Artificial Intelligence In Medicine 97 (2019) 152–167

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Profiling continuous sleep representations for better understanding of the dynamic character of normal sleep

Zuzana Rošťáková*, Roman Rosipal

Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

ARTICLE INFO

Keywords:
Sleep probabilistic model
Sleep probabilistic curves
Functional cluster analysis
Time alignment
Dynamic time warping
Daily life measures

ABSTRACT

The amount and quality of sleep substantially influences health, daily behaviour and overall quality of life. The main goal of this study was to investigate to what extent sleep structure, as derived from the polysomnographic (PSG) recordings of nocturnal human sleep, can provide information about sleep quality in terms of correlating with a set of variables representing the daytime subjective, neurophysiological and cognitive states of a healthy population without serious sleep problems. We focused on a continuous sleep representation derived from the probabilistic sleep model (PSM), which describes the microstructure of sleep by a set of sleep probabilistic curves representing a finite number of sleep microstates. This contrasts with approaches where sleep is characterised by a set of one-dimensional sleep measures derived from the standard discrete sleep staging. Considering this continuous sleep representation, we aimed to identify typical sleep profiles that represent the dynamic aspect of sleep during the night and that are associated with a set of studied daily life quality measures.

Cluster analysis of sleep probabilistic curves has proven to be a helpful tool when identifying specific sleep temporal profiles, but it faces problems when curves are complex and time misalignment is present. To overcome these problems, we proposed and validated a novel 2-step iterative clustering and time alignment method. We compared the quality of alignment and cluster homogeneity produced by the method with existing approaches in which (i) the time alignment of curves precedes the clustering step, and (ii) time alignment and clustering are performed simultaneously.

The obtained homogeneous clusters of *REM*, *Wake* and *Slow Wave Sleep* resembled the clustering structure of subjects with significantly different subjective scores of sleep quality and mood, as well as more objective cognitive test scores. Moreover, the sleep profiles associated with individual clusters help to better understand the existing associations between the overnight dynamics of specific sleep states and daily measures.

1. Introduction

Sleep can be described as a continuous process in time in which a finite set of sleep states are entered during the night. Sleep quality, length and structure influence human health, mood and performance in daily life [1].

A better understanding of the relationship between sleep structure and daily life outcomes, either in normal subjects or subjects with sleep disorders, is at the centre of interest in sleep research. Current studies often investigate this relationship by considering one-dimensional sleep characteristics derived from the standard clinical sleep models; models either based on Rechtschaffen and Kales sleep scores (R&K) [2] or their novel version produced by the American Academy of Sleep Medicine (AASM) [3].[1] Examples of such sleep characteristics are the total sleep time, sleep efficiency, sleep latency, time awake after sleep onset or the number of awakenings. The relationships between the computed sleep characteristics and daily life measures are then investigated using standard statistical approaches; for example the Spearman correlation coefficient [4–7], linear regression model [8] or cluster analysis [9]. The other group of studies is focused on predicting outcomes of the daily life measures using the extracted sleep characteristics. Linear regression models [10] or machine learning methods have been used in this context [11].

In contrast, the focus of this study is to represent sleep as a continuous profile, not as a one-dimensional measure compressing all time information into a single value. Within the set of individual continuous sleep profiles representing the sleep of each subject, we aim to identify specific sleep profiles, that is, continuous sleep biomarkers or sleep bioprofiles, associated with important physiological aspects of sleep. An important property of these sleep biomarkers would be their

* Corresponding author.
E-mail addresses: zuzana.rostakova@savba.sk (Z. Rošťáková), roman.rosipal@savba.sk (R. Rosipal).

[1] According to the R&K manual, people usually pass through five stages of sleep: *S*1, *S*2, *S*3, *S*4, and REM (rapid eye movement) sleep.

https://doi.org/10.1016/j.artmed.2018.12.009
Received 2 July 2018; Received in revised form 12 December 2018; Accepted 27 December 2018

relationship with different physiological, demographic or daily life measures. This may include physiological factors (as blood pressure, pulse rate), the results of questionnaires about subjective sleep quality, mood, drowsiness or the results of neuropsychological tests focused on, for example attention, fine motor activity or short-term memory [4].

To meet this goal we consider an alternative way of sleep modelling – the probabilistic sleep model (PSM) [12,4]. The model characterises sleep with the probability values of a set of sleep microstates that evolve over the whole night and allows a description of the dynamic attributes of the sleep process. By considering the probability values as a function of time, sleep probabilistic curves can be obtained. These curves then represent the individual sleep profile of each subject. In turn, the continuous sleep representation allows us to use advanced methods of mathematical statistics – functional data analysis (FDA) – to analyse the structure of sleep [13].

To extract and detect associations between representative sleep profiles and available results of cognitive tests, questionnaires or physiological measures, we applied methods of functional data cluster analysis on a set of all available sleep probabilistic curves. This approach allowed us to divide the set of curves into subgroups with homogeneous profiles. Then, we tested the significance of differences in daily measures among subjects that belonged to the formed subgroups.

Unfortunately, many methods for functional data clustering encounter a problem when analysed curves are misaligned in time. Due to misalignment, the profiles of clusters might not be well defined and important features can be missing. This can lead to incorrect clustering and interpretation of the formed clusters. For example, due to incorrect clustering, existing relationships between sleep profiles and daily measures may remain hidden.

In the literature, curve alignment is also known as curve registration or time warping and several methods to solve this problem were developed. A solution when the curve alignment precedes the clustering step usually leads to inferior results, which were also pointed out by Tang and Müller [14] "…However, in the presence of multiple shape patterns, time warping is often unidentifiable due to the confounding of time and shape variation… The dilemma is that we cannot identify the shape clusters without first removing time variation; on the other hand, we cannot remove time variation without causing some degree of shape distortion if the cluster structure is unknown."

A natural solution of this problem would be a procedure that combines both the cluster analysis and alignment steps. We applied and validated the performance of several such existing clustering and time alignment methods on our sleep curves data; however, we observed unsatisfactory performance.

Therefore, to tackle the problem of a proper time alignment and clustering of strongly heterogeneous sleep curves, we proposed and validated a novel 2-step iterative approach. The method iteratively combines the dynamic time warping (DTW) [15] based clustering step with the curve alignment step separately applied to each cluster. The flexibility of the approach is increased by the possibility to apply a preferred curve alignment method in the alignment step that operates on the formed clusters. For datasets with many heterogeneous profiles we recommend to use methods with restrictions to alignment, for example [16,17]. The approaches described in [16,18,19] are appropriate for data observed over the same time interval or when the misalignment has a nonlinear character. When the curves are also allowed to be shifted beyond the observation interval or the misalignment is close to linear, then alignment techniques from [20] and [21] performed well.

On generated functional data that mimic the sleep process we carried out a series of comparisons when (i) clustering and alignment steps are applied separately, and (ii) simultaneous curve alignment and clustering is applied. This validation study confirmed the good performance of the 2-step iterative approach, which successfully overcomes the problem of curve alignment in the case of a dataset with many heterogeneous and complex profiles.

The description and validation of this novel clustering and alignment method represents the additional contribution of the paper.

Finally, returning to our major study objective, we applied the proposed 2-step approach to sleep data with the aim to detect associations between different profiles of sleep probabilistic curves and daily life measures.

The major findings include the observed relationship between the improved subjective feeling scores in the morning and the increased probability for the rapid-eye movement (*REM*) related sleep microstates and the decreased probability of awakenings during the night. Moreover, *Wake* and *S2* related sleep microstates resembled the clustering structure of subjects with significantly different scores of cognitive tests focused on fine-motor activity, concentration and attention. The results we obtained are in line with the conclusions of the existing studies. However, the major advantage of the sleep probabilistic curves analysis is a more thorough insight into the changes of sleep dynamics over the whole period of night. Such a representation cannot be obtained from approaches operating on measures that compress all the sleep dynamics information into a single value. This makes the proposed sleep study methodology unique.

The structure of the paper is as follows. Section 2 introduces the sleep dataset used in this study. Section 3 gives a brief overview of the probabilistic sleep model [12] followed by an introduction to the curve misalignment problem. Then, the methodology for curve alignment and clustering that leads to the proposed 2-step iterative approach is described in detail. In Section 4.1 the proposed iterative method is validated and compared with the existing approaches on the generated data that mimic the character of the sleep dataset but have known cluster memberships. In Section 4.2 the 2-step approach is applied to the sleep dataset. Section 5 provides the discussion and concluding remarks. A detailed description of the DTW method and three chosen curve alignment approaches are included in Appendices A and B. Finally, Appendix C provides details of the simulation study in Section 4.1.

2. Data

2.1. Sleep dataset

Polysomnographic (PSG) recordings of 146 healthy subjects (66 men and 80 women; average age 51 years) from the European Sleep Database SIESTA [22] were used in this study. The subjects had no serious sleep problems and they spent two consecutive nights in the sleep lab. To avoid possible misinterpretation of the results due to the first night effect, only data recorded during the second night are considered in this study.

The PSG measurement started at the subject's usual bedtime after switching the lights off and terminated at their usual time of getting up in the morning.

The electroencephalogram (EEG) signal was measured by three pairs of electrodes – frontal (Fp1-M2, Fp2-M1), central (C3-M2, C4-M1) and occipital (O1-M2, O2-M1). The reference electrodes M1 and M2 were placed on the mastoid and their average was used for referencing all EEG channels. Two electrodes for monitoring the submental electromyogram (EMG) were used and placed above the chin of a subject with the reference electrode placed below the chin. For more details about other types of the recorded PSG signals see [22].

EEG data were band-pass filtered into the frequency range from 0.4 to 40 Hz by using the Butterworth filter of order 8 and down-sampled to 100 Hz. The artefact detection procedure of the Somnolyzer24×7 was applied for detecting eye, muscle, sweat and EEG amplitude related artefacts [23]. The detected artefacts were excluded from the analysis.

2.2. Daily measures

After awakening, the subjects filled out several questionnaires that scored their sleep and awakening quality. Subjective sleep and awakening quality were assessed in the morning utilising a standardised Self-rating Scale (SSA) [24]. The SSA consists of 20 items that yield in three sub-scores (sleep quality, awakening quality and somatic

Table 1
The list of daily measures and their abbreviations used in the article.

Daily measure	Abbreviation
Self-rating Questionnaire for	
Sleep quality	*s_qua*
Awakening quality	*a_qua*
Somatic Complaints	*s_com*
Visual Analogue Scale test for	
Drive	*drive*
Mood	*mood*
Affectivity	*aff*
Drowsiness	*drows*
Well-being Self Assessment Scale	
morning/evening	*wb_m, wb_e*
Diastolic Blood Pressure	
morning/evening	*dia_m, dia_e*
Pulse rate	
morning/evening	*pul_m, pul_e*
Systolic Blood Pressure	
morning/evening	*sys_m, sys_e*
Alphabetical cross-out test	
Total score	*ad_ts*
Variability	*ad_sv*
% of errors	*errp*
Fine motor activity test	
right/left hand	*fma_r, fma_l*
Numerical Memory Test	*num_m*

complaints). Four 100 mm visual analogue scales [25] for *drive*, *mood*, *affectivity* and *drowsiness* were also used. The self-assessment questionnaire of well-being [26] consisting of 28 items was filled out by the subjects in the evening and morning sessions.

Moreover, the subjects performed several neuropsychological tests for an assessment of attention, attention variability, concentration, short-term memory and fine motor activity [27]. Finally, the evening blood pressure and pulse values were recorded in less than two hours before bedtime and in the morning after sleep. The list of all daily measures used in the article and their abbreviations can be found in Table 1.

The daily measures were preprocessed in the following way. First, they were transformed so that low values indicated good sleep. The results from several neuropsychological tests and physiological factors were significantly correlated with age. The age effect was adjusted by subtracting the nth order polynomial curve ($n = 1, 2, 3$) fitted to the data with the least squares approach.

Moreover, following the research lines presented in our previous work [4], we considered three factor scores – factor of subjectively scored sleep and awakening quality (*FA1*), physiological factor (*FA2*) and neuropsychological factor (*FA3*) - that were obtained as the dominant factors after applying factor analysis to the set of daily measures listed in Table 1.

3. Methods

3.1. The probabilistic sleep model (PSM) and sleep probabilistic curves

The PSM [12] characterises the sleep process by probability values of a finite number of sleep states, called sleep microstates. Lewandowski, Rosipal and Dorffner [12] determined the number of microstates as 20.

The original PSM is based on the EEG signal recorded from the central pairs of electrodes [12]. In this study we extended the model by using the EEG signal from three pairs of symmetrically placed left and right hemisphere electrodes in the frontal, central and occipital regions. In addition the chin EMG signal was used with the aim to better discriminate the wakefulness and *REM* sleep stages.

The model training and estimation of the probability values of sleep microstates were performed in the same way as described in [12]. To improve the physiological interpretation of the extracted sleep microstates, the probability values that represent a link between each sleep microstate and the standard sleep stages *Wake, S1, S2, Slow Wave Sleep (SWS, stages S3* and *S4)* and *REM* were estimated [2,12].

By considering the probability values of a given sleep microstate as a function of time, the sleep probabilistic curve is obtained (Fig. 1). These probabilistic curves then represent the subjects' overnight sleep profiles.

Before analysis, the extracted sleep probabilistic curves were preprocessed in the following way. The starting point of each curve was set to the sleep latency defined as the first occurrence of the three consecutive periods of the *S1* sleep stage or the first period of the *S2* sleep stage, whichever comes first. Within each sleep microstate separately, the sleep probabilistic curves were smoothed by functional principal component analysis with the smoothing covariance surface approach [28] (Fig. 1). An important by-product of the method is that the profiles of the smoothed curves are predicted at the end of night and therefore the curves are defined over the same time interval.

3.2. Curve cluster analysis and time alignment

The goal of functional data cluster analysis is to detect subgroups or clusters of curves such that the within-cluster similarity is as high as possible and the curves from different cluster are as dissimilar as possible [13,29]. Standard clustering methods such as k-means, k-medoids or hierarchical clustering can be applied to functional data by replacing the Euclidean distance between vectors by an appropriate distance measure between the curves [30]. However, the majority of clustering techniques encounter problems when the time misalignment is present between curves (Fig. 2, left).

Let us consider a pair of curves X_1, X_2 defined over a common time interval T. Without loss of generality we assume $T = [0, 1]$. The assumption of smoothness and differentiability of X_1, X_2 over the whole interval T is necessary for the majority of methods that are considered in this article.

To temporally align (or register) the pair of curves X_1, X_2 means to find a warping function $h^\star$: $T \to T$ from a set H of all strictly increasing bijective functions defined on the interval T such that

$$h^\star \in \operatorname{argmin}_{h \in H} S(X_1, X_2, h)$$

under the condition of a common start and end point

$$h^\star(0) = 0 \quad \text{and} \quad h^\star(1) = 1$$

and where S is an arbitrary curve distance criterion. The area under the squared difference of aligned curves

$$S(X_1, X_2, h) = \int_T (X_1(t) - (X_2 \circ h)(t))^2 \mathrm{d}t \tag{1}$$

is one such criterion that is often used [18,16,31]. The operator $\circ$ represents the composition of two functions

$$(X \circ h)(t) = X(h(t)), \quad t \in T.$$

An example of a pair of smoothed sleep probabilistic curves, their aligned version using the criterion (1), and the estimated warping function is depicted in Fig. 2.

3.3. Existing methods for curve clustering and alignment

Two general approaches exist for clustering misaligned curves in time:

(A) the time alignment of the whole set of curves preceding the clustering step,
(B) simultaneous time alignment and clustering.

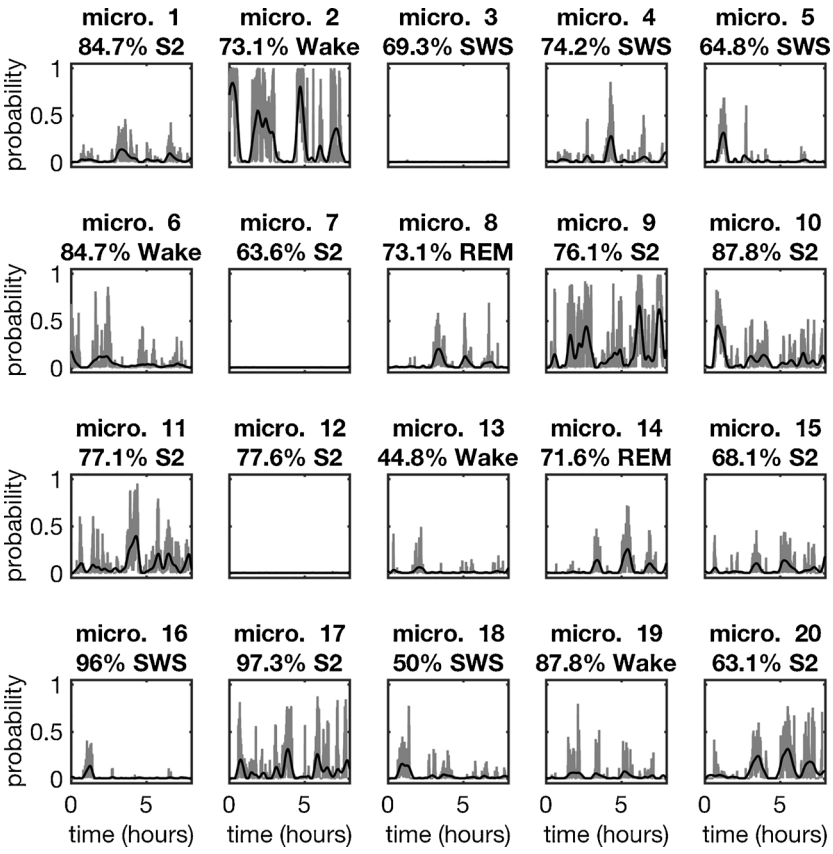

Fig. 1. An example of the sleep probabilistic curves (grey) for 20 microstates of a 41-year-old healthy woman. The smoothed version of the sleep probabilistic curves is depicted in black. In the title of each subfigure the leading probability value (transformed to percentage) representing a link between the sleep microstate and the standard sleep stages is listed.

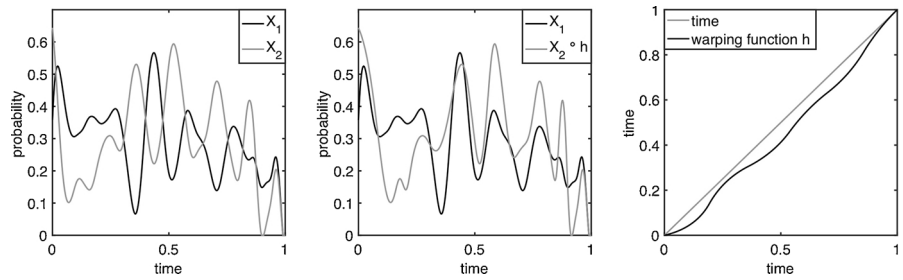

Fig. 2. An example of two smoothed sleep probabilistic curves (left plot), their time aligned representation (middle plot) and the corresponding warping function (right plot). The original time was transformed into the interval [0,1].

(A) The representative methods belonging to the first approach are (i) self-modelling time warping (SMTW) [18], (ii) pairwise curve synchronisation (PCS) [16], and (iii) elastic time warping (ETW) [19]. SMTW and ETW align a set of curves to one target curve. The PCS method aligns each pair of curves separately. The technical details of these methods are summarised in Appendix B.

In Section 3.2 the assumption of a strictly increasing warping function was stressed. In its original version, the SMTW method does not guarantee that the estimated warping function is strictly increasing, only nondecreasing. Gervini and Gasser [18] solved this problem by repeating several steps of the algorithm until the condition of a strictly increasing warping function is met.[2] However, considering the real-world sleep data or simulated data presented in the next section, we observed that the estimated warping function did not become strictly monotonic. This occurred although we set the number of repetitions to 1000 or more to evaluate monotonicity.

To solve this problem, we introduced a penalty to the first derivative of the warping function close to zero (Appendix B). Moreover, to avoid the alignment of too distant segments of two curves we also considered a restriction to the distance between the warping function and the real time. Thus the SMTW method with restriction is denoted as rSMTW. A similar restriction is included in the PCS method and can also be added to the ETW algorithm; denoted rETW.

The three discussed methods align all curves at once and then curves are clustered by the k-means or another preferred clustering method. However, when curves with heterogeneous profiles are present, time alignment applied to a whole set of curves usually fails or leads to unsatisfactory results. This problem is depicted in Fig. 3 and is also demonstrated in Appendix C.

(B) Three approaches that simultaneously combine clustering and time alignment were proposed in the literature and are discussed in this study. These are (i) the k-mean alignment for curve clustering (KMACC) method [21], (ii) the joint probabilistic curve clustering and alignment (JPCCA) [20], and (iii) the truncated version of PCS (tPCS) followed by the k-means clustering [14].

KMACC and JPCCA assume a linear transformation of time

$$h(t) = \mathrm{ct} + b, \quad c > 0,\ b \in \mathbb{R},\ t \in T$$

[2] See step (a) – updating warping function, part (ii) of the SMTW algorithm in [18] or [32].

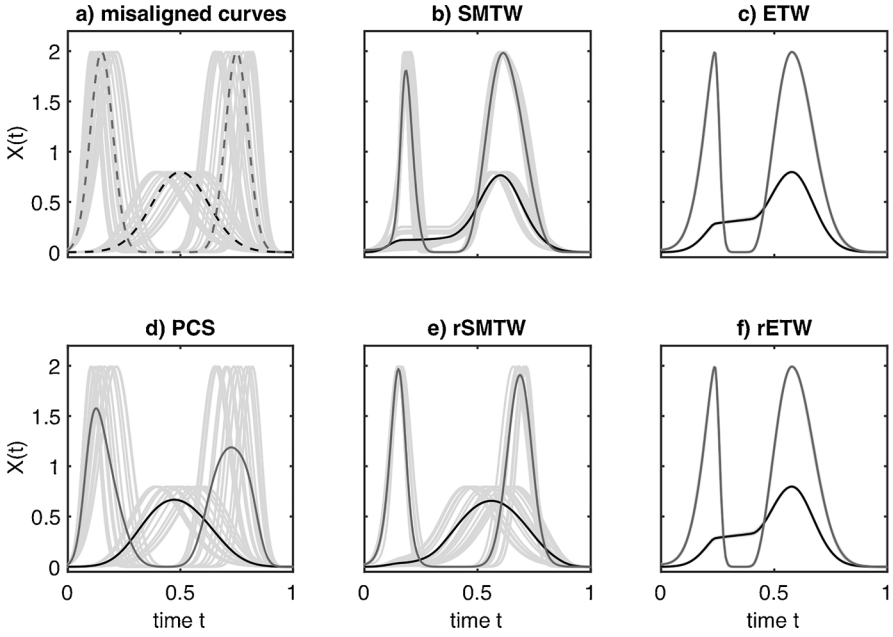

Fig. 3. When a dataset includes misaligned curves (plot a), grey) that follow two different profiles (plot a), dashed lines), the alignment of all curves simultaneously leads either (i) to unsatisfactory results, when the curves are only slightly transformed in time; that is, the results of the PCS (pairwise curve synchronisation) and rSMTW (self-modelling time warping with restriction) methods or (ii) to rapid distortions of several curve profiles; that is, the results of SMTW and elastic time warping with (rETW) and without restriction (ETW). The average of the aligned curves for each of the profiles is colour-coded.

when solving the curve misalignment problem. This limits the flexibility of the methods to address situations where a nonlinear transformation of time is needed. In addition, in our study we consider the same time interval for all sleep curves and therefore the only possible choices for the constants c and b are $c = 1$ and $b = 0$, effectively producing no alignment. Other values of c or b would annul the property of the common time interval. Therefore these two methods were not further considered in this study.

The only method that belongs to this category and is applicable to sleep curves is tPCS. In contrast to PCS, the tPCS method takes into account only pairs of the most similar curves when computing the global warping function.

3.4. 2-step iterative approach for clustering and alignment

To address the problems of the existing methods for combined clustering and alignment and to introduce an algorithm with a higher flexibility of algorithmic choices in the alignment step, we propose a 2-step approach that iteratively combines the clustering and within clusters alignment steps, more specifically

- *Initial clustering:* Because at the beginning the curves are misaligned, the standard k-means or k-medoids algorithm does not lead to reasonable results. Therefore, in the initial clustering step we apply the dynamic time warping method (DTW) with the aim to obtain the distance matrix M_{dtw} (Eq. (A.1) in Appendix A), and then to apply the k-medoids clustering operating on M_{dtw}.
- kth *step:* The misaligned curves are aligned separately in each cluster. In practice, one of the above mentioned three algorithms rSMTW, PCS or rETW can be used. In this study, we used only the curve alignment methods with restriction to the warping function (Section 3.3). According to the choice of the curve alignment method, the algorithm is denoted as 2DTW-rSMTW, 2DTW-PCS or 2DTW-rETW.
- $(k+1)$th *step:* Re-clustering of the aligned curves using the same clustering approach as in the initial step. With the aim to evaluate how well the curves are clustered and aligned, the following L-criterion was considered in this study

$$L = \frac{1}{N}\sum_{i=1}^{K}\sum_{j:X_j\in C_i}\int_T ((X_j\circ h_j)(t) - \mu_i(t))^2 \mathrm{d}t \qquad (2)$$

$$\mu_i(t) = \frac{1}{|C_i|}\sum_{j:X_j\in C_i}(X_j\circ h_j)(t)$$

The criterion is a modified version of the criterion used in the standard k-means algorithm [33] and represents a cumulative average of the squared differences between the aligned curves $X_1\circ h_1$, ...,$X_N\circ h_N$ and clusters C_1, ...,C_K representatives μ_1, ...,μ_K, defined as the cluster means. $|C_i|$ is the cardinality of the cluster C_i.

These two steps, clustering and alignment, are iteratively repeated until one of the following stopping criteria is met

- the number of iterations exceeds a given threshold; in this study set to 100,
- the L-criterion (2) is lower than a given small constant,
- clusters in the kth, $(k-1)$th and $(k-2)$th steps do not change.

The chosen stopping criteria mimic those used in the standard clustering techniques [33,34]. Finally, the cluster membership and aligned curves that belong to the iteration step with the smallest L-criterion are considered as the final result. A graphical schema of the proposed approach is depicted in Fig. 4.

An important property of the method is its convergence. We found it difficult to prove theoretical convergence. However, in each iteration of the 2-step algorithm, the curves within the clusters are closer and more similar to each other and therefore the average L-criterion decreases with iterations. In this sense, the convergence is guaranteed. In all experiments of this study, we observed decreasing values of the L-criterion with iterations.

The MATLAB code can be obtained upon request from the corresponding author.

4. Results

4.1. Simulated data

To objectively compare the performance of the existing curve alignment and clustering approaches, we generated 100 datasets that mimic the character of the sleep probabilistic curves, but have a known cluster structure. In each of 100 trials, five template sleep probabilistic

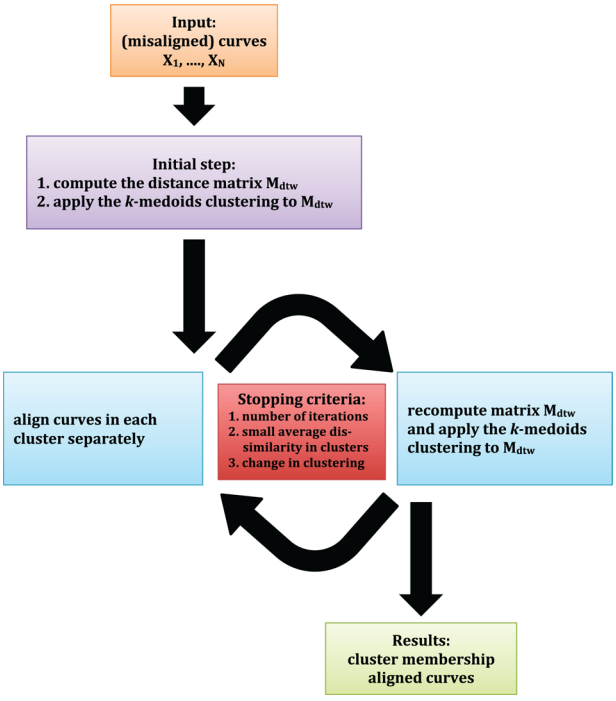

Fig. 4. Schema of the proposed 2-step approach.

curves Y_1, Y_2, Y_3, Y_4, Y_5 were selected at random from a set of all subjects' curves generated by the PSM. By considering these five template curves, we generated a set of 90 curves by the following formula

$$X_i(t) = Y_{U_i}(t) \circ h_i(t), \quad i = 1, ..., 90,$$
$$h_i(t) = c_i g_i^{\star}(t) + d_i,$$

$$g_i^{\star}(t) = \begin{cases} g_i(t), & \text{if } b_i = 1, \\ \frac{g_i(t) - g_i(0)}{g_i(1) - g_i(0)}, & \text{if } b_i = 0, \end{cases}$$
$$g_i(t) = t + \alpha_1 e^{t - \alpha_2},$$

where $b_i \sim \text{Alt}(0.5)$, $\alpha_{1i} \sim \mathcal{N}(1, 0.01^2)$, $\alpha_{2i} \sim \mathcal{N}(0, 4)$ and c_i, d_i are normalising constants that guarantee $h_i(0) = 0$ and $h_i(1) = 1$ and $U_i \in \{1, 2, 3, 4, 5\}$. For each of the first two template curves we generated 10 curves, for the templates Y_3 and Y_4 we generated 20 curves and finally for Y_5 we generated 30 curves. An example of the dataset and original template curves is depicted in Fig. 5. To ensure the reproducibility of our results, the generated data and analysis results can be obtained upon request from the corresponding author.

Using the generated curves we validated and mutually compared the performance of (i) the SMTW, rSMTW, PCS, ETW and rETW methods which align the whole set of curves before the k-means clustering step, (ii) the tPCS method which simultaneously clusters and aligns the curves, (iii) the three versions of the 2-step approach; 2DTW-rSMTW, 2DTW-PCS and 2DTW-rETW, (iv) the k-means clustering of the misaligned curves which serves as a reference method. We used the following criteria to evaluate the clustering and time alignment performance

- the proportion of curves assigned into the true cluster (quality of clustering) measured by the Rand index $R \in [0, 1]$ [35], and
- the quality of alignment evaluated by the L-criterion (2) and the average silhouette (AS) [36]. The silhouette represents the tightness and separation of each cluster. It has values in the interval $[-1, 1]$ and shows how well a curve lies within the cluster and which curves lay in between clusters. The closer AS is to the value 1, the more compact clusters are formed.

4.1.1. Summary of the results

Next, the summary of the main results is provided, and the details can be found in Appendix C.

First, as expected, we observed that applying the curve alignment to the whole dataset before clustering produces inferior results. However, the simultaneous curve alignment and clustering represented by tPCS also did not improve this inferior performance. Qualitatively, this was reflected by non-significant differences in the Rand indices produced by the k-means clustering of misaligned curves and the curve alignment preceding the clustering step or tPCS. The quality of alignment was also low.

In contrast, the ETW without restriction to the warping time that preceded the clustering step, the 2DTW-rSMTW and the 2DTW-rETW methods produced nearly true clustering and satisfactory alignment. However the ETW method aligned curves with different profiles to one target curve and therefore it shifted the curves too far in time. Consequently, this may lead to an incorrect interpretation of the clustered curves, as graphically demonstrated in Appendix C. The problem with the interpretation of the aligned curves was not observed in the case of 2DTW-rETW and 2DTW-rSMTW, favouring the proposed 2-step approach.

We conclude that on non-trivial generated data that mimic the character of the sleep probabilistic curves the iterative combination of clustering and curve alignment either with rETW or rSMTW, that is

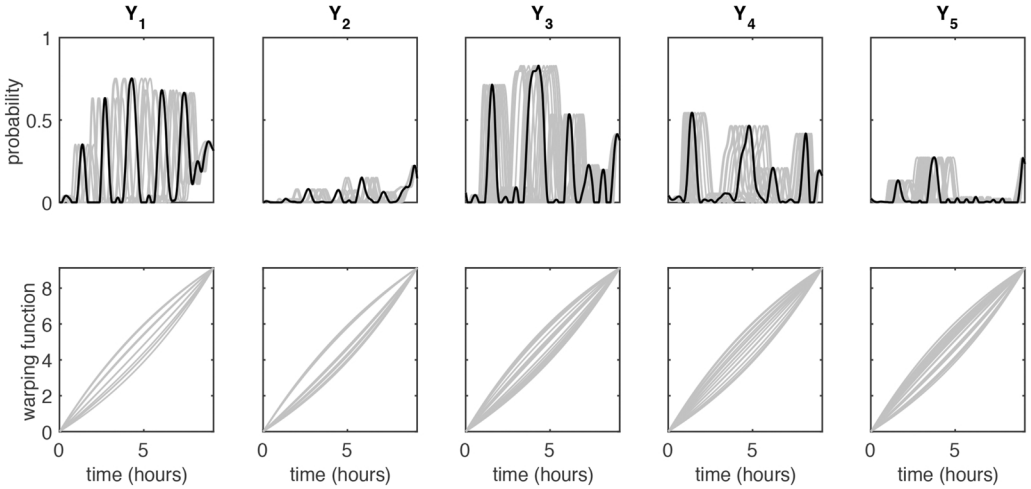

Fig. 5. Example of the generated dataset (top, grey) that mimics the character of the sleep probabilistic curves. The original template curves (top, black) were misaligned in time by using randomly generated time transformations (bottom).

Table 2
List of daily measures in which significant differences between formed clusters were detected.

	k-means	2DTW-rSMTW	2DTW-PCS	2DTW-rETW
Microstate 16 (8 clusters)	*FA2, age*	*FA2, age, wb_m, ad_ts*	*FA2, age, drows*	*FA2, age, drows*
Microstate 8 (9 clusters)	*pul_e, dia_m, dia_e*	*pul_e, dia_e, FA1,FA2, age, num_m drive, mood, aff, drows*	*pul_e, dia_m, age*	*pul_e, pul_m*
Microstate 14 (3 clusters)	*s_com, drive*	*s_com, drive, drows*	*s_com, drows, s_qua, FA2, dia_e*	*s_com, drows, s_qua, age*
Microstate 1 (2 clusters)	*ad_sv, aff*	*ad_sv, aff errp, mood*	*ad_sv*	–
Microstate 6 (3 clusters)	*FA1, aff, drows*	*FA1, aff, mood*	– *aff, mood*	– – *sys_m*
Microstate 13 (4 clusters)	*FA3, age, s_qua, drows*	*FA3, age, s_qua, drows, s_com, drive, FA2, sys_e, ad_sv, errp fma_l, fma_r*	*FA3, age, s_qua, drows*	*FA3, age, s_qua, drows, s_com, drive, FA2, pul_e*
Microstate 19 (2 clusters)	*age*	*age, drive*	*age, aff, FA2*	*age, FA2, fma_r, s_qua, s_com*

2DTW-rETW and 2DTW-rSMTW, outperformed

- approaches in which the curve alignment precedes the clustering step,
- the tPCS approach for simultaneous curve alignment and clustering.

4.2. Sleep data

PSM models the sleep process by considering a set of 20 sleep microstates. Considering each of these microstates separately, the sleep probabilistic curves of 146 subjects were aligned and clustered by the 2-step approach with rSMTW or rETW algorithms. Due to the strong individuality that can be present in each subject's curve, choosing a proper number of clusters is a challenging task. Kodinariya and Makwana [37] present an overview of criteria for choosing an appropriate number of clusters k in multivariate vector data. Following this work, in this study we used the

- *L-criterion* [37], where the L-criterion (2) values for the chosen clustering method are plotted as a function of the number of clusters. This dependence is visually inspected and searched for a rapid drop that is followed by a visible plateau. This rapid drop point is then selected for setting the number of clusters k,
- *average silhouette* [36], where the AS values are plotted as a function of the number of clusters for a chosen clustering method. Recall that the closer AS is to the value 1, the more compact clusters are formed. The number of clusters k is then set to the point where a rapid increment in the AS values that is followed by approximately constant values is observed.

After the alignment and clustering of sleep curves, we investigated whether differences in the sleep profiles between clusters are also mirrored in differences in daily measures. That is, cluster membership was used to split each daily measure into corresponding subgroups and the non-parametric Kruskal–Wallis test was used to test the significance of the formed subgroup differences.

Recall that during the training process of the PSM, a link of each sleep microstate to one of the standard sleep stages *Wake, S1, S2, SWS* and *REM* is estimated. In the next section, we sort our major results following this link.

4.2.1. Microstates similar to SWS

The relationship between the structure of Microstate 16 and age or the physiological factor *FA2* (Table 2) was visible for both in time misaligned and aligned curves and for an arbitrary number of clusters varying between 2 and 20.

Elderly people were assigned into clusters with lower probability values of *SWS* as represented by Microstate 16 (Fig. 6, cluster 7). For these clusters higher values of the physiological factor *FA2* that reflected higher values of the blood pressure were also typical. On the other hand, clusters with clearly visible periods of high probabilities of Microstate 16 were formed by younger people with significantly lower values of the physiological factor *FA2*.

Using the subjects' class membership obtained by 2DTW-PCS or 2DTW-rETW, the Kruskal–Wallis test detected significant differences in the level of *drowsiness* (*drows*) in the morning. The subjects with a clear periodic pattern of Microstate 16 and a decreased amplitude of the local

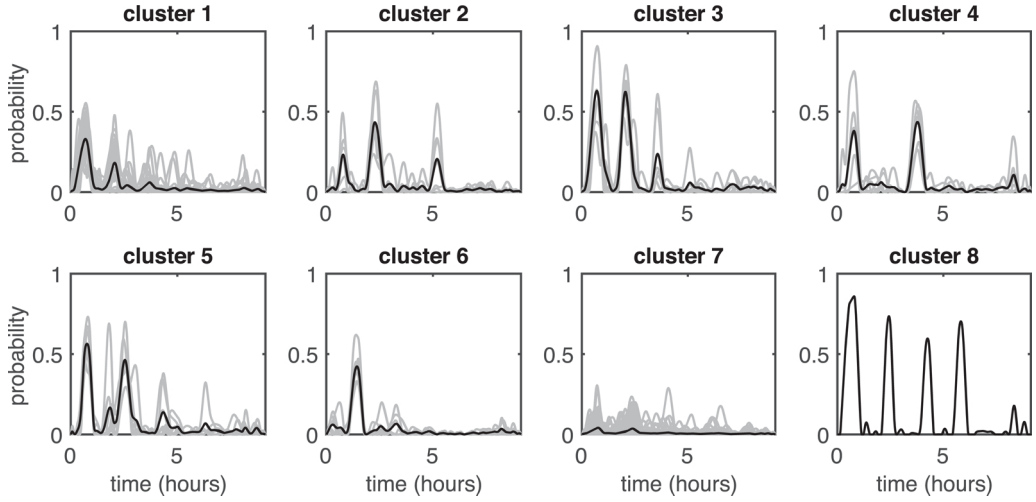

Fig. 6. Microstate 16. Clustering of 146 sleep probabilistic curves into 8 clusters by using the 2-step approach with the rSMTW algorithm used in the alignment step and the k-medoids in the clustering step (2DTW-rSMTW). The cluster representatives (black) were computed as an average of the aligned curves (grey) from the corresponding cluster.

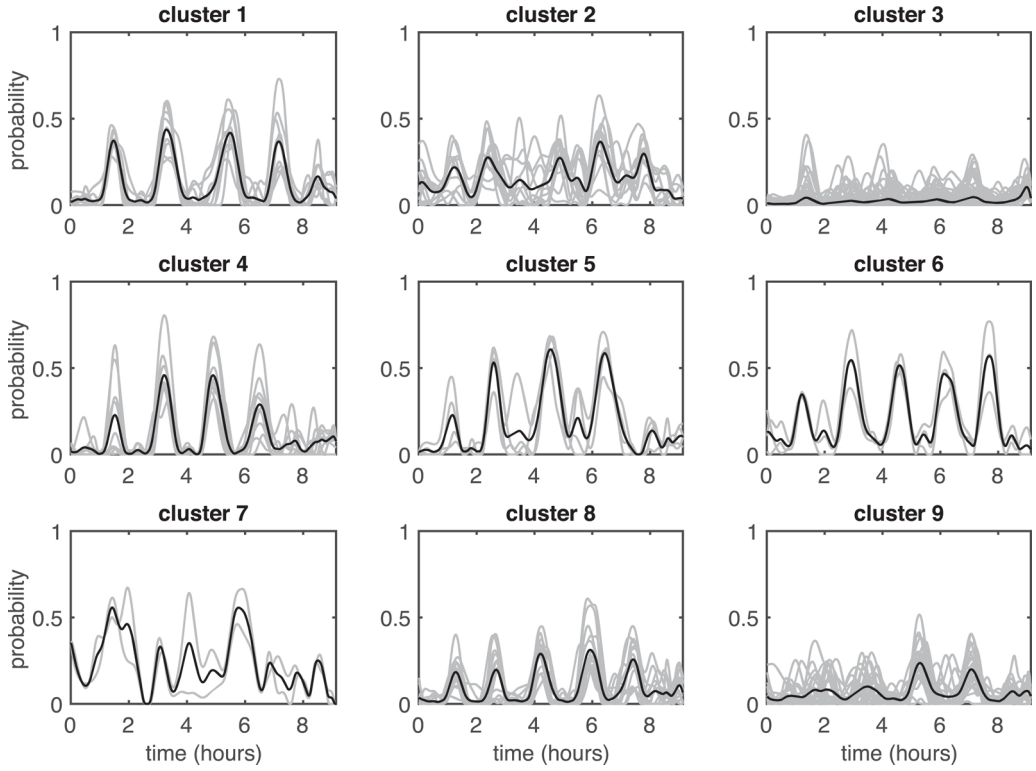

Fig. 7. Microstate 8. Clustering of 146 sleep probabilistic curves into 9 clusters using the 2DTW-rSMTW method. The cluster representatives (black) were computed as an average of the aligned curves (grey) from the corresponding cluster.

maxima of the probability values reported a higher drowsiness in the morning when compared with the subjects with low probability values for the microstate. Moreover, using the 2DTW-rSMTW algorithm, a significant difference in the morning *well-being* (*wb_m*) was detected between clusters 4 and 5 (Fig. 6).

4.2.2. Microstates related to REM

By considering the *k*-means clustering of misaligned curves, we observed that the increased probability of Microstate 8 (73% *REM*) is associated with the increased value of the diastolic blood pressure, both in the morning (*dia_m*) and in the evening (*dia_e*), and the higher pulse rate in the evening (*pul_e*). This observation is consistent with Rosipal, Lewandowski and Dorffner [4] who observed positive correlations between the time spent in the *REM* stage and physiological measures.

After applying the 2-step algorithm, the observed difference in the evening pulse rate between the formed clusters remained, but the significance of the difference in the morning or evening diastolic blood pressure varied with the choice of the alignment method (Table 2). Using the 2DTW-rSMTW approach, we detected a significant difference in the values of the physiological factor *FA2* (which represents the systolic and diastolic blood pressure both in the morning and in the evening) between clusters 2, 8, and 9 (Fig. 7). For subjects assigned into cluster 2, higher values of *FA2* together with an increased probability of Microstate 8 are typical in comparison to the subjects from cluster 8 or 9.

Using 2DTW-rSMTW, significant differences in the level of *drive, mood, affectivity* (*aff*) and *drowsiness* (*drows*) were also observed. Subjects showing a clearly visible oscillatory pattern of high and low probability values of Microstate 8 (cluster 4 in Figure 7) achieved lower values of *drive, mood, affectivity, drowsiness* or the factor of subjectively scored sleep and awakening quality *FA1* in the morning, indicating their better subjective feeling in contrast to the subjects with low probability values of the microstate (clusters 8, 9 in Fig. 7).

Considering the second *REM*-related Microstate 14 (72% *REM*), a strong relationship with the subjective assessment of the somatic complaints (*s_com*) and the level of *drive* or *drowsiness* in the morning was visible for both aligned and misaligned curves (Fig. 8).

The clusters formed by the 2DTW-rETW method also significantly differ in the subjectively scored sleep quality (*s_qua*). This effect was not observed using the *k*-means clustering of the misaligned curves. Similarly to Microstate 8, the increased probability for Microstate 14 is related to improved sleep quality and subjective feelings in the morning. Moreover, the subjects assigned into the cluster with higher probability values of Microstate 14 showed significantly fewer somatic

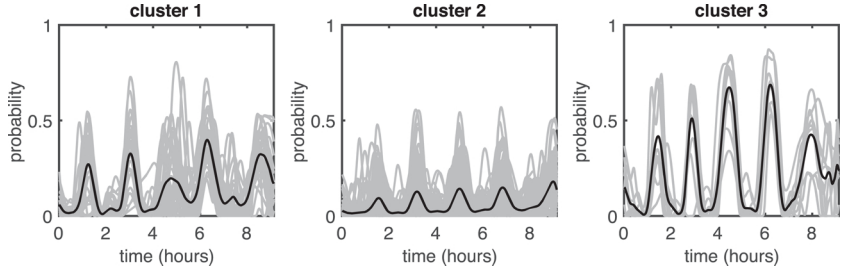

Fig. 8. Microstate 14. Clustering of 146 sleep probabilistic curves into three clusters using the 2DTW-rSMTW method. The cluster representatives (black) were computed as an average of the aligned curves (grey) from the corresponding cluster.

complaints (*s_com*) than the subjects with lower probability values, see for example clusters 1 and 2 in Fig. 8.

4.2.3. Microstates similar to S2

Using the Kruskal–Wallis test, clusters of the sleep probabilistic curves of Microstate 1 (85% *S2*) significantly differed in the attention variability scored by the difference between the extreme scores of the alphabetical cross-out test [27] (*ad_sv*) when misaligned curves were considered. After applying the 2DTW-rSMTW algorithm, this difference remained significant, but in addition, a significant difference in the percentage of errors of the test (*errp*) between formed clusters was detected. A higher variability in attention and an increased percentage of errors were related to an increased probability for Microstate 1.

Clusters formed by the k-means algorithm significantly differed in the level of *affectivity* (*aff*) in the morning. On other hand, clusters formed by 2DTW-rSMTW significantly differed not only in the level of *affectivity* but also of *mood*. We observed that an increased probability for Microstate 1 results into an impairment of *mood* and *affectivity* in the morning.

4.2.4. Microstates related to Wake

Following the results presented in Table 2 we hypothesised that the structure of Microstate 6 (85% *Wake*) influences the subjects' subjective feelings in the morning. Subjects with approximately one hour of wakefulness before the final awakening (cluster 3 in Fig. 9) scored higher values of *mood* or *affectivity* (*aff*) and they felt drowsier in contrast to the subjects from cluster 2 in Fig. 9. A similar phenomenon was also observed for the factor of subjectively scored sleep quality (*FA1*). Considering the 2-step approach, new results were not observed in comparison to the misaligned case. By applying 2DTW-rETW, the existing relationship with *FA1* or *drowsiness* diminished after the alignment (Table 2).

The relationship between the increased probability values of Microstate 13, the sleep microstate representing the boundary between wakefulness and sleep (45% *Wake*, 44% *S1*), and the worst subjectively scored sleep quality (*s_qua*) or increased *drowsiness* in the morning was strong and it was visible for the misaligned curves as well as for the curves aligned in time. Moreover, the subjects above 60 years of age had higher probability values for Microstate 13 in contrast to the younger people less than 40 years of age. On other hand, the better performance in the cognitive tests represented by the neurophysiological factor (*FA3*) was typical for clusters with lower probability values of Microstate 13. After applying the 2DTW-rSMTW algorithm, a similar relationship was also observed by considering the individual cognitive tests (*fma_l, fma_r, ad_sv, errp*).

2DTW-rSMTW and 2DTW-rETW formed clusters that significantly differed in the level of *drive* in the morning, as well as in the physiological factor *FA2* and evening pulse rate (*pul_e*) (Table 2). Moreover, the Kruskal–Wallis test detected significant differences for several cognitive tests (*fma_l, fma_r, ad_sv, errp*) (Table 2) between clusters formed by the 2DTW-rSMTW algorithm. Subjects with the periods of increased probability of Microstate 13 reached lower scores in the tests when compared to the subjects with increased probability of this microstate towards the end of the night.

Microstate 19 is also linked to wakefulness (88% *Wake*). Considering the misaligned curves, the clusters formed by the k-means algorithm significantly differ only in age (Table 2). As expected, an increased probability for this microstate was typical for elderly people above 60 years. After the alignment and clustering produced by the 2-step approach, the formed clusters also differ in daily measures that represent subjective feelings in the morning, *s_qua, s_com* and *drive*. Similarly to Microstates 13 or 6, subjects from the cluster of a higher probability of wakefulness scored their subjective sleep quality and somatic complaints worse than the subjects from clusters associated with low probability values of Microstate 19.

5. Conclusions and discussion

In this study we aimed to identify specific continuous sleep profiles that are significantly associated with daily life behaviour. In contrast to one-dimensional sleep variables, we represent the sleep process through a set of sleep probabilistic curves obtained from the previously proposed probabilistic sleep model [12]. The curves were assigned into subgroups (clusters) according to the similarity in their overall profiles and the subjects' membership to each cluster was obtained. Using the obtained membership values, the significance of differences in daily life measures between formed clusters was tested by the Kruskal–Wallis test.

5.1. Clustering and alignment

However, when the studied curves are misaligned in time, standard clustering techniques without alignment are not able to properly detect curves with similar shape and therefore misalignment can lead to incorrect clustering.

We experimentally observed that approaches where the time alignment precedes the clustering step led either to the rapid distortion of the curve shapes that caused their misinterpretation or to poor alignment. This was already pointed out by Tang and Müller [14] and demonstrated in this article on generated data that mimic the character of sleep probabilistic curves. Moreover, existing methods for simultaneous curve alignment and clustering were either not appropriate for the analysis of sleep curves or they led to unsatisfactory results.

To overcome these problems, we proposed and validated the novel 2-step iterative approach for clustering and time alignment of curves with different profiles. Within this 2-step approach, we proposed to use the DTW based distance matrix for the initial cluster identification and for the k-medoids clustering in each iterative step. We observed that it led to better results in comparison to k-means. In the second step of the proposed approach, the curves of every cluster are separately aligned by the selected alignment method. This provides the flexibility of choosing a preferred alignment method that can be driven by experience, properties of studied curves, etc. The clustering and registration steps are repeated until stopping criteria are met.

We also modified the SMTW method. To guarantee that the warping function of SMTW is strictly increasing, we applied a penalty term to

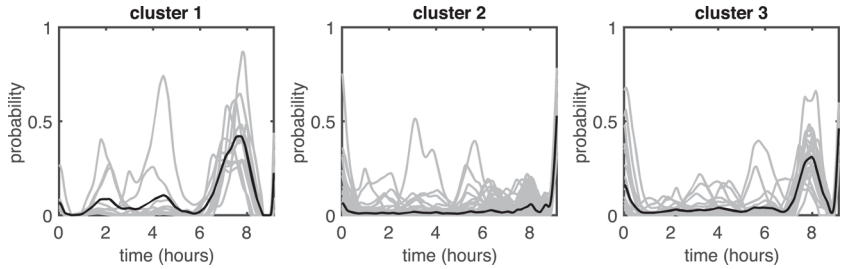

Fig. 9. Microstate 6. Clustering of 146 sleep probabilistic curves into three clusters by the k-means algorithm applied to misaligned curves. The cluster representatives (black) were computed as an average of the aligned curves (grey) from the corresponding cluster.

the slope of the warping function. This penalty avoids the situation in which the first derivate of the warping functions is close to zero. We also applied penalisation to the distance between the warping and real time, which eliminates the strong discrepancy between these two times. This step was motivated by the original PCS method where a similar penalty is incorporated.

We validated the approach on 100 generated datasets that closely mimic the character of the sleep probabilistic curves. In terms of the Rand index (quality of clustering), L-criterion and average silhouette (quality of alignment and clustering), the proposed 2-step iterative combination of clustering and alignment outperformed the tPCS algorithm, as well as the methods where the time alignment is applied to the whole set of curves and precedes the k-means clustering.

Three different time alignment procedures were used and compared within the 2-step clustering and alignment approach. On the same generated dataset, we observed that the 2DTW-PCS version sometimes led to inferior results in comparison to 2DTW-rSMTW and 2DTW-rETW.

5.2. Sleep profiles

Following the experimental findings on the generated data set, we applied the 2-step iterative approach to a set of the probabilistic curves representing the sleep process. In the first step an optimal number of clusters k for each sleep microstate should be chosen. The choice was based on the elbow diagram with the L-criterion (2) and the average *sillhouette*. However due to the strong individuality that can be present in each subject's curve, choosing a proper number of clusters was a challenging task in some microstates. In this case a natural question arises: how does the choice of k affects the results?

When k is small ($K = 2, 3$), the sleep probabilistic curves for a given sleep microstate are assigned into clusters mainly according to their differences in amplitude (curves with low probability values and high probability values). A significant difference in a selected daily measure, denoted here as D, between these clusters means that values of D are associated with the presence/absence of the sleep microstate.

Increasing k leads to clusters that differ not only in amplitude, but also in the other aspects of their profile, for example, the number of periods of higher probability for the *REM* sleep. This means that there are still at least two clusters (with low and high amplitude) for which D significantly differs, but new significant differences may occur between clusters.

However, when k is high, the significant differences in D between clusters may occur only due to the presence of small clusters or clusters with outlier profiles, but they miss the physiological interpretation.

In summary, the relationships between the structure of a sleep microstate and selected daily measures, which were observed for a smaller number of clusters, remain significant, but new significant differences may occur between clusters when the number of clusters increases. However, it is appropriate to set an upper bound to k, for example based on the character of the data, and to check the physiological interpretation of the results. In the case of the 146 sleep probabilistic curves, we therefore restrict k to be at most 20.

Once the number of clusters was fixed, the 2-step approach was applied.

The relationship between the decreased probability for wakefulness and better subjectively scored sleep quality was strong and it was already clearly visible when considering clusters of the original, in time misaligned, sleep curves. This result is consistent with [38,39,7]. Moreover, similar to [40] we observed the age-dependent changes in the sleep patterns, more specifically, the clusters with an increased probability for wakefulness and a decreased probability of *SWS* were formed by elderly subjects above 60 years of age.

The same relationships were observed for sleep structure determined by the 2-step approach. However, and importantly, after applying the 2-step approach, new correlations between the sleep structure and daily measures were found.

First, we observed, that the higher probability values of Microstate 16 (sleep microstate linked to *SWS*) are reflected in the increased drowsiness in the morning. From an initial perspective, this may appear to be an opposite result to our expectation and also to [41,42]. However, it is important to note that Microstate 16 is only one of several microstates linked to *SWS* and does not represent the whole *SWS* process as defined by the R&K rules. Investigating the power spectral density of this microstate, we observed that the microstate can be characterised by a higher level of delta activity when compared to other microstates. We can therefore hypothesise that entering this microstate with a higher probability during the night can cause sleep inertia in the morning, which is reflected by a higher level of reported drowsiness in the morning.

The clustering in time misaligned sleep probabilistic curves of Microstates 8 and 14, both linked to *REM*, resulted only in a visible relationship with the subjective feelings in the morning for Microstate 14. After applying the 2-step approach, we observed that the increased probability of Microstate 8 is related to a better subjectively scored level of *mood, drive, drowsiness* or *affectivity* in the morning. This is consistent with the negative correlation between the time spent in the *REM* stage and the total scores of the SSA questionnaire (lower values of SSA indicate better sleep) observed by Rosipal, Lewandowski and Dorffner [4] or Goelema et al. [7] and the relationship between higher probability values of Microstate 14 (72% *REM*) and better subjectively scored sleep quality (s_qua) or fewer somatic complaints (s_com).

The relationship between increased wakefulness during the night and worse subjective feelings in the morning is well known [38,39,4]. For Microstate 13, both the k-means clustering of misaligned sleep curves as well as the 2-step approach mirrored this expected relationship. However, for Microstate 19 this expected relationship was observed only when the curves were aligned and clustered by 2DTW-rETW.

In Rošťáková and Rosipal [43] we observed that the curve alignment produced by ETW can be counter-productive when predicting daily measures outcomes using the structure of the *Wake* stage. Similar results were observed for Microstate 6 that represented the full wakefulness during the night. Significant differences in *FA1* or the level of *affectivity* between clusters of misaligned curves diminished after the alignment. This indicates that the exact timing of the periods of increased probability of Microstate 6 during the night is the important factor when exploring existing relationships with the considered daily measures.

Finally, we conclude that the sleep probabilistic curves create a promising representation of the sleep process and its structure. In contrast to the one-dimensional sleep characteristics extracted either from the Rechtschaffen and Kales [2] or AASM [3] based sleep hypnogram, the sleep curves represent the temporal dynamics of sleep changes over the whole night. The advanced techniques of mathematical statistics – functional data analysis – can be applied to sleep curves when the goal of detecting relationships between sleep structure and daily life measures is in the focus.

The proposed 2-step approach has proven to be a useful tool in the analysis of sleep probabilistic curves. Recently, we also applied the method to analyse the sleep structure of patients after ischemic stroke. Although preliminary, the results confirm the validity of the approach [44]. Moreover, the observed good performance of the method indicates that it can be successfully used in a wider range of functional data clustering tasks.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgments

The work presented in this paper would not have been possible without the support of the Slovak Research and Development Agency (grants APVV-0668-12 and APVV-16-0202) and the VEGA grant 2/0011/16.

Appendix A. Dynamic time warping as a clustering step

Dynamic time warping (DTW) is a method that was a priori developed and used for aligning curves observed, in general, at different sets of time points [45]. Let us suppose that two curves X_1, X_2 defined on the time interval $T = [0, 1]$ are observed at a finite number of time-points

$$x_1 = \{X_1(t_1), ..., X_1(t_{n_1}),\ 0 = t_1 < t_2 < \cdots < t_{n_1} = 1\},$$
$$x_2 = \{X_2(s_1), ..., X_2(s_{n_2}),\ 0 = s_1 < s_2 < \cdots < s_{n_2} = 1\}$$

It is not required that the sets of time-points $\{t_i\}_{i=1}^{n_1}$ and $\{s_i\}_{i=1}^{n_2}$ are equal.

The main goal of the DTW method is to find the best match between curves X_1 and X_2 by constructing the *warping path*

$$w = \{(i_l, j_l),\ i_l \in \{1, ..., n_1\},\ j_l \in \{1, ..., n_2\},\ l = 1, ..., W_L\}$$

which minimises the cost function

$$Q(X_1, X_2, w) = \sum_{(i_l, j_l) \in w} (X_1(t_{i_l}) - X_2(t_{j_l}))^2$$

where W_L is the length of the warping path w. This problem can be solved by using dynamic programming [15].

There are a few natural constraints for the warping path w of DTW

- $i_{l-1} \leq i_l$ and $j_{l-1} \leq j_l$ for all $l \in \{1, ..., W_L\}$ (monotonicity),
- $i_1 = 1,\ j_1 = 1,\ i_{W_L} = n_1,\ j_{W_L} = n_2$ (common starting and end point),
- $i_l - i_{l-1} \leq 1,\ j_l - j_{l-1} \leq 1$ (continuity).

Finally, there is an optional argument $r > 0$ in the DTW method, which controls the distance between the real time and the warping path

$$|i_l - j_l| \leq r, \quad l = 1, ..., W_L.$$

An interesting by-product of the DTW method is a distance measure between curves X_1, X_2

$$\mathrm{dtw}(X_1, X_2) = \min_w Q(X_1, X_2, w) = \min_w \sum_{(i_l, j_l) \in w} \left(X_1(t_{i_l}) - X_2(t_{j_l})\right)^2,$$

where w represents a warping path. In the case of misaligned curves there is an evident advantage of such a similarity measure which represents a discrete version of the criterion (1).

In our study, we use dtw for constructing a distance matrix $M_{\mathrm{dtw}} \in \mathbb{R}^{N \times N}$ for a set of curves $X_1, ..., X_N$

$$M_{\mathrm{dtw}} = \{\mathrm{dtw}(X_i, X_j)\}_{i,j=1,...,N} \tag{A.1}$$

which can be used in the hierarchical or k-medoids functional data clustering algorithms [30].

Appendix B. Existing curve alignment methods

Self-modelling time warping The SMTW method [18] assumes that a set of curves X_i: $T \to \mathbb{R}$, $i = 1, ..., N$ follows the model

$$X_i(t) = \alpha_i \mu(h_i^{-1}(t)) + \varepsilon_i(t),\ i = 1, ..., N.$$

In other words, each X_i is an α_i-multiple of a common mean function μ (target curve) transformed in time, ε_i represents the error term. The warping functions h_i, $i = 1, ..., N$ are defined as an argument of the minimum of the formula

$$S_{\mathrm{SMTW}}(h_1, ..., h_N) = \sum_{i=1}^{N} \int_T (X_i(h_i(t)) - \alpha_i \mu(t))^2 h_i'(t) \mathrm{dt}. \tag{B.1}$$

However, the method itself does not guarantee that the estimated h_i, $i = 1, ..., N$ are strictly increasing functions, only nondecreasing. Constant segments in the warping function implicate that the first derivative of h is zero within these intervals. To avoid this situation, we add a penalty term to the slope of the warping function h. Now the criterion (B.1) changes to

$$S_{\mathrm{rSMTW}}(h_1, ..., h_N) = S_{\mathrm{SMTW}}(h_1, ..., h_N) + \lambda_1 \sum_{i=1}^{N} \int_T \left(\frac{1}{h_i'(t)} - 1\right)^2 \mathrm{dt} \tag{B.2}$$

where λ_1 is the penalty weight. The penalty helps to avoid alignment of in time too distant segments and therefore restricts close to ideal synchronisation of possibly dissimilar data. This version of SMTW was denoted as rSMTW in the main text.

With the aim to find an optimal penalty weight, we varied λ_1 between 0 and 0.5[3] and for each case the mean squared error MSE of aligned curves

[3] $\lambda_1 > 0.5$ led to poor or no alignment.

$$\mathrm{MSE}(X_1 \circ h_1, ..., X_N \circ h_N) = \frac{1}{N}\sum_{i=1}^{N}\int_T (X_i(h_i(t)) - \bar{X}(t))^2 \mathrm{dt},$$
$$\bar{X}(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(h_i(t))$$

was computed. As expected, the lowest MSE was obtained for $\lambda_1 = 0$. The optimal λ_1[4] was then selected as the value after which the MSE was approximately constant or only slightly increasing.

Pairwise curve synchronisation The PCS method [16] solves the curve misalignment problem by aligning all possible pairs of curves X_i, X_j, $i \neq j$ separately

$$S_{\mathrm{PCS}}(X_i, X_j, h) = \int_T (X_i(h(t)) - X_j(t))^2 \mathrm{dt} + \lambda_2 \int_T (h(t) - t)^2 \mathrm{dt},$$
$$h_{\mathrm{ij}} \in \mathrm{argmin}_h S_{\mathrm{PCS}}(X_i, X_j, h).$$

Then, for each curve X_i, $i = 1, ..., N$ a global warping function h_i is estimated

$$h_i(t) = \left(\frac{1}{N}\sum_{j=1}^{N} h_{\mathrm{ji}}(t)\right)^{-1}, \tag{B.3}$$

where -1 denotes the inverse function. The exact methodology for finding an appropriate value of the penalty parameter λ_2 can be found in [16,46].

Elastic time warping The main idea of the ETW method [19] is the alignment of the *square root slope functions* (SRSFs) $q_{X_1}, ..., q_{X_N}$

$$q_{X_i}(t) = \mathrm{sign}(X_i'(t))\sqrt{|X_i'(t)|}, \quad t \in T,\ i = 1, ..., N.$$

to a target function called the *Karcher mean* q_μ [19], instead of alignment of the original curves. The warping functions $h_1, ..., h_N$ are found by minimising the criterion

$$S_{\mathrm{ETW}}(h_i) = \int_T (q_\mu(t) - (q_{X_i} \circ h_i)(t)\sqrt{h_i'(t)})^2 \mathrm{dt}, \quad i = 1, ..., N. \tag{B.4}$$

In contrast to the other algorithms, criterion (B.4) is symmetric and invariant to a random warping [17] which makes ETW one of the most powerful methods for curve alignment. Alternatively, since there are no restrictions to the distance between the real time and the warping function, ETW may produce close to ideal alignment of possibly dissimilar curves. Fortunately, it is possible to add a penalty to the cost function (B.4), for example in the form

$$S_{\mathrm{rETW}}(h_i) = S_{\mathrm{ETW}}(h_i) + \lambda_3 \int_T (1 - \sqrt{h_i'(t)})^2, \quad i = 1, ..., N. \tag{B.5}$$

The optimal value for the penalty parameter λ_3 was selected in a similar way as in the rSMTW algorithm. Other types of penalties can be found in [46,Chapter 8].

Appendix C. Simulated data results

On generated curves with known cluster memberships, defined in Section 4.1, we validated and mutually compared the following methods:

- the k-means clustering with the functional version of the Euclidean distance (the L^2 norm of curves)

  $$d(X, Y) = \int_T (X(t) - Y(t))^2 \mathrm{dt}$$

  applied to misaligned curves. This method serves as a reference allowing us to compare the obtained results with the clustering that operates on in-time misaligned curves.
- the SMTW, rSMTW, PCS, ETW and rETW algorithms applied to the whole dataset and followed by the k-means clustering, more specifically:
  - The PCS method with the automatic selection of the restriction parameter λ_2 as implemented in the PACE package [46].
  - The SMTW method implemented in the MATLAB routine and available on Gervini's webpage [32]. We also modified this SMTW version by adding a penalty term (B.2) to the original MATLAB routine (rSMTW).
  - The restricted rETW and non-restricted ETW method was taken from the R package *fdasrvf* [47], version 1.7.0.[5]

  By using the procedure described in Appendix B, the parameter λ_1 was set to 0.1 for rSMTW and λ_3 varied in the interval $[0.04, 0.08]$ for the rETW method.
- The tPCS algorithm, where each curve was pairwise aligned only with the most similar curves (10% of the whole dataset[6]).
- The 2-step approach with the rSMTW, PCS and rETW algorithms used in the registration step; denoted as 2DTW-rSMTW, 2DTW-PCS and 2DTW-rETW, respectively.

As mentioned in Section 4.1, the quality of clustering was compared with the true clustering by using the Rand index [35], while the quality of alignment among different methods was evaluated by the average silhouette [36] and the L-criterion (2). Significant differences in the performance of methods were measured by the Wilcoxon rank-sum test applied to the Rand indices with the Bonferroni's correction ($\alpha = \frac{0.05}{55} \approx 0.0009$).

C.1 Time alignment of the whole dataset followed by the k-means clustering

Due to different profiles of the template curves (Fig. C.10a), the alignment of the whole dataset did not yield satisfactory results. Specifically,

[4] We would like to highlight that there is no direct relationship between the parameter λ_1 and the warping window r which is mentioned in Appendix A.
[5] Higher versions of the package do not allow restrictions to the distance between the real time and the warping function.
[6] We observed that the higher proportion (>10%) of the most similar curves led to inferior results.

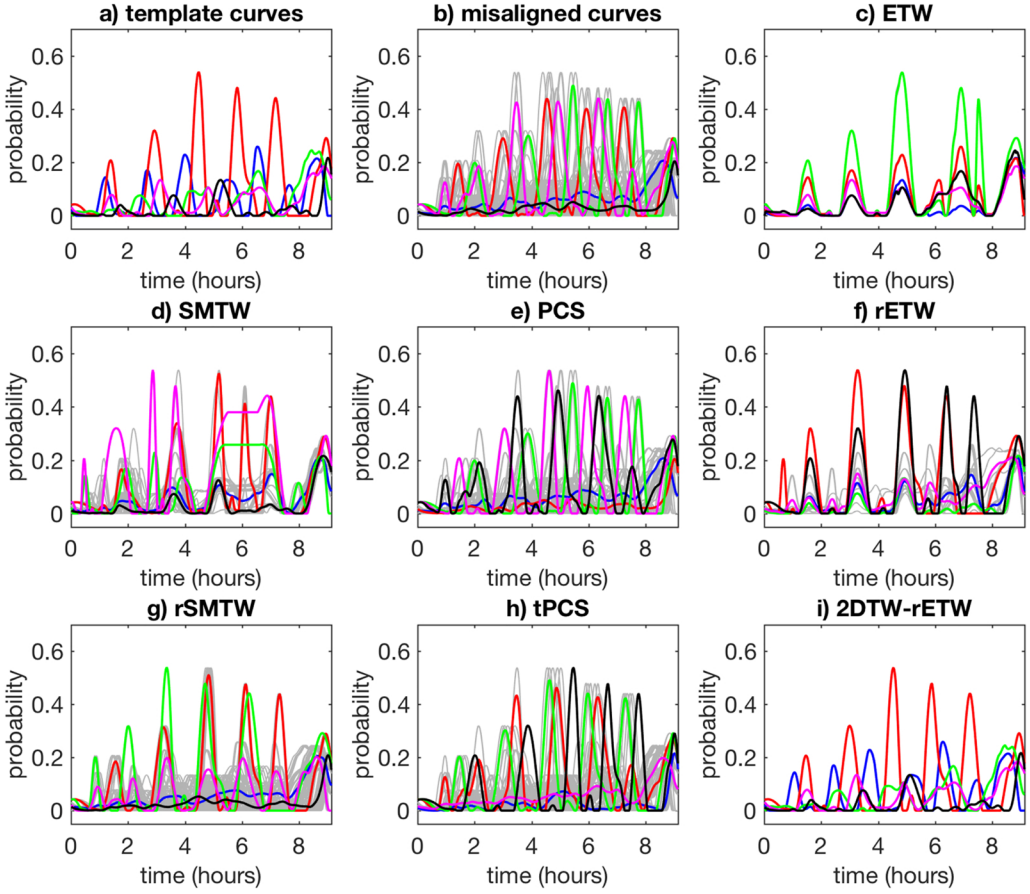

Fig. C.10. An example of the generated dataset (grey) that mimics the character of the sleep probabilistic curves. First, the misaligned curves were assigned into clusters by the k-means algorithm. Then before the k-means clustering the whole dataset was aligned by the (i) pairwise curves synchronisation algorithm (PCS), (ii) self-modelling time warping method (SMTW) and its suggested penalised version (rSMTW) and (iii) elastic time warping with (rETW) and without (ETW) constraints to the distance between the real time and warping function. Finally, the truncated version of the PCS algorithm (tPCS) and the 2-step approach with rETW (2DTW-rETW) were applied. The cluster representatives (average curves) are colour-coded. (For interpretation of the references to colours in this figure legend, the reader is referred to the web version of the article).

- the PCS method aligns each pair of curves separately and then for each curve the global warping function is computed. However, when applied to generated data that mimic the character of the sleep curves, the obtained global warping functions were approximately equal to the real time and therefore the method produced visually poor alignment (Fig. C.10e). The Rand index values ($\bar{R} \approx 0.7$, Table C.3) were not significantly different from the values produced by the k-means clustering of misaligned curves (p-value = 0.88). The quality of alignment was low (Table C.4).
- by considering the original SMTW algorithm, unexpected flat segments occurred in several curves (Fig. C.10d). These flat segments are caused by improper penalisation of nondecreasing warping functions in SMTW and due to the alignment of the curves with different profiles to one target. After applying the above proposed modified version of the SMTW algorithm (rSMTW), the flat segments diminished, but at the cost of poorer alignment of the curves (Fig. C.10g). This is also reflected by the significantly decreased average Rand index (p-value $<10^{-5}$), decreased AS and the increased L-criterion average in comparison to the unconstrained SMTW method (Tables C.3 and C.4).
- the ETW method visually aligns the curves well (Fig. C.10c), which was confirmed also by the average silhouette close to 1 and the L-criterion

Table C.3
The average and median Rand index [35] (R) values for methods used to validate the alignment and clustering performance on 100 generated datasets that mimic the character of the sleep probabilistic curves.

	Average R	Median R	#{R: $R > 0.75$}
k-means	0.70	0.71	34
ETW + k-means	0.999	1.00	100
rETW + k-means	0.83	0.83	88
PCS + k-means	0.70	0.72	32
SMTW + k-means	0.84	0.85	84
rSMTW + k-means	0.78	0.79	70
tPCS	0.73	0.74	47
2DTW-rETW	0.997	1.00	100
2DTW-PCS	0.71	0.72	37
2DTW-rSMTW	0.80	0.81	85

Table C.4
The average silhouette (AS) and the L-criterion (Eq. (2)) values for methods used to validate the alignment and clustering performance on generated data that mimic the character of the sleep probabilistic curves. The values were averaged across all 100 trials.

	Average silhouette	L-criterion
k-means	0.525	0.098
ETW + k-means	0.996	143×10^{-1}
rETW + k-means	0.752	0.022
PCS + k-means	0.556	0.089
SMTW + k-means	0.653	0.035
rSMTW + k-means	0.590	0.052
tPCS	0.566	0.082
2DTW-rETW	0.999	1.19×10^{-1}
2DTW-PCS	0.581	0.079
2DTW-rSMTW	0.805	0.022

average close to 0 (Table C.4). Considering the quality of clustering, on average, only one or two curves were assigned into an incorrect cluster ($\bar{R} \approx 1$). However, the method tries to align curves with different profiles to one target and therefore sometimes important curve elements are shifted too far in time or they are diminished during the alignment process, which may cause misinterpretation of the obtained results. An example is demonstrated in Fig. C.11. The curves in the top plot represent cluster representatives of the original data aligned by the ETW algorithm and the original template curves are depicted in the bottom plot. There are evident differences in the profiles of templates 2 and 3 and the corresponding cluster representatives.
Considering the rETW method with $\lambda = 0.05$ (Fig. C.10f), the method achieved the highest average silhouette and the lowest L-criterion average value among all methods with the restriction to the distance between the real time and warping functions. This result is also confirmed by visually better alignment of curves produced by the rETW method than by the PCS or rSMTW methods. On the other hand, the typical sleep profiles are difficult to detect and on average only 75% of curves were assigned into correct clusters. As expected, in comparison to the non-restricted version of the algorithm, the L-criterion average increased and AS decreased.

We can conclude, that these results underline the need for the approach that combines clustering and curve alignment steps.

C.2 Combination of curve clustering and alignment

In this section, we compare the performance of the tPCS algorithm followed by the k-means clustering with the three versions of the 2-step approach (2DTW-rSMTW, 2DTW-PCS, 2DTW-rETW). We observed that,

- From statistical point of view, the tPCS (Fig. C.10h) produced clustering of the same quality as the simple k-means clustering (Fig. C.10b) or PCS method followed by the k-means clustering (p-values >0.007) but was outperformed by the 2DTW-rSMTW and 2DTW-rETW approaches (p-values $<10^{-12}$). The quality of clustering was measured by the Rand index. This observation was confirmed also when considering the quality of clustering and alignment represented by the AS and L-criterion average (Table C.4).
- The 2DTW-rSMTW version of the 2-step algorithm outperformed or produced similar results in comparison to tPCS and the approaches when the alignment preceded the clustering step. Significantly better results were obtained only by the ETW algorithm followed by the k-means clustering and the 2DTW-rETW method. However, as we demonstrated above, the ETW approach without restrictions shifts the curves too far in time and consequently the correct interpretation of the curves can be lost.
- The variant of the 2-step approach with the PCS algorithm (2DTW-PCS) produces similar results to the tPCS algorithm or PCS followed by the

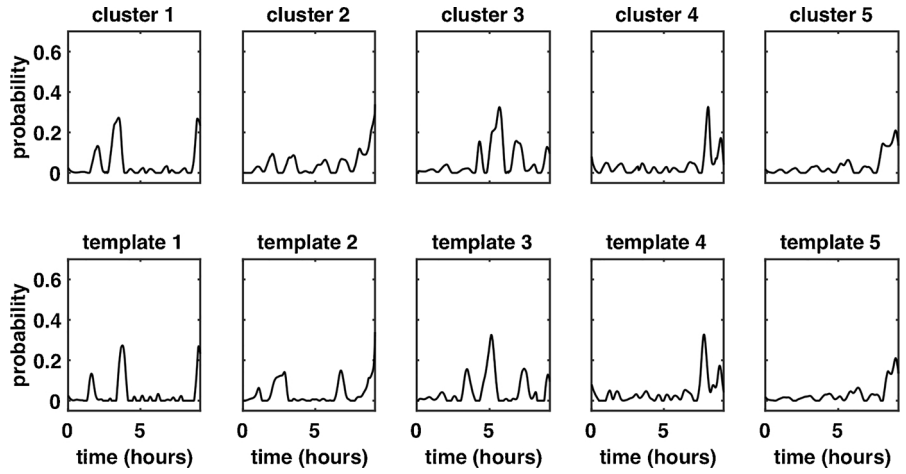

Fig. C.11. Comparison of the cluster representatives (the mean curve of a cluster) of generated data (top row) and the original template curves (bottom row). Before clustering, data were aligned by the elastic time warping algorithm without constraints.

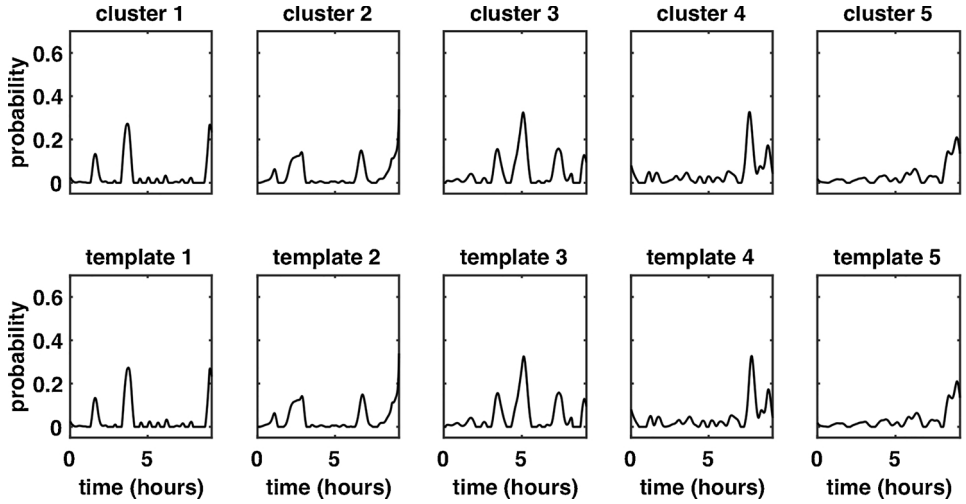

Fig. C.12. Comparison of the cluster representatives (the mean curve of a cluster) of generated data (top row) and the original template curves (bottom row). Data were aligned and clustered by the 2-step approach with the elastic time warping algorithm in the registration step (2DTW-rETW).

k-means clustering. The obtained results are significantly inferior in comparison to 2DTW-rSMTW and 2DTW-rETW (Tables C.3 and C.4).

- the best results were obtained by the 2DTW-rETW algorithm (Fig. C.10i), either when measuring its performance by the Rand index, L-criterion or AS. In contrast to the ETW method preceding the k-means clustering, in the case of 2DTW-rETW the problem with the interpretation of the aligned curves is not present (Fig. C.12).

References

[1] Dement W, Vaughan C. The promise of sleep: the scientific connection between health, happiness, and a good night's sleep. Pan; 2001. URL: https://books.google.sk/books?id=cgF_HAAACAAJ.

[2] Rechtschaffen A, Kales A. A manual of standardized terminology techniques and scoring system for sleep stages of human subjects. Bethesda, MD: U.S. Dept. of Healthy, Education and Welfare; 1968.

[3] 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.

[4] Rosipal R, Lewandowski A, Dorffner G. In search of objective components for sleep quality indexing in normal sleep. Biol Psychol 2013;94(1):210–20. https://doi.org/10.1016/j.biopsycho.2013.05.014.

[5] O'Donnell D, Silva EJ, M anch M, Ronda JM, Wang W, Duffy JF. Comparison of subjective and objective assessments of sleep in healthy older subjects without sleep complaints. J Sleep Res 2009;18(2):254–63. https://doi.org/10.1111/j.1365-2869.2008.00719.x.

[6] Åkerstedt T, Hume K, Minors D, Waterhouse J. The meaning of good sleep: a longitudinal study of polysomnography and subjective sleep quality. J Sleep Res 1994;3(3):152–8. https://doi.org/10.1111/j.1365-2869.1994.tb00122.x.

[7] Goelema M, Leufkens T, Haakma R, Markopoulos P. Determinants of self-reported sleep quality in healthy sleepers and patients. Cogent Psychol 2018;5(1):1–14. https://doi.org/10.1080/23311908.2018.1499197.

[8] Laffan A, Caffo B, Swihart B, Punjabi N. Utility of sleep stage transitions in assessing sleep continuity. Sleep 2010;33(12):1681–6.

[9] Buysse DJ, Hall ML, Strollo PJ, Kamarck TW, Owens J, Lee L, Reis SE, Matthews KA. Relationships between the Pittsburgh sleep quality index (PSQI), Epworth sleepiness scale (ESS), and clinical/polysomnographic measures in a community sample. J Clin Sleep Med 2008;4(6):563–71.

[10] Edinger J, Means MK, Carney C, Krystal A. Psychomotor performance deficits and their relation to prior nights' sleep among individuals with primary insomnia. Sleep 2008;31:599–607.

[11] Kaplan KA, Hirshman J, Hernandez B, Stefanick ML, Hoffman AR, Redline S, Ancoli-Israel S, Stone, Friedman L, Zeitzer JM. When a gold standard isn't so golden: lack of prediction of subjective sleep quality from sleep polysomnography. Biol Psychol 2017;123:37–46. https://doi.org/10.1016/j.biopsycho.2016.11.010.

[12] Lewandowski A, Rosipal R, Dorffner G. Extracting more information from EEG recordings for a better description of sleep. Comput Methods Programs Biomed 2012;108(3):961–72. https://doi.org/10.1016/j.cmpb.2012.05.009.

[13] Ramsay JO, Silverman BW. Functional data analysis. 2nd ed. New York: Springer Series in Statistics; 2005. https://doi.org/10.1007/b98888.

[14] Tang R, Müller HG. Time-synchronized clustering of gene expression trajectories. Biostatistics 2009;10(1):32–45. https://doi.org/10.1093/biostatistics/kxn011.

[15] Wang K, Gasser T. Alignment of curves by dynamic time warping. Ann Stat 1997;25(3):1251–76. https://doi.org/10.1214/aos/1069362747.

[16] Müller HG, Tang R. Pairwise curve synchronisation for functional data. Biometrika 2008;95(4):875–89. https://doi.org/10.1093/biomet/asn047.

[17] Srivastava A, Klassen EP. Functional and shape data analysis. New York: Springer-Verlag; 2016. https://doi.org/10.1007/978-1-4939-4020-2.

[18] Gervini D, Gasser T. Self-modeling warping functions. J R Stat Soc Ser B 2004;66(4):959–71. https://doi.org/10.1111/j.1467-9868.2004.B5582.x.

[19] Tucker JD, Wu W, Srivastava A. Generative models for functional data using phase and amplitude separation. Comput Stat Data Anal 2013;61:50–66. https://doi.org/10.1016/j.csda.2012.12.001.

[20] Gaffney S, Smyth P. Joint probabilistic curve clustering and alignment. Advances in neural information processing systems, vol. 17 2004:473–80. MIT Press.

[21] Sangalli L, Secchi P, Vantini S, Vitelli V. K-mean alignment for curve clustering. Comput Stat Data Anal 2010;54(5):1219–33. https://doi.org/10.1016/j.csda.2009.12.008.

[22] Klösch G, Kemp B, Penzel T, Schlögl A, Rappelsberger P, Trenker E, Gruber G, Zeitlhofer J, Saletu B, Herrmann W, Himanen S, Kunz D, Barbanoj M, Röschke J, Varri A, Dorffner G. The SIESTA project polygraphic and clinical database. Med Biol Mag 2001;20(3):51–7. https://doi.org/10.1109/51.932725.

[23] 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 24x7 utilizing the SIESTA database. Neuropsychobiology 2005;51(3):115–33. https://doi.org/10.1159/000085205.

[24] Saletu B, Kindshofer G, Anderer P, Grünberger J. Short-term sleep laboratory studies with cinolazepam in situational insomnia induced by traffic noise. Int J Clin Pharmacol Res 1987;7(5):407–18.

[25] Aitken RCB. Measurement of feelings using visual analogue scales. Proc R Soc Med 1969;62(10):989–93.

[26] von Zerssen D, Köller D, Rey E. Die befindlichkeitsskala (b-s): Ein einfaches instrument zur objektivierung von befindlichkeitsstoerungen, inbesondere im rahmen von laengsschnittuntersuchungen. Arzneimittelforschung (Drug Res) 1970;20:915–8.

[27] Grünberger J. Psychodiagnostik des Alkoholkranken. Methodischer Beitrag zur Bestimmung der Organizität in der Psychiatrie, für Ärzte, Juristen und Sozialhelfer. Wien: Maudrich; 1977.

[28] Yao F, Müller HG, Clifford AJ, Dueker SR, Follett J, Lin Y, Buchholz BA, Vogel JS. Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. Biometrics 2003;59(3):676–85. https://doi.org/10.1111/1541-0420.00078.

[29] Jacques J, Preda C. Functional data clustering: a survey. Adv Data Anal Classif 2014;8(3):231–55. https://doi.org/10.1007/s11634-013-0158-y.

[30] Montero P, Vilar JA. TSclust: an R package for time series clustering. J Stat Softw 2014;62(1):1–43. https://doi.org/10.18637/jss.v062.i01.

[31] Parodi A, Patriarca M, Sangalli L, Secchi P, Vantini S, Vitelli V. Functional data analysis: K-mean alignment. 2015 [online, accessed 06.04.2016]. https://cran.r-project.org/web/packages/fdakma/fdakma.pdf.

[32] Gervini D. Self-modeling registration. 2004 [Online, accessed 10.03.2015]. http://people.uwm.edu/gervini/software/.

[33] Lloyd SP. Least squares quantization in PCM. IEEE Trans Inf Theory 1982;28:129–37. https://doi.org/10.1109/TIT.1982.1056489.

[34] Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. New York: John Wiley; 1990. https://doi.org/10.1002/9780470316801.

[35] Rand WM. Objective criteria for the evaluation of clustering methods. J Am Stat

Assoc 1971;66(336):846–50. https://doi.org/10.2307/2284239.
[36] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20(1):53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
[37] Kodinariya TM, Makwana PR. Review on determining number of cluster in *k*-means clustering. Int J Adv Res Comput Sci Manage Stud 2013;1(6):90–5.
[38] Baekeland F, Hoy P. Reported vs recorded sleep characteristics. Arch Gen Psychiatry 1971;24(6):548–51. https://doi.org/10.1001/archpsyc.1971.01750120064011.
[39] Hoch C, Reynolds CF, Kupfer D, Berman S, Houck P, Stack J. Empirical note: self report versus recorded sleep in healthy seniors. Psychophysiology 1987;24(3):293–9. https://doi.org/10.1111/j.1469-8986.1987.tb00298.x.
[40] Ohayon M, Carskadon M, Guilleminault C, Vitiello M. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep 2004;27(7):1255–73. https://doi.org/10.1093/sleep/27.7.1255.
[41] Riedel B, Lichstein K. Objective sleep measures and subjective sleep satisfaction: how do older adults with insomnia define a good night's sleep? Psychol Aging 1998;13(1):159–63. https://doi.org/10.1037/0882-7974.13.1.159.
[42] Westerlund A, Lagerros Y, Kecklund G, Axelsson J, Åkerstedt T. Relationships between questionnaire ratings of sleep quality and polysomnography in healthy adults. Behav Sleep Med 2014;10:1–15.
[43] Rošťáková Z, Rosipal R. Time alignment as a necessary step in the analysis of sleep probabilistic curves. Meas Sci Rev 2018;18(1):1–6. https://doi.org/10.1515/msr-2018-0001.
[44] Rošťáková Z, Rosipal R. Relationship between sleep structure of patients after ischemic stroke and daily measures. J Sleep Res 2018;27(Suppl 1):144–5.
[45] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. Readings in speech recognition. Morgan Kaufmann Publishers Inc.; 1990. p. 159–65. https://doi.org/10.1016/b978-0-08-051584-7.50016-4.
[46] Müller HG. PACE: principal analysis by conditional expectation. 2012 [Online, accessed 10.03.2015]. http://www.stat.ucdavis.edu/PACE/.
[47] Tucker JD. Package fdasrvf. 2016 [Online, accessed 11.02.2016]. https://cran.r-project.org/web/packages/fdasrvf/fdasrvf.pdf.