Multilevel functional clustering analysis of probabilistic sleep microstate curves

Roman.Rosipal@savba.sk, Department of Theoretical Methods, Slovak Academy of Sciences

Introduction

We introduced and validated a probabilistic EEG data-based model of the sleep process with an arbitrary number of different sleep states and a high time resolution [1]. The proposed model describes sleep via posterior probabilities of a finite number of microstates. In the current work we are investigating functional data clustering methods applied to sleep 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 combined with clustering methods on extracted scores [2, 3].

Probabilistic sleep model (PSM)

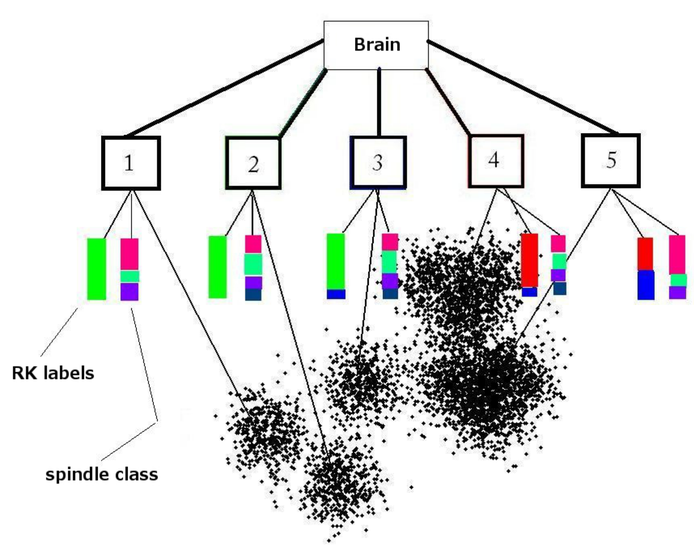
We use the C3-M2 (and C4-M2 as a substitute) EEG channel. For each 3 sec segment a AR(10) model is fitted:

$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_{10} X_{t-10} + e_t$

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 RK labels c are assigned by the automatic sleep scoring system to each 30 sec long data segment.

The probabilistic sleep model (PSM, a form of Gaussian Mixture Model) assumes the existence of a latent variable Z with K possible states and

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



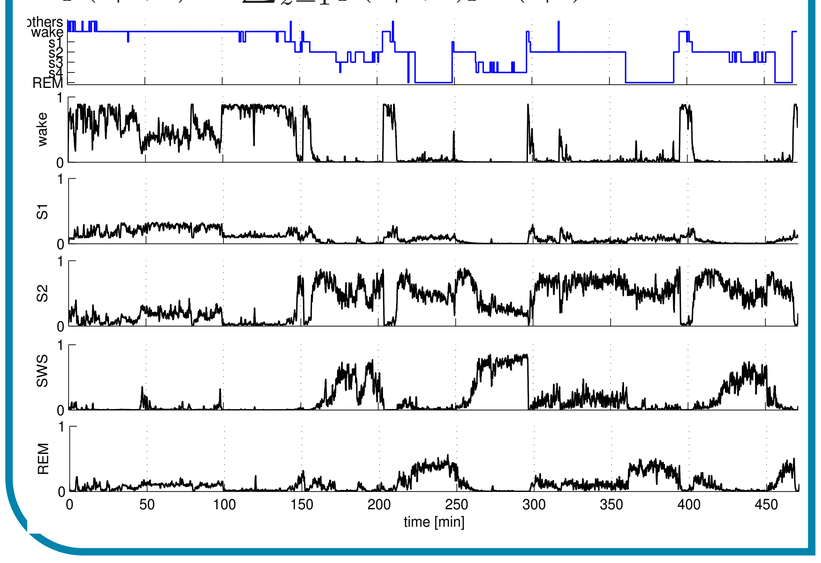
Two night PSG recordings of 175 healthy subjects between 20 and 95 were used. Models were fitted using the EM algorithm.

Applying a model means to present an observation (a, s) and to calculate the posteriors

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

or the RK posteriors

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



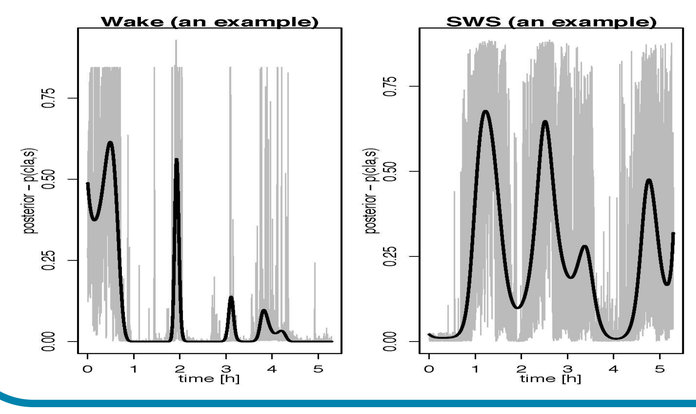
Two-way functional ANOVA

We consider the two-way functional ANOVA model for our; i.e. sleep profiles of subjects $i = 1, \ldots, 175$ 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.

Pre-processing: posteriors curves were positive smoothed by using B-splines basis (20) and downsampled to 30 sec intervals. Nights were aligned to 5.3 hours starting by sleep latency.



Multilevel functional clustering

We use multilevel functional PCA [2] for extracting intra- and inter- subject specific component scores ξ, ζ and eigenfunctions $\phi(t)$

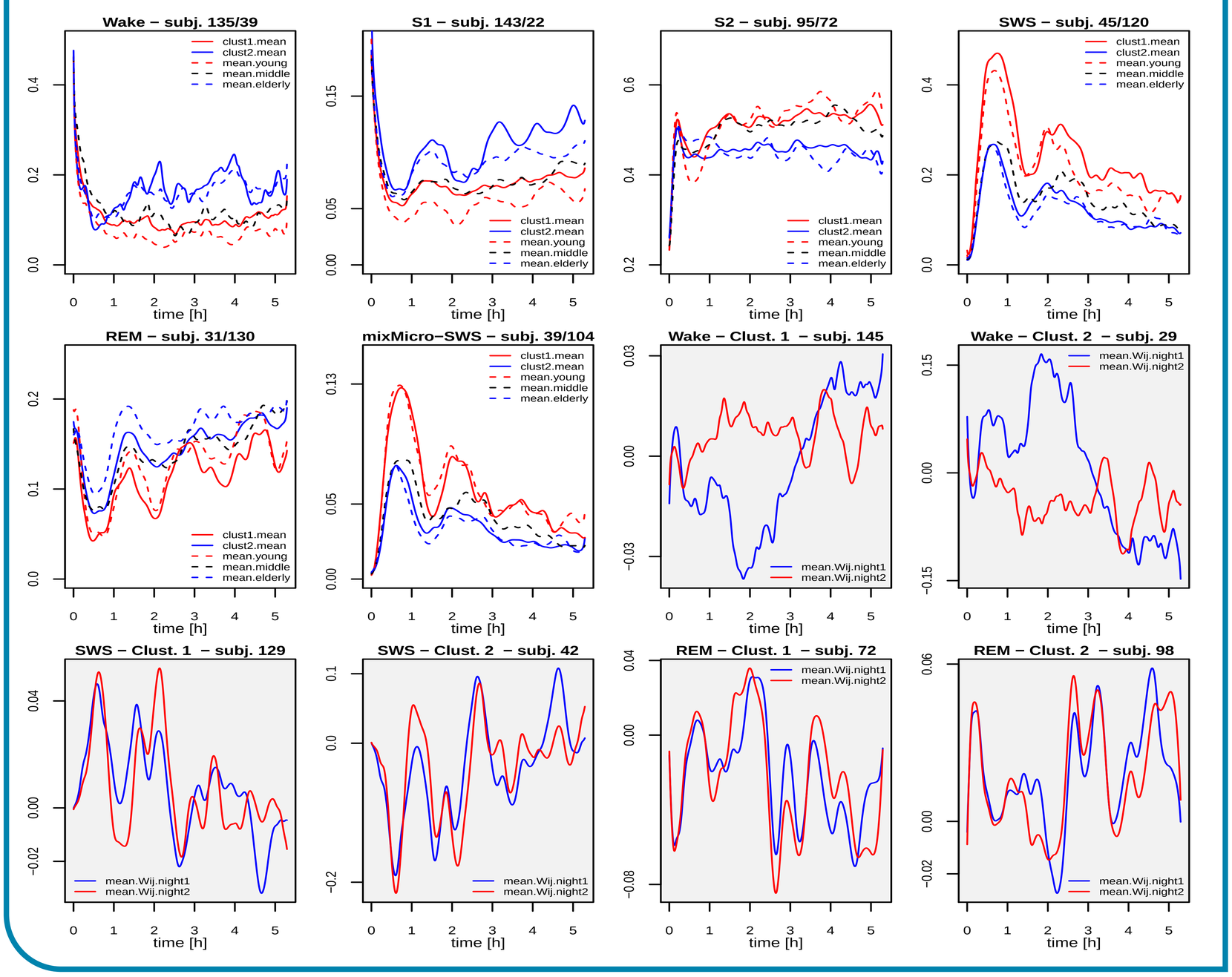
$$Y_i(t) = \sum_k \xi_{ik}\phi_k^{(1)}(t), \quad W_{ij}(t) = \sum_l \zeta_{ijl}\phi_l^{(2)}(t)$$

$\phi_k^{(1)}(t) \sim K_B(s,t)$, $\phi_l^{(2)}(t) \sim K_W(s,t)$
where $K_T(s,t)$, $K_W(s,t) := K_T(s,t) - K_B(s,t)$, and $K_B(s,t)$ are the *total*, *within*, and *between* subjects covariance functions. The number of selected eigenfunctions $\phi_k^{(1)}(t)$ and $\phi_l^{(2)}(t)$ was determined by the explained variance of 90%.

Level-1:
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 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.

Application: Age related clustering & night specific effects



Conclusions

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 are focused on correlating extracted cluster information and subjects' day-time performance.

References

[1] Lewandowski A., Rosipal R., Dorffner G. Extracting more information from EEG recordings for ... *Comp. Meth. and Progr. in Biomed.*, 108:961–72, 2012.

[2] Ch.-Z. Di, et al. Multilevel functional PCA. *The Annals of App. Stat.*, 3:458–88, 2009.

[3] Serban N., Jiang H. Multilevel functional clustering analysis. *Biometrics*, 68:805–14, 2012.

Funding

The work was supported by the Slovak Research and Development Agency under the contract No. APVV-0096-10.