cut-off frequency $f_2 = 0.001$.

Entropy. Entropy characterises the level of uncertainty of a signal. For each microstructure of sleep stages it was computed as follows

$$ent = -\int_T \frac{X(t)}{\int_T X(s)ds} \log \frac{X(t)}{\int_T X(s)ds} dt, \qquad \text{where } \log 0 = 0 \text{ by definition.} \quad (2)$$

Log-Spaced Power Spectral Density. The power spectral density of the microstructure of sleep EEG was estimated through the modified covariance method. This method fits a p^{th} order autoregressive (AR) model to a signal, which is assumed to be the output of an AR system driven by white noise and minimising the forward and backward prediction errors. The order of the AR model was empirically selected as 13.

To characterise the output vector X_{PSD} of the normalised estimate of the AR system parameters with only one number we chose the common logarithm of its mean value

$$psd = \log_{10} \overline{X_{PSD}}. \tag{3}$$

Moving Window Features. Two statistical features were extracted by using a moving window through the probabilistic profile of sleep EEG. The moving window has two kinds of parameters – a height h and a length l (the extension of the window). Each of the non–overlapping windows of length l was represented by the number of probability values which were higher than the height parameter. This procedure results in a sequence MWS (*moving window sequence*) of length L that resemble a smoothed version of the entire profile.

The moving window features were calculated as the arithmetic mean (Am) and skewness (Sm) of MWS [2,3]

$$Am = \sum_{m=1}^{L} \frac{MWS(m)}{L} = \overline{MWS}, \tag{4}$$

$$Sm = \frac{\frac{1}{L}\sum_{m=1}^{L}\left(MWS(m) - \overline{MWS}\right)^3}{\left(\sqrt{\frac{1}{L}\sum_{m=1}^{L}\left(MWS(m) - \overline{MWS}\right)^2}\right)^3}. \tag{5}$$

The optimal window parameters were set as $l = 140$ and $h = 0.22$.

Arithmetic Mean (AM), Geometric Mean (GM), Median (Med) were considered as features of the discrete observation of a probabilistic profile X.

Area Under a Curve. In the case of PSM the area under a probabilistic sleep profile X forms an analogy to the time spent in a given sleep stage

$$AUC(X) = \int_T X(t)dt. \tag{6}$$

Moments of a Curve. The first moment mom_1 of a curve X characterises the expected value of the curve according to time [8]

$$mom_1 = \int_T t\frac{X(t)}{AUC(X)}dt. \tag{7}$$

Higher order central moments

$$mom_k = \int_T (t-\mu)^k \frac{X(t)}{AUC(X)} dt, k = 2, 3, \ldots. \tag{8}$$

describe the variability in X. In this study only the second order central moment mom_2 was considered.

Moments of a Feature Function of a Curve. A feature function is a strictly positive transformation of a curve which highlights a set of curves features [8], for example

$$I_{max}(t) = c\left(X(t) - \min_{t \in T} X(t)\right)^r, \ r \in \mathbb{R}, \tag{9}$$

which concentrates its weight to the local maxima of the curve X or

$$I_m(t) = c|X^{(m)}(t)|, X^{(m)} \text{ is the } m^{th} \text{ derivative of } X, \ m = 1, 2, \tag{10}$$

which highlights global characteristics. The constant c guarantees that the area under a feature function is equal to 1. For all three feature functions the first order moment mom_1 was computed.

Curve Length. The curve length characterises changes of a curves profile over a time interval T

$$cl = \int_T \sqrt{1 + (X'(t))^2} dt. \tag{11}$$

4 Methodology

To relate objective and subjective measures of sleep we performed a linear regression for modelling the total score from the SSA scale by a linear combination of variables describing the sleep architecture.

The whole dataset was divided into two parts. The first part served for modelling the SSA scores as a linear combination of sleep features. Because of presence of either redundant or irrelevant sleep features, right before models fitting the feature selection procedure was performed in order to simplify the model. More details about the procedure are given in the next section.

The second part (testing dataset) was used for checking the models' quality by computing the Spearman's correlation coefficient between the real and predicted total SSA scores for the testing dataset.

To avoid problems caused by randomly splitting the dataset into two parts, a 10-fold cross-validation was considered.

The procedure was performed separately for the variables from the PSM and hypnogram as well as for joined datasets of sleep features. Furthermore, this was done for the sleep recordings of the first night and second night separately.

Finally, to detect whether the differences in the correlations estimated using different sets of sleep features are significant, the Student t–test and the Wilcoxon rank–sum test were considered.

4.1 Feature Selection

High number of features considered for the sleep models (25 standard sleep features and 75 features for PSM) may cause inaccuracies in estimation of parameters in the linear regression model. Moreover, some of the features are redundant – they are either highly correlated with other features or include only small information about the sleep process.

Dimensionality reduction is important in machine learning. It leads not only to a decrease in computational time, but may also increase the comprehensibility and the performance of the model. It includes two main approaches – feature extraction methods and feature selection techniques.

Feature extraction methods transform high–dimensional data into a vector space with lower dimension by designing new variables expressed as a linear combination of the original ones. Principal Component Analysis, Factor Analysis and other techniques are typical representatives.

On the other hand, feature selection methods work only with original features and they aimed to find the smallest subset of features with the most informative features.

For the Rechtschaffen and Kales sleep model there is a standard set of sleep features used in the majority of sleep studies. We aimed to design similar set of features for PSM and therefore the feature selection approach is more appropriate for our case.

In this study, we performed sequential feature selection procedure which executes a sequential search among each candidate feature subset in order to find out the smallest subset of features which is able to predict the SSA scores in the best way. This algorithm is implemented for example in the function *sequentialfs* in the MATLAB environment [10].

Similarly to the previous case the 20–fold cross–validation was used. In each of 20 trials, one fold served as a validation dataset, the remaining part of 19 folds formed a training dataset. A linear regression model was fitted to the training dataset using a candidate subset of features. The mean squared error (MSE) between the real and predicted SSA scores for the validation dataset measured the quality of the model. The algorithm started with an empty feature set and then candidate feature subsets were created by sequentially adding each of the features not yet selected until there was no improvement in prediction.

The process resulted into 20 possibly different linear regression models. The model with the lowest MSE in all was used in the further analysis.

5 Results and Discussion

In Table 2 average Spearman's correlation coefficients between original and predicted total SSA scores are listed. Regarding the first night, the highest correlations were obtained by considering joined datasets of sleep features for the PSM and standard hypnogram ($\approx$0.38). Using the standard sleep features only the average correlation ($\approx$0.37) was higher than in the case of the PSM sleep

features (≈0.34). On other hand, because of high standard deviations the differences in results between all three cases were not significant.

Comparing results for the hypnogram features and PSM based sleep variables separately or results of joined datasets of features (Table 2) for the second night the correlations are higher for the first two cases. However, the differences between results were still not significant.

Similar values of mean squared error (Table 3) obtained in the last iteration of the feature selection step confirmed similar performance of the standard hypnogram and PSM.

These results contradict our expectations, that PSM includes more information about the sleep process. A possible reason for no or only slight improvement in the correlations may be that the designed sleep features for the PSM model do not describe the dynamical behaviour of the probabilistic sleep curves properly. The features extracted from the PSM (Sect. 3.2) were chosen so that they are natural counterparts of the standard hypnogram features or they highlight specific properties of a sleep probabilistic curve. However, it is difficult to say whether our features are the most appropriate. The major task for future research is to design new features which would improve the results. Another idea is to use the whole probabilistic sleep curves instead of their one–dimensional characteristics for modelling the results of the subjectively scored sleep and awakening quality.

Table 2. Average Spearman's correlation coefficients and their standard deviations for the PSM based sleep measures, the standard hypnogram sleep features as well as for joined datasets of sleep measures.

	1. Night	2. Night
PSM		
Wake stage	0.3551 ± 0.1732	$0.2224 \pm 0,2500$
S1 stage	0.2024 ± 0.2134	0.2886 ± 0.1778
S2 stage	0.3145 ± 0.2202	0.1921 ± 0.2053
SWS stage	0.2390 ± 0.2158	0.1653 ± 0.1892
REM stage	0.1358 ± 0.2025	-0.0048 ± 0.1836
All stages together	0.3396 ± 0.1800	0.2954 ± 0.1626
Standard features	0.3679 ± 0.2230	0.2762 ± 0.1349
PSM + standard features	0.3809 ± 0.2496	0.2688 ± 0.2238

On the other hand, considering a single sleep feature for predicting the SSA scores in the testing dataset, the correlations were significantly lower or approximately equal to the case when a linear combination of features was used. This was true for both sleep models as well as for joined datasets of sleep features.

In addition, we were interested in sleep features which were selected in the majority of trials. Regarding to the standard sleep features, only the time spent in the S1 stage (total or relative) was selected in more than 7 trials (Table 4). In the case of PSM especially sleep variables related to the Wake and S1 stages

Table 3. An example of the mean squared error between real and predicted values of the Self–rating questionnaire for sleep and awakening quality and sequences of chosen sleep features in the feature selection step for the first night. For both sleep models first three and the last iteration of a randomly chosen trail are described.

Model	1. Iteration	2. Iteration	3. Iteration	Last iteration
Standard features	76.64	67.20	62.75	61.86
	-	S1p	eff, S1p	eff, S1, S1p, S4p
PSM	76.11	66.5	64.17	58.47
	-	Am (Wake)	Am (Wake), psd (Wake)	Am (Wake), psd (Wake), psd (S1), mom_2 (Wake), mom_1 (REM), ...

dominated. Moreover, using the PSM sleep variables related only to the Wake stage the third highest average correlation was observed for the first night.

Considering the joined datasets of sleep features, approximately the same sleep variables related to wakefulness or light sleep were selected.

Finally we observed that the average correlations for the first night are higher in comparison to the results of the second night. We hypothesise this is caused by the "first night effect" – subjects usually sleep poor in a new environment and feel tired in the morning. This is reflected by higher variability in values of sleep features related to the Wake or S1 stage which were the most selected sleep variables in the feature selection process as well as by possibly increased variability in values of the SSA scores. Because of visible changes in sleep features and SSA scores which are typical for the majority of the first nights data, the observed correlations are expected to be higher than in the case of the second night where this effect is diminished.

Table 4. List of features selected in at least 7 of 10 trials for each sleep model and each night separately. In the case of joined datasets of the hypnogram and PSM based sleep features the first mentioned are accentuated with italics.

1. Night	PSM	mom_2 (Wake), AM (Wake), Am (Wake), Sm (S1)
	Standard features	S1, S1p
	PSM + standard features	mom_2 (Wake), Sm (S1), psd (SWS), mom_1 (REM), *S2*, *FS*
2. Night	PSM	Sm (S1), AM (Wake), Med (Wake), ent (REM)
	Standard features	S1p
	PSM + standard features	AM (Wake), Sm (S1), *S1p*

6 Conclusion

In this study we compared the quality of prediction of subjective sleep quality scores using sleep features extracted from the standard hypnogram, the probabilistic sleep model or both sleep models together.

PSM based sleep features led to approximately equal or only slightly higher average correlation coefficients in comparison to the case when a subset of standard hypnogram sleep features was considered. Because of high standard deviations the Student t-test and Wilcoxon test were not able to reject the hypothesis that the correlations are equal. The sleep features from both sleep models together did not lead to significant improvement in prediction of the SSA scores. From this point of view the standard hypnogram and PSM seem to be of equivalent quality.

On the other hand, correlation coefficients estimated when the SSA scores were modelled by a single sleep feature were either significantly lower or approximately equal to the correlations between SSA scores and a linear combination of sleep features.

In the feature selection step, sleep features representing the wakefulness or light sleep were selected in the majority of trials for the hypnogram or PSM related features separately as well as joined datasets of sleep variables. This indicates that the subjectively scored sleep quality is influenced mainly by the amount of the time spent awake or in the light sleep during the night.

Acknowledgements. This research was supported by the Ernst Mach Stipendien der Aktion Österreich–Slowakei ICM-2016-03516, the Slovak Research and Development Agency (grant number APVV–0668–12), the Ministry of Health of the Slovak Republic (grant number MZ 2012/56–SAV–6) and by the VEGA 2/0011/16 grant. Dr. Aydemir's contribution was supported by a scholarship from The Scientific and Technological Research Council of Turkey (TUBITAK).

References

1. Buysse, D.J., et al.: Relationships between the Pittsburgh Sleep Quality Index (PSQI), Epworth Sleepiness Scale (ESS), and clinical/polysomnographic measures in a community sample. J. Clin. Sleep Med. **4**(6), 563–571 (2008)
2. Edinger, J.D., et al.: Psychomotor performance deficits and their relation to prior nights' sleep among individuals with primary insomnia. Sleep **31**(5), 599–607 (2008)
3. Rosipal, R., Lewandowski, A., Dorffner, G.: In search of objective components for sleep quality indexing in normal sleep. Biol. Psychol. **94**(1), 210–220 (2013)
4. Lewandowski, A., Rosipal, R., Dorffner, G.: Extracting more information from EEG recordings for a better description of sleep. Comput. Methods Programs Biomed. **108**(3), 961–972 (2012)
5. Saletu, B., et al.: Short-term sleep laboratory studies with cinolazepam in situational Insomnia induced by traffic noise. Int. J. Clin. Pharmacol. Res. **7**(5), 407–418 (1987)
6. Klosch, G., et al.: The SIESTA project polygraphic and clinical database. IEEE Eng. Med. Biol. Mag. **20**(3), 51–57 (2001)
7. Jain, A., Zongker, D.: Feature selection: evaluation, application, and small sample performance. IEEE Trans. Pattern Anal. Mach. Intell. **19**(2), 153–158 (1997)
8. James, G.M.: Curve alignment by moments. Ann. Appl. Stat. **1**(2), 480–501 (2007)

9. Anderer, G., Gruber, S., Parapatics, M., Woertz, T., Miazhynskaia, G., Klösch, B., Saletu, J., Zeitlhofer, M., Barbanoj, H., Danker-Hopfe, S., Himanen, B., Kemp, T., Penzel, M., Grőzinger, D., Kunz, P., Rappelsberger, A., Schlögl, G., Dorffner, G.: An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24 × 7 utilizing the SIESTA database. Neurophysiology **51**, 115–133 (2005)
10. MATLAB, version 8.3.0 (R2014a), The MathWorks Inc., Natick, Massachusetts (2014)