Analysing Complex Life Sequence Data with Hidden Markov Modelling

Satu Helske, Jouni Helske, and Mervi Eerola

Abstract When analysing complex sequence data with multiple channels (dimensions) and long observation sequences, describing and visualizing the data can be a challenge. Hidden Markov models (HMMs) and their mixtures (MHMMs) offer a probabilistic model-based framework where the information in such data can be compressed into hidden states (general life stages) and clusters (general patterns in life courses).

We studied two different approaches to analysing clustered life sequence data with sequence analysis (SA) and hidden Markov modelling. In the first approach we used SA clusters as fixed and estimated HMMs separately for each group. In the second approach we treated SA clusters as suggestive and used them as a starting point for the estimation of MHMMs.

Even though the MHMM approach has advantages, we found it to be unfeasible in this type of complex setting. Instead, using separate HMMs for SA clusters was useful for finding and describing patterns in life courses.

1 Introduction

In social science applications, sequence analysis (SA) has gained more and more interest since its introduction in the mid-80s. It is now central to the life course perspective where it has been used to understand various trajectories and crucial transitions (Gauthier et al., 2014).

Satu Helske
University of Jyvaskyla, PO Box 35, FI-40014, University of Jyväskylä, Finland, e-mail: satu.helske@jyu.fi

Jouni Helske
University of Jyvaskyla, Finland

Mervi Eerola
University of Turku, Finland

Often the goal in SA is to find a typology of life sequences described as categorical time series data. Dissimilarities between each pair of sequences is determined using some criterion. Common choices have been optimal matching (McVicar and Anyadike-Danes, 2002) and Hamming distances (Hamming, 1950; Lesnard, 2010), but many modifications to these and also more fundamentally different methods have been developed (see, e.g., Aisenbrey and Fasang, 2010; Elzinga and Studer, 2014). Usually these dissimilarities are then grouped using cluster analysis such as Ward's agglomerative algorithm.

Life course data often consists of not only one sequence per subject, but multiple parallel sequences, one for each life domain of interest. We refer to *complex sequence data* for data which consist of multiple subjects and long multichannel (multidimensional) sequences.

One option for studying such data is to combine the sequences of each subject time point by time point by extending the state space of observations. This approach is simple if the number of possible combinations is moderate, but the combined state space grows rapidly as the number of domains and/or states grows. Multichannel sequence analysis (Gauthier et al., 2010) has been used for computing pairwise dissimilarities and finding clusters in complex sequence data (see, e.g., Eerola and Helske, 2016; Müller et al., 2012; Spallek et al., 2014). However, the dissimilarities are largely affected by the chosen dissimilarity metric and the cluster allocation may not be well suited to borderline cases. Also, describing, visualizing, and comparing such data is difficult. We use hidden Markov modelling for gaining a probabilistic descriptions of complex sequence data.

Hidden Markov models (HMMs) have been widely used in biological sequence analysis (Durbin et al., 1998) and speech recognition (Rabiner, 1989). Typically, the interest is in one long time series or another type of sequence. In social sciences this approach has been called latent Markov modelling. Typically, the data consists of a few measurements for multiple subjects.

Mixture hidden Markov model (MHMM) is a generalization of the HMM. There we assume that the data consists of latent subpopulations with different model structures. In the context of social sciences, the mixture hidden Markov model approach was formulated by van de Pol and Langeheine (1990) as the mixed Markov latent class model and later generalized to include time-constant and time-varying covariates by Vermunt et al. (2008) (who named the resulting model as the mixture latent Markov model, MLMM).

Multidimensional responses are included in the formulation of the MLMM but, to our knowledge, there are no empirical studies with complex life sequence data. Few studies use (M)HMMs for multichannel social science data. Helske and Helske (2016) have illustrated HMMs and MHMMs for multichannel data but do not conduct actual analyses with real data. Bartolucci et al. (2007) have studied criminal trajectories using HMMs with multiple binary sequences per subject. The data were large in the number of subjects (684 000 individuals), but sequences were short (6 age categories) and they had fixed groups (men and women) instead of latent clusters. Crayen et al. (2012) have used a hierarchical MLMM for two-channel categorical sequences to model dynamics of mood regulation of university students

during one week. The sequences were longer (56 time points) but the number of subjects is moderate (164) and they used only three states in both channels. In their hierarchical model there were two parallel latent structures; one between the days and the other within the days.

We study two approaches to analysing complex sequence data. The first is to use sequence analysis and cluster analysis for finding a few sets of clusters and then, separately for each cluster, to estimate an HMM. In this approach, hidden Markov modelling is used to compress and describe life course information within the clusters and to help choosing the number of clusters.

The second approach is to estimate a mixture model. Now the clustering is not fixed but we get a probability of each individual belonging to each cluster. For large data, estimating the MHMM with the maximum likelihood can be a complex and time-consuming task unless the set of candidate models is restricted. We study the option of using SA clusters and simple HMMs as a starting point for mixture modelling.

2 Interpretation of hidden Markov models for life sequences

One rationale behind using the HMM approach for life sequence analysis was the attempt to identify similar life course patterns based on similar hidden state trajectories. The similarity of hidden state sequences can be attributed to both external factors, which are common to groups of populations, or to internal behavioural similarities between individuals with similar features. Finding hidden dynamics is thus important for analysing and grouping life courses and also for understanding relationships between factors that are measured. The significance of hidden states in life sequence data is dependent on the chosen structure of the model. The goals of our analysis were two-folded:

1. to group individuals with similar life course patterns (clusters) and
2. to compress information in observed states across life domains to capture patterns and dynamics within a group (hidden states)

The aim was to find hidden states that compress the information across several life domains into more general life stages. These life stages could be either stable episodes between two transitions (e.g., employed and married without children) or characterized by transitions in some of the life domains (e.g., moving between unemployment and short-term jobs). We restricted to left-to-right models where transitions back to previous hidden states are not possible. Such representation makes it easier to comprehend the overall dynamics within a group and is also natural from the life course perspective: even though individuals may be in similar states at different times, the second time has a different history compared to the first time. E.g., there could be a group where, at some points of their lives, individuals are married with children, then divorced for a while, and later again married with children (but with the history of having experienced a divorce).

3 Data

We illustrate the analysis of complex life sequence data using a subsample of the German National Educational Panel Survey (NEPS) (Blossfeld et al., 2011).

We restricted to life courses of an age cohort born in 1955–1959. Only individuals who were born in Germany or moved there before age 14 were included.

The data consisted of monthly life statuses of 1731 individuals in three life domains (career, partnerships, and parenthood) from age 15 to age 50. For each individual, there were three parallel sequences of length 434, which made altogether 2,253,762 data points. Using the monthly time scale allowed for detecting also smaller fluctuations in life courses, e.g. recurrent transitions between unemployment and employment.

3.1 Sequences

The sequences in three life domains were constructed as follows:

Career with 4 states:

- Studying (in school, vocational training, or vocational preparation)
- Employed (full-time or part-time)
- Unemployed
- Else (parental leave, military or non-military service, voluntary work, or other gap in employment history)

Partnerships with 4 states:

- Single (never lived with a partner)
- Cohabiting
- Married/in a registered partnership
- Divorced/separated/widowed

Parenthood with 2 states:

- No children
- Has (had) children (biological, adopted, or foster children)

The coding for parenthood was very simple. A practical reason was that this record was available for most individuals, whereas more detailed information was often missing. On the other hand, we can argue that specifically the experience of becoming a parent is relevant as one step in the developmental process into adulthood.

For the latter two life domains, the status of each month was usually determined from the latest event. An exception was made for the rare partnerships that lasted for less than a month; there separation was coded from the following month onward. In a case of multiple records per month in the career domain, the final status was

given according to assumed importance: school and vocational training came before employment, which in turn dominated over vocational preparation, unemployment, and other non-employment statuses.

Altogether 306 individuals (17.7%) had some missing information in one or two life domains. Thus, at each time point we have at least some information from each individual.

4 Hidden Markov models

In the context of hidden Markov models, observed states are determined via a Markov process of hidden states. These hidden states cannot be observed directly, but only through the sequence(s) of observations, since hidden states generate ("emit") observations on varying probabilities.

Assume we have multichannel sequence data for N individuals with C parallel sequences of length T. Naturally, the following applies for single-channel data (subjects with one sequence only) by setting $C = 1$. Let us denote the observation in channel $c,\ c = 1,\ldots,C$, of individual $i,\ i = 1,\ldots,N$, at time $t,\ t = 1,\ldots,T$, with y_{itc} and the corresponding hidden state with z_{it}. A discrete first order hidden Markov model $\mathscr{M}$ is characterized by the following parameters:

- Initial probability of hidden state s:

$$\pi_s = P(z_{i1} = s);\ s \in \{1,\ldots,S\}, \text{for all } i = 1,\ldots,N.$$

- Transition probability from hidden state s to hidden state r:

$$a_{sr} = P(z_{it} = r | z_{i(t-1)} = s);\ s,r \in \{1,\ldots,S\}, \text{for all } i = 1,\ldots,N.$$

- Emission probability of observed state m_c in channel c given the hidden state s:

$$b_s(m_c) = P(y_{itc} = m_c | z_{it} = s);\ s \in \{1,\ldots,S\},\ m_c \in \{1,\ldots,M_c\}, \text{ for all } i = 1,\ldots,N. \quad (1)$$

The (first order) Markov assumption states that the hidden state transition probability at time t only depends on the hidden state at the previous time point $t-1$:

$$P(z_{it} | z_{i(t-1)},\ldots,z_{i1}) = P(z_{it} | z_{i(t-1)}). \quad (2)$$

Also, the observed states at time t are independent of all other observations and hidden states given the hidden state at t. For multichannel sequence data, we assume the same latent structure applies for all channels, i.e., the hidden state at time t for individual i generates the observed state y_{itc} in all channels c. Observations $y_{it1},\ldots,y_{itC}$ are assumed independent of each other given the hidden state z_{it}, i.e.,

$P(\mathbf{y}_{it}|z_{it}) = P(y_{it1}|z_{it}) \cdots P(y_{itC}|z_{it})$. Fig. 1 illustrates an HMM with a hidden state sequence and two channels.

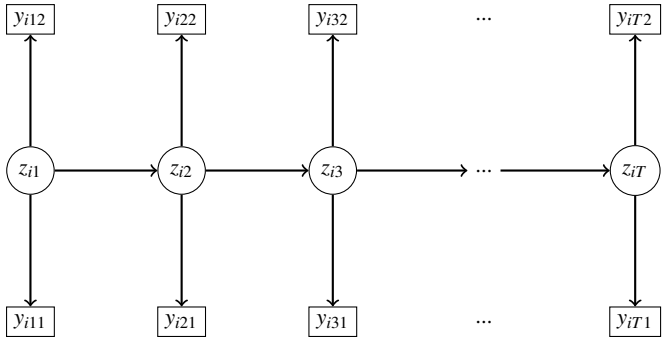


Fig. 1 Illustration of the hidden Markov model structure for two-channel sequence data for individual i with hidden states $z_{i1} \ldots, z_{iT}$ and observed states $y_{i1c} \ldots, y_{iTc}, c = 1, 2$.

The log-likelihood for the HMM is written as

$$\log L = \sum_{i=1}^{N} \log P\left(Y_i|\mathscr{M}\right), \tag{3}$$

where Y_i are the observed sequences in channels $1, \ldots, C$ for subject i and $\mathscr{M}$ describes the model and its parameters $\{\pi, A, B_1, \ldots, B_C\}$, where $A = \{a_{sr}\}$ is a matrix of transition probabilities and $B_c = \{b_s(m_c)\}$ is a matrix of emission probabilities for channel c. The probability of observation sequences for subject i given the model is

$$\begin{aligned} P(Y_i|\mathscr{M}) &= \sum_{\text{all } z} P\left(Y_i|z, \mathscr{M}\right) P\left(z|\mathscr{M}\right) \\ &= \sum_{\text{all } z} P(z_1|\mathscr{M}) P(\mathbf{y}_{i1}|z_1, \mathscr{M}) \prod_{t=2}^{T} P(z_t|z_{t-1}, \mathscr{M}) P(\mathbf{y}_{it}|z_t, \mathscr{M}) \\ &= \sum_{\text{all } z} \pi_{z_1} b_{z_1}(y_{i11}) \cdots b_{z_1}(y_{i1C}) \prod_{t=2}^{T} \left[a_{z_{t-1}z_t} b_{z_t}(y_{it1}) \cdots b_{z_t}(y_{itC})\right], \end{aligned} \tag{4}$$

where the hidden state sequences $z = (z_1, \ldots, z_T)$ take all possible combinations of values in the hidden state space $\{1, \ldots, S\}$ and where $\mathbf{y}_{it}$ are the observations of subject i at t in channels $1, \ldots, C$; π_{z_1} is the initial probability of the hidden state at time $t = 1$ in sequence z; $a_{z_{t-1}z_t}$ is the transition probability from the hidden state at time $t-1$ to the hidden state at t; and $b_{z_t}(y_{itc})$ is the probability that the hidden state of subject i at time t emits the observed state at t in channel c.

4.1 Mixture hidden Markov model

The mixture hidden Markov model is, by definition, a mixture of simple hidden Markov models. We assume that the population consists of subpopulations of individuals (latent classes or clusters) with different life patterns. Respectively, the mixture model consists of varying submodels that characterize the clusters. Transitions from one cluster to another are not allowed.

Assume that we have a set of HMMs $\mathscr{M} = \{\mathscr{M}^1, \ldots, \mathscr{M}^K\}$, where $\mathscr{M}^k = \{\pi^k, A^k, B_1^k, \ldots, B_C^k\}$ for clusters $k = 1, \ldots, K$. We denote $P(\mathscr{M}^k) = w_k$ as the prior probability that an arbitrary observation sequence is generated by the submodel $\mathscr{M}^k$ such that $\sum_{k=1}^K w_k = 1$.

The log-likelihood of the MHMM is of the form

$$\begin{aligned}
\log L &= \sum_{i=1}^{N} \log P(Y_i|\mathscr{M}) \\
&= \sum_{i=1}^{N} \log \left[\sum_{k=1}^{K} P(\mathscr{M}^k) \sum_{\text{all } z} P\left(Y_i|z, \mathscr{M}^k\right) P\left(z|\mathscr{M}^k\right) \right] \\
&= \sum_{i=1}^{N} \log \left[\sum_{k=1}^{K} w_k \sum_{\text{all } z} \pi_{z_1}^k b_{z_1}^k(y_{i11}) \cdots b_{z_1}^k(y_{i1C}) \prod_{t=2}^{T} \left[a_{z_{t-1} z_t}^k b_{z_t}^k(y_{it1}) \cdots b_{z_t}^k(y_{itC}) \right] \right].
\end{aligned} \tag{5}$$

For more detailed description of MHMMs, see Helske and Helske (2016) or Vermunt et al. (2008).

4.2 Model estimation

The log-likelihoods of (4) and (5) are efficiently calculated with the *forward–backward algorithm* (Baum and Petrie, 1966; Rabiner, 1989). A common maximum likelihood estimation method is the Baum–Welch algorithm, i.e., the expectation–maximization (EM) algorithm in the HMM context.

The Baum–Welch algorithm requires starting values for model parameters. In order to reduce the risk of being trapped in a poor local optimum, a large number of initial values should be tested. Simpler models with few parameters are fast to estimate; therefore, it is possible to fit the model numerous times with varying random starting values for finding the model with the best likelihood. When the model is large, estimation is more time-consuming and good starting values for model parameters are useful or even essential.

The most probable path of hidden states for each subject given their observations and the model can be computed using the *Viterbi algorithm* (see, e.g., Rabiner, 1989). This path maximizes the probability of $P(z|Y_i, \mathscr{M})$.

The forward–backward algorithm can also be used for computing posterior cluster probabilities (the probability that subject i belongs to a certain cluster) for MHMMs. These can be used for classifying subjects into different groups.

4.3 Model comparison

Models with the same number of parameters can be compared with the value of the log-likelihood function. For choosing between models with a different number of hidden states, we need to take account of the number of parameters.

Bayesian information criterion (BIC) is the usual criterion for comparing (M)HMMs. We define it as

$$BIC = -2\log(L) + p\log\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{C}\sum_{c=1}^{C}\mathrm{I}(y_{itc}\text{ observed})\right), \tag{6}$$

where L is given in equation 3, p is the number of estimated parameters, I is the indicator function, and the summation in the logarithm is the size of the data. If data are completely observed, the summation is simplified to $N \times T$. The smaller the BIC, the better the model.

When computing the log-likelihood for the combined model with fixed SA clusters we simply sum the log-likelihoods of the cluster-wise HMMs. BIC of the combined model is determined as

$$BIC = -2 \times \sum_{k=1}^{K}\log(L_k) + \sum_{k=1}^{K} p_k \log\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{C}\sum_{c=1}^{C}\mathrm{I}(y_{itc}\text{ observed})\right), \tag{7}$$

where L_k is the likelihood of the HMM of cluster k, p_k is the number of estimated parameters in the HMM for cluster k, and the summation in the logarithm is the size of the full data set.

5 Visualizing sequence data and models

Visualization is an important tool throughout the analysis process from the first glimpses into the data to presenting the results. As an example, we consider the data and the HMM for one of the preliminary clusters described “Long education and later family” (from the ten-cluster solution).

Fig. 2 illustrates a five-state HMM with the following life stages:

1. Single and (mostly) studying
2. Cohabiting, separated, or divorced; studying or employed
3. Married, studying or employed
4. Married with children, non-employed

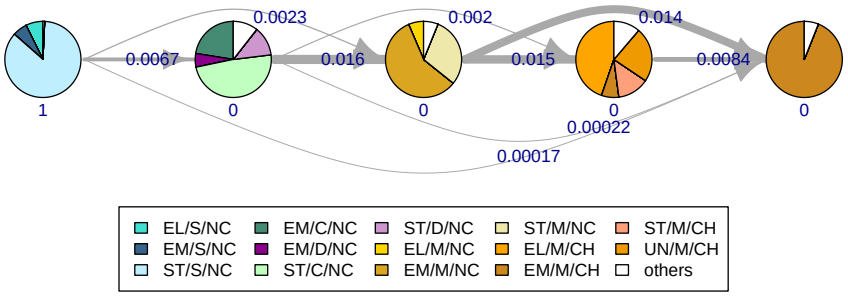


Fig. 2 Illustrating the hidden Markov model for the cluster of individuals with long education and later family. Pies present five hidden states, with slices showing the emission probabilities of combinations of observed states. States with emission probability less than 0.05 are combined into one slice for easier interpretation. The edges show the transtion probabilities – the thicker the edge, the higher the probability. Initial probabilities of the hidden states are given below the pies. The descriptions of the combined states show career/partnership/parenthood statuses: ST=studying, EM=employed, UN=unemployed, EL=else; S=single, C=cohabiting, M=married, D=divorced/separated; NC=no children, CH=has child(ren).

5. Married with children, employed

The hidden states are described by the most probable emitted observations, but there are also less probable states that are omitted from the plot for readability. E.g., the second state also emits marriages with a small probability—from the most probable hidden state paths in Fig. 3 we can see that these are marriages which end in divorce relatively fast. We could interpret that the second hidden state describes a life stage of searching for a partner before forming a long-lasting marriage.

All subjects start from the first state at age 15. At the start of the follow-up they are all single and mostly studying. The most common transition is to the second state, but the third state is quite probable also. Due to the monthly data, the transition probabilities are small—individuals usually spend years in each state.

Most individuals move to the third hidden state which describes childless marriage. It is the hidden state where individuals spend the least time on average. Transitions to the fourth and the fifth hidden state are almost as common. These both describe parenthood; some move out of workforce for a while or until the end of the follow-up, while some continue working.

6 Analysis

Estimating a large MHMM for complex sequence data can be difficult and time-consuming unless the structure of the model is fixed or known, even approxim-

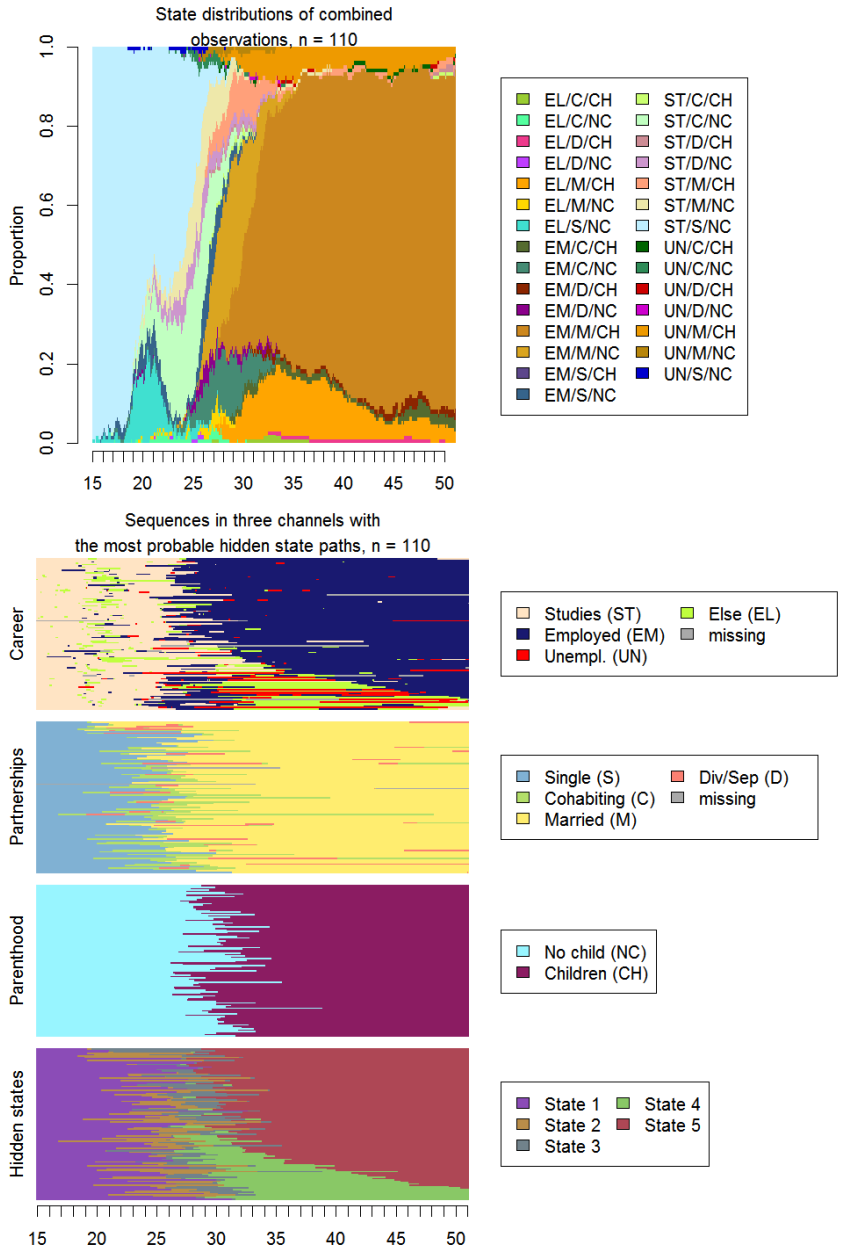


Fig. 3 State distributions of combined observations (top) and sequences of observations in each channel as well as the most probable paths of hidden states (bottom). Sequences are ordered by multidimensional scaling scores. States 1–5 correspond to the hidden states presented in Fig. 1.

ately. In other cases, the set of candidate models must be somehow restricted. In this case we had little prior knowledge on the structure of the model; hence, how many clusters to choose and how many hidden states to include in each cluster? As transitions were frequent in some of the trajectories and infrequent in others, it was clear that some of the clusters should contain more hidden states than others, leading to an unfeasible large number of possible model structures.

We compared two different approaches for the analysis of complex sequence data, of which both were conducted in a stepwise manner. The first two steps applied for both approaches, whereas step 3 was different (denoted as 3a and 3b). More detailed descriptions of the analysis process are given in the following sections.

1. **Sequence analysis**. Computing the dissimilarities between the subjects with the Hamming distance. Using Ward's hierarchical method for clustering individuals with similar life courses. Choosing a set of reasonable clustering solutions for preliminary analysis.
2. **Hidden Markov models.** Separately for each SA cluster, fitting simple HMMs with a different number of hidden states. Choosing the best model for each preliminary cluster.

3a. **Combined HMMs.** Constructing a combined model from separate HMMs (from step 2), keeping parameters fixed. Computing the likelihood and BIC for combined models with 7–12 clusters for determining the number of clusters. Computing the most probable path of hidden states for each individual.

3b. **Mixture hidden Markov models.** For each clustering solution (7–12 clusters), estimating an MHMM by using parameters of the corresponding HMMs (from step 2) as starting values. Computing the likelihood and BIC of the MHMMs for determining the number of clusters. Computing the most probable path of hidden states for each individual.

6.1 Step 1: Sequence analysis and preliminary clustering

We started by applying multichannel sequence analysis and computed the dissimilarities between the sequences. These were then used in cluster analysis.

6.1.1 Sequence dissimilarities

We compared a few dissimilarity metrics that are suitable for multichannel data: optimal matching (OM), generalized Hamming distance (HAM), and dynamic Hamming distance (DHD) (Lesnard, 2010). We chose the generalized Hamming distance with theory-driven substitution costs (see Table 1). The metric compares observed states time point by time point and gives a cost for mismatches. It generally works relatively well in a problem where timing is important and also here resulted in meaningful clusters with high goodness-of-fit (see Sect. 6.1.2).

Table 1 Substitution costs for Hamming distances.

Career status	→ ST	→ EM	→ UN	→ EL	→ *
Studying (S) →	0	3	2	1	0
Employed (EM) →	3	0	2	2	0
Unemployed (UN) →	2	2	0	1	0
Else (EL) →	1	2	1	0	0
Missing (*) →	0	0	0	0	0

Partnership status	→ S	→ C	→ M	→ D	→ *
Single (S) →	0	2	2	3	0
Cohabiting (C) →	2	0	1	2	0
Married (M) →	2	1	0	2	0
Divorced/sep. (D) →	3	2	2	0	0
Missing (*) →	0	0	0	0	0

Parenthood status	→ NC	→ CH	→ *
No children (NC) →	0	3	0
Has children (CH) →	3	0	0
Missing (*) →	0	0	0

6.1.2 Cluster analysis

Ward's method was chosen for clustering since it typically produces usable and relatively even-sized clusters compared to most of the other clustering methods (Aassve et al., 2007; Helske et al., 2015). We chose six clustering solutions with 7–12 clusters for further examination. The choice was based on the dendrogram and interpretability of the clusters. Ward's method is agglomerative, so when two smaller clusters are merged, all other clusters remain the same. This means that within the six sets of clustering results there were only $7+2+2+2+2+2=17$ distinct clusters (see Fig. 4 for an illustration).

Table 2 shows the goodness-of-fit statistics for different clustering results and dissimilarity metrics, as measured by the proportion of the variation explained by the clusters (pseudo coefficient of determination (R^2); see Studer et al., 2011). Here, generalized Hamming distances resulted in meaningful clusters with a relatively high goodness-of-fit. OM resulted in clusters with as high goodness-of-fit while DHD resulted in somewhat lower values of R^2 (though not by much). OM clusters were similar to HAM clusters in many ways but had more variation in the timings of first transitions into employment, partnerships, and parenthood.

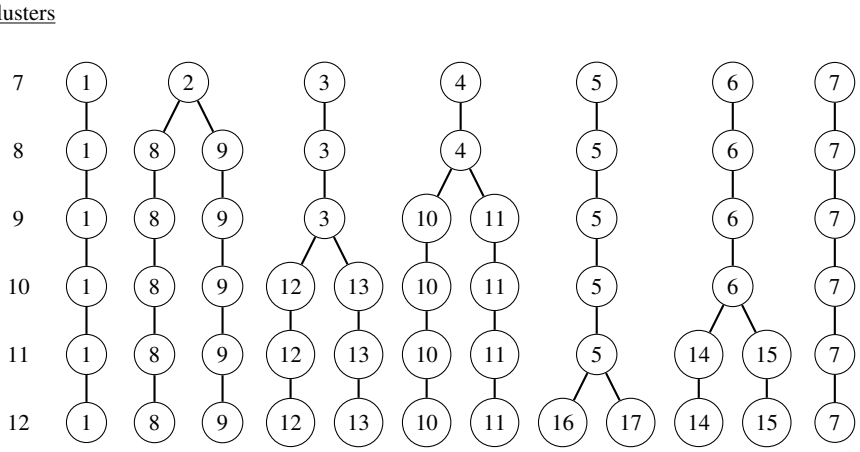


Fig. 4 Clustering structure for Ward's agglomerative method shown for six sets of clustering results with 7–12 clusters.

Table 2 Proportion of variation covered by 7–12 clusters. Clustering was based on different dissimilarity metrics; generalized Hamming distance (HAM), optimal matching (OM), and dynamic Hamming distance (DHD)).

Clusters	HAM	OM	DHD
7	0.38	0.38	0.35
8	0.40	0.40	0.37
9	0.42	0.42	0.38
10	0.43	0.43	0.40
11	0.44	0.44	0.41
12	0.44	0.45	0.42

6.2 Step 2: Simple hidden Markov models for clusters

At the next step, we estimated five HMMs with 4–9 hidden states separately for each of the 16 clusters—fewer hidden states for simpler clusters, more for more complex ones. Since the goal was to find life stages between adolescence and middle age, having too few or too many hidden states was not plausible nor interpretational.

6.2.1 Model estimation

We set starting values for parameters by determining candidate hidden states from observed data and re-estimated the model numerous times by altering these values as follows. At first, we estimated the model 10,000 times with a large variation in starting values. For each re-estimation step we added noise from the $N(0, 0.3^2)$ distribution to the the original starting values (with proper scaling and correction of

signs). The aim of this estimation was to broadly explore the parameter space and to get closer to the global maximum.

To make sure that we were at or near the global optimum, we re-estimated the model by using the model with the highest likelihood as a stating point, now adding noise from the $N(0, 0.15^2)$ distribution. If the model with the highest likelihood was found only a few times, similar estimation was repeated (again using the best model as the new starting point) in order to be fairly certain to have found the global optimum. For clusters with fewer members and models with fewer hidden states, the first estimation step was often enough for finding the (assumed) global maximum.

6.2.2 Model comparison

For each cluster, the HMMs with a different number of hidden states were compared to find the best model to use in the mixture models. BIC and other information criteria are common choices for comparison of HMMs with different numbers of hidden states. Another common option for model selection is cross-validation.

We chose to use BIC as it generally selects parsimonious models. BIC has been proven consistent for ergodic stationary HMMs (Whiting and Pickett, 1988), but not to left-to-right HMMs. Here, also BIC consistently chose models with more hidden states and clusters than is interpretational or plausible.

A likely reason for poor performance of information criteria in this problem was that we were comparing models which all were considerably simple compared to the complexity of real life. The goal was to simplify and describe the overall patterns and dynamics in life trajectories, not to find data-generating models.

However, we did use BIC as one source of information for choosing the number of hidden states by looking for turning points in BIC after which additional hidden states were not as profitable. In addition to BIC, the choice of the number of hidden states was based on interpretability of the model and the prevalence of an additional hidden state in the most probable hidden state paths—if a hidden state was "visited" only rarely it was regarded as unnecessary.

6.3 Step 3 a: Combined HMMs

At this step we used the separate cluster-specific HMMs to construct combined models with 7–12 clusters. For each combined model, we computed the likelihood and BIC to determine the best number of clusters.

The combined model with the smallest BIC was used for determining the best number of clusters. Given the best clustering, we computed the most probable paths of hidden states for each individual.

6.4 Step 3 b: Mixture hidden Markov models

At this step we constructed six MHMMs with 7–12 clusters. We used the estimated parameters of respective cluster-wise HMMs as starting values for mixture models. To avoid non-structural zeros in starting values, we added a small amount of 0.001 to each starting value (with proper scaling). We estimated models in a similar manner to the previous step, by using randomized starting values—first with a larger noise and, after getting closer to the optimum, again with a smaller noise.

6.5 Software

Analyses were conducted with the R software (R Core Team, 2015) by using packages TraMineR (Gabadinho et al., 2011) for sequence analysis, cluster (Maechler et al., 2015) for cluster analysis, and seqHMM (Helske and Helske, 2016) for hidden Markov modelling.

7 Results

The number of hidden states per cluster varied between six and eight. We applied both the combined model and the mixture model approach for describing data and determining the best number of clusters.

7.1 Combined model approach

Table 3 shows the BICs for models with 7–12 clusters. The model with eight clusters resulted in smallest BIC (even the highest likelihood) and was chosen as the best model. The model with seven clusters was almost as good; the only difference was that the two childless clusters (see Fig. 6) were combined into one.

Table 3 Number of parameters, log-likelihood, and BIC for combined models with 7–12 clusters. The smallest value of BIC is shown in bold.

Clusters	Parameters	Log-likelihood	BIC
7	533	−369075.7	745059.4
8	**595**	**−364825.9**	**743368.2**
9	643	−370746.2	755208.7
10	705	−368985.0	751686.5
11	767	−368977.5	751671.5
12	800	−373550.3	760817.0

Fig. 5 and Fig. 6 illustrate the HMM structure for each of the eight clusters. More detailed visualizations with observed sequences and most probable hidden state paths are shown in the Appendix.

The clusters were well separated from each other by the timing and occurrence of career and family states. The two largest clusters were characterized by (mostly) short education and family. They differed in the timing of partnership and parenthood transitions which occurred either earlier in life (cluster A with 461 members of which 59% were females) or later (cluster B, 403 members, 54% males) The third largest cluster (cluster C, 266 members, 68% males) mostly consisted of individuals with long education and later family. Another cluster with early family transitions (cluster D, 159 members, 96% females) was characterized with a long career break for mostly taking care of children.

Two clusters were characterized by no or very late parenthood. They differed in timing of the partnerships; the larger cluster (cluster E, 177 members, 51% males) had earlier first partnerships while in the smaller cluster (cluster F, 116 members, 59% males) partnerships were delayed or omitted altogether.

The two smallest clusters consisted of single parents (cluster G, 47 individuals, 72% females) or parents living divorced or separated (cluster H, 102 individuals, 61% females).

7.2 Mixture model approach

The estimation of ordinary HMMs can be challenging due to multiple local optima in likelihood surfaces, since typical parameter estimation algorithms often only find these suboptimal solutions. Therefore, multiple starting values for the estimation are needed to ensure that the global optimum is found. The same problem is even more prevalent in complex MHMM settings with a large amount of parameters and mixture components. In addition, when the structure of the model (the number of mixture components and/or hidden states) is unknown, the amount of required computing resources naturally multiplies.

Therefore, even after using allegedly reasonable starting values (from simple HMMs), parallel computation, and extensive computing resources, we were not able reach satisfactory results. With different starting values the estimation always resulted in a different solution, so finding the global optimum would have required an unfeasible amount of computing time and/or resources.

Even though we were not able to find optimal MHMMs, we did study some of the suboptimal solutions. To study the differences of SA and MHMM clusters, we estimated a mixture model by keeping the initial, transition, and emission parameters of the submodels fixed (i.e., estimating only prior cluster probabilities, later referred to as the "non-estimated MHMM"). This approach was similar to the combined model approach, but instead of keeping the cluster memberships fixed we allowed individuals to switch clusters. Each individual was assigned to the cluster with the highest posterior cluster probability given their observed sequences.

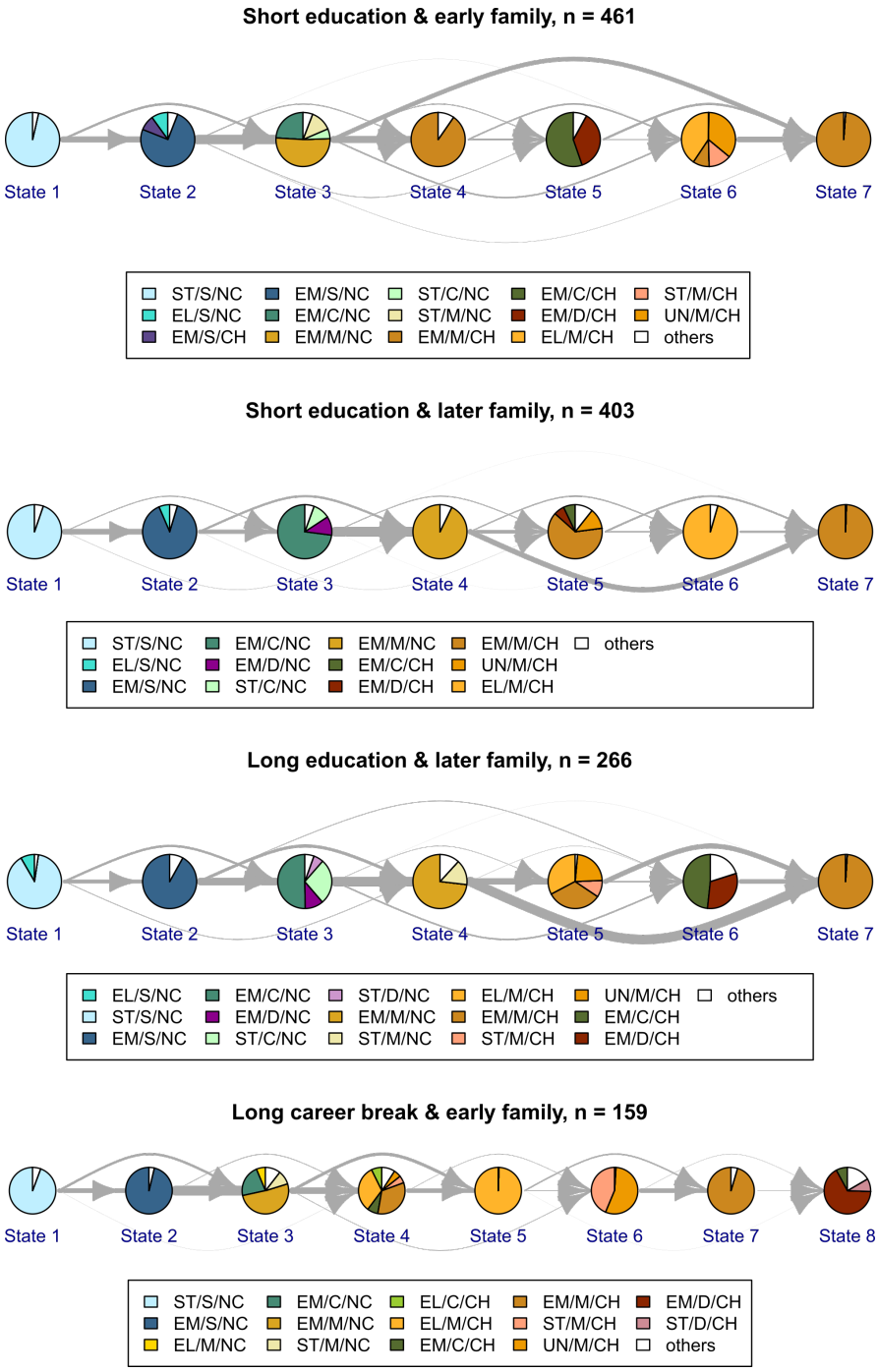


Fig. 5 HMM graphs for the eight cluster solution (clusters A–D). State abbreviations show career/partnership/parenthood statuses: ST=studying, EM=employed, UN=unemployed, EL=else; S=single, C=cohabiting, M=married, D=divorced/separated; NC=no children, CH=has child(ren).

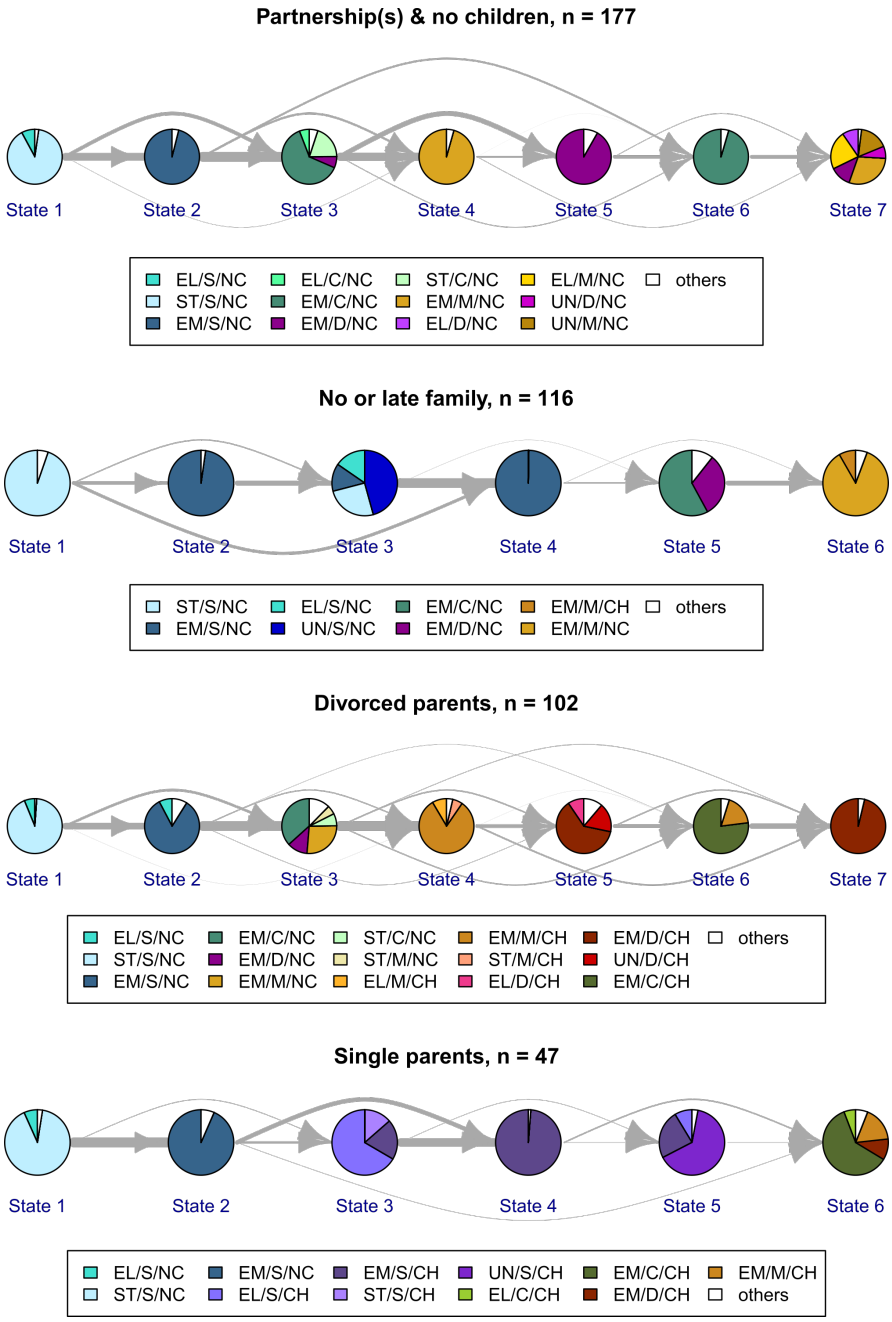


Fig. 6 HMM graphs for the eight cluster solution (clusters E–H). State abbreviations show career/partnership/parenthood statuses: ST=studying, EM=employed, UN=unemployed, EL=else; S=single, C=cohabiting, M=married, D=divorced/separated; NC=no children, CH=has child(ren).

Many individuals switched clusters compared to the SA solution (see Table 4). Some cluster were more stable; close to 90% of the members of the SA clusters "Single parents" and "Partners and no children" stayed in the same cluster in the MHMM solution. Others had many switchers; less than half of the members of SA clusters "Short education and early family" and "Long education and later family" stayed in their original clusters in the MHMM solution.

Table 4 Comparison of SA cluster memberships (left) to most probable cluster memberships from the non-estimated MHMM (top). Probabilities of staying in the same cluster are shown in bold.

	MHMM clusters								
SA clusters	A	B	C	D	E	F	G	H	Members
Short educ. & early fam. (A)	**0.32**	0.35	0.15	0.11	0.00	0.00	0.06	0.01	461
Short educ. & later fam. (B)	0.09	**0.64**	0.16	0.09	0.00	0.00	0.03	0.00	403
Long educ. & later fam, (C)	0.06	0.32	**0.43**	0.13	0.00	0.00	0.07	0.00	266
Career break & early family (D)	0.04	0.39	0.03	**0.54**	0.00	0.00	0.00	0.00	159
Partnership(s) & no child (E)	0.00	0.05	0.03	0.00	**0.87**	0.05	0.00	0.01	177
No or late family (F)	0.00	0.03	0.01	0.03	0.32	**0.60**	0.00	0.01	116
Divorced parents (G)	0.04	0.00	0.03	0.16	0.00	0.00	**0.77**	0.00	102
Single parents (H)	0.00	0.00	0.02	0.00	0.00	0.00	0.04	**0.94**	47
Number of cluster members	207	577	260	228	191	79	138	51	1731

If the MHMM parameters were estimated jointly, the differences compared to the SA clusters were even larger (we do not report the findings as we were not able to find the globally optimal model). In both MHMM approaches, the order and occurrence of states were generally more determining for the cluster memberships than the timing and duration of states. Fig. 7 illustrates this difference seen in the cluster "Short education and early family", showing the observed and hidden state sequences of members of the SA cluster and the cluster from the non-estimated MHMM. One can easily see that the variation in the timing of transitions between states (both observed and hidden) is much larger in the MHMM cluster compared to the SA cluster.

8 Discussion

When analysing complex sequence data with multiple channels, describing and visualizing the data can be a challenge. Hidden Markov models and their mixtures offer a probabilistic model-based framework where the information in data can be compressed into hidden states (different life stages) and clusters (general patterns in life courses). Hidden states can capture general life stages that include not only rather stable episodes (as the fifth hidden state of work, marriage, and children in Fig. 2) but also life stages characterized by change (as the second hidden state of searching for a partner in Fig. 2).

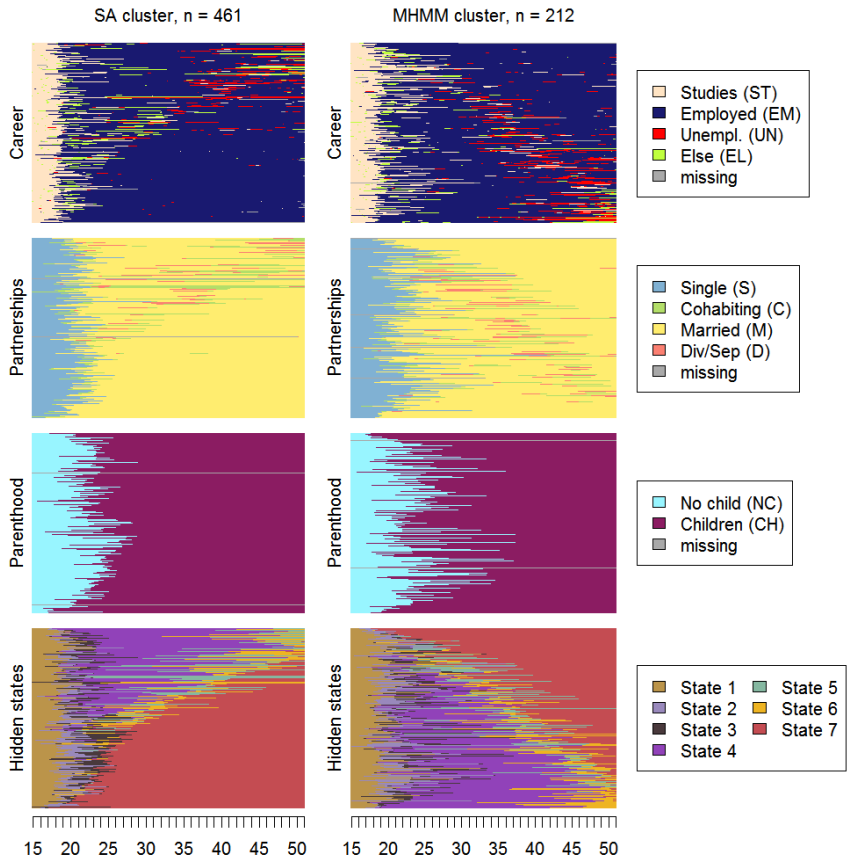


Fig. 7 Comparison of a cluster (Short education and early family) given by SA and the non-estimated MHMM.

Mixture hidden Markov modelling has several advantages. With posterior cluster probabilities we get information on certainty of the clustering for each individual and a measure for the goodness of the classification. We can also extend the model by adding covariates for explaining cluster memberships or transitions between hidden states. The MHMM approach has been used successfully in simpler settings, e.g., for accounting for measurement error and for finding clusters of "movers" and "stayers" between two hidden states.

The downsides of MHMM analysis are related to computational issues. Maximum likelihood estimation of parameters of a complex MHMM is computationally heavy. Due to multimodality of the likelihood surface we need to estimate the model numerous times with different starting values. Also, often the structure of the model (in terms of the number of hidden states and/or clusters) is not known and in general

selecting the best structure is a nontrivial task. Thus, finding the globally optimal MHMM can become unfeasible without constraining the problem.

Using sequence analysis and cluster analysis as a starting point might be useful by providing preliminary classification and by limiting the set of candidate models for a complex MHMM setting. In our study we were not able to reach satisfactory results. Our data was much more complex than in a typical MHMM analysis where sequences often come from panel data with a moderate number of measurement points. The multichannel structure, long sequences, and the relative large number of individuals in our data was a challenging combination for parameter estimation. Also, typically the number of candidate models is rather limited; when HMMs are used for accounting for measurement error, the number of hidden states is known in advance and usually the state space is very limited (e.g., poor/nonpoor or drug user/nonuser). In our study the model structure was unknown and we expected to find several clusters, each with an unknown number of hidden states.

Instead of using mixture models, we treated the SA clusters as fixed and estimated HMMs separately for each cluster (the combined model approach). With SA we found clusters that were adequately well separated by the timing and duration of life states. Hidden Markov models were used for choosing the number of clusters and for describing the overall dynamics within clusters.

Clusters found using SA and the MHMM were different in several ways. When defining sequence dissimilarities, we considered the timing of the events very important and used Hamming distances. In the MHMM analysis many individuals switched clusters; the order of states was generally more determining than their timing and duration. Further research is needed in order to determine distance metrics that result in SA clusters which capture similar features as HMMs. Metrics that weight the order of states instead of their timing such as the number of matching subsequences or the subsequence vectorial representation metric (Studer and Ritschard, 2016), might produce clustering results that are better suited for the starting point of MHMM estimation. Unfortunately, using these metrics with multichannel data is not a straightforward task.

Another topic for further research is model selection of left-to-right HMMs and MHMMs. In our study, BIC performed poorly. Further theoretical and empirical studies are needed for detecting the reasons for its failure and for discovering selection criteria that are better suited for finding parsimonious HMMs.

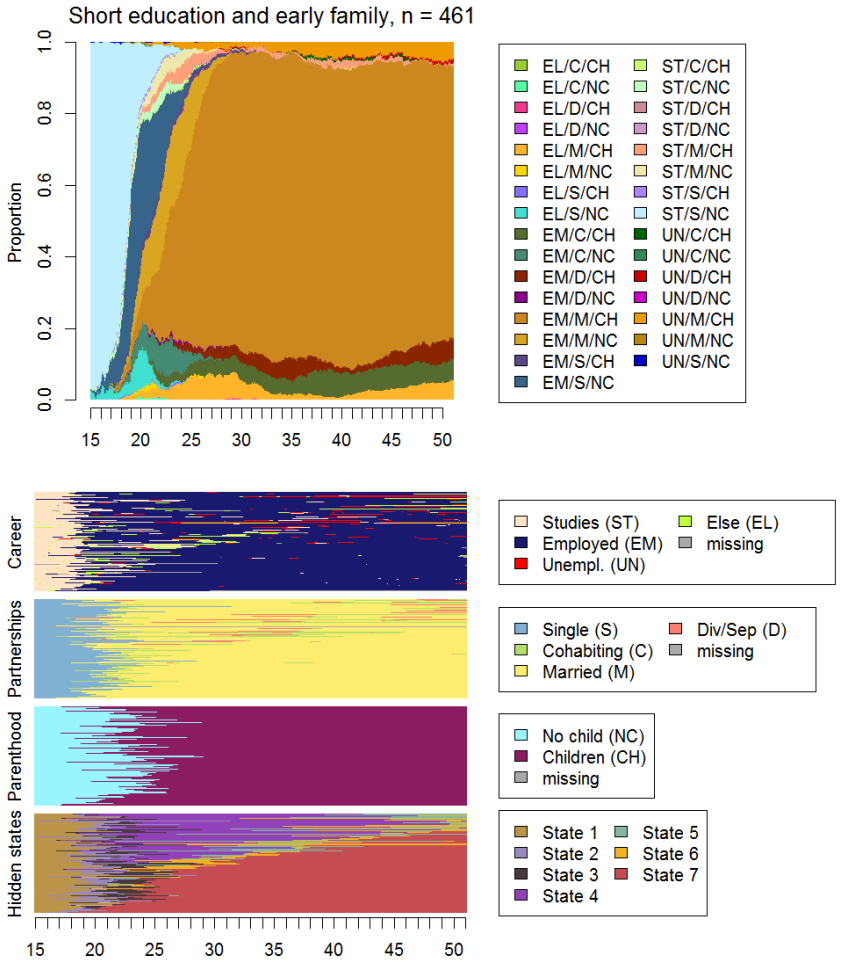
The aim of our study was to describe complex life sequence data. For that goal, SA and the combined HMM approach gave satisfactory results in a reasonable time. We were able to find meaningful clusters and to visualize their complex life course information by using stacked sequence plots, combined state distributions, and HMM graphs.

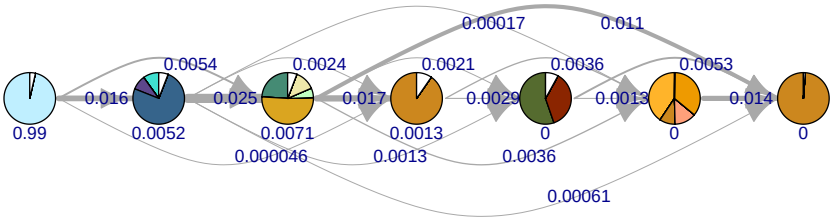
9 Acknowledgements

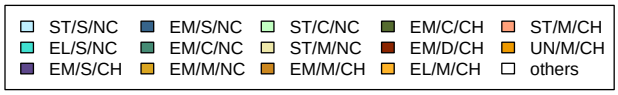
This paper uses data from the National Educational Panel Study (NEPS) Starting Cohort 6–Adults (Adult Education and Lifelong Learning), doi:10.5157/NEPS:SC6:3.0.1. From 2008 to 2013, the NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research and supported by the Federal States. As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi).

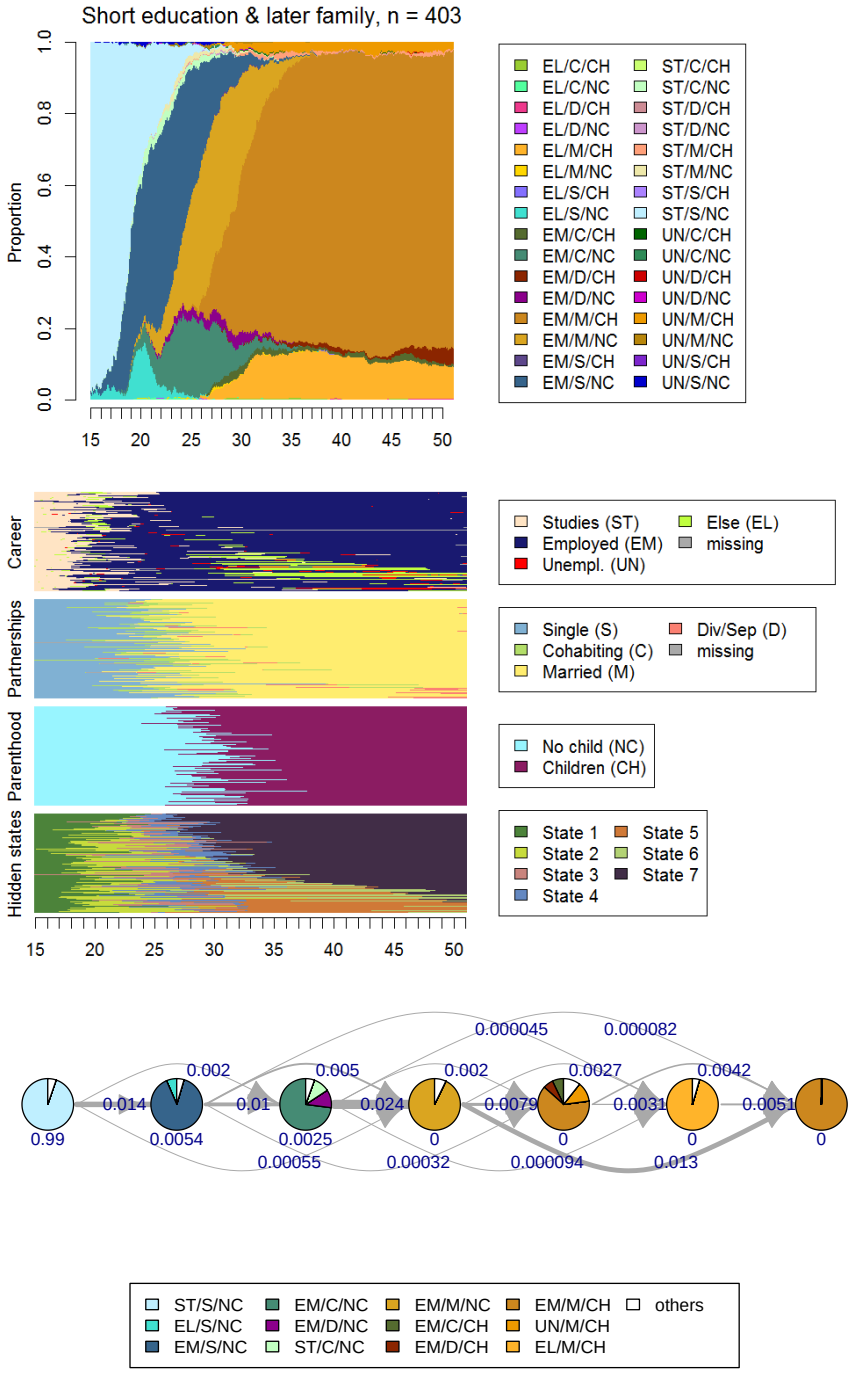
Appendix

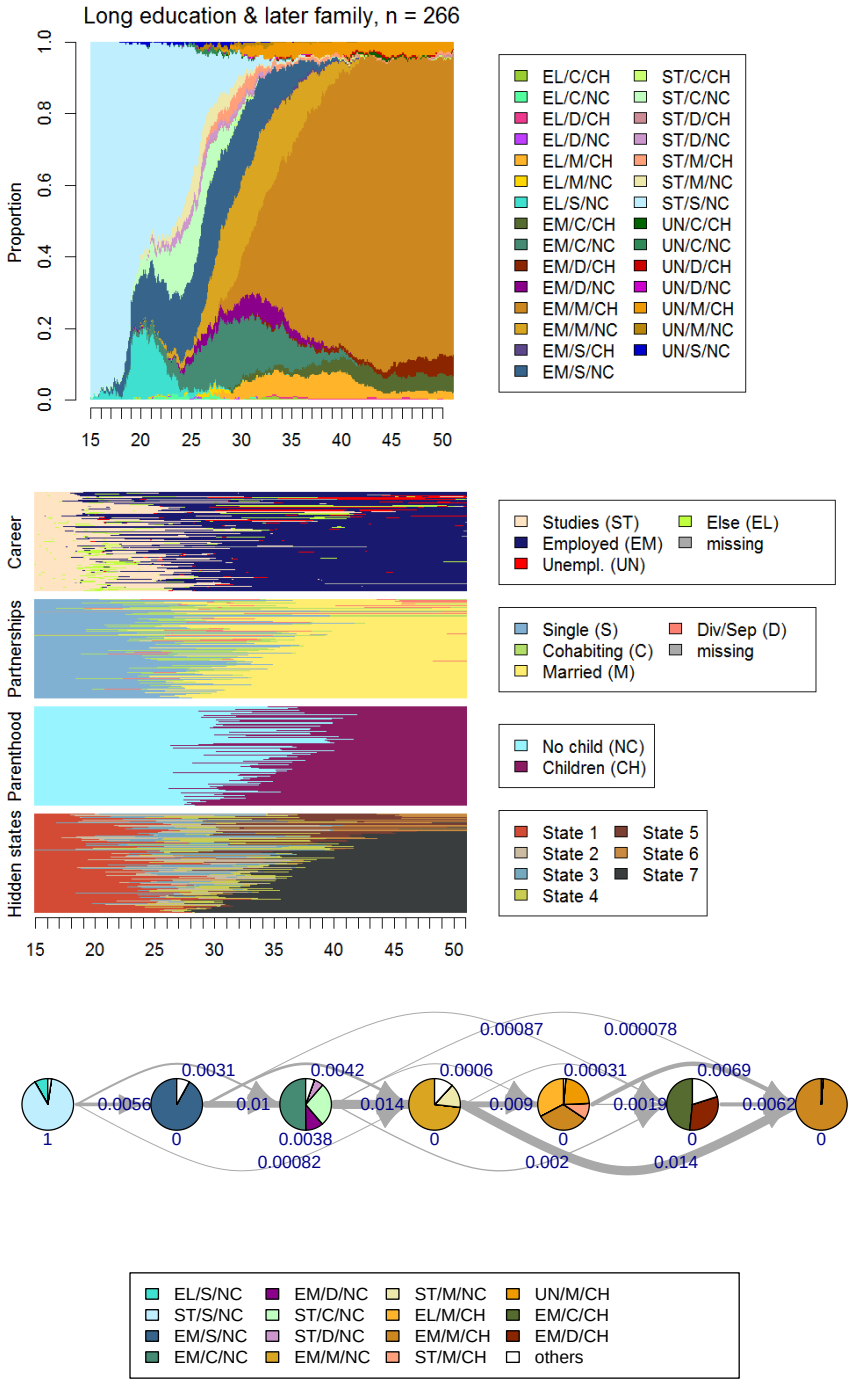
Detailed visualizations of the eight SA clusters and the respective HMMs. Figures show state distributions of combined observations at each time point (top), observed sequences in three life domains and the most probable hidden state paths given the HMM (middle), as well as HMM graphs with initial and transition probabilities (bottom). See Sect. 5 for more information on how to interpret the visualizations.

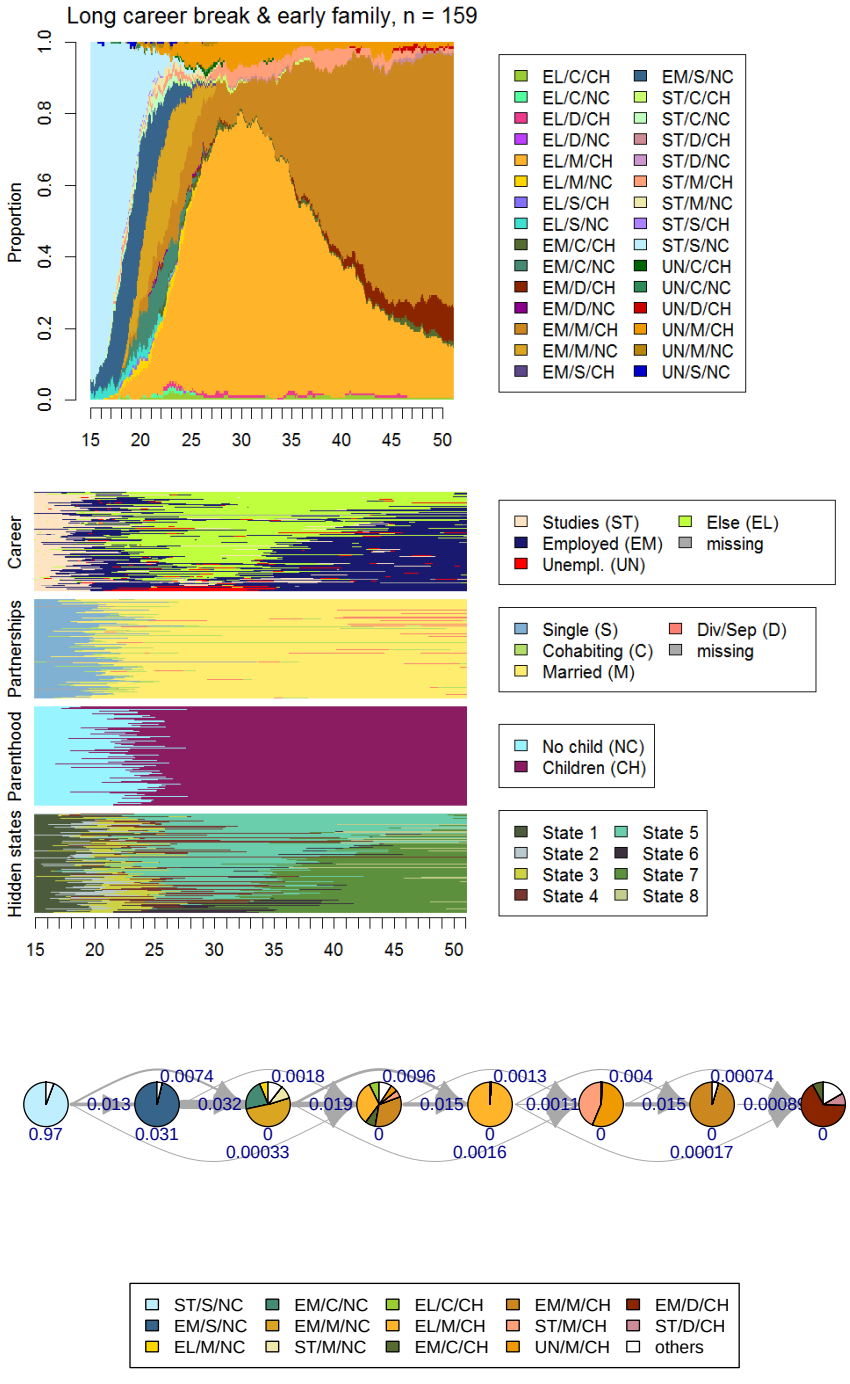

Short education and early family, n = 461
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
50
EL/C/CH
EL/C/NC
EL/D/CH
EL/D/NC
EL/M/CH
EL/M/NC
EL/S/CH
EL/S/NC
EM/C/CH
EM/C/NC
EM/D/CH
EM/D/NC
EM/M/CH
EM/M/NC
EM/S/CH
EM/S/NC
ST/C/CH
ST/C/NC
ST/D/CH
ST/D/NC
ST/M/CH
ST/M/NC
ST/S/CH
ST/S/NC
UN/C/CH
UN/C/NC
UN/D/CH
UN/D/NC
UN/M/CH
UN/M/NC
UN/S/NC
Career
Partnerships
Parenthood
Hidden states
Studies (ST)
Employed (EM)
Unempl. (UN)
Else (EL)
missing
Single (S)
Cohabiting (C)
Married (M)
Div/Sep (D)
missing
No child (NC)
Children (CH)
missing
State 1
State 2
State 3
State 4
State 5
State 6
State 7

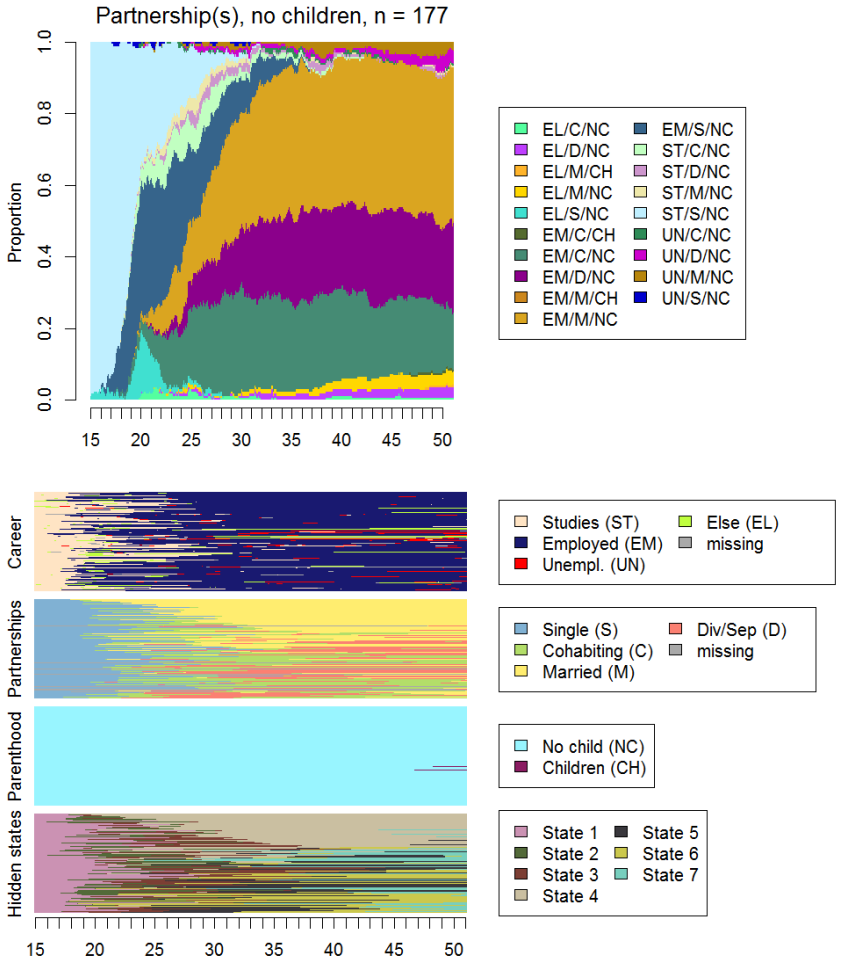

0.99
0.016
0.0054
0.0052
0.025
0.0024
0.0071
0.000046
0.017
0.0021
0.0013
0.0013
0.00017
0.0029
0.0036
0
0.0036
0.011
0.0013
0.0053
0
0.014
0
0.00061

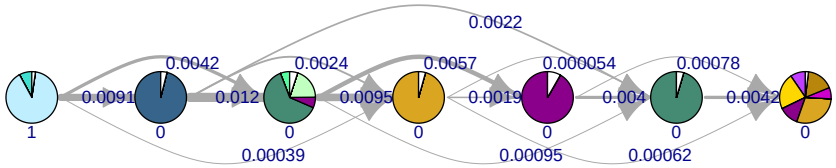

ST/S/NC
EL/S/NC
EM/S/CH
EM/S/NC
EM/C/NC
EM/M/NC
ST/C/NC
ST/M/NC
EM/M/CH
EM/C/CH
EM/D/CH
EL/M/CH
ST/M/CH
UN/M/CH
others


Short education & later family, n = 403
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
50
EL/C/CH
EL/C/NC
EL/D/CH
EL/D/NC
EL/M/CH
EL/M/NC
EL/S/CH
EL/S/NC
EM/C/CH
EM/C/NC
EM/D/CH
EM/D/NC
EM/M/CH
EM/M/NC
EM/S/CH
EM/S/NC
ST/C/CH
ST/C/NC
ST/D/CH
ST/D/NC
ST/M/CH
ST/M/NC
ST/S/CH
ST/S/NC
UN/C/CH
UN/C/NC
UN/D/CH
UN/D/NC
UN/M/CH
UN/M/NC
UN/S/CH
UN/S/NC
Career
Partnerships
Parenthood
Hidden states
Studies (ST)
Employed (EM)
Unempl. (UN)
Else (EL)
missing
Single (S)
Cohabiting (C)
Married (M)
Div/Sep (D)
missing
No child (NC)
Children (CH)
State 1
State 2
State 3
State 4
State 5
State 6
State 7
0.99
0.014
0.002
0.0054
0.01
0.005
0.0025
0.00055
0.024
0.002
0
0.00032
0.0079
0.000045
0.0027
0
0.000094
0.0031
0.000082
0.0042
0
0.013
0.0051
0
ST/S/NC
EL/S/NC
EM/S/NC
EM/C/NC
EM/D/NC
ST/C/NC
EM/M/NC
EM/C/CH
EM/D/CH
EM/M/CH
UN/M/CH
EL/M/CH
others

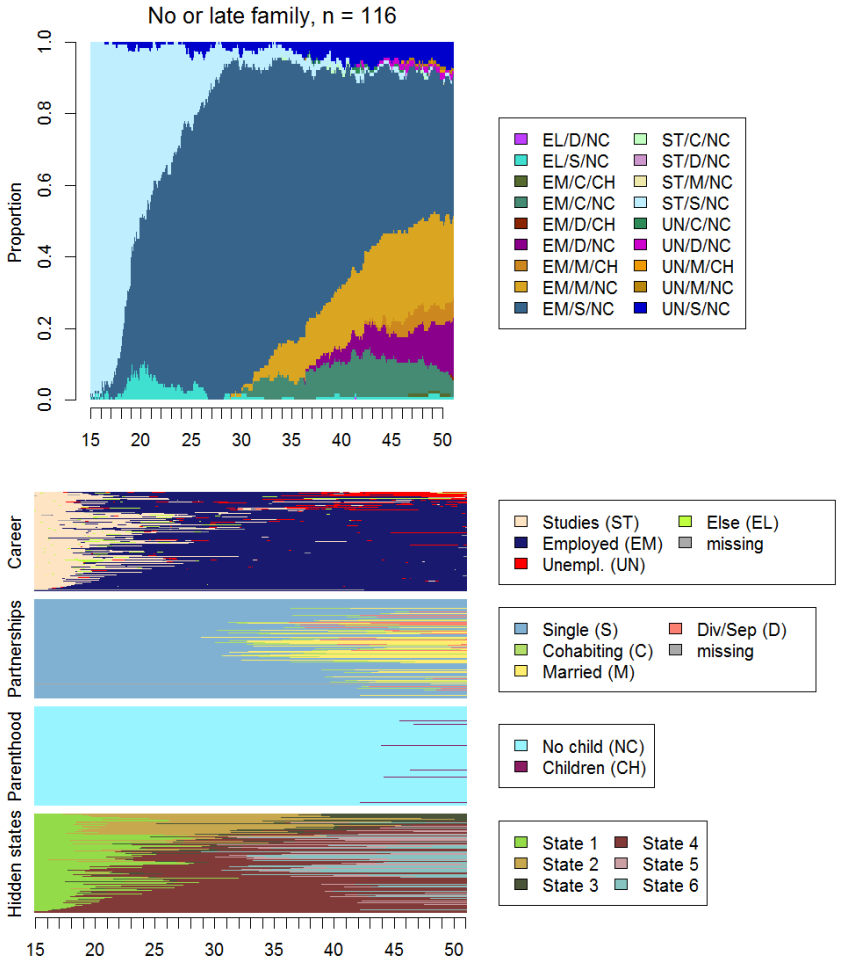

Long education & later family, n = 266
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
50
EL/C/CH
EL/C/NC
EL/D/CH
EL/D/NC
EL/M/CH
EL/M/NC
EL/S/NC
EM/C/CH
EM/C/NC
EM/D/CH
EM/D/NC
EM/M/CH
EM/M/NC
EM/S/CH
EM/S/NC
ST/C/CH
ST/C/NC
ST/D/CH
ST/D/NC
ST/M/CH
ST/M/NC
ST/S/NC
UN/C/CH
UN/C/NC
UN/D/CH
UN/D/NC
UN/M/CH
UN/M/NC
UN/S/NC
Career
Partnerships
Parenthood
Hidden states
Studies (ST)
Employed (EM)
Unempl. (UN)
Else (EL)
missing
Single (S)
Cohabiting (C)
Married (M)
Div/Sep (D)
missing
No child (NC)
Children (CH)
State 1
State 2
State 3
State 4
State 5
State 6
State 7
0.00087
0.000078
0.0031
0.0042
0.0006
0.00031
0.0069
0.0056
0.01
0.014
0.009
0.0019
0.0062
1
0
0.0038
0
0
0
0
0.00082
0.002
0.014
EL/S/NC
ST/S/NC
EM/S/NC
EM/C/NC
EM/D/NC
ST/C/NC
ST/D/NC
EM/M/NC
ST/M/NC
EL/M/CH
EM/M/CH
ST/M/CH
UN/M/CH
EM/C/CH
EM/D/CH
others

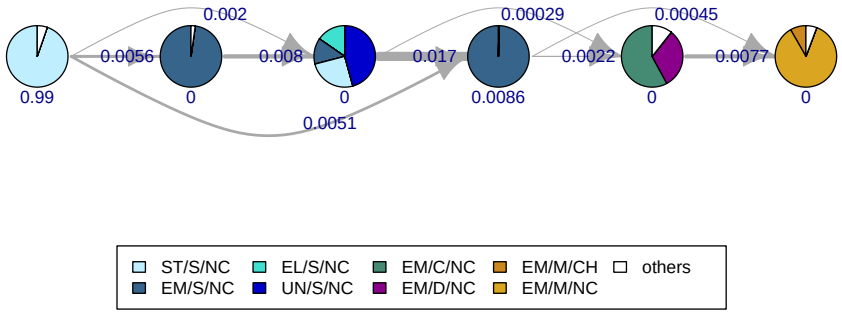

Long career break & early family, n = 159
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
50
EL/C/CH
EL/C/NC
EL/D/CH
EL/D/NC
EL/M/CH
EL/M/NC
EL/S/CH
EL/S/NC
EM/C/CH
EM/C/NC
EM/D/CH
EM/D/NC
EM/M/CH
EM/M/NC
EM/S/CH
EM/S/NC
ST/C/CH
ST/C/NC
ST/D/CH
ST/D/NC
ST/M/CH
ST/M/NC
ST/S/CH
ST/S/NC
UN/C/CH
UN/C/NC
UN/D/CH
UN/M/CH
UN/M/NC
UN/S/NC
Career
Partnerships
Parenthood
Hidden states
Studies (ST)
Employed (EM)
Unempl. (UN)
Else (EL)
missing
Single (S)
Cohabiting (C)
Married (M)
Div/Sep (D)
missing
No child (NC)
Children (CH)
State 1
State 2
State 3
State 4
State 5
State 6
State 7
State 8
0.97
0.013
0.0074
0.031
0.032
0.0018
0
0.00033
0.019
0.0096
0
0.015
0.0013
0
0.0016
0.0011
0.004
0
0.015
0.00074
0
0.00017
0.00089
0
ST/S/NC
EM/S/NC
EL/M/NC
EM/C/NC
EM/M/NC
ST/M/NC
EL/C/CH
EL/M/CH
EM/C/CH
EM/M/CH
ST/M/CH
UN/M/CH
EM/D/CH
ST/D/CH
others

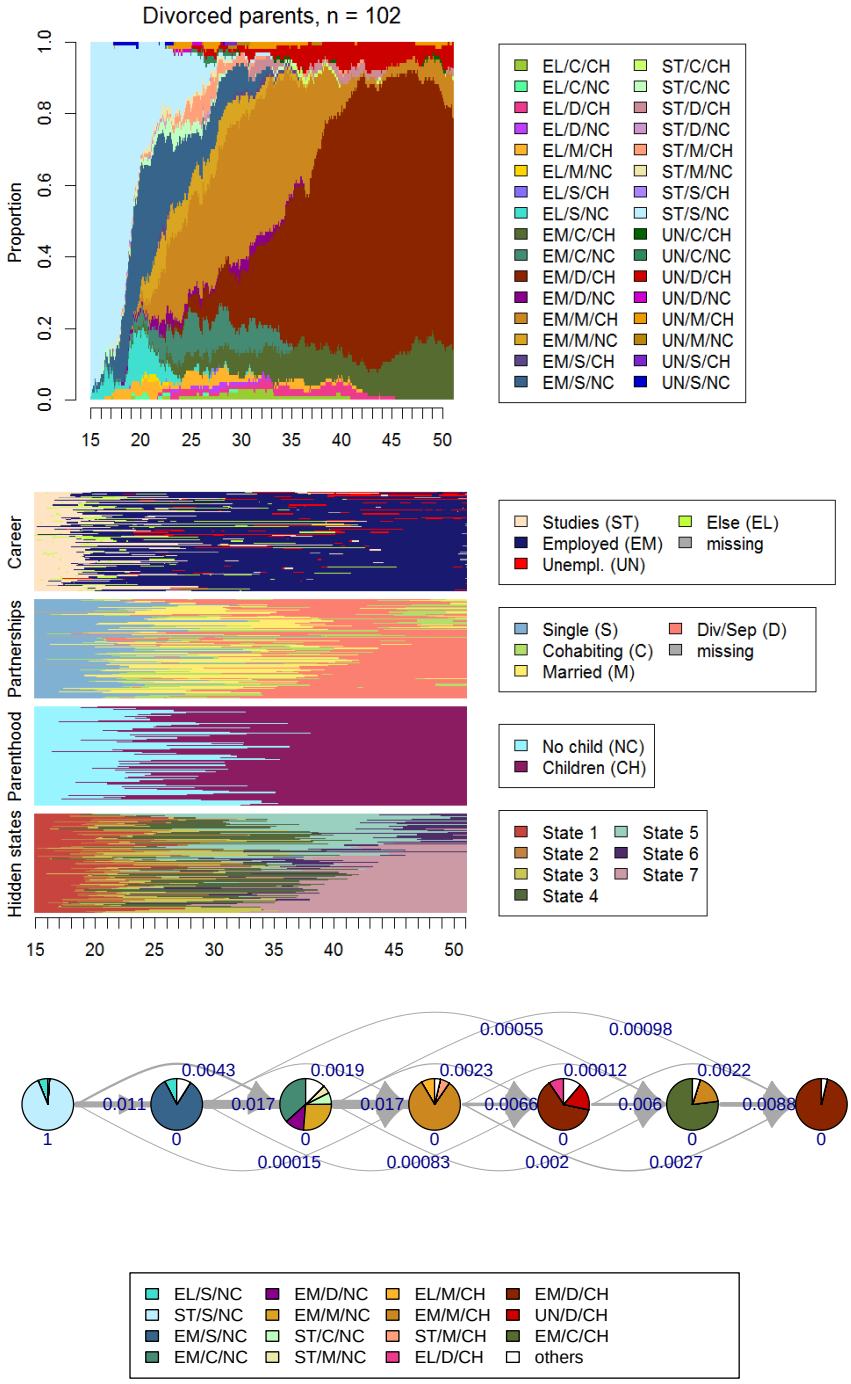

Partnership(s), no children, n = 177
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
50
EL/C/NC
EL/D/NC
EL/M/CH
EL/M/NC
EL/S/NC
EM/C/CH
EM/C/NC
EM/D/NC
EM/M/CH
EM/M/NC
EM/S/NC
ST/C/NC
ST/D/NC
ST/M/NC
ST/S/NC
UN/C/NC
UN/D/NC
UN/M/NC
UN/S/NC
Career
Partnerships
Parenthood
Hidden states
Studies (ST)
Employed (EM)
Unempl. (UN)
Else (EL)
missing
Single (S)
Cohabiting (C)
Married (M)
Div/Sep (D)
missing
No child (NC)
Children (CH)
State 1
State 2
State 3
State 4
State 5
State 6
State 7

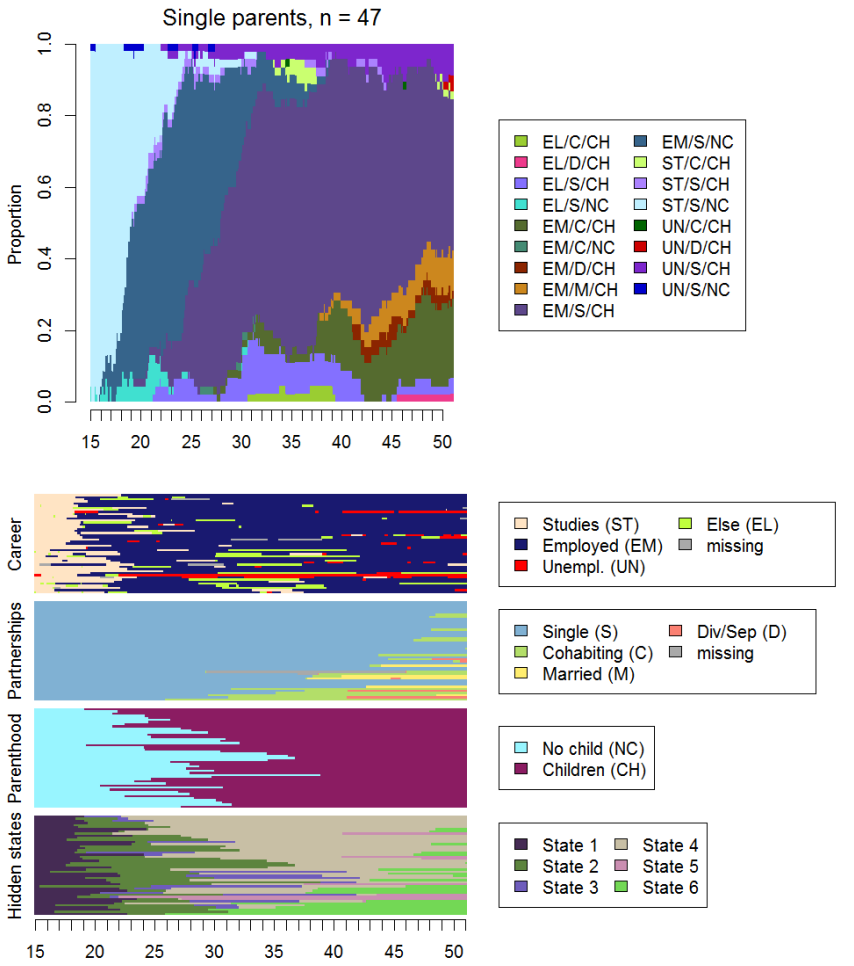

0.0022
0.0042
0.0024
0.0057
0.000054
0.00078
0.0091
0.012
0.0095
0.0019
0.004
0.0042
1
0
0
0
0
0
0
0.00039
0.00095
0.00062

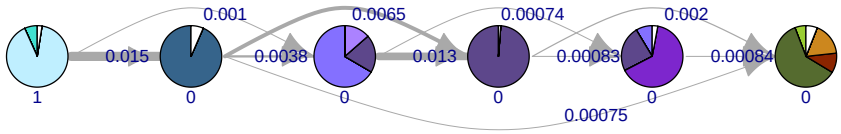

EL/S/NC
ST/S/NC
EM/S/NC
EL/C/NC
EM/C/NC
EM/D/NC
ST/C/NC
EM/M/NC
EL/D/NC
EL/M/NC
UN/D/NC
UN/M/NC
others

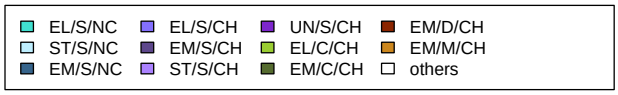

No or late family, n = 116
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
50
EL/D/NC
EL/S/NC
EM/C/CH
EM/C/NC
EM/D/CH
EM/D/NC
EM/M/CH
EM/M/NC
EM/S/NC
ST/C/NC
ST/D/NC
ST/M/NC
ST/S/NC
UN/C/NC
UN/D/NC
UN/M/CH
UN/M/NC
UN/S/NC
Career
Partnerships
Parenthood
Hidden states
Studies (ST)
Employed (EM)
Unempl. (UN)
Else (EL)
missing
Single (S)
Cohabiting (C)
Married (M)
Div/Sep (D)
missing
No child (NC)
Children (CH)
State 1
State 2
State 3
State 4
State 5
State 6


0.99
0.0056
0
0.002
0.008
0
0.0051
0.017
0.0086
0.00029
0.0022
0
0.00045
0.0077
0
ST/S/NC
EM/S/NC
EL/S/NC
UN/S/NC
EM/C/NC
EM/D/NC
EM/M/CH
EM/M/NC
others


Divorced parents, n = 102
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
50
EL/C/CH
EL/C/NC
EL/D/CH
EL/D/NC
EL/M/CH
EL/M/NC
EL/S/CH
EL/S/NC
EM/C/CH
EM/C/NC
EM/D/CH
EM/D/NC
EM/M/CH
EM/M/NC
EM/S/CH
EM/S/NC
ST/C/CH
ST/C/NC
ST/D/CH
ST/D/NC
ST/M/CH
ST/M/NC
ST/S/CH
ST/S/NC
UN/C/CH
UN/C/NC
UN/D/CH
UN/D/NC
UN/M/CH
UN/M/NC
UN/S/CH
UN/S/NC
Career
Partnerships
Parenthood
Hidden states
Studies (ST)
Employed (EM)
Unempl. (UN)
Else (EL)
missing
Single (S)
Cohabiting (C)
Married (M)
Div/Sep (D)
missing
No child (NC)
Children (CH)
State 1
State 2
State 3
State 4
State 5
State 6
State 7
0.00055
0.00098
0.0043
0.0019
0.0023
0.00012
0.0022
0.011
0.017
0.017
0.0066
0.006
0.0088
1
0
0
0
0
0
0
0.00015
0.00083
0.002
0.0027
EL/S/NC
ST/S/NC
EM/S/NC
EM/C/NC
EM/D/NC
EM/M/NC
ST/C/NC
ST/M/NC
EL/M/CH
EM/M/CH
ST/M/CH
EL/D/CH
EM/D/CH
UN/D/CH
EM/C/CH
others


Single parents, n = 47
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
50
EL/C/CH
EL/D/CH
EL/S/CH
EL/S/NC
EM/C/CH
EM/C/NC
EM/D/CH
EM/M/CH
EM/S/CH
EM/S/NC
ST/C/CH
ST/S/CH
ST/S/NC
UN/C/CH
UN/D/CH
UN/S/CH
UN/S/NC
Career
Partnerships
Parenthood
Hidden states
Studies (ST)
Employed (EM)
Unempl. (UN)
Else (EL)
missing
Single (S)
Cohabiting (C)
Married (M)
Div/Sep (D)
missing
No child (NC)
Children (CH)
State 1
State 2
State 3
State 4
State 5
State 6


0.001
0.0065
0.00074
0.002
0.015
0.0038
0.013
0.00083
0.00084
0.00075
1
0
0
0
0
0


EL/S/NC
ST/S/NC
EM/S/NC
EL/S/CH
EM/S/CH
ST/S/CH
UN/S/CH
EL/C/CH
EM/C/CH
EM/D/CH
EM/M/CH
others

References

Aassve, A., Billari, F. C., and Piccarreta, R. (2007). Strings of adulthood: A sequence analysis of young British women's work-family trajectories. *European Journal of Population/Revue européenne de Démographie*, 23(3-4):369–388.

Aisenbrey, S. and Fasang, A. (2010). New life for old ideas: The "second wave" of sequence analysis – bringing the "course" back into the life course. *Sociological Methods & Research*, 38(3):420–462.

Bartolucci, F., Pennoni, F., and Francis, B. (2007). A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(1):115–132.

Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 67(6):1554–1563.

Blossfeld, H.-P., Roßbach, H.-G., von Maurice, J., Schneider, T., Kiesl, S. K., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S., Prenzel, M. S., et al. (2011). Education as a lifelong process-the German National Educational Panel Study (NEPS). *Age*, 74(73):72.

Crayen, C., Eid, M., Lischetzke, T., Courvoisier, D. S., and Vermunt, J. K. (2012). Exploring dynamics in mood regulation—mixture latent Markov modeling of ambulatory assessment data. *Psychosomatic Medicine*, 74(4):366–376.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge, UK: Cambridge University Press.

Eerola, M. and Helske, S. (2016). Statistical analysis of life history calendar data. *Statistical Methods in Medical Research*, 25(2):571–597.

Elzinga, C. H. and Studer, M. (2014). Spell sequences, state proximities, and distance metrics. *Sociological Methods & Research*, pages 3–47.

Gabadinho, A., Ritschard, G., Müller, N. S., and Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4):1–37.

Gauthier, J.-A., Bühlmann, F., and Blanchard, P. (2014). Introduction: Sequence analysis in 2014. In *Advances in Sequence Analysis: Theory, Method, Applications*, pages 1–17. Springer.

Gauthier, J.-A., Widmer, E. D., Bucher, P., and Notredame, C. (2010). Multichannel sequence analysis applied to social science data. *Sociological Methodology*, 40(1):1–38.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2):147–160.

Helske, S. and Helske, J. (2016). Mixture hidden Markov models for sequence data: the seqHMM package in R. *Submitted*.

Helske, S., Steele, F., Kokko, K., Räikkönen, E., and Eerola, M. (2015). Partnership formation and dissolution over the life course: applying sequence analysis and event history analysis in the study of recurrent events. *Longitudinal and Life*

Course Studies, 6(1):1–25.

Lesnard, L. (2010). Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociological Methods & Research*, 38(3):389–419.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2015). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.0.3.

McVicar, D. and Anyadike-Danes, M. (2002). Predicting successful and unsuccessful transitions from school to work by using sequence methods. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(2):317–334.

Müller, N. S., Sapin, M., Gauthier, J.-A., Orita, A., and Widmer, E. D. (2012). Pluralized life courses? an exploration of the life trajectories of individuals with psychiatric disorders. *International Journal of Social Psychiatry*, 58(3):266–277.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Spallek, M., Haynes, M., and Jones, A. (2014). Holistic housing pathways for Australian families through the childbearing years. *Longitudinal and Life Course Studies*, 5(2):205–226.

Studer, M. and Ritschard, G. (2016). What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2):481–511.

Studer, M., Ritschard, G., Gabadinho, A., and Müller, N. (2011). Discrepancy analysis of state sequences. *Sociological Methods & Research*, 40(3):471–510.

van de Pol, F. and Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, 20:213–247.

Vermunt, J. K., Tran, B., and Magidson, J. (2008). *Latent Class Models in Longitudinal Research*, pages 373–385. Handbook of Longitudinal Research: Design, Measurement, and Analysis. Elsevier, Burlington, MA.

Whiting, R. and Pickett, E. (1988). On model order estimation for partially observed Markov chains. *Automatica*, 24(4):569–572.