MEASUREMENT 2013, Proceedings of the 9th International Conference, Smolenice, Slovakia

Clustering Probabilistic Sleep Microstate Curves: a Functional Data Analysis Approach

R. Rosipal

Department of Theoretical Methods, Institute of Measurement Science
Slovak Academy of Sciences, Bratislava, Slovak Republic
Email: roman.rosipal@savba.sk

Abstract. *We introduced and validated an EEG data-based model of the sleep process with an arbitrary number of different sleep states and a high time resolution allowing modelling of sleep microstructure. The proposed probabilistic sleep model describes sleep via posterior probabilities of a finite number of microstates. Using the model, we extracted objective sleep parameters describing quantitative and qualitative characteristics of the probabilistic sleep microstate curves and proved their usefulness when assessing selected aspects of sleep quality. In the current work we are investigating functional data clustering methods applied to sleep microstate posterior curves. The hierarchical structure of the data given by the repeated visits of subjects in the sleep lab motivates our focus on recently proposed multilevel functional clustering analysis approaches. We are applying the multilevel functional principal component analysis to the sleep posterior curves. Preliminary results show promising potential of the approach to separate age-related sleep profiles and extracting subjects' specific night deviations from the mean sleep profiles.*

Keywords: Probabilistic Sleep Model, Functional Data Clustering, Multi-level Functional Principal Component Analysis

1. Introduction

In spite of the fact that sleep accounts for about one-third of human live, the purposes and mechanisms of sleep are only partially clear and are the subject of substantial research. One of the long-term open questions in sleep research is how to define and objectively measure 'normal' sleep and its quality. A different but not less important question is how sleep structure relates to selected daytime quality of life measures, including cognitive, emotional, psychometric or physiological tests and measures. The conventional description of sleep architecture from polysomnographic (PSG) recordings is carried out through applying the standardized Rechtschaffen and Kales (R&K) scoring manual [1] or the recently published update of the rules [2]. However, some aspects of the R&K sleep staging rules were criticized in the past and new ways of analysing sleep have been discussed [3]. With the aim of avoiding known drawbacks of R&K, an alternative computerized sleep model was introduced [4]. The model based on solid probabilistic principles allows describing sleep on an arbitrarily fine time scale and allows considering sleep as a continuous process of transitions between a larger number of sleep sub-states (microstates). Microstates can be combined into subsets. This feature allows defining new sleep states whose physiological interpretation and specific task-related performances can be studied. In the current study we considered posterior values of combined microstates as continuous time curves and we used advanced functional data analysis tools [5,6] to cluster curves into a meaningful structure. The validity of the approach is demonstrated on tasks of clustering age-dependent sleep profiles and clustering specific night deviations observed on subjects spending two consecutive nights in the sleep lab.

2. Subject and Methods

Data of 175 subjects (94 males and 81 females), age between 20 and 95 (mean 51 years and

standard deviation 20 years), from the sleep database created during the EU Siesta project were used [7]. One aim of the Siesta project was to create a normative database of healthy sleepers and sleep-disturbed patients. According to the Siesta recording protocol all subjects had to document their sleep habits over 14 nights. Subjects spent two consecutive nights (7th and 8th night) in the sleep laboratory during which PSG recordings were obtained. Subjects used in this study were classified as healthy sleepers without sleep related disorders [7].
The PSG recording protocol specified 16 channels of biosignals: 6 EEG channels with mastoid as reference (Fp1-M2, C3-M2, O1-M2, Fp2-M1, C4-M1, O2-M1), an additional EEG channel (M1-M2) for re-referencing, 2 EOG channels, submental EMG and EMG recorded from electrodes placed at the musculus anterior tibialis of the left and right leg (electrodes were linked), electrocardiogram and respiratory signals (airflow; movements of the chest wall and abdomen and O2 saturation of arterial blood).
First, using the PSG recording the sleep structure was analysed in 30 sec epochs according to the standard R&K scoring rules for sleep staging [1]. To this aim, the computerized system Somnolyzer 24x7 was used [8]. Next, a new probabilistic sleep model (PSM) was used to represent sleep as a continuum [4].

Probabilistic Sleep Model (PSM)
We use the C3-M2 (or C4-M2 as a substitute) EEG channel. For each 3 sec segment a AR(10) parameter vector $a = (a_1, \dots, a_{10})$ is fitted: $X_t = a_1 X_{\{t-1\}} + a_2 X_{\{t-2\}} + \cdots + a_{10} X_{\{t-10\}} + e_t$
To each 3 sec interval a value $s \in \{0,1,2,3\}$ is assigned (0: no spindle, 1,2,3: spindle with increasing certainty). The R&K labels c are assigned by the automatic sleep scoring system to each 30 sec long data segment. The PSM assumes the existence of a latent variable Z with K possible states (Fig. 1, left)

$$p(a,c,s) = \sum_{z=1}^{K} p(z)p(a|z)p_R(c|z)p_S(s|z)$$

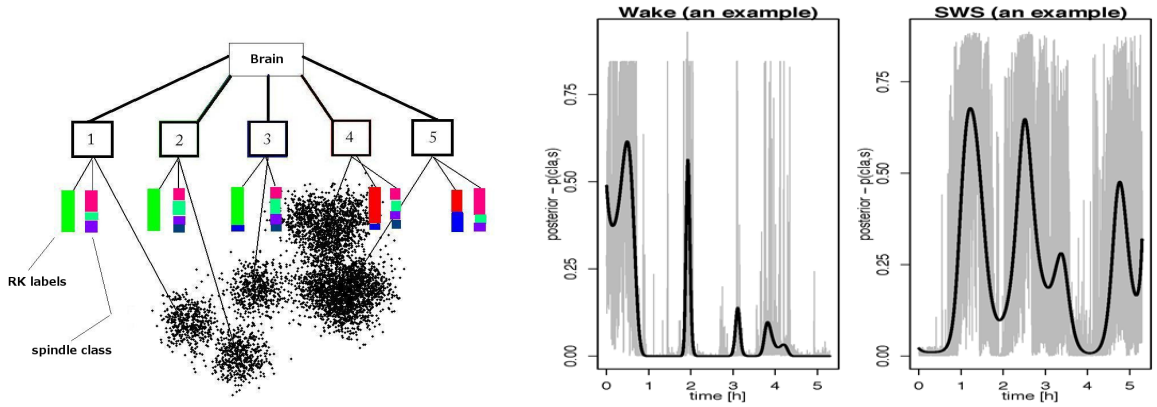

Fig. 1. *Left:* The schematic structure of the probabilistic sleep model (PSM). *Right:* An example of the PSM posterior values (light grey) and the B-spline smoothed curve (black). SWS – slow wave sleep.

The PSM with $K = 20$ sleep microstates was fitted using the EM algorithm [4]. Applying a model means to present an

$$p(z|a,s) = \frac{p(z)p(a|z)p_S(s|z)}{\sum_k p(k)p(a|k)p_S(s|k)}$$

or the R&K posteriors

$$p(c|a,s) = \sum_{z=1}^{K} p(z|a,s)p_R(c|z) \qquad (1)$$

Two-way Functional ANOVA & Multilevel Functional Clustering

We consider the two-way functional ANOVA model for our sleep profiles of subjects $i = 1, \dots, 175$ and nights $j \in \{1,2\}$

$$X_{ij}(t) = \alpha(t) + \beta_j(t) + Y_i(t) + W_{ij}(t) + \epsilon_{ij}(t)$$

where $\alpha(t)$ and $\beta_j(t)$ are fixed functional means specifying the global and night specific functional trends; $Y_i(t)$ is the subject-specific deviation from the night-specific mean and $W_{ij}(t)$ is the subject- and night-specific deviation from the subject-specific mean. Posteriors curves were positive smoothed by using B-splines with 20 basis functions and down-sampled to 30 sec intervals (Fig. 1, right). Nights were aligned to 5.3 hours starting by sleep latency. We use multilevel functional PCA [5] for extracting intra- and inter- subject specific component scores ξ, ζ and eigenfunctions $\phi(t)$

$$Y_i(t) = \sum_k \xi_{ik}\phi_k^{(1)}(t), \quad W_{ij}(t) = \sum_l \zeta_{ijl}\phi_l^{(2)}(t)$$

where $\phi_k^{(1)}(t) \sim K_B(s,t)$, $\phi_l^{(2)}(t) \sim K_W(s,t)$

and $K_T(s,t), K_B(s,t)$ and $K_W(s,t) := K_T(s,t) - K\_B(s,t)$ are the *total, between*, and *within* subjects covariance functions. The number of used eigenfunctions $\phi_k^{(1)}(t)$ and $\phi_l^{(2)}(t)$ was determined by the explained variance of 90%. Finally, we adapted clustering scheme proposed in [6]:

Level-1 Clustering:

Clustering of subject-specific means: subjects i_1 and i_2 will be in the same cluster if their subject-specific deviations $Y_{i1}(t)$ and $Y_{i2}(t)$ are similar in shape.

Level-2 Clustering:

Clustering of night-specific deviations: subjects i_1 and i_2 will be in the same cluster if their deviations from the subject-specific means $W_{i1,j}(t)$ and $W_{i2,j}(t)$, $j \in \{1,2\}$ are similar.

3. Results

First, 20 sleep microstates of the PSM were combined into the R&K representation (wake, S1, S2, SWS – slow wave sleep, and REM – rapid-eye movement) using eq. (1). In addition, microstates representing SWS ($p(\text{SWS}|z) > 0.5$) were combined into a sleep state, which we denote mixMicro-SWS. Using the Level-1 clustering and considering the existence of three clusters we observed for all sleep states a structure with two dominant clusters. The percentage of samples falling into these two clusters was for wake – 99%, S1 – 94%, S2 – 95%, SWS - 94%, REM – 92% and mixMicro-SWS – 94%.

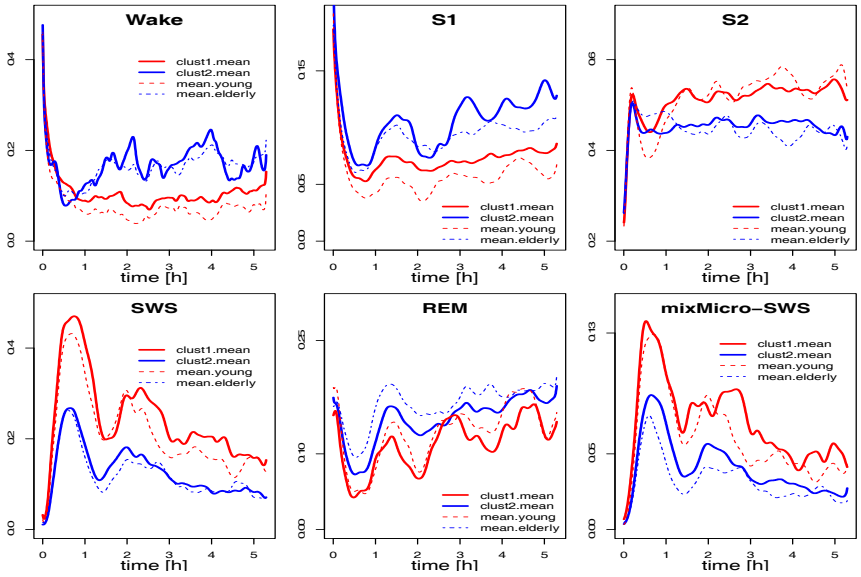

Fig. 2. The averaged sleep profiles of the two dominant clusters (thicker-solid lines) and the averaged sleep profiles of the young and elderly subjects. Curves represent 5.3 hours of sleep aligned by sleep latency. SWS - slow wave sleep, REM - rapid-eye movement.

Following our previous studies [9], we considered three age-related groups of subjects: young (< 40 years old), middle-age (40 to 60) and elderly (≥ 60). We compared averaged posterior curves computed for each age group with averages obtained by averaging posteriors curves falling into each of the two dominant clusters (Fig. 2). Close match between a mean curve of each cluster and averaged sleep profile either for the young or elderly subjects can be observed for every sleep state. For clarity, the averaged sleep profile of the middle-age subjects is not plotted, however, for all six sleep states this profile lies between the averaged sleep profiles of the young and elderly subjects. Finally, by considering the Level-2 clustering we observed strong night effects for wake state reflecting known phenomenon of the first-night effect associated with a higher level of wakefulness and longer sleep latency. Nigh effects were also observed for other sleep states and these are subject of further studies and validation.

4. Discussion

Multilevel functional data clustering was applied to smoothed posterior sleep profiles of the PSM. Preliminary results show promising potential of the approach to separate age-related sleep profiles and extracting subjects' specific night deviations from the mean sleep profiles. Further studies will focus on correlating extracted cluster information and subjects' daytime performance.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0096-10 and by the Scientific Grant Agency of the Ministry of Education of the Slovakia Republic and the Slovak Academy of Sciences VEGA 1/0503/13 and VEGA 2/0043/13 grants.

References

[1] Rechtschaffen A, Kales A. A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subject. U.S. Dept. of Health, Education, and Welfare, Bethesda, MD, 1968.

[2] Iber C, Ancoli-Israel S, Chesson A, Quan S. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, 2007 .

[3] Himanen S, Hasan J. Limitations of Rechtschaffen and Kales. *Sleep Medicine Reviews*, 4, 149-167, 2000.

[4] Lewandowski A, Rosipal R, Dorffner G. Extracting more information from EEG recordings for a better description of sleep. *Computer Methods and Programs in Biomedicine*, 108, 961-972, 2012.

[5] Di Ch-Z, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel Functional Principal Component Analysis, *Annals of Applied Statistics*, 3, 458-488, 2009.

[6] Serban N, Jiang H. Multilevel functional clustering analysis. *Biometrics*, 68: 805-814, 2012.

[7] Klösch et al. The SIESTA project polygraphic and clinical database. *IEEE Engineering in Medicine and Biology Magazine*, 20, 51-57, 2001.

[8] Anderer et al. An E-Health Solution for Automatic Sleep Classification according to Rechtschaffen and Kales: Validation Study of the Somnolyzer 24×7 Utilizing the Siesta Database. *Neuropsychobiology*, 5, 115-133, 2005.

[9] Rosipal R, Lewandowski A, Dorffner G. In search of objective components for sleep quality indexing in normal sleep. *Biological Psychology*, under review.