Online resource to

Hidden Markov model segmentation to demarcate trajectories of residual apnoea-hypopnoea index in CPAP-treated sleep apnoea patients to personalize follow-up and prevent treatment failure

Alphanie Midelet[1,2,*], MD, Sébastien Bailly[1,3,*], PharmD, PhD, Renaud Tamisier[1,3], MD, PhD, Jean-Christian Borel[1,4], PhD, Sébastien Baillieul[1,3], MD, PhD, Ronan Le Hy[2], PhD, Marie-Caroline Schaeffer[2], PhD, Jean-Louis Pépin[1,3], MD, PhD

Affiliations:

[1]HP2 Laboratory, Inserm U1300, Grenoble Alpes University, Grenoble, 38000 France

[2]Probayes, Montbonnot-Saint-Martin, France

[3]EFCR Laboratory, Grenoble Alpes University Hospital, Grenoble, 38000, France

[4]AGIR à dom. Homecare charity, Meylan, 38240, France

*The two first authors contributed equally to the work

Corresponding author: Jean-Louis Pépin

1. METHODS SUPPLEMENT

Hidden Markov Models

HMMs model the interaction between two sequences: 1) an observable sequence corresponding to the observed time-series and referred to as the observation sequence which can take its values in $R+$ and 2) an unobservable sequence, the discrete-time Markov chain referred to as the "hidden state" which can take K distinct discrete values, where K is the number of states. In sequences described by a first-order Markov chain, each value depends exclusively on the previous value.

A HMM is fully characterized by the following parameters:

- the transition matrix A, governing the sequence of hidden states, given that the value of z_{t+1} depends exclusively on the current value z_t, the value of the hidden state at time t: $A(t) = (a_{ij})$ with $a_{ij} = P(z_{t+1} = j \vee z_t = i)$
- the vector of probabilities π associated with the value of the first hidden state: $\Pi = [\Pi_1, \ldots, \Pi_i, \ldots, \Pi_k]$ with $\Pi_i = P(z_1 \vee z_i)$
- the state emission density, corresponding to the state-specific probability distribution of the observed variable: $P_t(x_t \vee z_t)$

The HMM is described by the parameter-set $\lambda = A, \pi, b_j$, where$\{b_1, \ldots, b_k\}$are the parameters characterizing the state emission densities.

In the present data set, the observation sequence corresponds to the measured rAHI and the hidden state corresponds to the CPAP treatment efficacy state. From the set of observable sequences *X*, the rAHI time series, the goal is to learn the best parameter-set which maximizes *P(X|λ).* Then, from the sequences of observations and the parameter-set of the HMM, we can estimate the most probable sequence of states by using the Viterbi algorithm.[1]

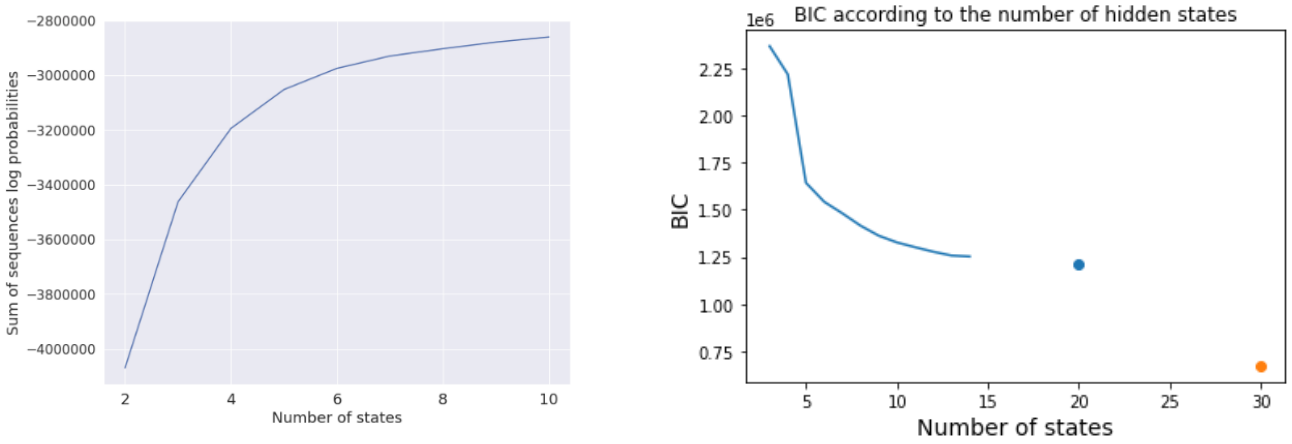
Except in the presence of some missing records, our observations are regularly-sampled temporal data, and the time between consecutive observations is not exponentially distributed, this is why we did not model the sojourn time in each state as a random variable following an exponential distribution with a continuous time HMM (CT-HMM).

HMM - Number of states

The more difficult problem when learning HMMs is determining the number of states K. Too few states cannot take into account occasional states (in which the rAHI is not well controlled) whereas too many states may cause overfitting. In the present context, we ignore how many plateaus and how many levels of variance we can find in the time series of a patient. Also, as the final goal is to cluster the patients, we do not know how much we should discretize the space of states to enable discrimination between relevant time series. In such an unsupervised scenario without any *a priori* information on the number of states, one method consists in fitting HMMs with differing numbers of states and then comparing the goodness-of-fit of the resulting HMMs. Once the model has been built, we can compute the log-probability of each sequence generated by the model, $P(sequence \vee model)$,

using the forward algorithm. To compute this log-probability, we sum over all possible paths instead of just the single most likely path.

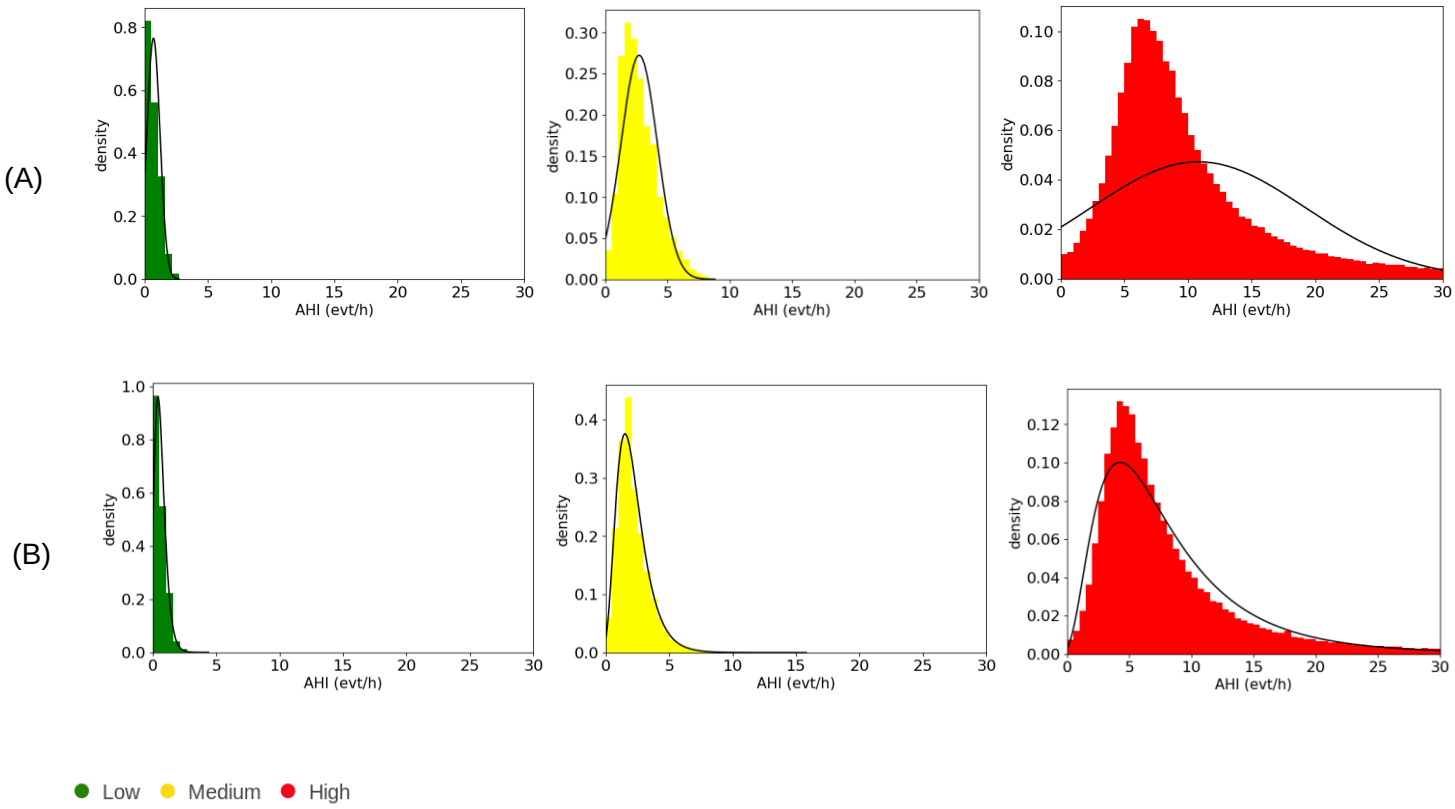
We compared the goodness-of-fit of HMMs with a number of states between 2 and 10. Figure S1 shows the likelihood value as a function of the number of states, with a higher value indicating a better fit with the observations. We observed a first rupture of slope when the number of states equaled 3 and a plateau starting around 5, meaning that, above 5 states, adding more states would not give much better modeling of the data. In the present paper, we focus on the simplest model in order to avoid overfitting, limiting the number of states to 3.



Supplement figure S1: Likelihood and BIC according to the number of states of the HMM.

HMM - Choice of emission distribution

The choice of conditional model for the observed data is another parameter influencing the relevance of our HMM. The index studied in this paper, the rAHI, is the average of the number of apnoea and hypopnoea events per hour. Averages of counts are distributed as non-negative real numbers, are right skewed, have a variance that increases with the mean, and are thus commonly modelled using a log-normal distribution. We trained the HMM using log-normal emission densities and compared it to a HMM with Gaussian emission densities. As proposed by Rachel MacKay Altman,[2] graphical techniques can detect lack of fit. Indeed, plots of the estimated distribution against the empirical distribution show how well the models capture the behaviour of the data. Any observation can be associated with a state. Figure S2 shows, for each of the three hidden states, how well the state emission distribution fits the distribution of the empirical observations associated with this state. As the plots suggest, a log-normal distribution is best suited to our data, and this is due to the skewness towards zero of the rAHI distribution. The log-normal distribution was learnt on the data augmented by 1 to avoid introducing a big variance in the distribution of values inferior to 1 with the log-transform and to guard against numerical instability.



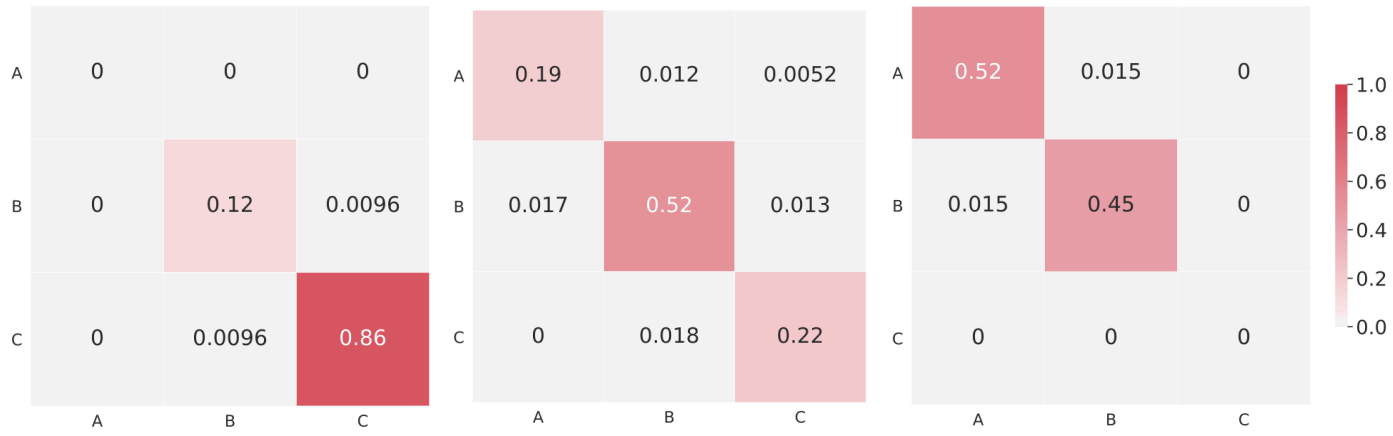
Supplement figure S2: Plots of the (A) Gaussian and (B) log-normal state emission densities (black curve) superimposed on the empirical density.

Extracted features – patient's transition matrices

With the 3-state HMM, each patient is characterized by both their rAHI time series and the corresponding sequence of states, which enables us to compute their transition matrix, built with the proportions of transitions from state *i* to state *j*, with *i* and *j* in *{A, B, C}*.

We obtain, for each patient, a matrix composed of 9 features, presented in Figure S3 for a sample of patients. The features corresponding to the elements $a(i,i)$carry information on the time spent in state i (the proportion of transitions from state i to state i) while features corresponding to the elements $a(i,j)$ with $i \neq j$ characterize the frequency of switches from one state to another.

The matrix of dimensions NxK^2 is the matrix of extracted features to which we apply K-means. The three matrices in Figure S3 present the transition matrices of three patients with different kinds of trajectories.



Supplement figure S3: Transition matrices of 3 patients, with $a(i,j) = P(z_{t+1} = j | z_t = i)$ *where* z_t *denotes the state at time* t *and (i, j)* ϵ *{A, B, C}. The left-hand patient spent 86% of his treatment in state C, 12% in state B, and transitioned 0·96% of the days from state B to state C.*

K-means - Number of clusters

Several criteria are typically used for validating a partition and they are all based on the idea that the higher the intra-class similarity and the lower the inter-class similarity, the better the partition is. The K-means algorithm aims to choose the centroids that minimize the inertia. Figure S4 shows the value of inertia, or within-cluster sum-of-squares criterion, against the number of clusters. Inertia is computed as the sum of squared distances to the closest cluster centroid for all observations:[3]

$$\text{Inertia} = \sum_{i=0}^{n} min_{\mu_j \in C} \left(\left| \left| x_i - \mu_j \right| \right|^2 \right)$$

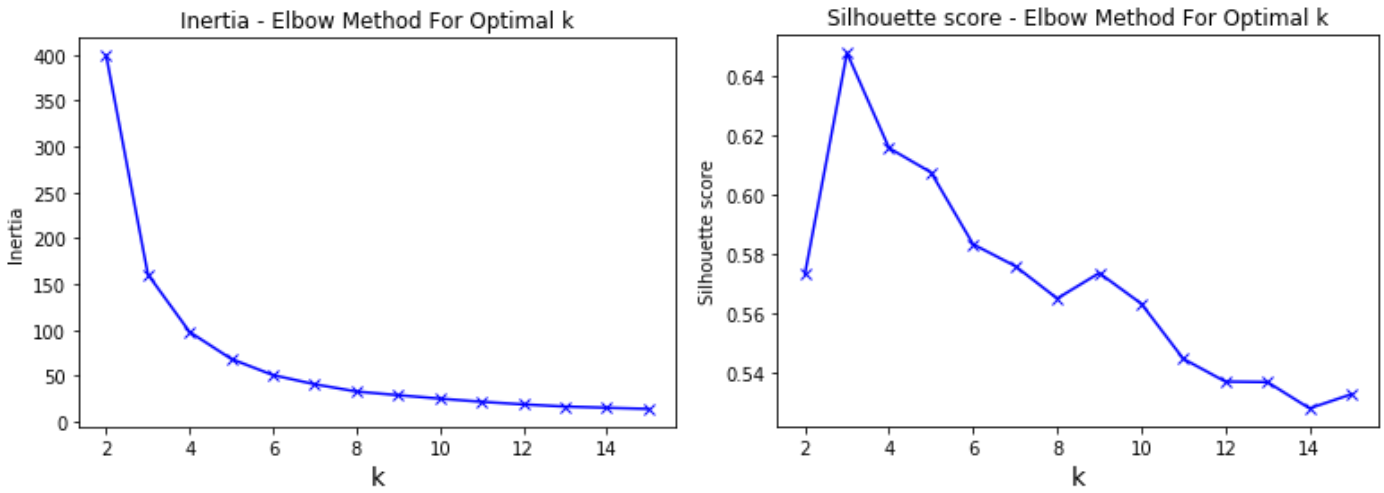
Another score, the silhouette coefficient is adapted to evaluate the clusters' relevance. The silhouette coefficient is computed for each individual as:[3]

$$\text{silhouette} = \frac{b-a}{max(a,b)}$$

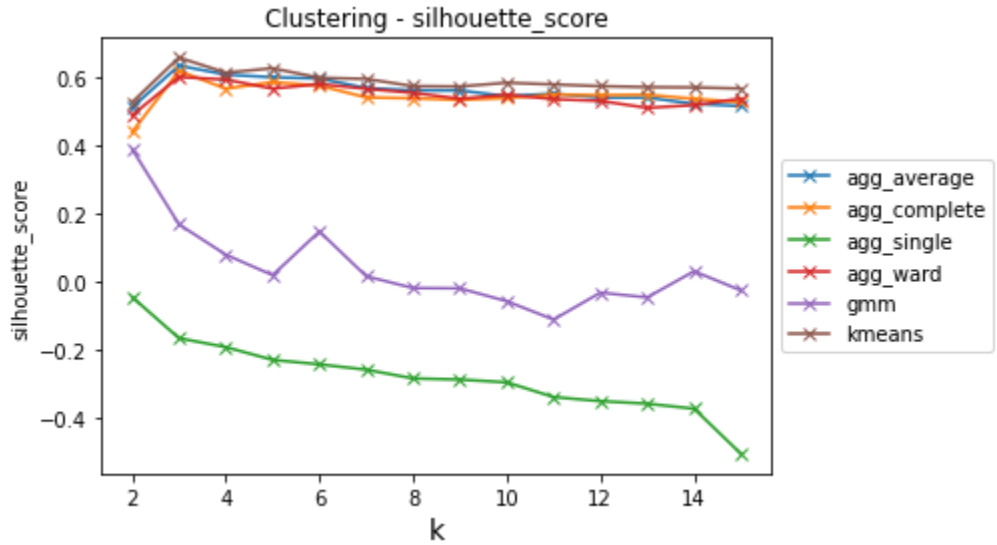
with *a*, the mean distance between the patient and all other patients belonging to the same cluster and *b*, the mean distance between the patient and all other patients in the next nearest cluster. A higher silhouette coefficient indicates a better partition. The score is bounded between -1 for incorrect clustering and 1 for dense and well separated clusters.

The elbow plot of inertia presented in Figure S4 shows at what value of *k* the distance between the mean of a cluster and the other data points in the cluster is at its lowest. We observe that the more clusters we take between 2 to 6, the better they are defined. On the contrary, above 6 clusters, it is not clear that the partition improves with the number of clusters. The silhouette plot shows that the best partition would be with 3 clusters. A 3-cluster

partition may offer the densest and best separated clusters but in our context, we aim at a finer differentiation of trajectories. Thus, we chose 6 clusters as offering a relevant compromise between a good geometric definition of clusters and clusters having informative meaning. With 6 clusters, we identified groups of trajectories alternating between stable and unstable periods: these profiles may have been impacted by external factors, which could be explored in future studies.



Supplement figure S4: Elbow plot of the inertia (left) and silhouette coefficient (right) against the number of clusters specified by the K-means algorithm.



Supplement figure S5: Comparison of hierarchical clustering (agglomerative with several linkage methods), Gaussian mixture models, and K-means algorithm.

2. RESULTS SUPPLEMENT

Results with three clusters

The choice of k=3 clusters is suggested by the silhouette index and results in clinically interpretable clusters, as presented in table S1. Adherence is significantly higher in cluster 0 compared to 1 and 2. Leaks are significantly higher in cluster 2 compared to 0 and 1 (p-value<0.05).

However, with only three clusters we group patients who are always stable, always quite stable and always unstable whereas our goal was to distinguish between more individualized r profiles, including trajectories presenting periods of stability and instability. Indeed, this study also aims at providing a tool for identifying interesting groups of trajectories which may have been impacted by external factors, which could be explored in future studies. Thus, we focus in the main document on the results obtained with 6 clusters, a partition providing more perspectives for future studies.

Cluster	N (%)	Mean age (years)	Sex (M/F)	Mean rAHI (evt/h)	Mean adherence (min)	Mean leaks (L/min)
0	975 (23%)	64.9 $\pm$ 12.3	614/361 (63%/37%)	0.80$\pm$ 0.85	378 $\pm$128	38 $\pm$5
1	1213 (11%)	66.2 $\pm$ 12.8	869/344 (72%/ 28%)	2.45$\pm$ 1.84	363 $\pm$131	39 $\pm$ 6
2	672 (11%)	71.9 $\pm$ 12.9	516/156 (77%/ 23%)	8.45$\pm$ 5.30	350 $\pm$ 134	43 $\pm$9

Table S1: Descriptive statistics of the 3 clusters.
Results are presented as mean ± SD

Sensitivity analysis for clustered data

As a sensitivity analysis to assess the clusters' robustness over time, we compared the partitions over two consecutive trimesters. A partition was deemed stable over time if most of the patients belonging to cluster i at time T_1 are still assigned to cluster$i$ at time T_2, provided they did not have a change of device of mask between T_1 and T_2.
The 6-month-long rAHI time series were split into two series of 3 months and clustering was performed on the two sets of 3-month-long series. Two HMMs were trained, one on each of the two sets of 3-month-long time series, and used to estimate the states associated with the rAHI records. The patients' state transition matrices were computed to perform a K-means clustering on each of the two sets.
We then compared the two resulting partitions: initial clusters over the first 3 months and final clusters over the last 3 months. To this end we computed, for a given initial cluster, the proportion of patients that belonged to each of the final clusters.

Of the 2860 patients, 370 had experienced a CPAP treatment period of at least 6 months without either mask change or device change. Figure S5 shows that cluster 0 (first row) and cluster 5 (last row) are robust over time with more than 80% of the clusters' participants staying in the same cluster over two consecutive trimesters. A

significantly higher rate of transitions from one cluster to another was found in intermediate clusters (clusters 1, 2, 3 and 4). Of the 370 patients, none has transitioned from cluster 0 to cluster 4 or 5, and vice versa.

Supplement figure S5: Heatmap of the cluster intersections

References

1. Viterbi A. Viterbi algorithm. Scholarpedia. 2009;4(1):6246.
2. MacKay Altman R. Assessing the Goodness-of-Fit of Hidden Markov Models. Biometrics 2004; 60: 444–450.
3. L. Hubert and P. Arabie, Comparing Partitions. Journal of Classification 1985. https://link.springer.com/article/10.1007%2FBF01908075