LIVES WORKING PAPER 2020 / 85

COMPARISON OF TWO APPROACHES IN MULTICHANNEL ANALYSIS

KEVIN EMERY AND ANDRÉ BERCHTOLD

RESEARCH PAPER

http://dx.doi.org/10.12682/lives.2296-1658.2020.85

ISSN 2296-1658

The National Centres of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation

Authors

Kevin Emery (1)

André Berchtold (1, 2)

Abstract

Sequence analysis is an established approach to study life courses. In this context, multichannel sequence analysis (MSA) and the extended alphabet (EA) approach are the most frequently used strategies when multiple life domains are considered simultaneously. We compare these two methods using real data composed of four life domains. We focus on clustering since sequence analysis usually aims to identify typical patterns in sequences. The analysis is first done on the full dataset. Then, since at least professional status trajectories proved to be different between men and women and, potentially, their link with the other domains, the same analyses are run separately by sex. Finally, two extreme cases of optimal matching, namely Levenstein and Hamming distance, are explored. Neither of the approaches is clearly superior. Indeed, the results of both methods are often close. Although MSA is generally easier to use and applies to a broader range of situations, EA can provide original typologies in some cases.

Keywords

Sequence analysis | multiple life domains | real data | clustering | multichannel sequence analysis |extended alphabet

Authors affiliation

(1) Swiss National Centre of Competence in Research LIVES

(2) Institute of Social Sciences, University of Lausanne, Switzerland

Correspondence to

kevin.emery@unil.ch

Acknowledgements

This paper benefited from the support of the Swiss National Centre of Competence in Research LIVES - Overcoming vulnerability: Life course perspectives (NCCR LIVES), which is financed by the Swiss National Science Foundation (grant number: 51NF40-185901). The authors are grateful to the Swiss National Science Foundation for its financial assistance.

Comparison of two approaches in multichannel analysis

Kevin Emery[1], André Berchtold[1,2]

[1]Swiss National Centre of Competence in Research LIVES
[2]Institute of Social Sciences, University of Lausanne, Switzerland

Corresponding author: kevin.emery@unil.ch (Kevin Emery)

November 2020

Abstract

Sequence analysis is an established approach to study life courses. In this context, multichannel sequence analysis (MSA) and the extended alphabet (EA) approach are the most frequently used strategies when multiple life domains are considered simultaneously. We compare these two methods using real data composed of four life domains. We focus on clustering since sequence analysis usually aims to identify typical patterns in sequences. The analysis is first done on the full dataset. Then, since at least professional status trajectories proved to be different between men and women and, potentially, their link with the other domains, the same analyses are run separately by sex. Finally, two extreme cases of optimal matching, namely Levenstein and Hamming distance, are explored. Neither of the approaches is clearly superior. Indeed, the results of both methods are often close. Although MSA is generally easier to use and applies to a broader range of situations, EA can provide original typologies in some cases.

Keywords: sequence analysis, multiple life domains, real data, clustering, multichannel sequence analysis, extended alphabet

1 Introduction

Life course analysis is concerned with the many events that punctuate the lives of individuals from birth to death. The focus is often on several supposedly interrelated domains, the idea being that resources, behaviours, and goals in one domain are linked with the resources, behaviours, and goals of other domains (Bernardi et al., 2019). Therefore, to fully understand a given life domain, its linked domains are considered simultaneously. One of the most striking examples is that of work and family, where numerous studies have demonstrated, for example, the impact of the birth of children on women's occupational trajectories (Piccarreta and Billari, 2007; Widmer and Ritschard, 2009; Aisenbrey and Fasang, 2017).

Sequence analysis, a central tool in the study of life courses, aims to determine their most important features. Possible situations occurring during the life course are represented by a finite set of mutually exclusive states, whose succession over time is called a sequence. These sequences are considered as a whole, with the idea that events cannot be isolated from each other (Piccarreta and Studer, 2019). A standard sequence analysis is typically conducted by computing the pairwise dissimilarities between the sequences of different individuals before applying clustering to identify a typology (Abbott and Tsay, 2000). Optimal matching, which was first applied to social sciences by Abbott and Forrest (1986), is often used to compute these pairwise dissimilarities. In this framework, the minimum effort necessary to change one sequence into another through the insertion, deletion, and substitution of states is determined. Since the introduction of optimal matching to social sciences, many variations of optimal matching and other types of dissimilarities have been provided (Studer and Ritschard, 2016). However, basic optimal matching remains the most often used approach. Sequence analysis has often been applied in the life course approach in such domains as transition into adulthood (Oris and Ritschard, 2014; Lorentzen et al., 2019), work pathways (Malin and Wise, 2018; Wahrendorf et al., 2018), and union trajectories (Jalovaara and Fasang, 2017).

Multiple sequences are sometimes considered simultaneously. This happens in broadly two types of situations (Studer, 2015). First, these sequences could be subdimensions of the same concept. For example, income and labour market positions are two indicators of careers trajectories (Mattijssen and Pavlopoulos, 2018). Then, sequences from different life domains could be considered simultaneously. The latter is either done to summarize the association between life domains (Spallek et al., 2014) or to reduce the information for a further analysis (Müller et al., 2012). With multiple sequences, joint sequence analysis, which is an extension of standard sequence analysis, can be used. Joint sequence analysis involves the computation of dissimilar-

ities based on all domains (Piccarreta, 2017). The two most common strategies are the extended alphabet (EA) approach and multichannel sequence analysis (MSA) (Gauthier et al., 2010). With the former, the states of each domain are combined to build a single set of super-states, called an extended alphabet, each super-state being defined by combining one state from each of the original domains. Then, to compute the pairwise dissimilarities between the sequences, optimal matching is usually used with substitution costs based on transition rates (Piccarreta and Billari, 2007; Lesnard, 2008). On the contrary, MSA extends optimal matching to the multidimensional case. Concretely, the substitution cost needed to align two multichannel sequences at a given timepoint is defined as the mean, possibly weighted, of the substitution costs needed to align each channel separately. Insertion/deletion (indel) costs are set as half the highest substitution cost for EA, while they are averaged over the different channels for MSA. Moreover, these two strategies are combinable: some channels can be first aggregated before applying MSA, as in Schwanitz (2017), where three channels relative to family where first aggregated before applying MSA to family and professional sequences. Whatever the approach, after the computation of pairwise dissimilarities, a clustering is often used to identify the most typical patterns in the data. However, the application of other sequence analysis tools such as pseudo-ANOVA (Studer et al., 2011) and regression trees (Studer, 2018) is also possible.

The goal of this paper is to compare empirically, via cluster analysis, EA with MSA. To the best of our knowledge, this has never been achieved using real data, the idea being to understand the differences in behavior of EA and MSA, and to determine which method is the most suited depending of the context. Concretely, we use four channels with various interrelations between them in order to consider a broad range of situations. We determine the ability of MSA and EA to produce clusterings that take the link between channels into account, when there is one, summarize individual channels efficiently, and have homogeneous clusters. Moreover, we look at the fundamental differences between EA and MSA clusterings, particularly in terms of duration, timing and sequencing. We use the last methodological developments in multichannel analysis for this purpose.

The remainder of this paper is organised as follows. The empirical dataset is described in Section 2 and the methodological tools used to compare EA with MSA are presented in Section 3. The results are presented in Section 4 and a discussion ends the paper.

2 Data

We used data from the Swiss Household Panel (Tillmann et al., 2016), a yearly panel study that started in 1999. People living in Switzerland are interviewed on different topics such as family, work, and health. In 2013, a third sample of 4093 households comprising 9945 individuals was added into the panel. In addition to the standard questionnaires, a retrospective life history calendar was used (Kühr et al., 2013). Life domains such as residential trajectory, residency, living arrangements, partner relationships and changes in civil status, family events, professional activities, and health issues are investigated from birth to 2013. Here, we considered sequences of individuals between the ages of 20 and 45 who had answered questions in the domains of professional activities, health issues, living arrangements, and family events and who responded without interruption, for a total of 1707 respondents. We then defined four independent channels with the following states:

- Child: *0 to 4 years old*, if at least one child between the age of 0 and 4 lives in the same household, *5 to 18 years old*, if there is at least one child between 5 and 18 but no child between 0 and 4, and *No* otherwise.
- Cohabitational status: living with *Both parents*, living with *One parent*, living with a *Partner*, living *Alone*, and *Other* situations.
- Professional status: *Education*, *Full-time* employment, *Part-time* employment, and *Non-working*.
- Health issues: *Yes* if the person has during the considered year suffered an illness/accident or undergone surgery or psychological issues and *No* otherwise.

Cohabitation and child trajectories are expected to be highly interrelated and are often considered in the literature as a single trajectory. They can be seen as two indicators of the family trajectory. The simultaneous analysis of occupational and family sequences is a typical application of MSA, especially for women, while we do not expect a link between health and any other domain.

Although we did not intend to draw conclusions about the Swiss population, the 1707 respondents were weighted to better account for selection bias and allow us to work with a more representative sample.

3 Methods

As pointed out by Gauthier et al. (2010), a joint analysis is suitable only if the domains are associated. Therefore, the first point is to check whether their degree

of association is sufficient to justify a joint analysis. Even when the channels are indicators of the same dimension, their supposed interrelation need to be verified. For that purpose, Piccarreta (2017) extended the Cronbach's α and principal component analysis (PCA) approaches. Consider p dissimilarity matrices $D_1, \ldots, D_p$ containing the pairwise dissimilarities computed on p domains. One can then consider $d_1, \ldots, d_p$, the vectors of the respective upper triangular values of the matrices $D_1, \ldots, D_p$, as p measurements of the same concept. Cronbach's α can then be applied to assess the similarity between the measurements. PCA can also be applied to the set of vectors $d_1, \ldots, d_p$ to detect the relations between domains by inspecting the loadings. Moreover, when a joint sequence analysis is selected and its related dissimilarity matrix D_{JSA} is computed, the correlations between the $d_s, s = 1, \ldots, p$, and d_{JSA} shed light on how well the domain-specific information is summarised by the joint method. All these measures depend on the choice of dissimilarity measure used to compute the pairwise dissimilarities on the individual domains.

To derive a typology, a hierarchical clustering with Ward linkage is commonly used with sequences. The average silhouette width (ASW) (Rousseeuw, 1987) and Hubert's C index (HC) (Hubert and Levin, 1976) are standard criteria for selecting the number of clusters. For each element, the silhouette value is a comparison between the cohesion of this element in its assigned cluster and its separation from other clusters. The ASW is the mean of these values. It ranges from -1 to 1; the higher the value, the better it is. When the data are weighted to account for selection bias, the weighted version of the ASW (ASWw) is used (Studer, 2013). The ASW can also be computed independently for each cluster to determine its internal cohesion. HC compares the sum of the obtained within-cluster distances with the minimum possible value with the same distance and number of groups. Contrary to the ASW, a smaller value is better for HC.

On the one hand, a clustering provides results even if the domains are fully dissociated. On the other hand, even with channels that are supposedly interrelated, some clusterings could take accordingly their link into account, while others may not. Therefore, to validate a joint typology, Studer (2019) extended the work of Hennig and Liao (2010) and Hennig and Lin (2015) to sequences to study the behaviour of a clustering quality measure with similar but unstructured data using permutation tests. More precisely, with multichannel sequences, one channel is kept fixed and the others are randomly permuted. A clustering is then applied and a cluster quality index is computed. This process is repeated many times, usually 1000 times, to build a bootstrap confidence interval for the desired quality index under the hypothesis that the domains are not associated. Comparing the value obtained by clustering the

empirical data with that from clustering the unstructured data, we can determine whether the clustering takes into account the interrelation between the domains or not. When more than two domains are involved, this procedure can also be applied to ensure that a given domain is taken into account by the clustering. In this context, the sequences of each domain are fixed except for those of the considered domain, which are randomly permuted. A bootstrap confidence interval for the desired cluster quality index is then built under the hypothesis that the domain of interest is not associated with the others. Comparing the bootstrap confidence interval with the value obtained from the empirical data allows us to deduce whether the clustering takes into account the individual domain or not.

Another useful tool to determine if a clustering takes into account each individual domain is the domain-specific R^2 (Piccarreta, 2017). This represents the share of the total pairwise dissimilarities of a domain explained by a clustering. If the R^2 computed on an individual domain is low, the clustering does not explain this domain satisfactorily.

We compared MSA and EA on their ability to produce meaningful results in different contexts. To do so, we first determined if the domains are linked using Cronbach's alpha and PCA. Since we did not have any clear cut value to decide that some domains are linked or not, we studied a large range of possibilities. This also gave a broad idea on the behavior of Cronbach's alpha and PCA with real data. Then, for each of the two methods, we determined the clusterings that were significant according to bootstrap validation. Even in the case of cohabitational status and child trajectories, which are supposedly intrinsically linked, the bootstrap procedure is useful to discard the clusterings that did not take accordingly the link into account. We then compared how the clusterings fundamentally differ between the two methods using chronograms and index plots. We particularly focused on the central aspects structuring sequences, namely timing (i.e the age of an individual in a specific state), sequencing (i.e the ordering of the states) and duration (i.e the age of an individual in a specific state). Moreover, when deriving a joint typology, it is desirable that each cluster is well-defined, that all channels are considered equally and efficiently, and that the result represents a decent part of the population. We determined how the clusterings satisfy these criteria.

We separated the analysis into three parts. First, the full dataset of 1707 individuals was analysed. Then, since at least the professional status trajectories proved to be different between men and women as well as, potentially, their relationship with the other domains, the same analyses were run separately by sex. Finally, we explored the two extreme cases of standard optimal matching, namely, when only

substitution costs are involved (Hamming distance) and when only indel costs are used (Levenstein distance). Indeed, as noted by Studer and Ritschard (2016), one can control the sensitivity of optimal matching to differences in timing and duration by controlling the trade-off between the substitution and indel costs. With the Hamming distance, the sensitivity is at its maximum for timing and minimum for duration, whereas the opposite is true with the Levenstein distance. Considering these two extreme cases of optimal matching thus provides a better comprehension of the behaviour of the EA and MSA methods. All the computations were performed within the R statistical environment (R Core Team, 2019). Specifically, the TraMineR package (Gabadinho et al., 2011) and WeightedCluster package (Studer, 2013) were used for most of the analyses.

4 Results

This section presents the comparison of the EA and MSA methods using our empirical data. First, we used the full dataset (see Section 4.1), then the analyses were performed separately by sex (4.2), and finally we considered two extreme cases of optimal matching based on the Hamming and Levenstein distances (4.3).

4.1 Full dataset

Figure 1 summarizes the full dataset by showing the chronograms of the four domains for the 1707 respondents, and we first used the tools developed by Piccarreta (2017) to assess the interrelation between the domains. We restricted ourselves to the standard features of the algorithms because we wanted to concentrate on the substantial differences between the EA and MSA approaches, without the risk of perturbations caused by subtle variations in the algorithms. Therefore, in each domain, optimal matching was used with a substitution cost of 1 and an indel cost of 0.5 to compute the dissimilarity between each pair of sequences. Considering all the domains together produced a Cronbach's α of 0.26, which is low. Removing the health issues or professional status domain slightly increased the value, while discarding the child or cohabitational status trajectories produced a Cronbach's α close to zero. Only considering the child and cohabitational status domains provided the best Cronbach's α (0.53). This was still below the usual acceptable threshold of 0.7; however, how the raw Cronbach's α value is interpretable in the context of sequences is unclear. The computation of Cronbach's α on each other pair of domains produced values lower than 0.1. Therefore, only the cohabitational status and child domains seemed potentially interrelated. These results were confirmed by

the outcome of the PCA, as shown in Table 1. Indeed, the first principal component was highly associated with the cohabitation and child domains, while the second was mainly related to the professional status domain and the third was almost totally related to health issues. We chose to focus only on the domains possibly suitable for a joint analysis according to the Cronbach's α and PCA, namely, the cohabitational status and child domains with the idea of deriving a family typology.

Table 1: Loadings of the PCA applied to the pairwise dissimilarities computed on the four domains.

Domain	PC1	PC2	PC3
Child	0.83	0.14	-0.04
Cohabitational	0.82	-0.16	0.05
Professional	-0.01	0.99	0.03
Health issues	0.01	0.03	1
Eigenvalues	1.36	1.02	1

In the next step, we performed a joint analysis using MSA and EA alternatively. We again restricted ourselves to the standard features of these algorithms. According to the correlation between the vectors containing the pairwise dissimilarities, both methods seemed to summarise the information from the single domains well, but the results were slightly better for MSA. Indeed, we found correlations of 0.811 and 0.814 between d_{MSA} and, respectively, d_{Child} and d_{Cohab}, while the respective correlations for d_{EA} with the individual domains were 0.752 and 0.749, respectively.

In both the MSA and the EA approaches, the optimal number of groups was somewhere between two and five. Figure 2 shows the considerable drop in ASWw and increase in HC between five and six groups, with these changes more pronounced for EA. In the case of MSA, both cluster quality indices were better for the two-cluster solution, whereas ASWw was the best for two clusters and HC was the best for five clusters in the case of EA. We used the procedure developed by Studer (2019). Figure 3 provides the values obtained with our data, represented by the dots, and the 95% bootstrap confidence intervals for both HC and ASWw under the null hypothesis that the domains are not significantly associated. In the case of EA, both HC and ASWw were significant between two and five groups. Since the relative distance between the dot and confidence interval is at its maximum for the separation in four and five groups, the most significant results were obtained with these two clusterings. The picture was less clear for MSA. Both cluster quality indices were significant for the clustering in two and four groups, but ASWw was not significant for any other number of clusters.

The final step consisted in analysing the different clusterings in two to five groups

and showing how they differ between the MSA and EA approaches. On the one hand, the two-cluster solutions for both MSA (Figure 4) and EA (Figure 5) were almost identical, and these methods mainly separated the individuals according to whether they lived with a child, with the exception of individuals living with a child at an older age after living for a long time alone or having short periods with a child in their household. Both approaches agreed on the assignment of 99% of the sequences. The first group contained about 73% of the individuals and the second one 27%. On the other hand, the results differed for the three-cluster solution. In comparison with the two-cluster solution, the child cluster was split according to the timing of childbirth for MSA (Figures 6 and 7), while for EA the non-child cluster was split with respect to cohabitational status (alone vs partner) (Figure 8). Thus, both approaches involved an additional cluster of individuals living with a child and other types of cohabitational statuses, which led to the four-cluster solutions (Figures 10, 11, 12, and 13). Finally, the separation into five groups provided by MSA involved two clusters of individuals living with a child that depended on the timing, two clusters of people not living with a child split according to cohabitational status, and one cluster of individuals living with a child and alternative cohabitational statuses (Figures 14 and 15). The five-cluster solution provided by EA (Figures 16 and 17) contained three groups of individuals without children. These three groups were split according to cohabitational status: living alone, living with a partner, and living neither alone nor with a partner. In addition, MSA tended to provide more balanced clusters than EA did. Indeed, the largest group in the five-cluster solution of MSA (group 1) represented about 48% of the total sample and three of the other groups represented more than 10% of respondents each. By contrast, group 1 of the five-cluster solution of EA represented about 68% of the total sample and only one of the four other groups comprised more than 10% of respondents.

On the basis of the MSA approach, one would select either the two- or the four-group solution. The separation into two groups was more significant according to bootstrap validation and the groups were relatively homogeneous with an ASWw of 0.48 for the first cluster and 0.38 for the second one (Table 2). The clustering was more driven by the child domain since the R^2 values of the individual domains were 0.8 (child domain) and 0.67 (cohabitational status domain). Although less significant according to the bootstrap validation, the separation into four groups allowed for a more detailed typology, and the clusters were balanced with ASWw values between 0.2 and 0.28. The last group is relatively small though. This is more problematic when the typology is the goal because the generalisability of the results is limited than when the typology is built to reduce the information for a further analysis.

For the EA approach, the clustering in five groups was the most suitable. It was, with the four-group clustering, the most significant one according to the bootstrap validation, and we observed a clear drop in terms of ASWw between the five-group and six-group solutions. Moreover, this explained almost the same share of the discrepancy by individual domain (0.8 for the child and 0.79 for the cohabitational status domain). However, the last group is ill-defined (ASWw by group of 0.02) and the first cluster contains almost 70% of the total weighted sample inducing small clusters. Therefore, the selected typologies differed depending on whether the MSA or EA approaches were used, thereby changing the conclusions drawn from the statistical analyses.

Table 2: Summary of the results obtained by clustering the child and cohabitational status channels with MSA and EA. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	Method	CQI		Group ASWw		Size clusters (%)		R^2 by channel	
		ASWw	HC	min	max	min	max	child	cohab
2	MSA	**0.45**	**0.09**	0.38	0.48	27	73	0.8	0.67
	EA	**0.42**	**0.09**	0.25	0.47	27	73	0.8	0.67
3	MSA	0.24	0.15	0.23	0.25	20	53	0.86	0.69
	EA	**0.4**	**0.08**	0.1	0.51	9	73	0.8	0.72
4	MSA	**0.26**	**0.12**	0.2	0.28	5	48	0.86	0.73
	EA	**0.36**	**0.06**	0.1	0.51	5	68	0.8	0.76
5	MSA	0.25	**0.1**	0.14	0.32	5	48	0.86	0.78
	EA	**0.37**	**0.05**	0.02	0.47	5	68	0.8	0.79
6	MSA	0.17	**0.11**	0.11	0.39	5	40	0.88	0.78
	EA	0.18	0.14	0.02	0.47	5	40	0.84	0.8
7	MSA	0.18	**0.1**	0.05	0.39	5	40	0.88	0.82
	EA	0.18	0.14	0	0.46	5	40	0.8	0.79
8	MSA	0.17	**0.1**	0.04	0.39	5	40	0.88	0.82
	EA	0.19	0.13	0	0.46	5	28	0.85	0.82
9	MSA	0.16	0.1	-0.06	0.35	5	26	0.89	0.83
	EA	0.19	0.13	-0.08	0.44	5	28	0.86	0.84
10	MSA	0.16	0.1	-0.06	0.45	5	26	0.9	0.84
	EA	0.19	0.12	-0.01	0.44	2	28	0.86	0.85

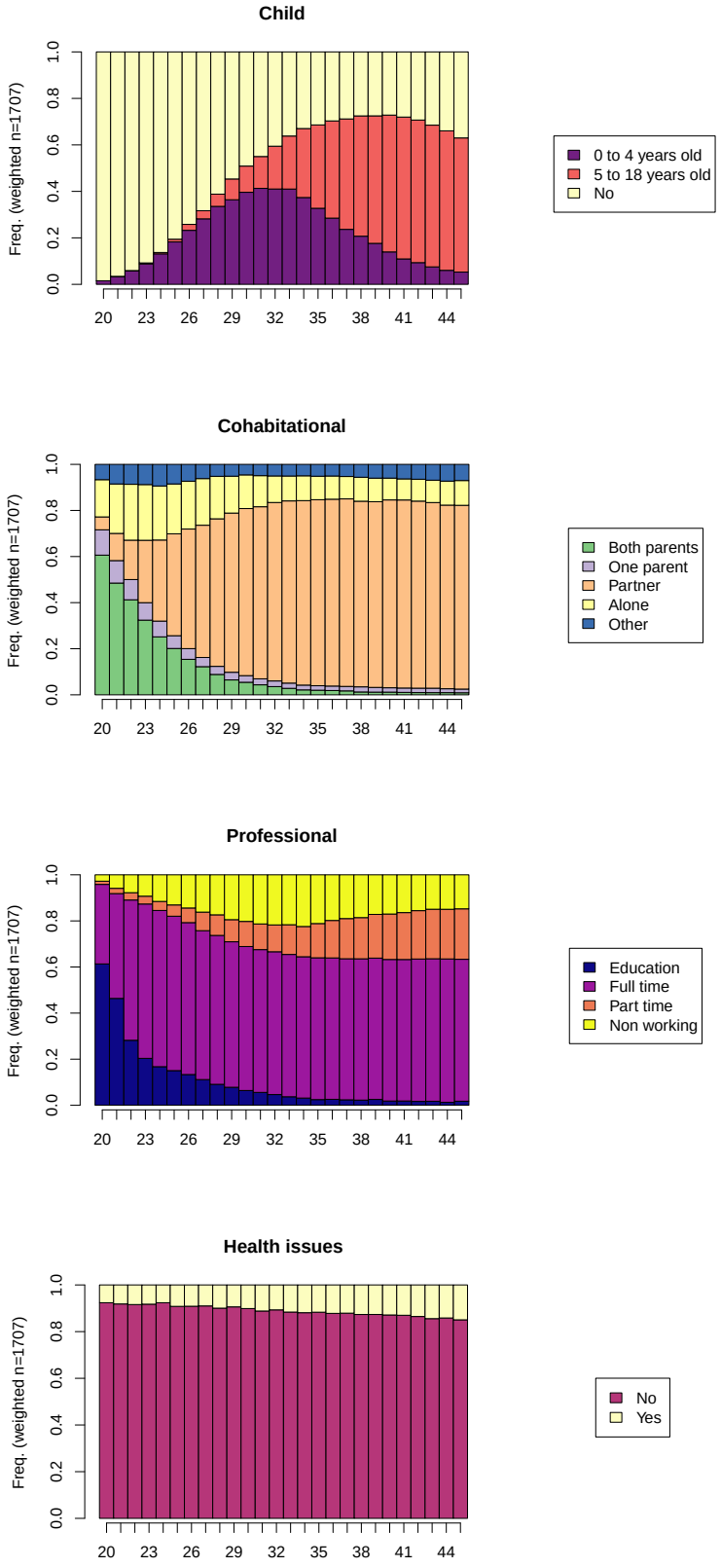

Figure 1: Chronograms for the child, cohabitational, professional and health issues trajectories over the entire dataset.

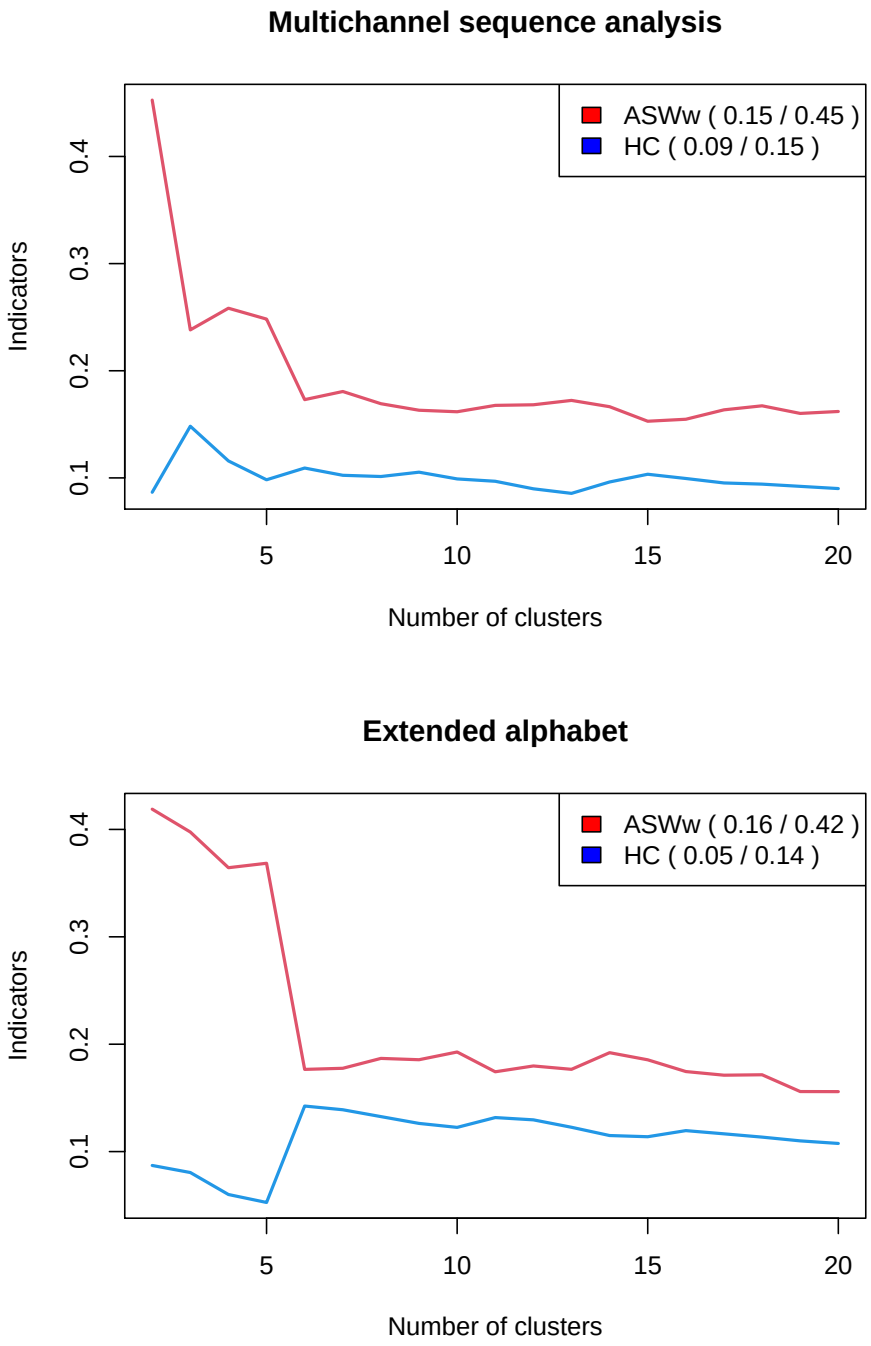

Figure 2: Evolution of the HC and ASWw cluster quality indices according to the number of clusters for MSA and EA.

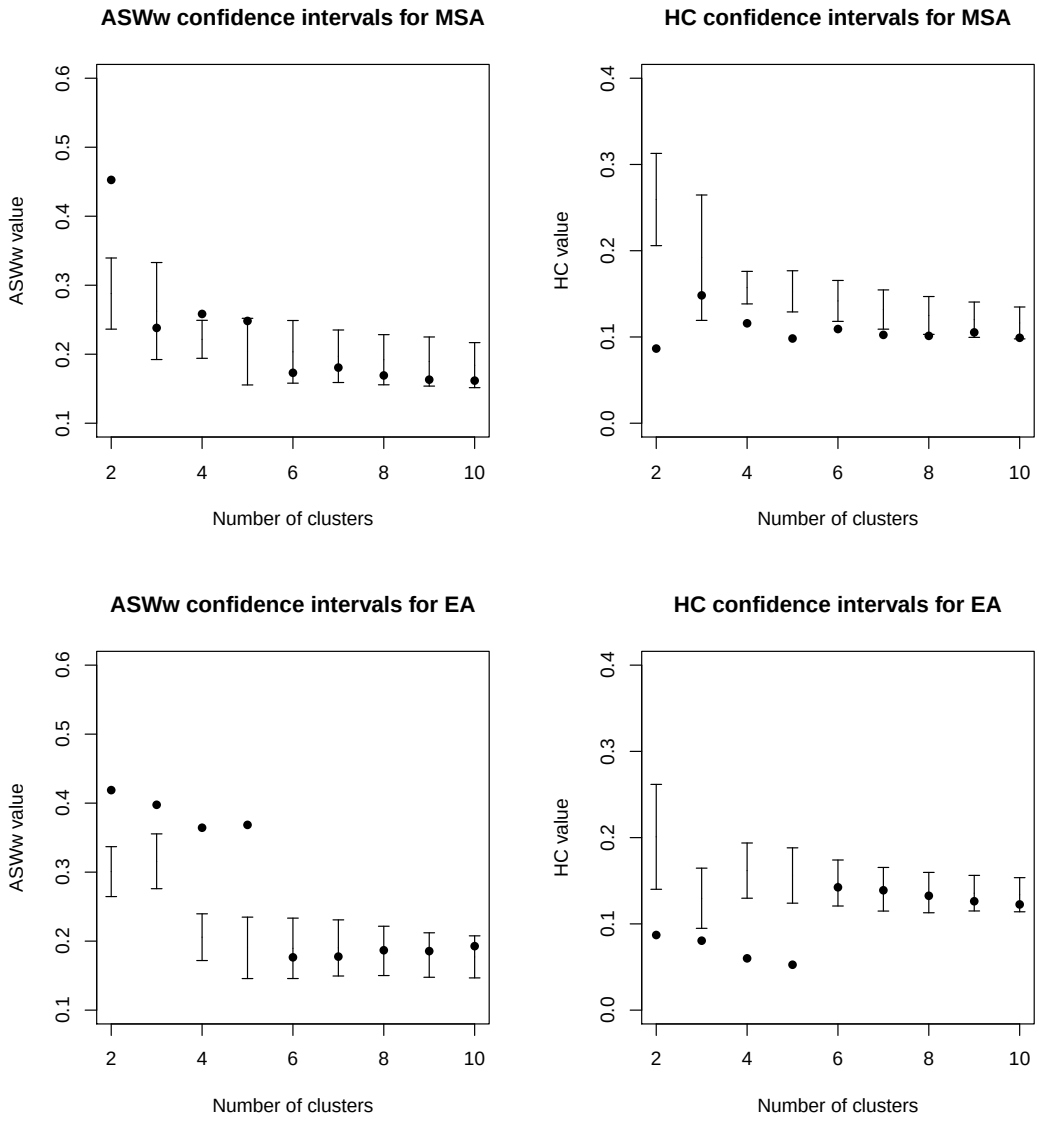

Figure 3: ASWw and HC values obtained by clustering the data, represented by the dots, together with confidence intervals built under the hypothesis that child and cohabitational domains are not associated.

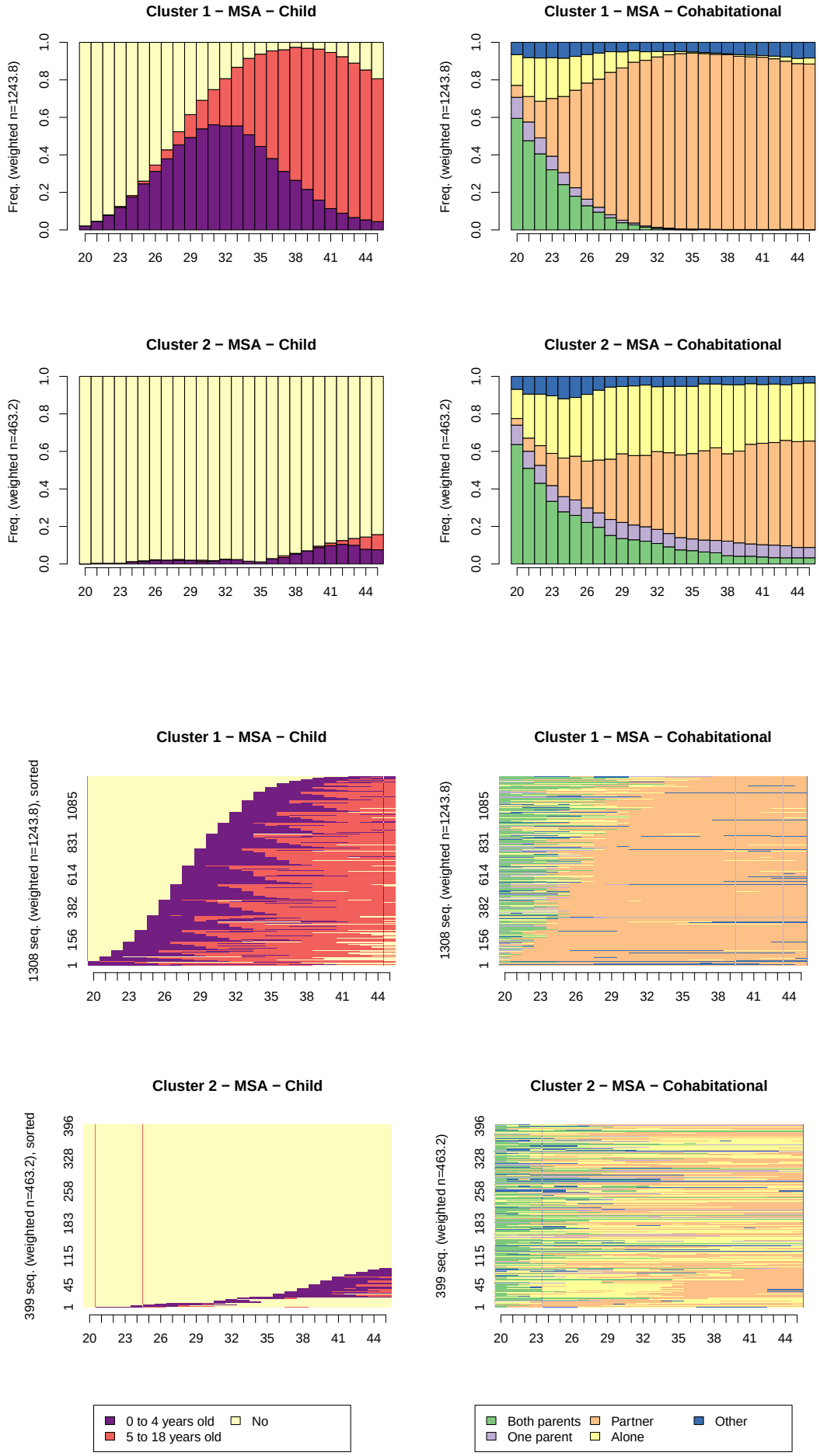

Figure 4: Chronograms (top) and index plots (bottom) of the two-group typology of cohabitational and child status obtained with MSA on the full dataset.

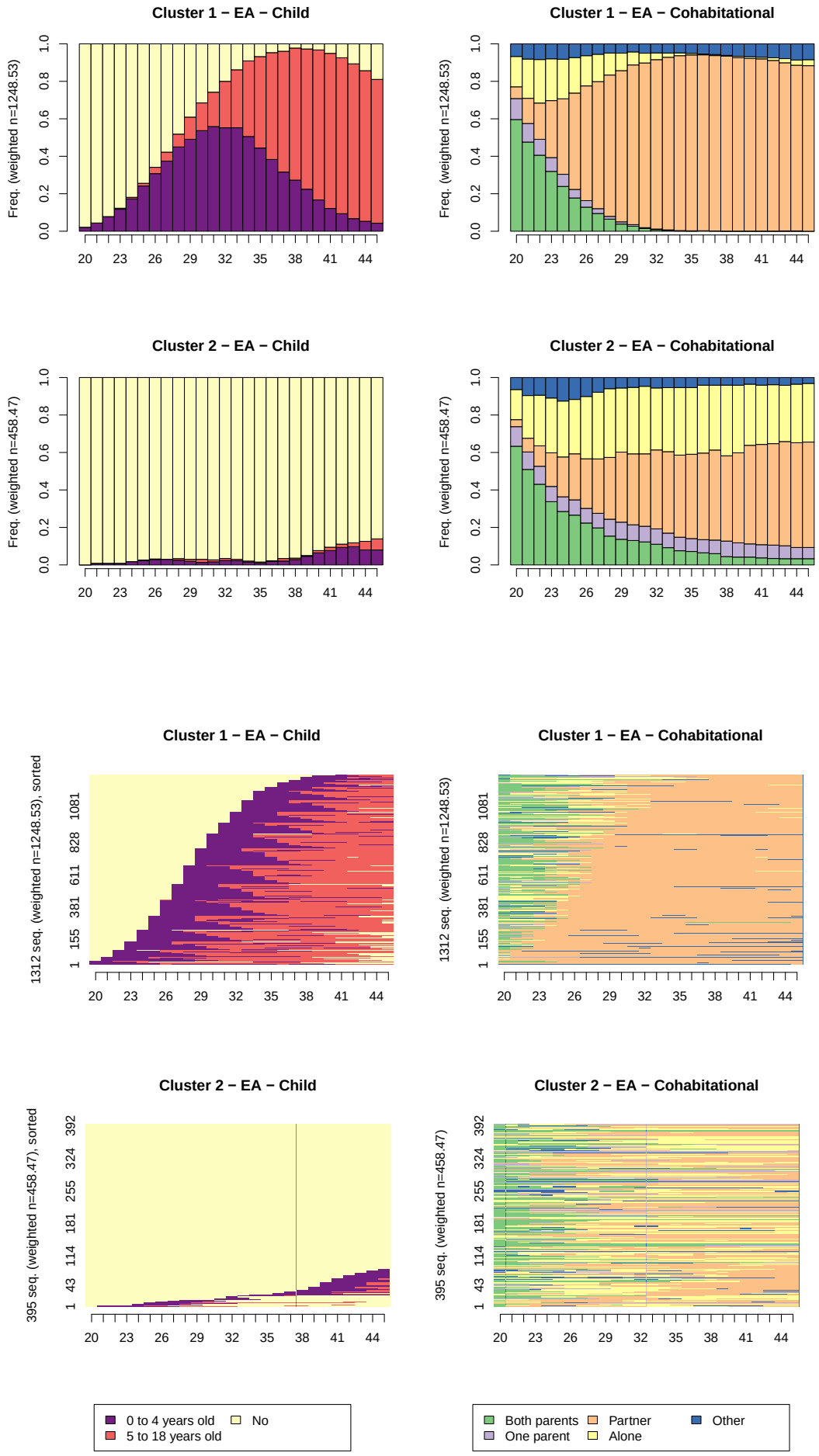

Figure 5: Chronograms (top) and index plots (bottom) of the two-group typology of cohabitational and child status obtained with EA on the full dataset.

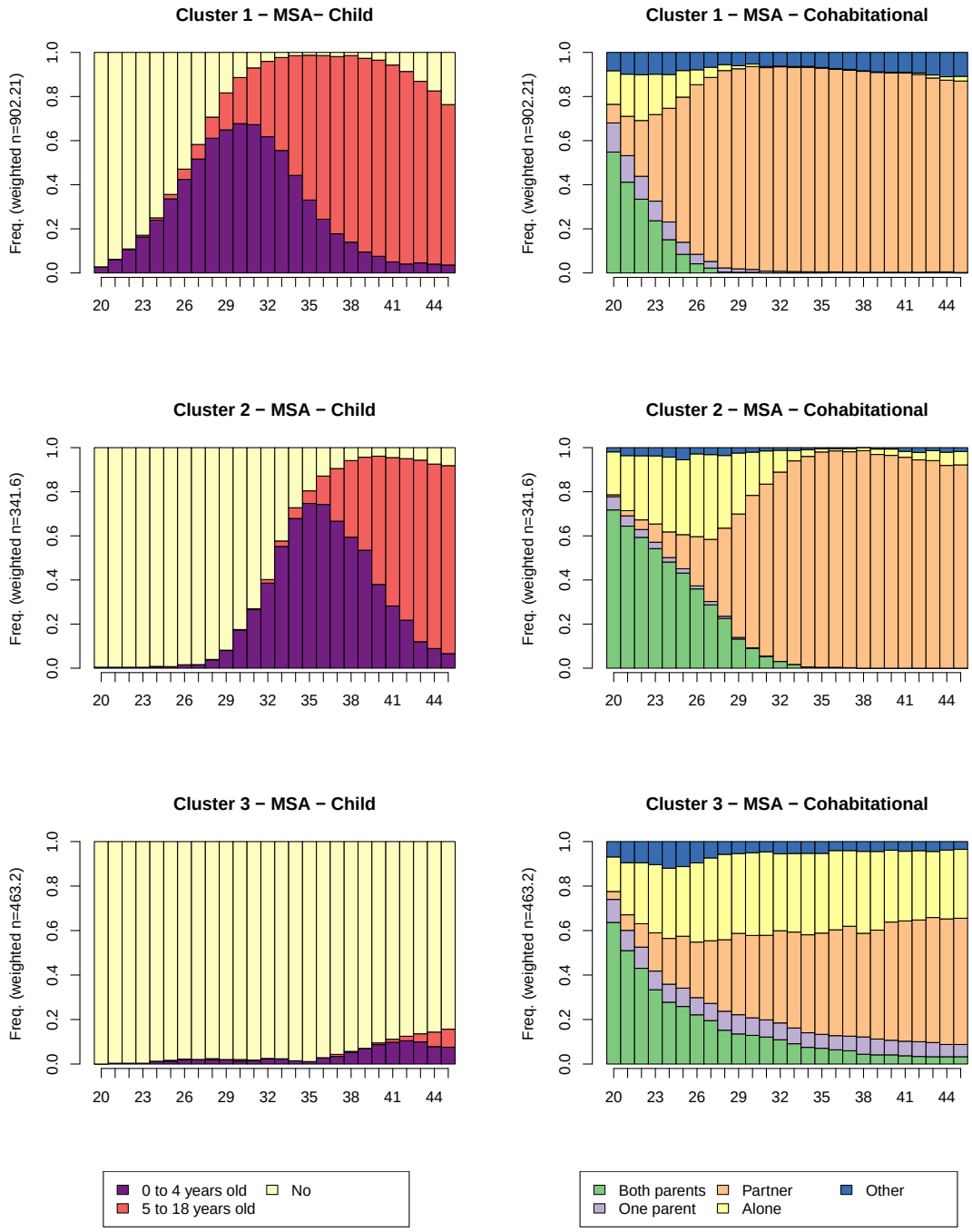

Figure 6: Chronograms of the three-group typology of cohabitational and child status obtained with MSA on the full dataset.

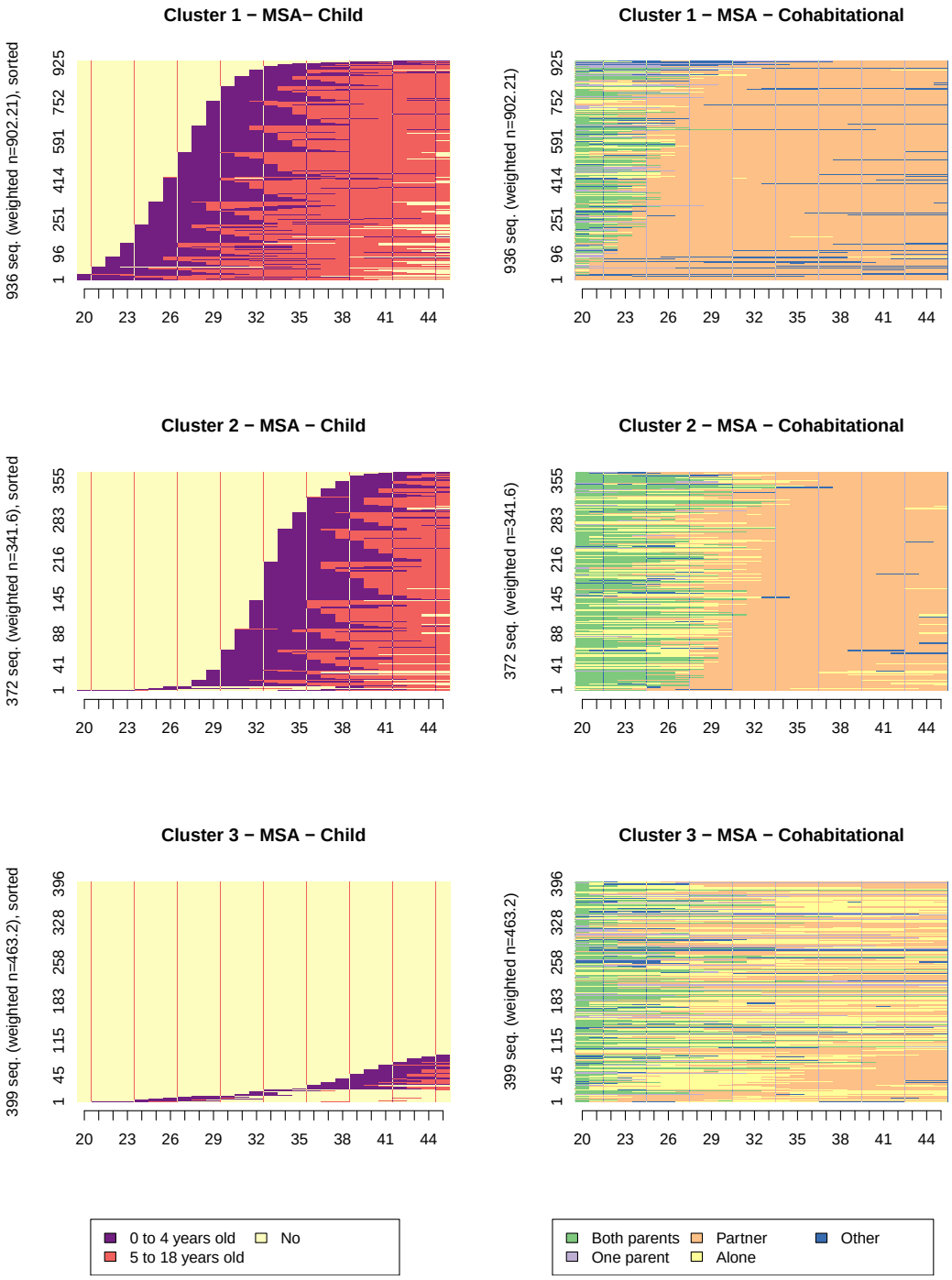

Figure 7: Index plots of the three-group typology of cohabitational and child status obtained with MSA on the full dataset.

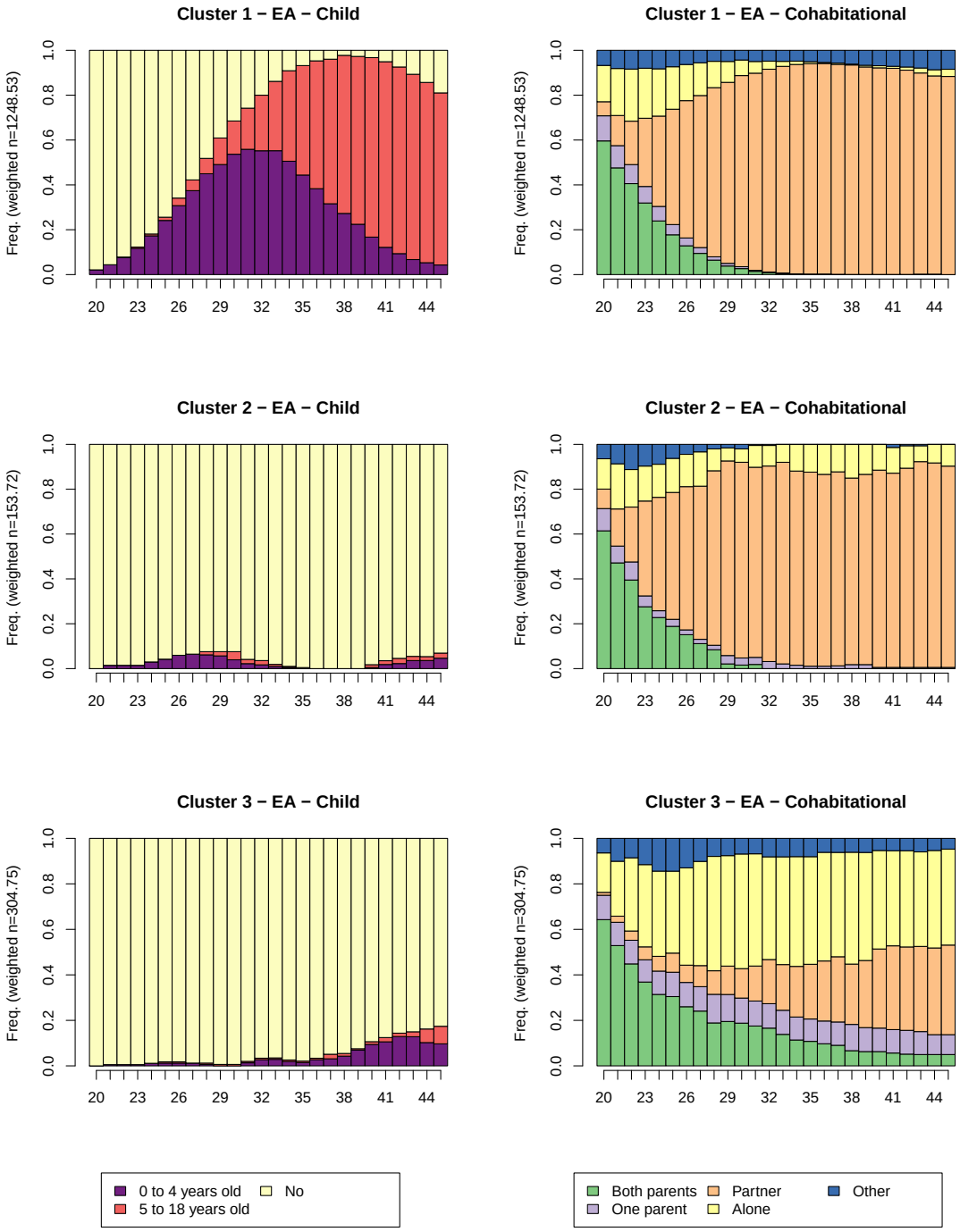

Figure 8: Chronograms of the three-group typology of cohabitational and child status obtained with EA on the full dataset.

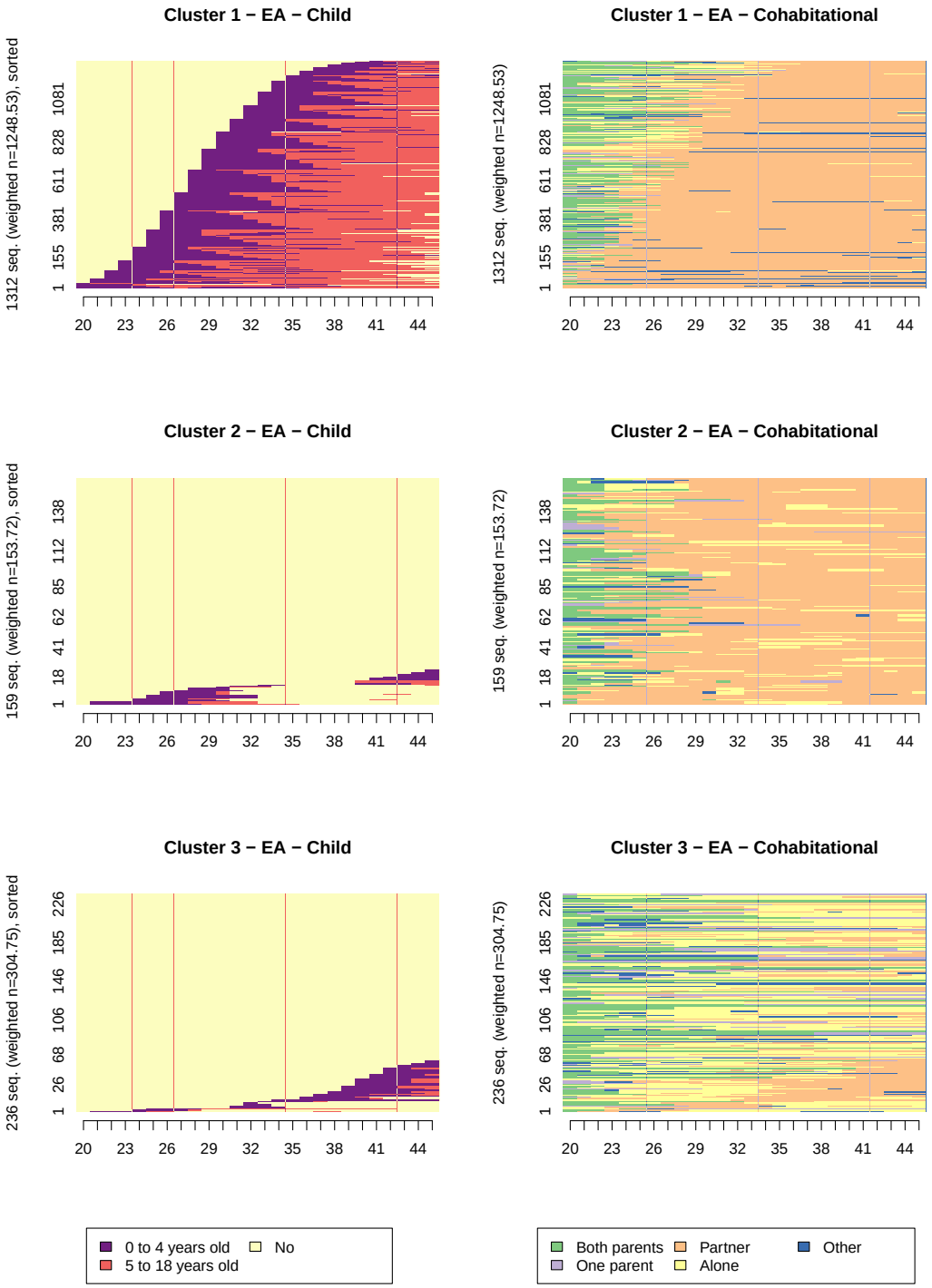

Figure 9: Index plots of the three-group typology of cohabitational and child status obtained with EA on the full dataset.

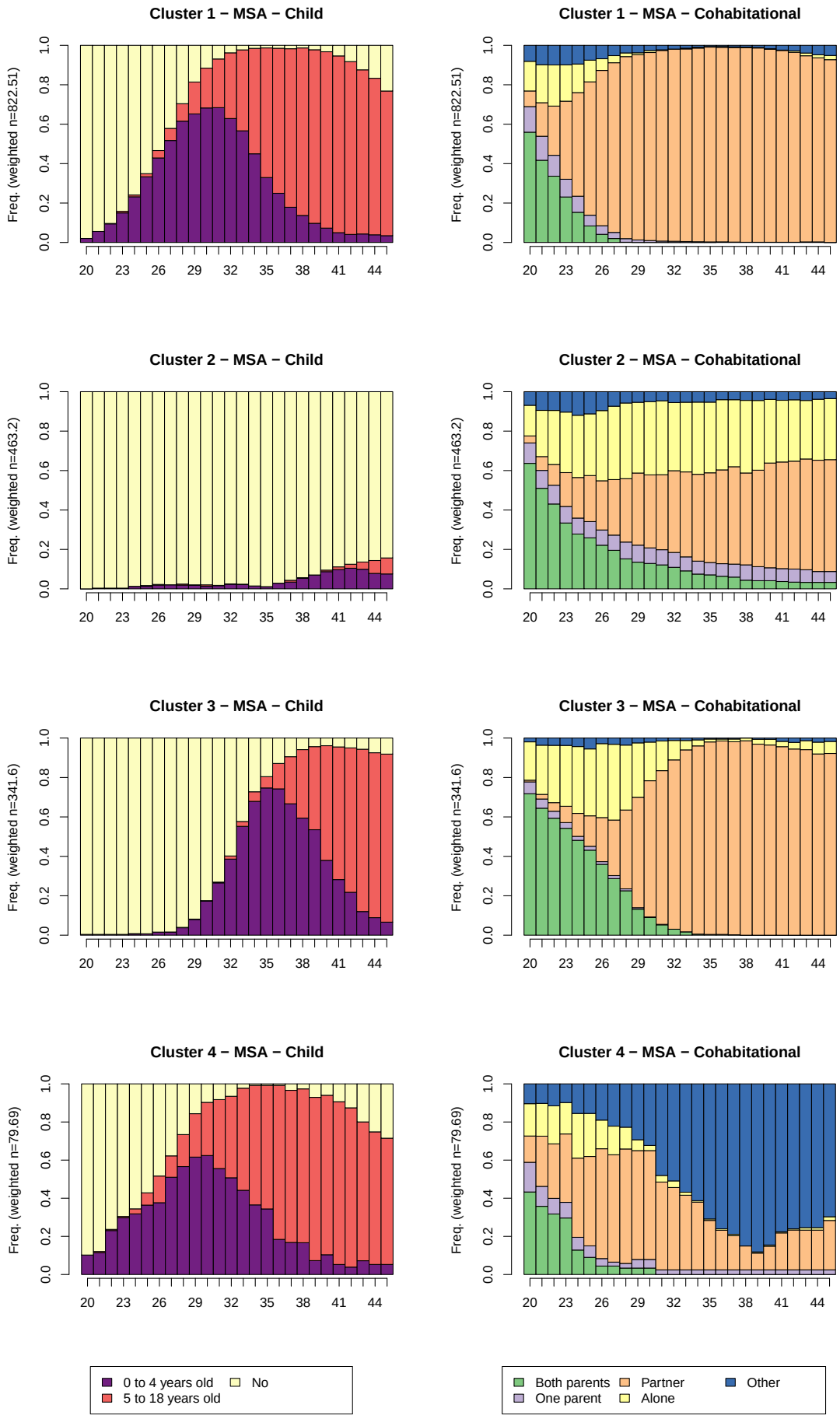

Figure 10: Chronograms of the four-group typology of cohabitational and child status obtained with MSA on the full dataset.

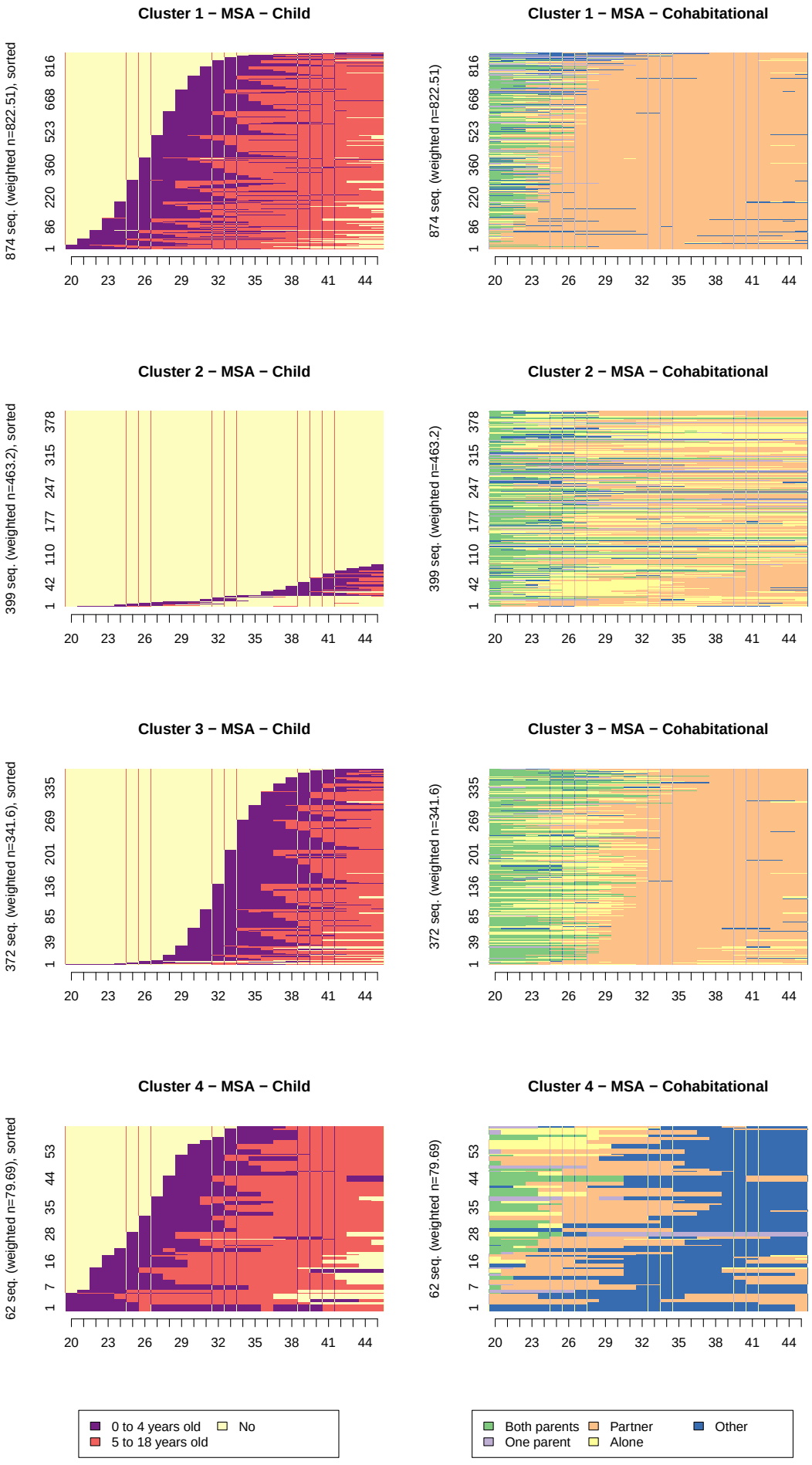

Figure 11: Index plots of the four-group typology of cohabitational and child status obtained with MSA on the full dataset.

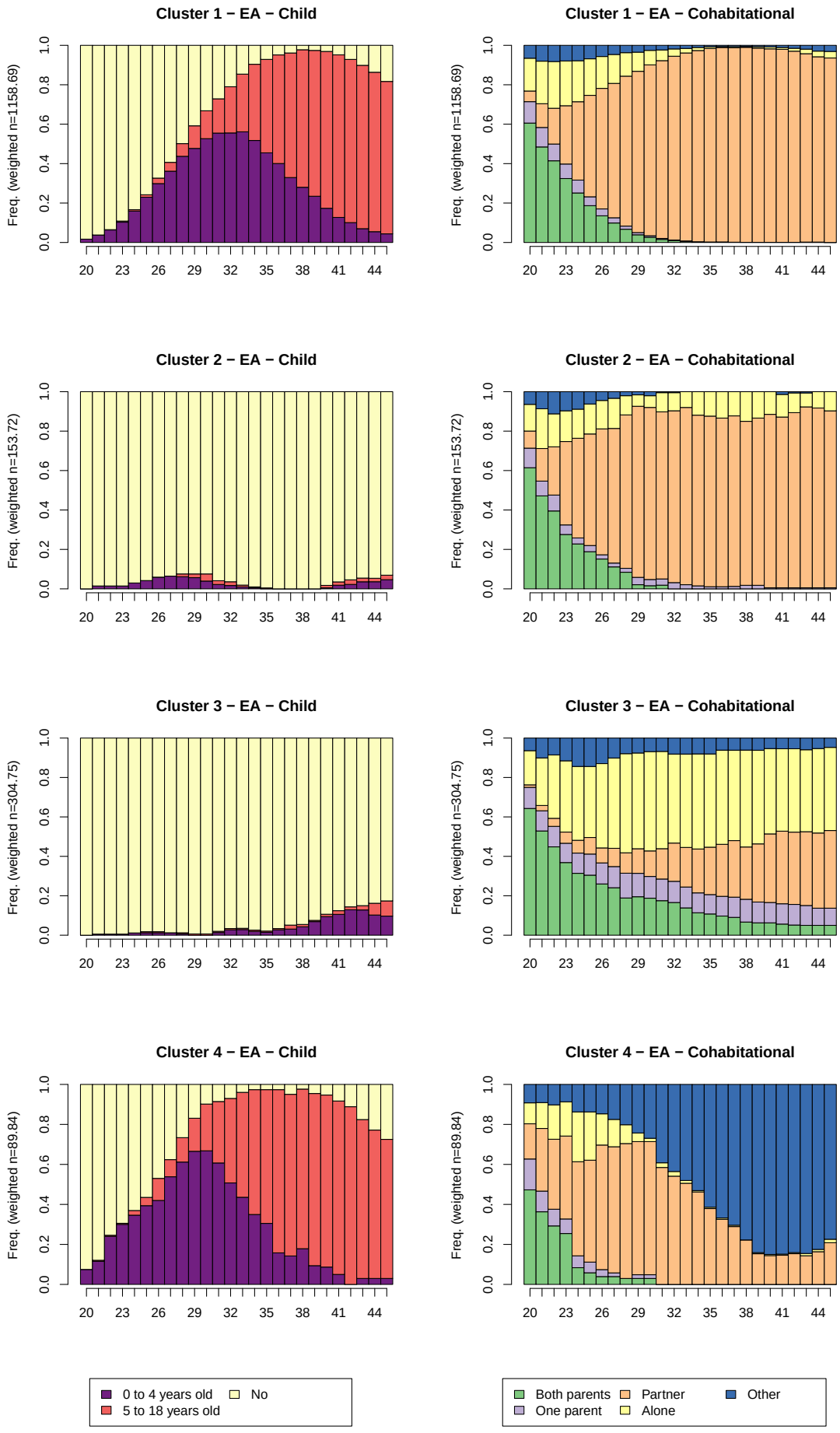

Figure 12: Chronograms of the four-group typology of cohabitational and child status obtained with EA on the full dataset.

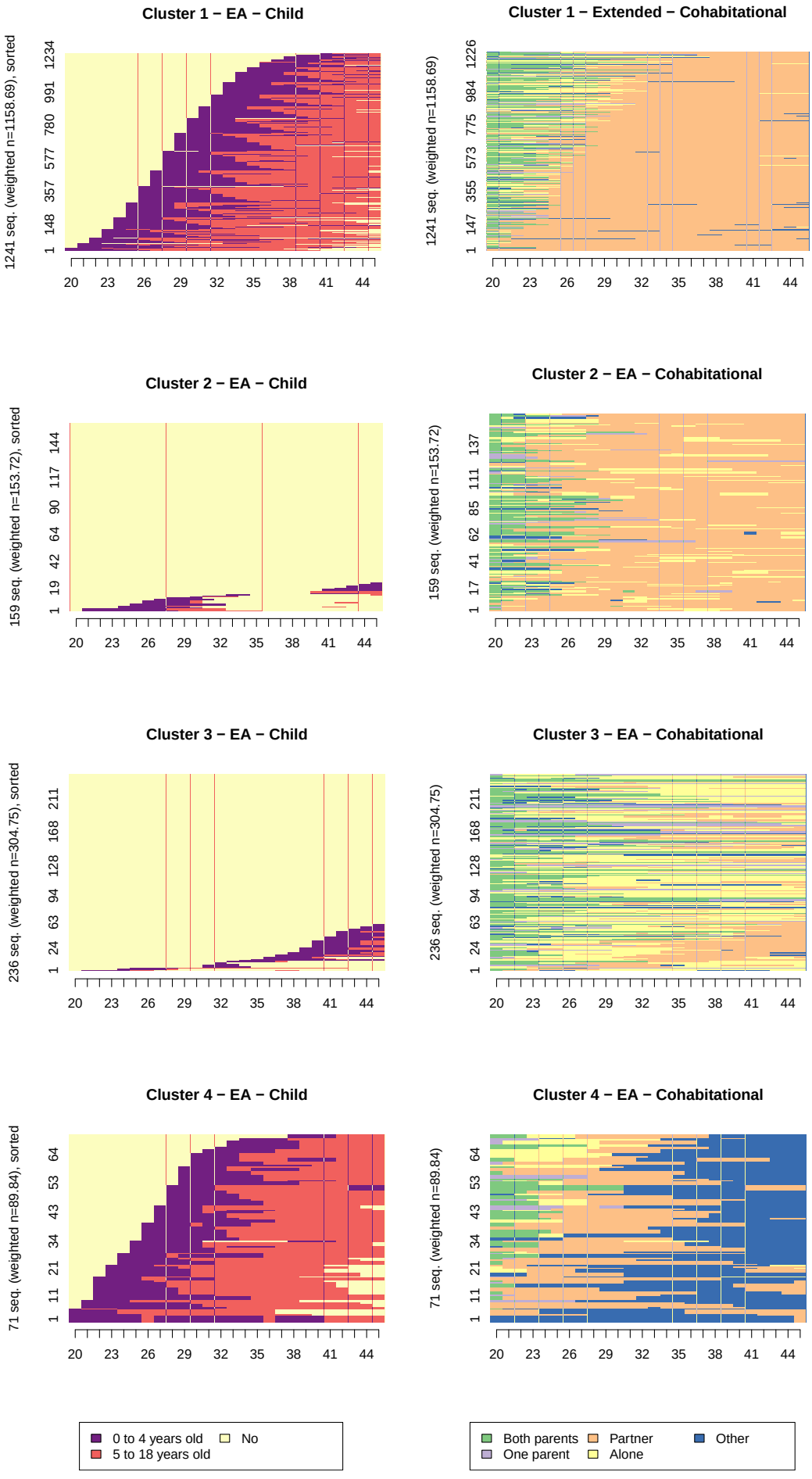

Figure 13: Index plots of the four-group typology of cohabitational and child status obtained with EA on the full dataset.

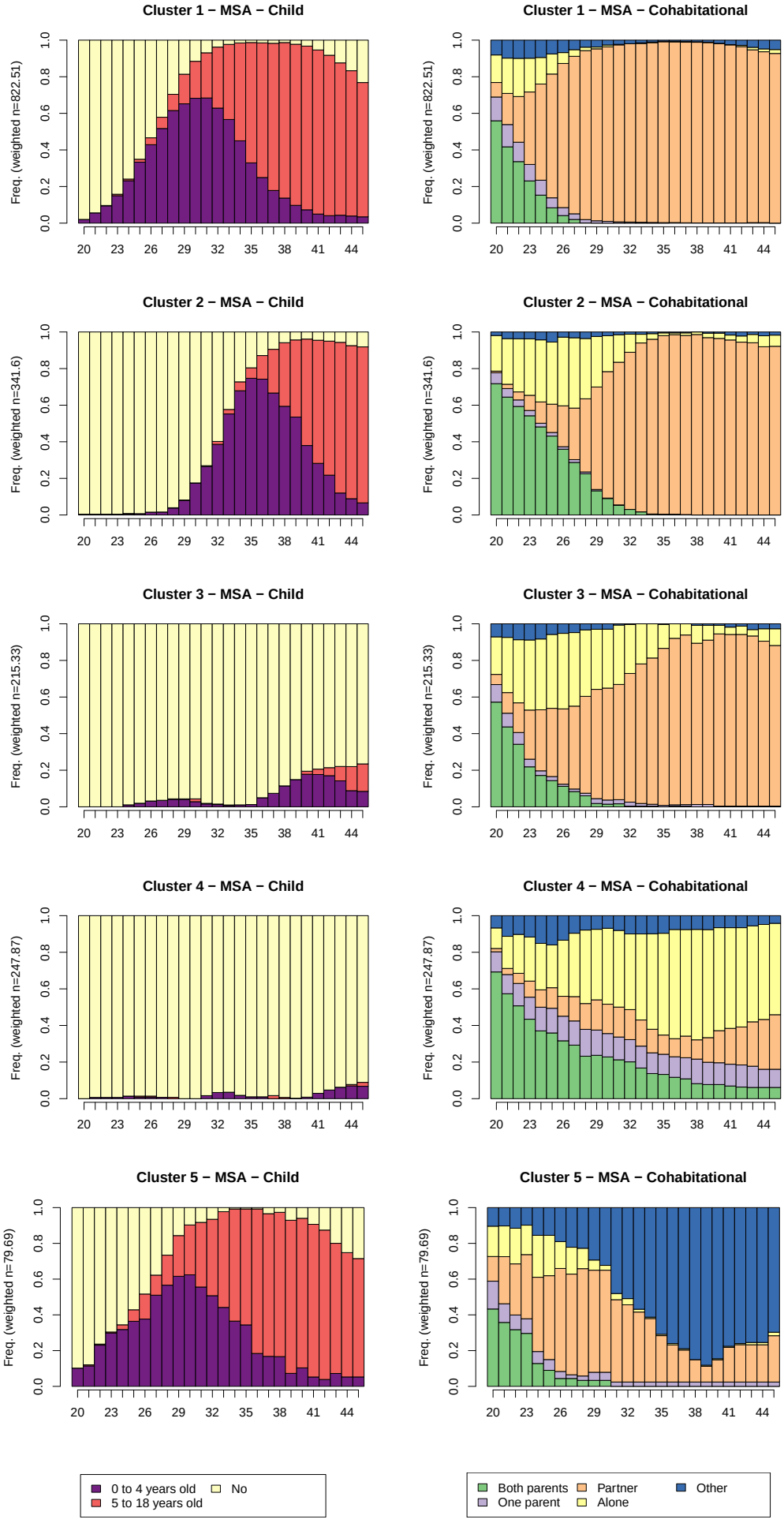

Figure 14: Chronograms of the five-group typology of cohabitational and child status obtained with MSA on the full dataset.

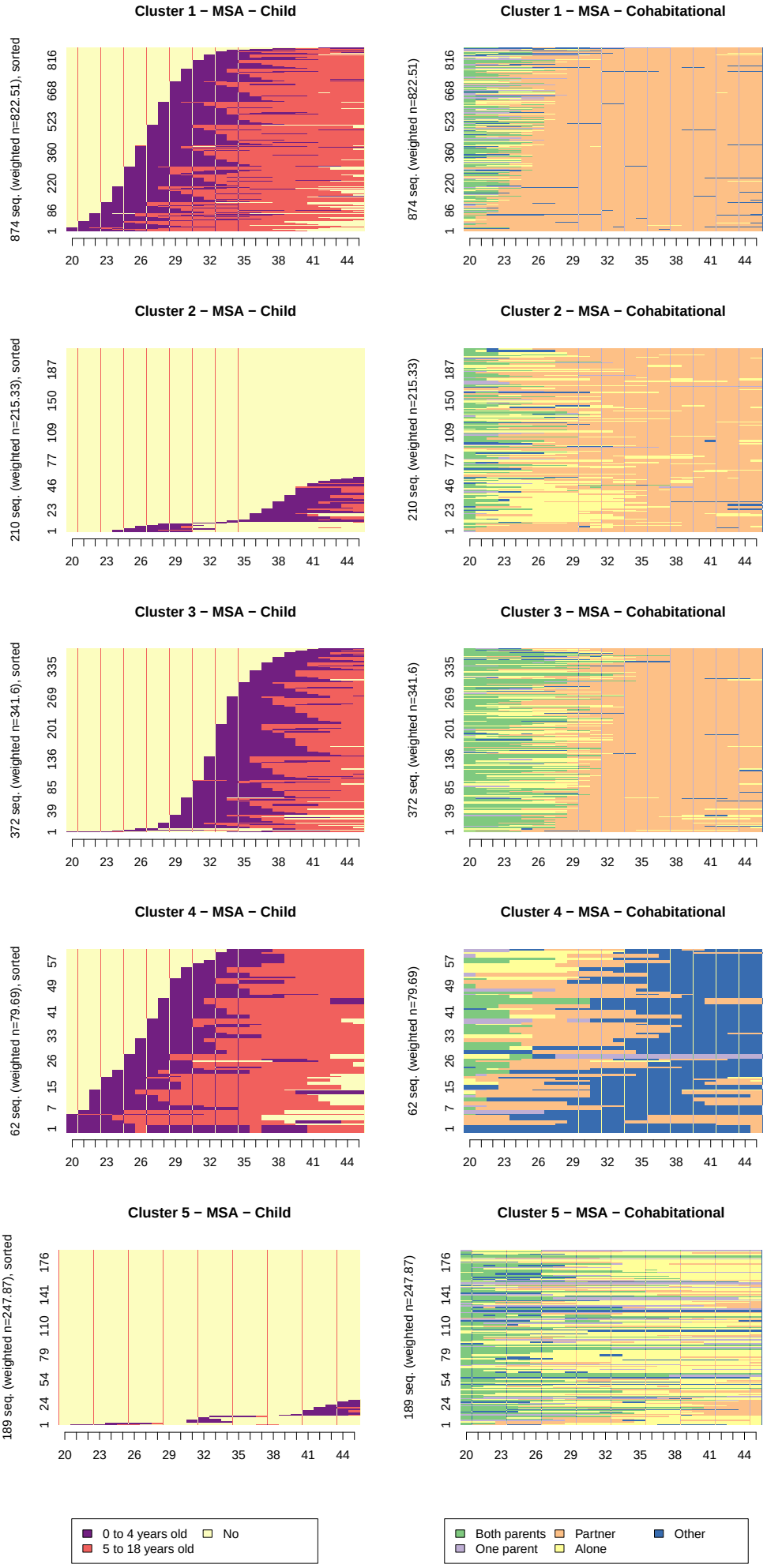

Figure 15: Index plots of the five-group typology of cohabitational and child status obtained with MSA on the full dataset. 26

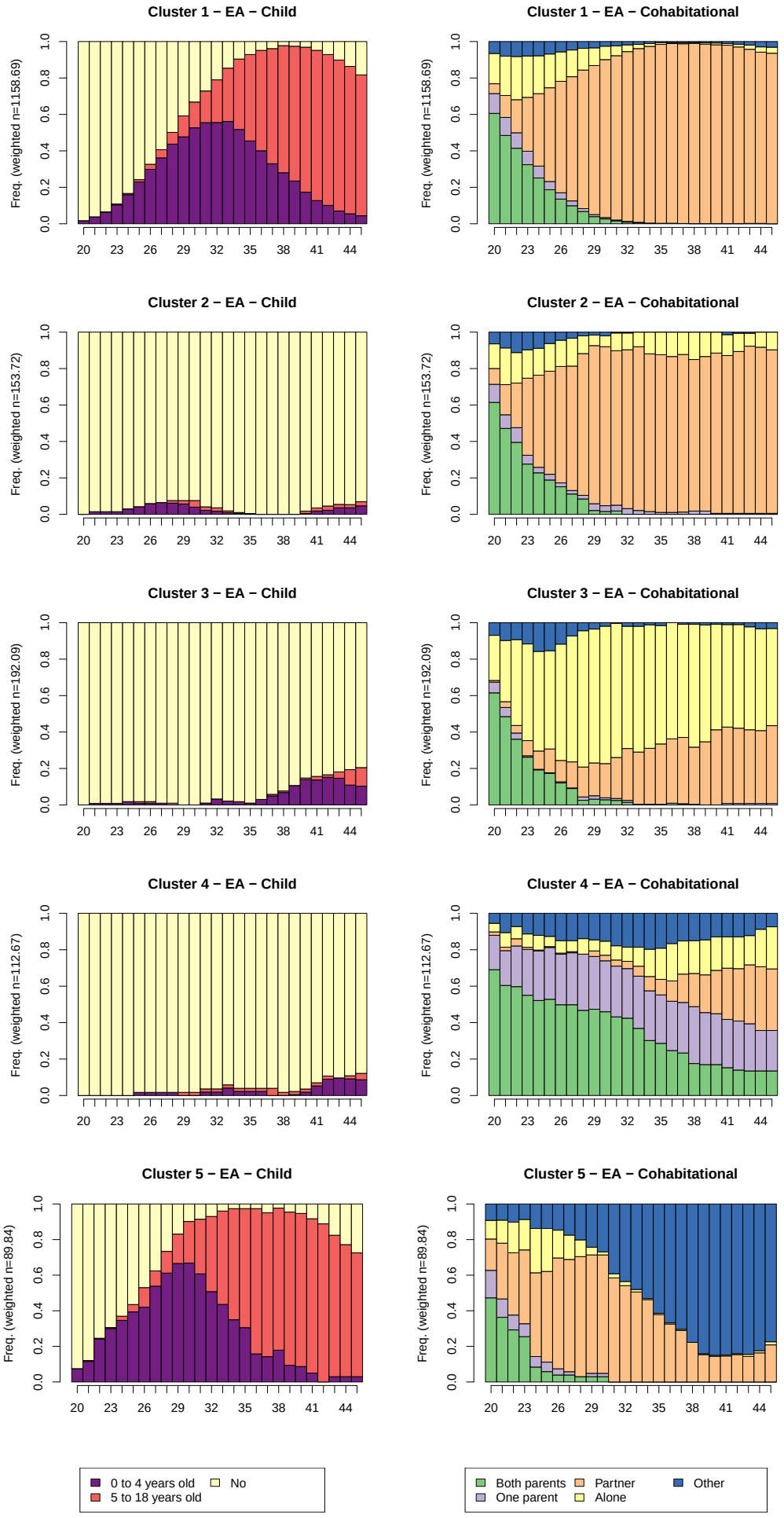

Figure 16: Chronograms of the five-group typology of cohabitational and child status obtained with EA on the full dataset.

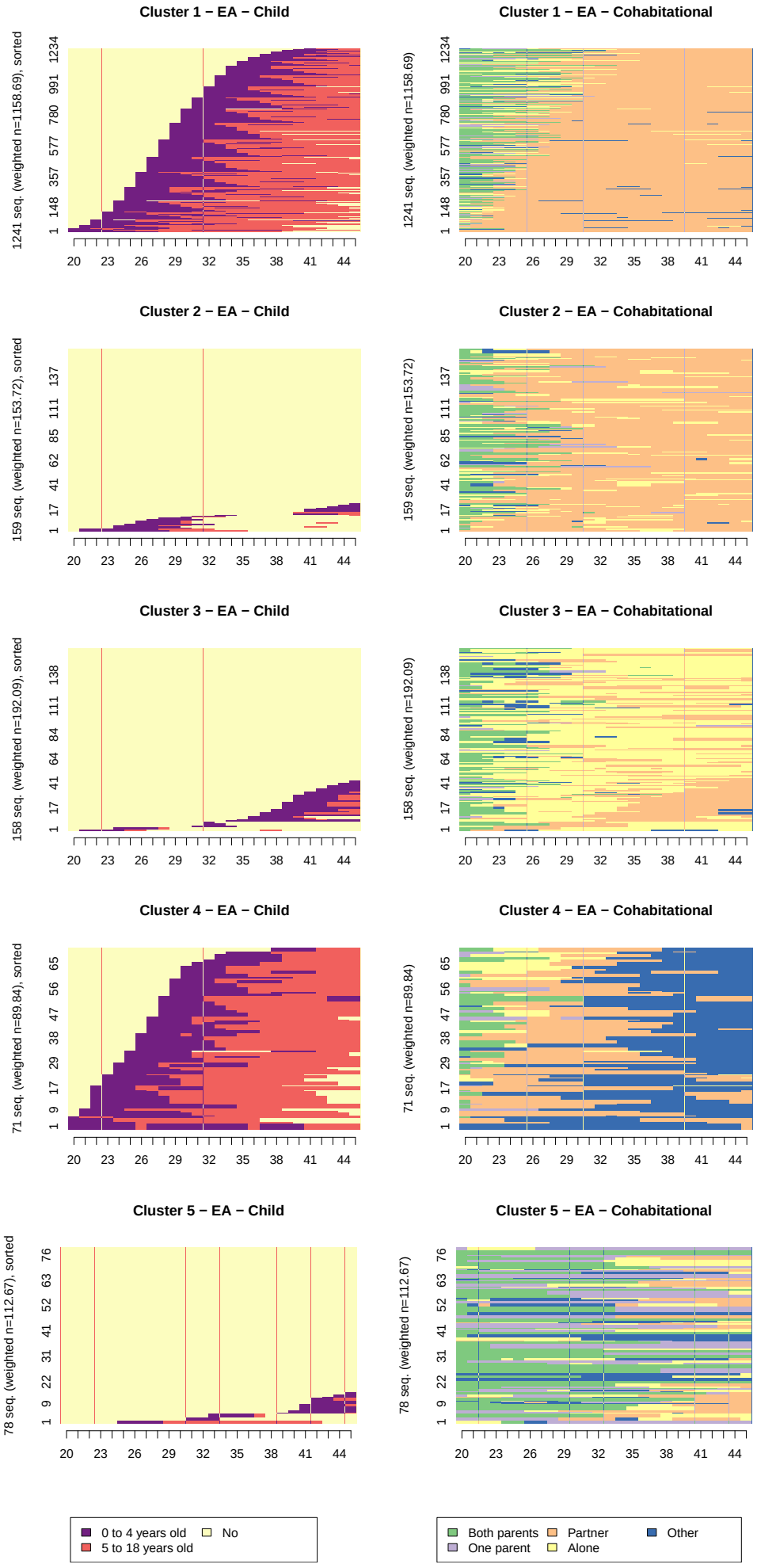

Figure 17: Index plots of the five-group typology of cohabitational and child status obtained with EA on the full dataset.

4.2 Analysis by sex

As pointed out by Levy et al. (2006) and Widmer and Ritschard (2009) among others, the professional status trajectories differ between men and women. An interrelation could exist between the child and professional status domains for women since their working rate often decreases when a child is born, whereas the same is barely observed for men. This is confirmed by the chronograms (Figure 18) and index plots (Figure 19), which are computed separately for each sex. Indeed, the professional status trajectories of men are mainly characterised by full-time work, while women are more prone to non-working and part-time work, and these states seem synchronised with the arrival of a child in the household. Moreover, women have children slightly earlier than men. These findings motivated us to re-run all the analyses presented in Section 4.1 separately by sex.

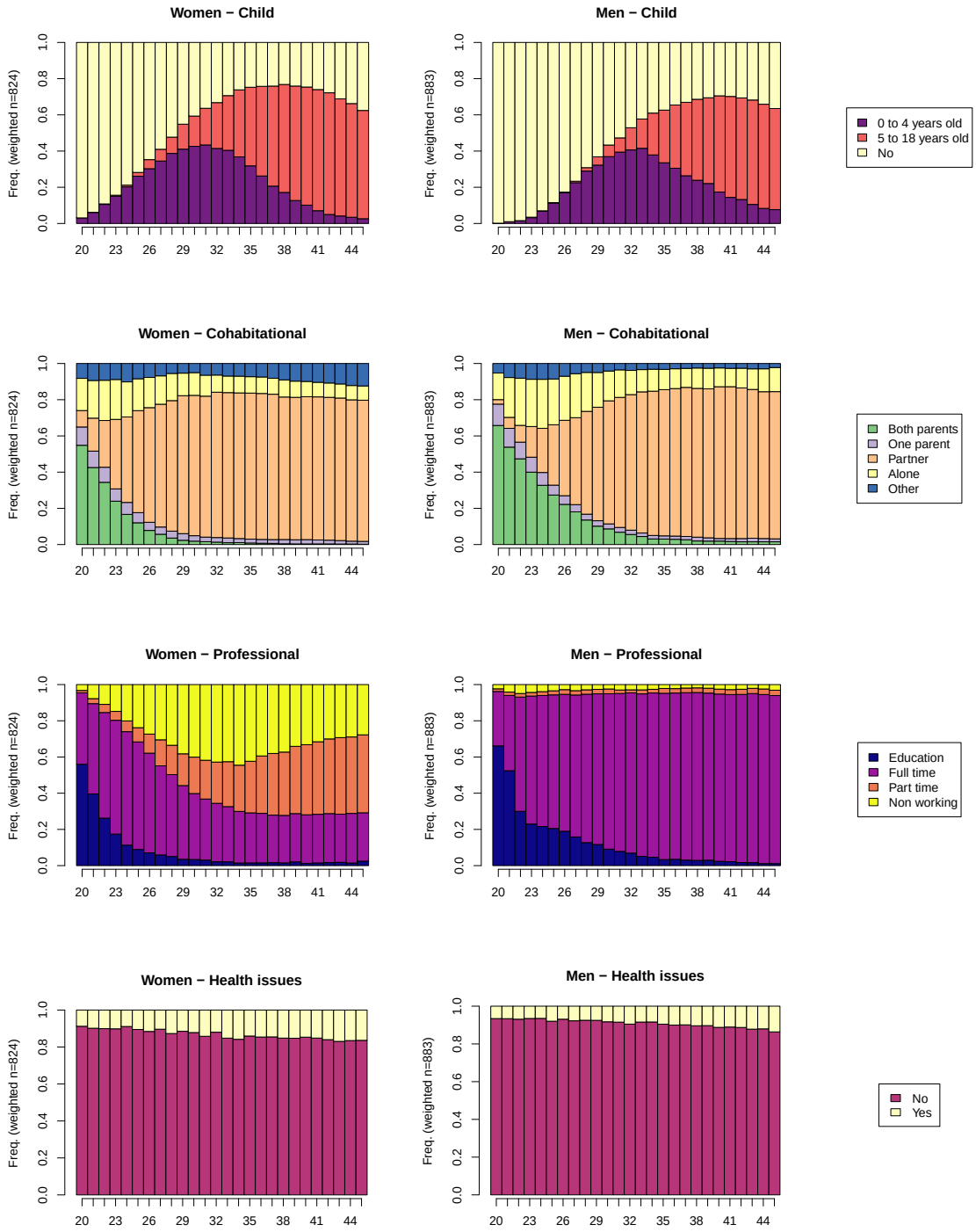

Figure 18: Chronograms of child, cohabitational, professional and health issues domains in function of sex.

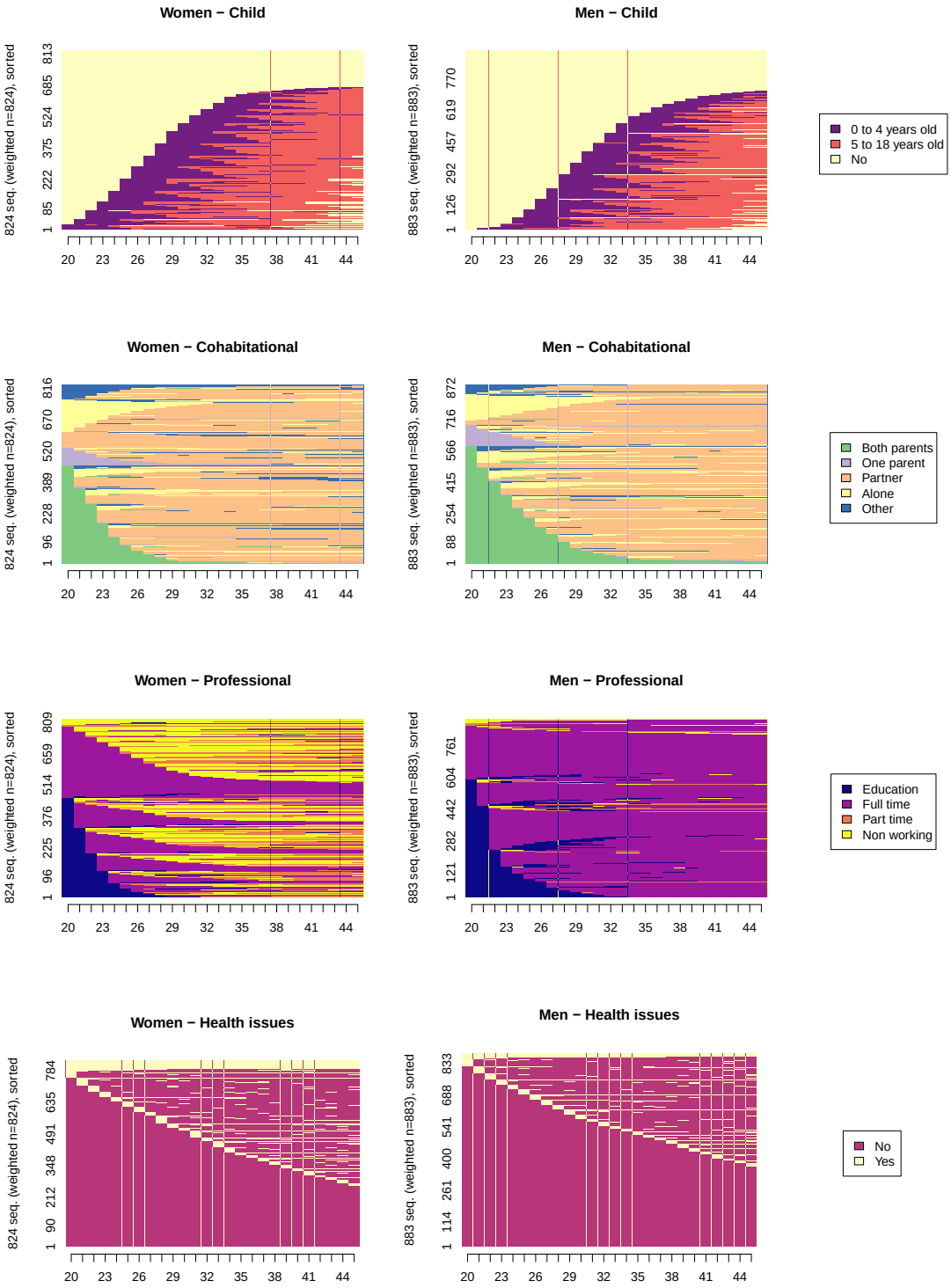

Figure 19: Index plots of child, cohabitational, professional and health issues domains in function of sex.

4.2.1 Men

We first determined which domains were associated. Considering all four domains together produced a Cronbach's α of 0.31. Discarding the professional status domain slightly increased the value to 0.34, while removing both the professional status and the health issues domains yielded a value of 0.5. Therefore, the cohabitational status and child domains were, as for the full dataset, the most interrelated ones according to Cronbach's α. This was confirmed by the PCA (Table 3) since the first principal component was highly correlated to the child and cohabitational status domains. We also analysed the Cronbach's α produced by each pair of domains to determine if any other pairs were linked. Table 4 presents the results. Unlike the full dataset, the pair composed of the health issues and cohabitational status domains as well as the pair of the professional status and cohabitational status domains produced Cronbach's α values larger than 0.1 (0.17 and 0.11, respectively). Although these values were still small, we chose to investigate these combinations of domains, as interpreting raw Cronbach's α values to evaluate joint domains can be unclear. Moreover, this provided information on how the two approaches (MSA and EA) behave when domains are weakly linked.

Table 3: Loadings of the PCA applied to the pairwise dissimilarities computed on the four domains in the case of men.

Domain	PC1	PC2	PC3
Child	0.83	-0.09	-0.06
Cohabitational	0.80	0.15	0.09
Professional	0.02	0.01	1
Health issues	0.03	0.99	0.01
Eigenvalues	1.34	1.01	1

Table 4: Cronbach's α of each pair of domains in the case of men.

Domains	Child	Cohabitational	Professional	Health issues
Child	1	0.5	0	0.02
Cohabitational		1	0.11	0.17
Professional			1	0.06
Health issues				1

Cohabitational status–Health issues In the first step, we computed the correlations between the vectors containing pairwise dissimilarities. The correlations between d_{EA} and, respectively, d_{Health} and d_{Cohab} were 0.45 and 0.91. For MSA,

the correlations were 0.68 and 0.77, respectively. Therefore, the values given by MSA were more balanced than those from EA. Moreover, d_{MSA} explained a larger proportion of health-based dissimilarities, while d_{EA} explained a larger proportion of cohabitational status-based dissimilarities. None of the clusterings realised with MSA were significant neither in terms of ASWw nor in terms of HC (Figure 20). For EA, only the three-cluster solution was significant both in terms of ASWw and in terms of HC. This grouping consisted of one large cluster of individuals mostly living with a partner and having at most small periods of health issues, a cluster of people mostly living with a partner and having health problems, and one cluster of individuals not living with a partner with diverse health statuses (Figures 21 and 22). The first two groups were relatively homogeneous, with ASWw values of 0.52 and 0.44, while the third group had a value of 0.13. The proportion of total pairwise dissimilarities explained by this grouping was 0.72 for the cohabitational status domain and 0.85 for the health issues domain (Table 5). Therefore, although d_{EA} was more correlated with d_{Cohab}, the clustering represented a more important share of the dissimilarities in the health issues domain. This is in line with that pointed out by Piccarreta and Studer (2019): when domains are not clearly interrelated, the clustering is driven by the less turbulent domain (health issues in our case).

Table 5: Summary of the results obtained by clustering the health issues and cohabitational status channels for men with MSA and EA. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	Method	CQI		Group ASWw		Size clusters (%)		R^2 by channel	
		ASWw	HC	min	max	min	max	cohab	health
2	MSA	0.55	0.16	0.32	0.56	6	94	0.59	0.87
	EA	0.41	**0.07**	0.02	0.57	18	82	0.7	0.74
3	MSA	0.47	0.06	0.02	0.55	6	78	0.72	0.88
	EA	**0.39**	**0.07**	0.13	0.52	4	82	0.72	0.85
4	MSA	0.20	0.14	-0.03	0.33	6	44	0.76	0.88
	EA	0.19	0.16	0.07	0.43	4	48	0.77	0.85
5	MSA	0.21	0.12	-0.12	0.42	6	44	0.8	0.88
	EA	0.20	0.14	-0.07	0.44	4	38	0.8	0.85

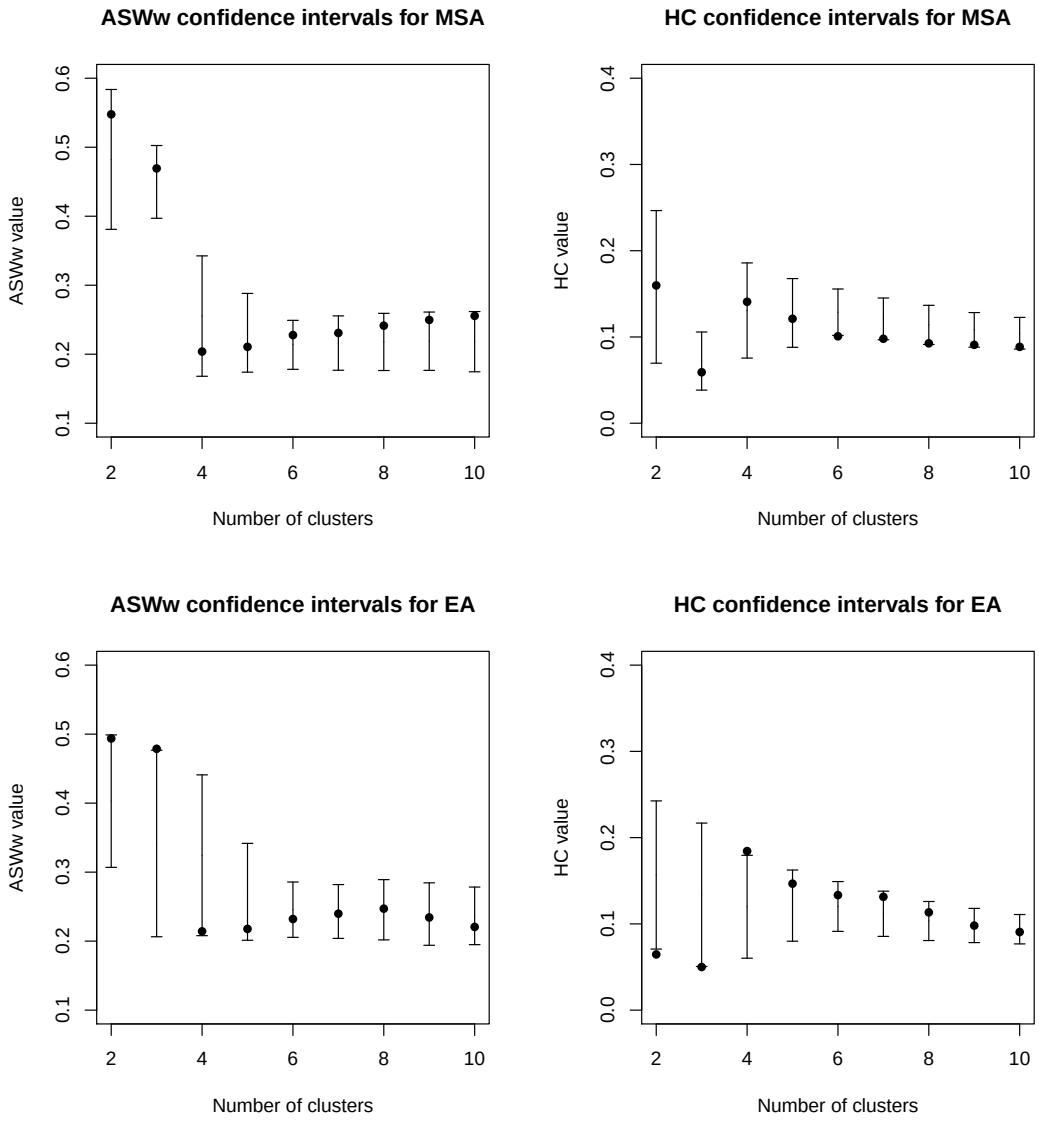

Figure 20: ASWw and HC values obtained by clustering the data, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that cohabitational and health issues domains are not associated.

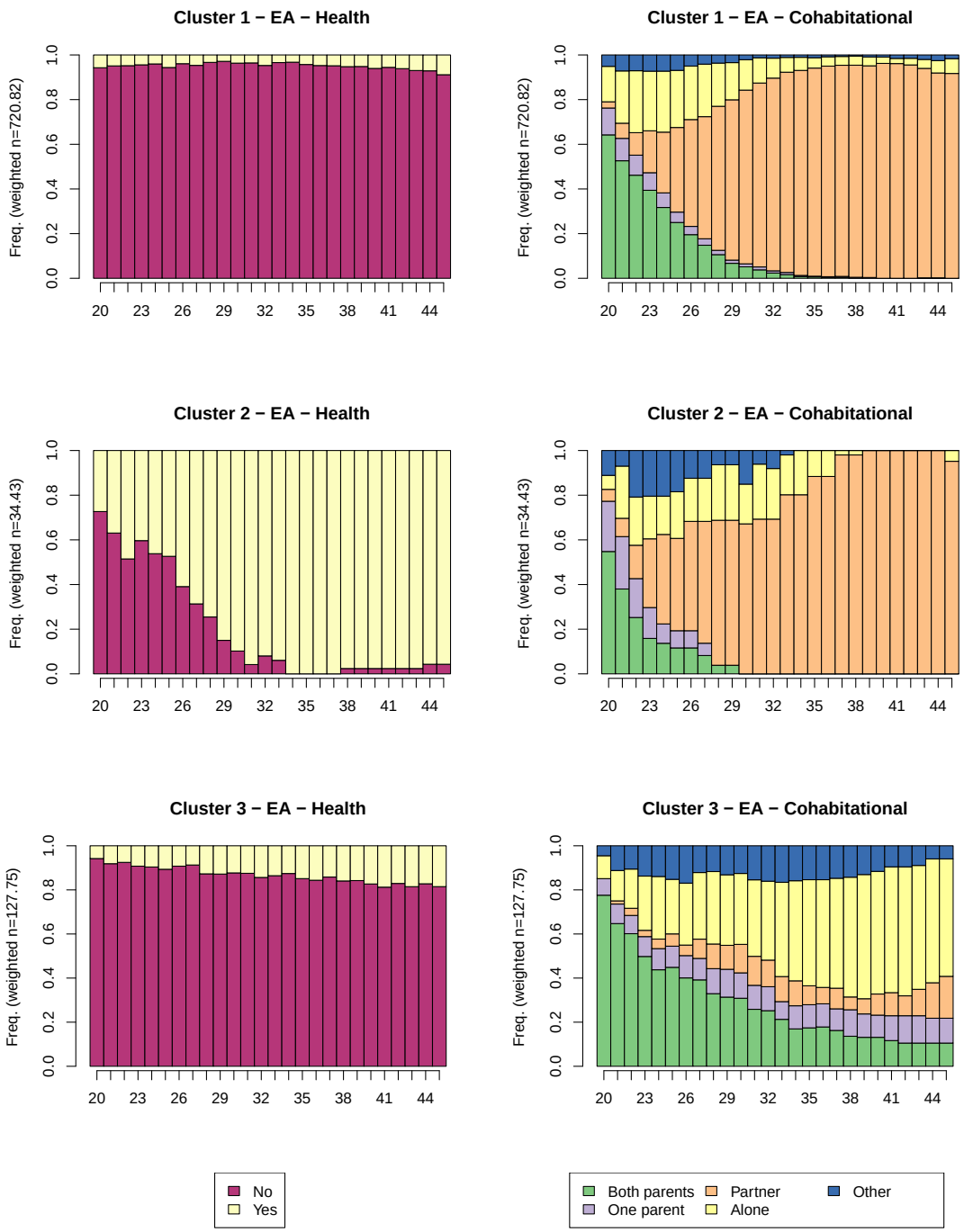

Figure 21: Chronograms of the health issues and cohabitational status typology in three groups obtained with EA for men.

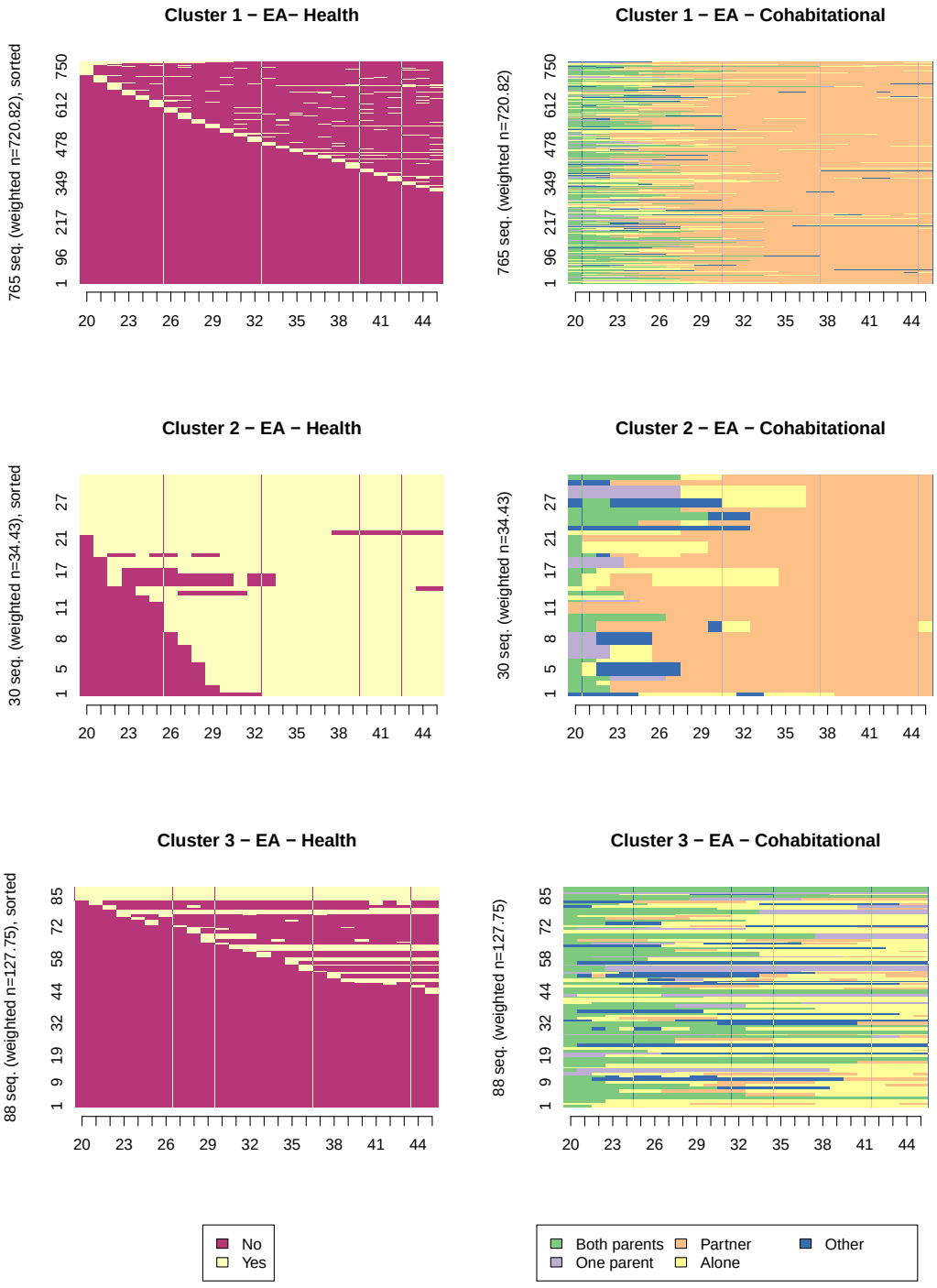

Figure 22: Index plots of the health issues and cohabitational status typology in three groups obtained with EA for men.

Cohabitational status–Professional status The case for these correlations was similar to the case for the cohabitational status and health issues domains. Indeed, the professional status trajectories, which are mainly characterised by full-time work, are relatively homogeneous, and the pairwise dissimilarities were therefore more correlated with the cohabitational status domain. The correlations between d_{EA} and, respectively, d_{Prof} and d_{Cohab} were 0.51 and 0.84, while, again, the correlations computed from MSA were more balanced (0.65 and 0.77). Concerning the extraction of a joint typology of the cohabitational status and professional status domains, none of the clusterings built by MSA and EA were significant (Figure 23). Since the Cronbach's α for these two domains (0.11) was even smaller than that of the cohabitational status and health issues domains (0.17), the link between the cohabitational status and professional status domains was probably too weak to allow for the extraction of a joint typology.

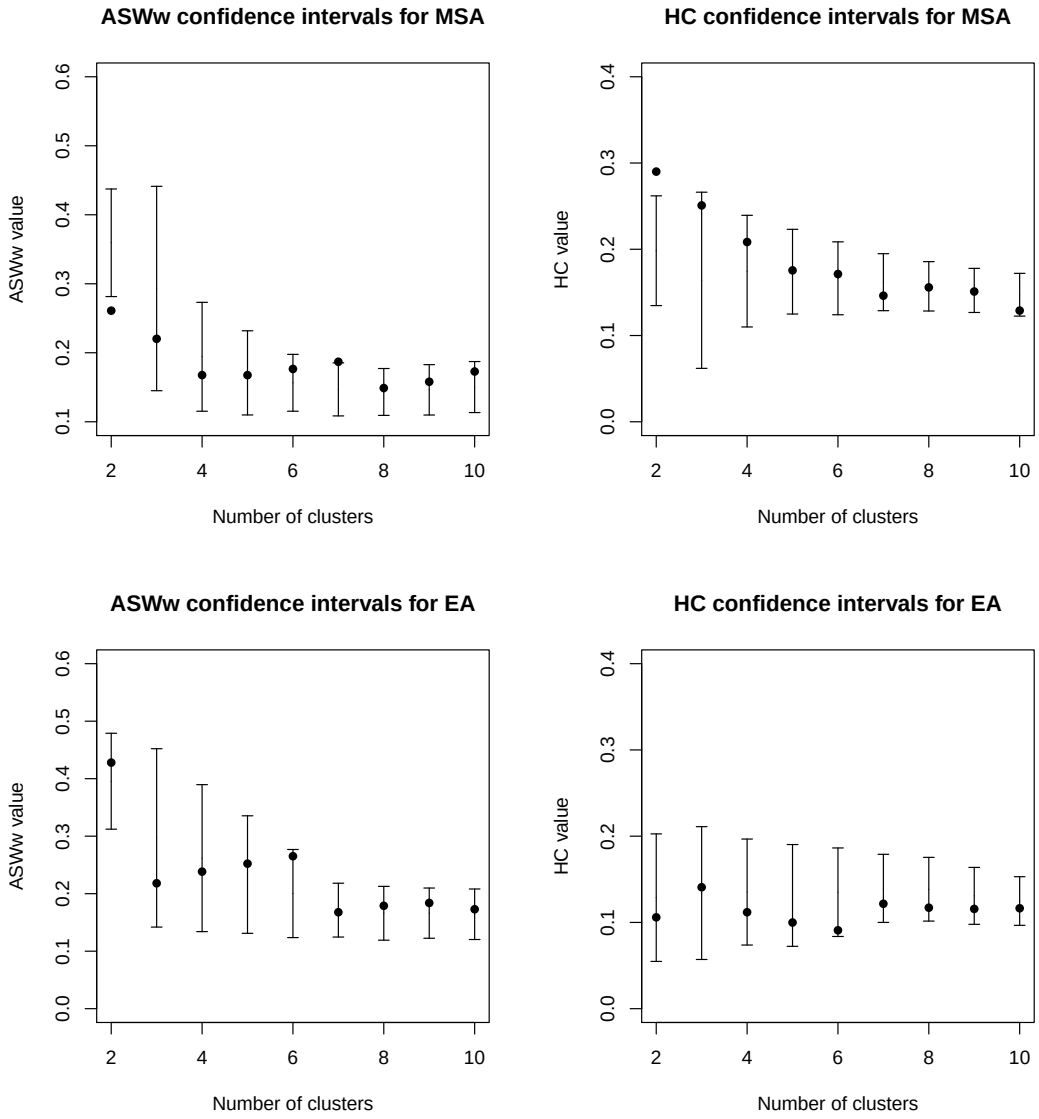

Figure 23: ASWw and HC values obtained by clustering the data, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that cohabitational and professional domains are not associated.

Cohabitational status–Child As for the full dataset, the cohabitational status and child domains were also the most linked in the case of men. We thus investigated whether the results obtained with men only differed from the results obtained using the full dataset. The correlations between d_{MSA} and, respectively, d_{Child} and d_{Cohab} were 0.8 and 0.81, while, for EA, both correlations were 0.74. Therefore, the individual domains were almost equally summarised by both approaches, whereas the values were slightly higher for MSA.

The most significant grouping for the two approaches was obtained by splitting the individuals mainly according to whether they had a child, with the exception

of people having a child at a late age or people living for short periods with a child. The groupings in the two clusters created by both approaches (Figure 25 and 26) were almost identical since they agreed on the assignment of 97% of the sequences. However, the solution provided by MSA was slightly more balanced in terms of ASWw by group. The logic behind the three-group clusterings was similar to those of the full dataset: MSA built two clusters of individuals having a child depending on the timing (Figures 27 and 28), while EA split the non-child cluster according to with whom a person is living (Figures 29 and 30). The third cluster of both clusterings were relatively ill-defined (ASWw values of 0.07 for MSA and 0.09 for EA). Then, the four-group solutions were based on the same idea for the two approaches: two clusters of childless people differentiated by the type of cohabitation (partner vs no partner) and two clusters of individuals having a child split according to timing. However, although the idea behind these two groupings was relatively similar (Figures 31, 32 and 33, and 34), the two approaches agreed on only 77% of the assignment of sequences. Moreover, by examining the ASWw computed by group, the second and last clusters built by EA, which had values of 0.05 and 0.07, respectively, were ill-defined, while only the third cluster built by MSA was so (ASWw of 0.04). According to the bootstrap validation (Figure 24), only the grouping obtained with MSA was significant, which may explain why some groups were ill-defined: there should not have been so many groups. According to the R^2 computed by domain (0.86 for the child and 0.75 for the cohabitational status domains), this clustering took account of the two domains relatively well; however, a larger share of the child domain was explained. Some clusterings in more groups were also significant for MSA, but not for EA. Figure 35 presents the seven-group clustering, which was, apart from the two-group clustering, the most significant for MSA. However, some clusters were ill-defined. For instance, the fifth cluster had an ASWw by group smaller than 0 (Table 6), while the value was even negative for the fourth one. Moreover, some clusters became small, limiting the generalisability of the results. This shows the limits of automatic clustering. Although significant from a statistical point of view, a given clustering can prove unsatisfactory from the point of view of thematic analysis. Therefore, it is always important to keep in mind the final goal of the analyses when looking for a typology.

Summary To summarise, as hypothesized, there were no clear link between professional and family domains for men. Even if the pair of cohabitational status and professional status domains provided a Cronbach's α value (0.11) slightly higher than on the full dataset, neither MSA nor EA could extract a joint typology of the

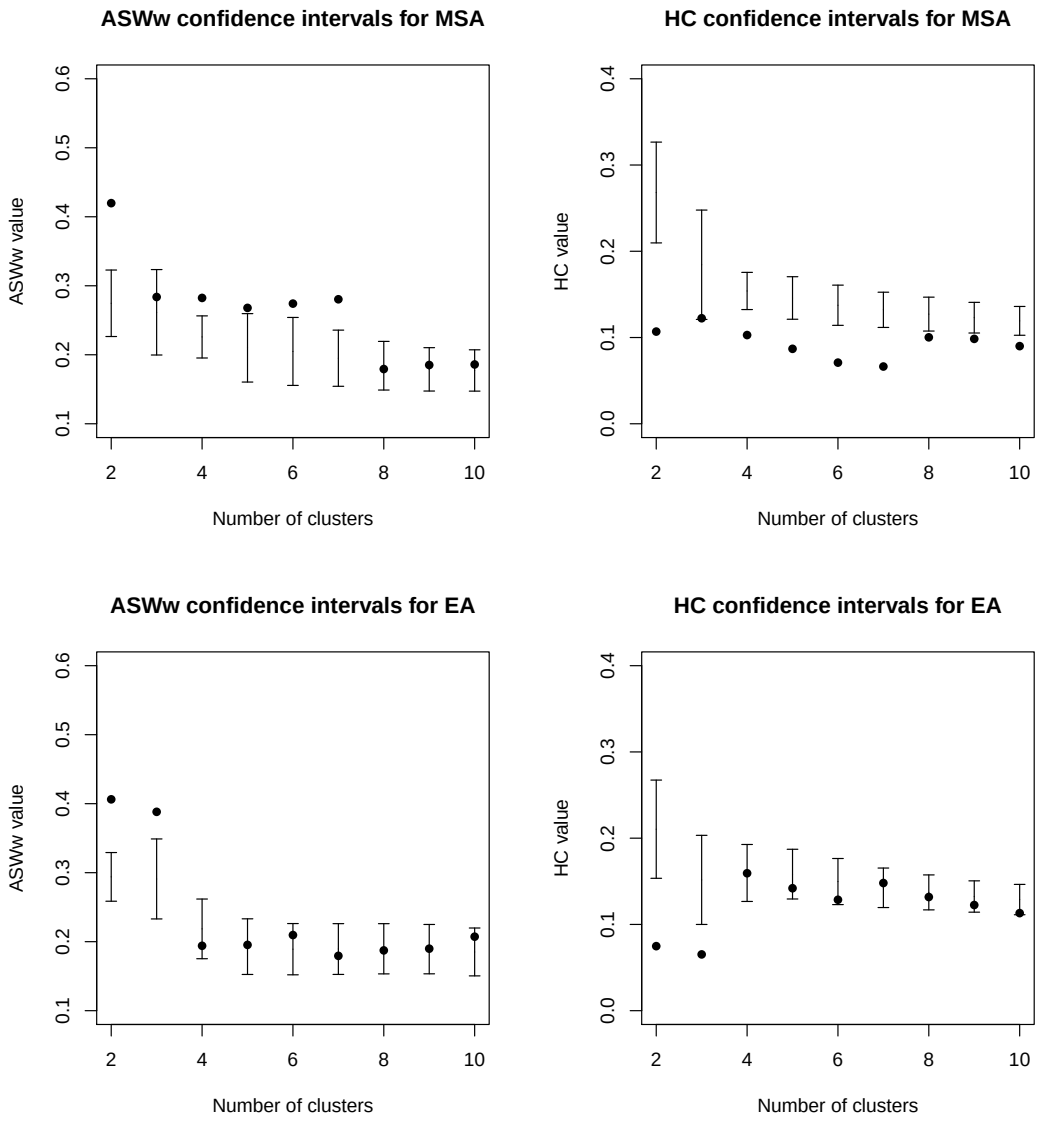

Figure 24: ASWw and HC values obtained by clustering the data, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that child and cohabitational domains are not associated.

professional status and cohabitational status domains. We hypothesised that the link between the two domains was too weak for that. Unexpectedly, the pair of cohabitational status and health issues domains were, exception of child and cohabitational domains, the most interrelated one according to Cronbach's α. Clusterings in two and three groups were slightly significant for EA, in line with our expectation: the weaker the link between domains, the harder it is to extract a joint typology. The derived typologies for the cohabitational status and child domains were slightly different from those extracted from the full dataset. The two-cluster solutions, which separated the dataset according to whether an individual had a

Table 6: Summary of the results obtained by clustering the child and cohabitational status channels for men with MSA and EA. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	Method	CQI		Group ASWw		Size clusters (%)		R^2 by channel	
		ASWw	HC	min	max	min	max	child	cohab
2	MSA	**0.42**	**0.11**	0.37	0.43	27	73	0.76	0.66
	EA	**0.41**	**0.07**	0.23	0.46	28	72	0.75	0.67
3	MSA	0.28	0.12	0.07	0.43	27	41	0.86	0.69
	EA	**0.39**	**0.07**	0.09	0.57	9	72	0.75	0.73
4	MSA	**0.28**	**0.10**	0.04	0.43	13	41	0.86	0.75
	EA	0.19	0.16	0.05	0.56	9	41	0.83	0.75
5	MSA	**0.27**	**0.09**	-0.07	0.38	10	41	0.86	0.78
	EA	0.20	0.14	0.02	0.53	9	41	0.83	0.77
6	MSA	**0.27**	**0.07**	-0.14	0.36	4	41	0.87	0.79
	EA	0.21	0.13	0.08	0.52	5	41	0.83	0.8
7	MSA	**0.28**	**0.07**	-0.14	0.47	4	41	0.87	0.81
	EA	0.18	0.15	0.03	0.52	5	21	0.84	0.81
8	MSA	0.18	**0.1**	-0.16	0.47	4	21	0.89	0.81
	EA	0.19	0.13	-0.05	0.51	5	21	0.85	0.83
9	MSA	0.19	**0.1**	0.02	0.46	1	21	0.9	0.82
	EA	0.19	0.12	-0.01	0.48	5	21	0.85	0.85
10	MSA	0.19	**0.09**	0.02	0.44	1	21	0.9	0.83
	EA	0.21	0.11	-0.07	0.48	5	20	0.85	0.86

child, were still among the most significant clusterings for both approaches. However, a broader range of clusterings was significant for MSA in comparison with the full dataset, while the opposite was true for EA. In general, timing was an important feature for classifying men's sequences into clusters since most clustering involved separations according to the arrival of a child in the household and MSA was better at taking that into account.

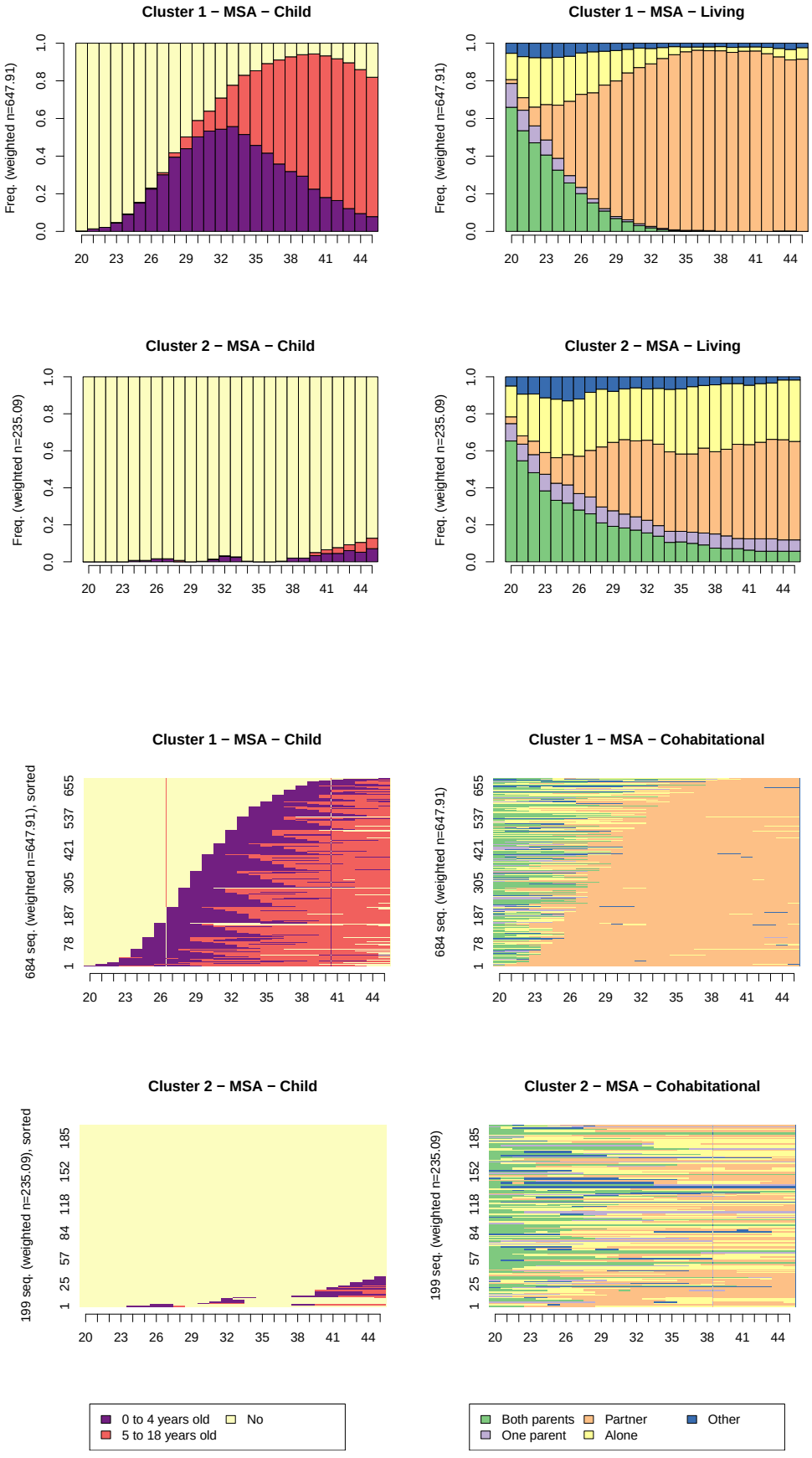

Figure 25: Chronograms (top) and index plots (bottom) of the child and cohabitational status typology in two groups obtained with MSA for men.

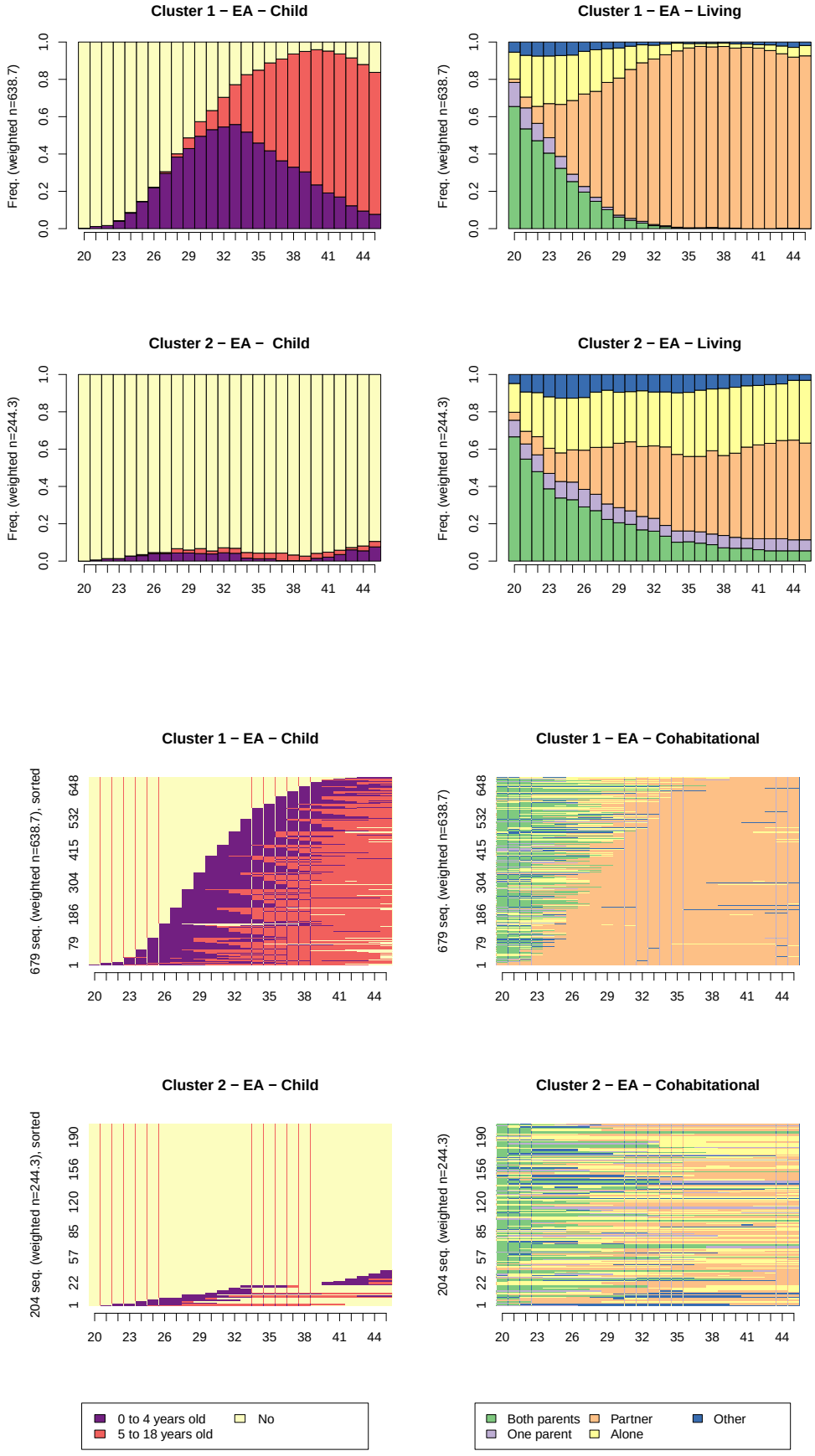

Figure 26: Chronograms (top) and index plots (bottom) of the child and cohabitational status typology in two groups obtained with EA for men.

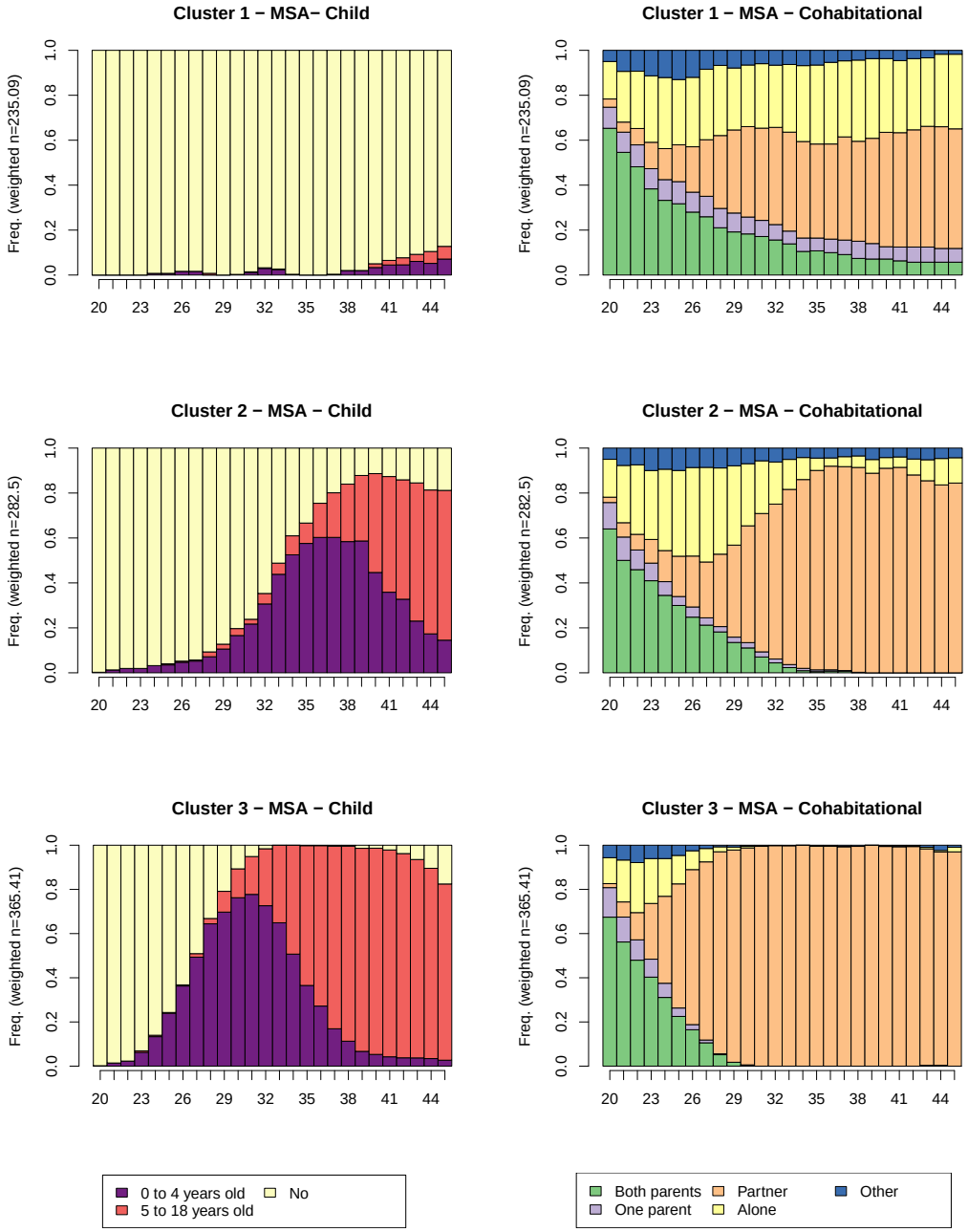

Figure 27: Chronograms of the child and cohabitational status typology in three groups obtained with MSA for men.

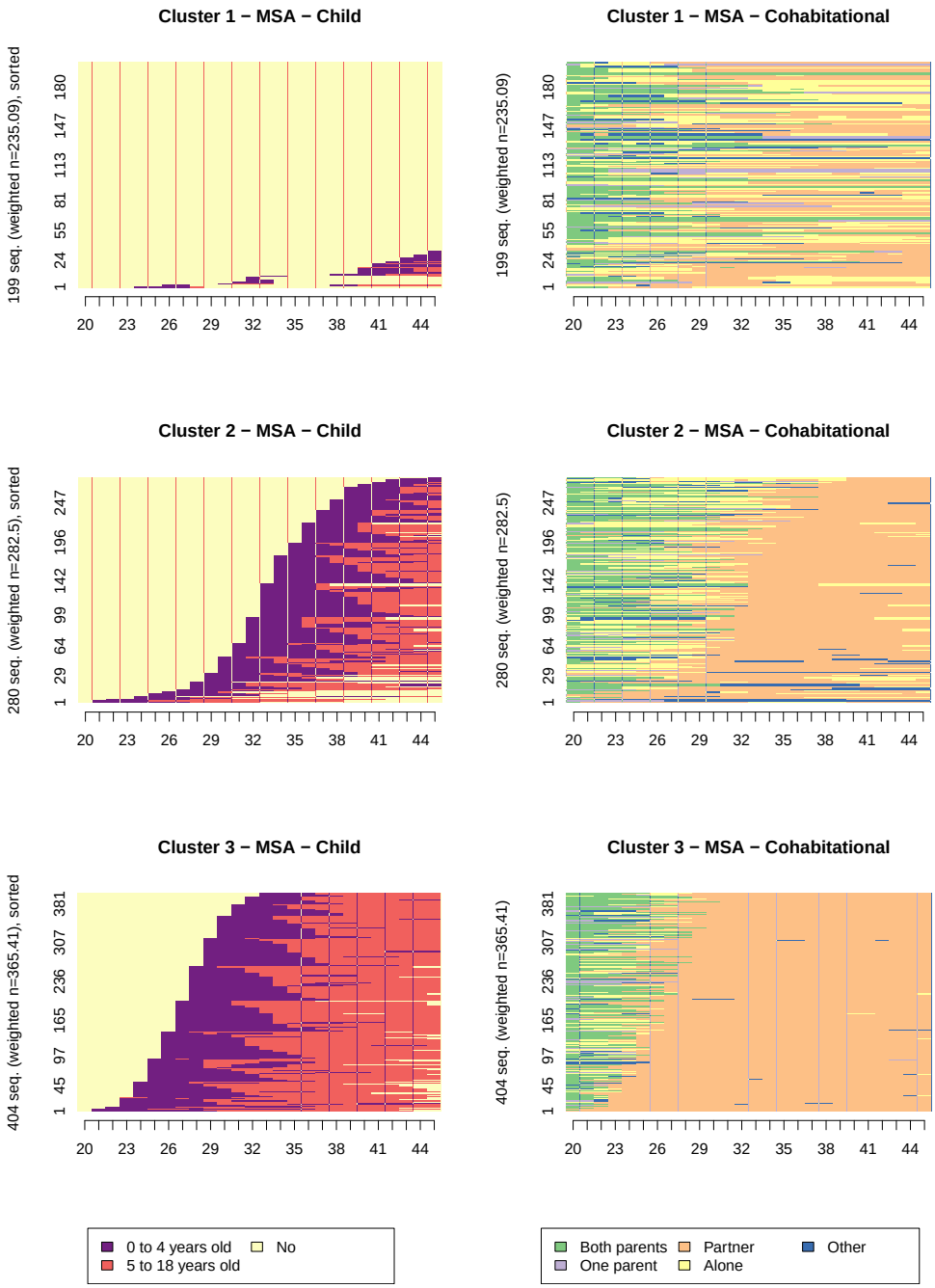

Figure 28: Index plots of the child and cohabitational status typology in three groups obtained with MSA for men.

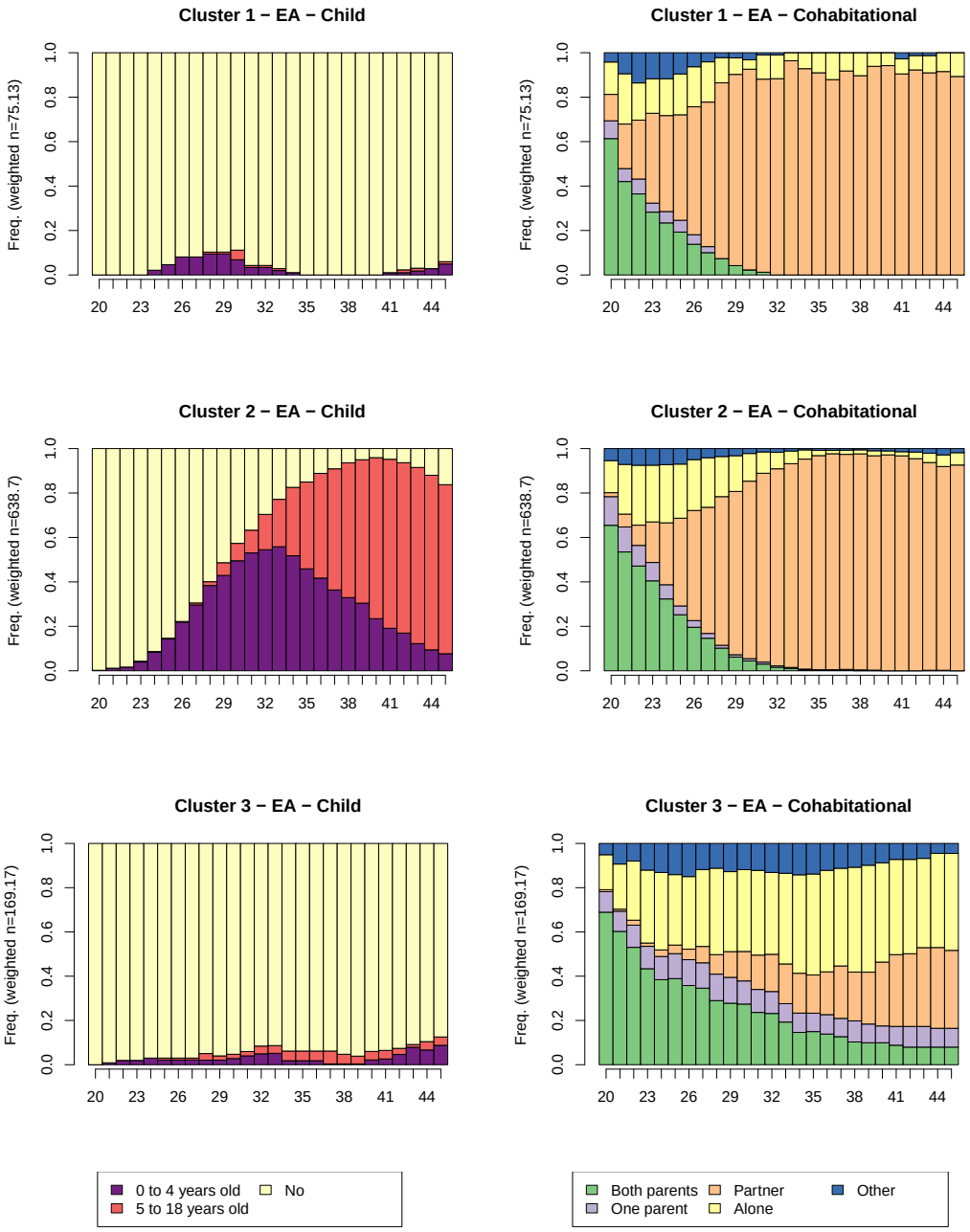

Figure 29: Chronograms of the child and cohabitational status typology in three groups obtained with EA for men.

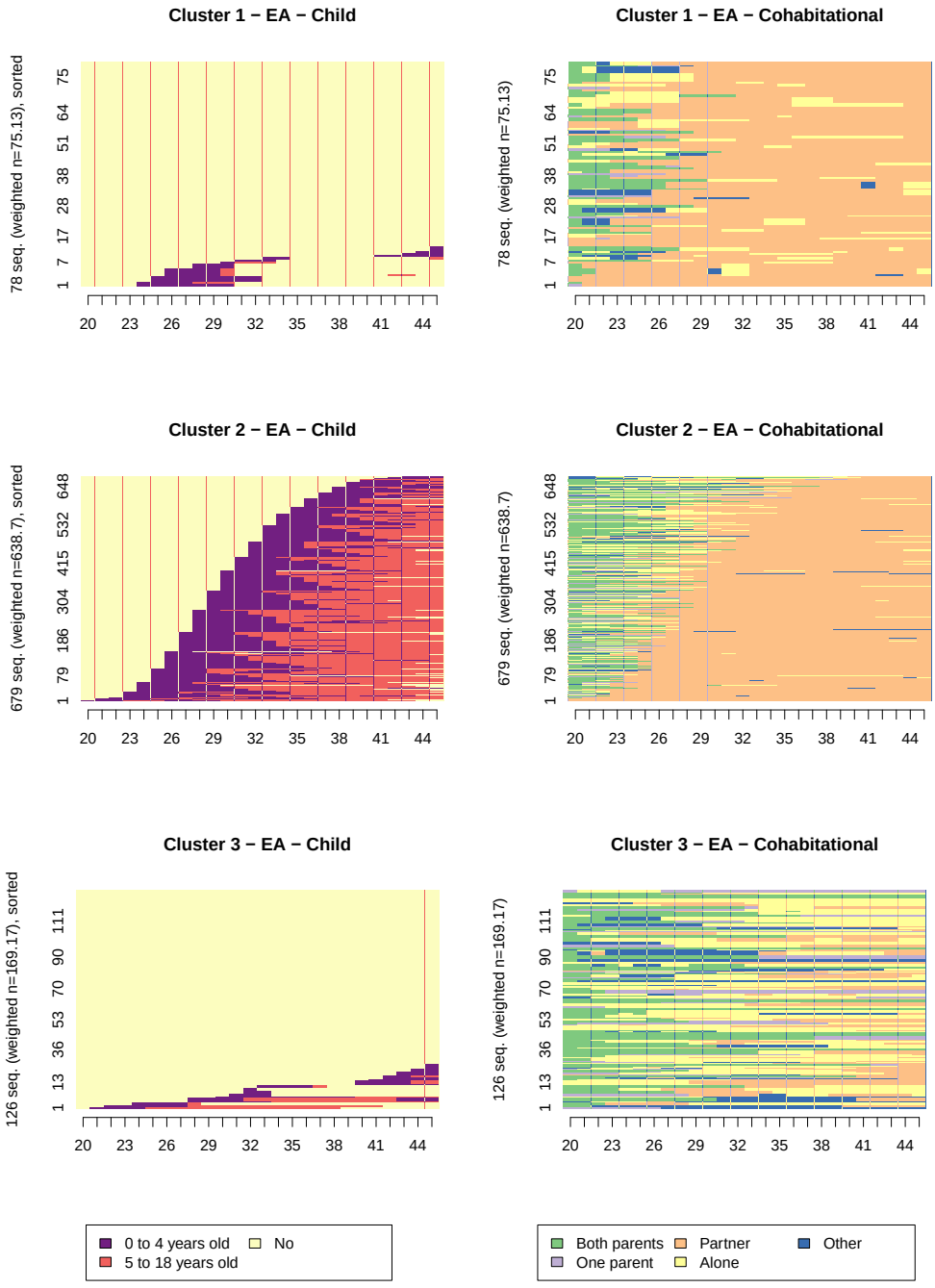

Figure 30: Index plots of the child and cohabitational status typology in three groups obtained with EA for men.

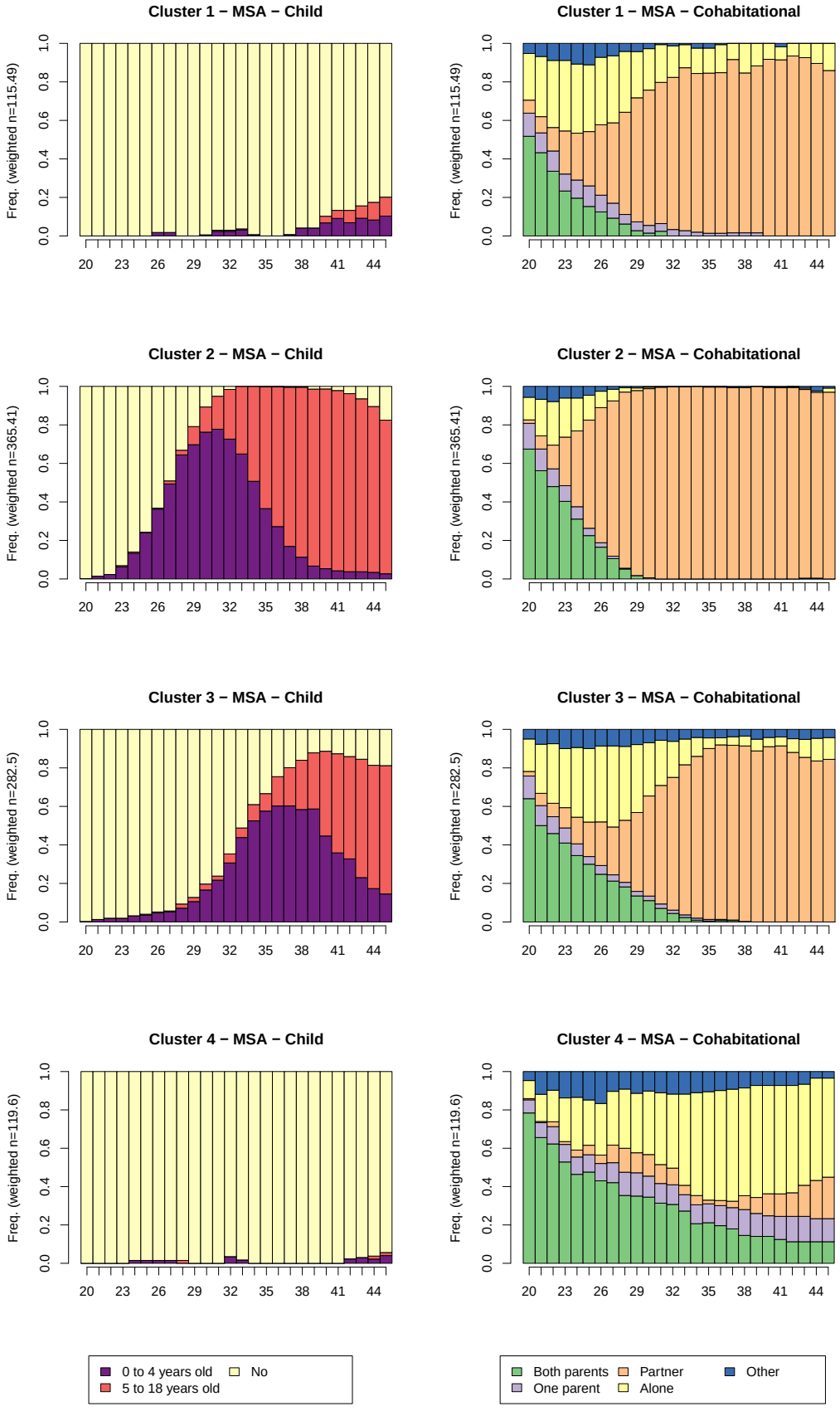

Figure 31: Chronograms of the child and cohabitational status typology in four groups obtained with MSA for men.

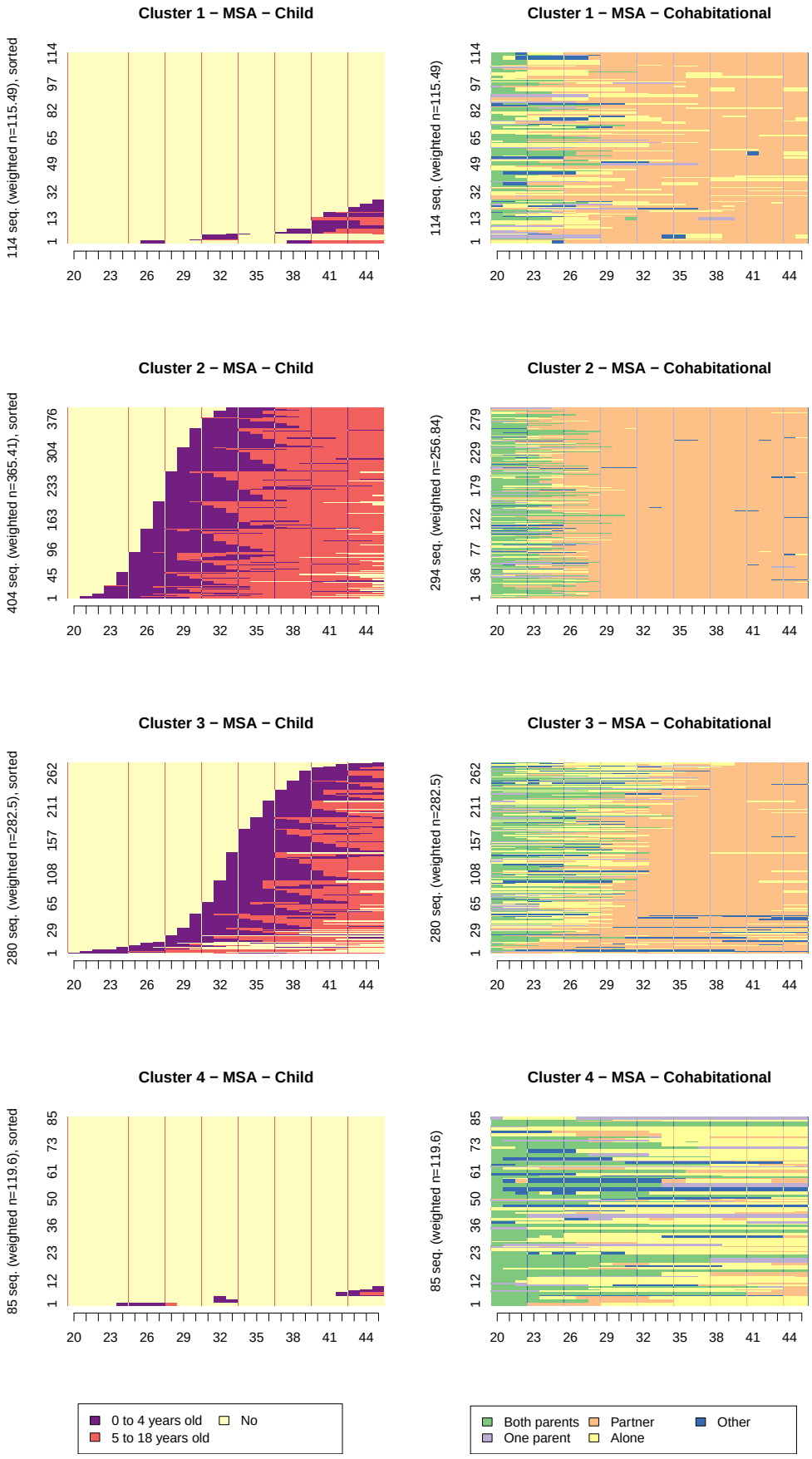

Figure 32: Index plots of the child and cohabitational status typology in four groups obtained with MSA for men.

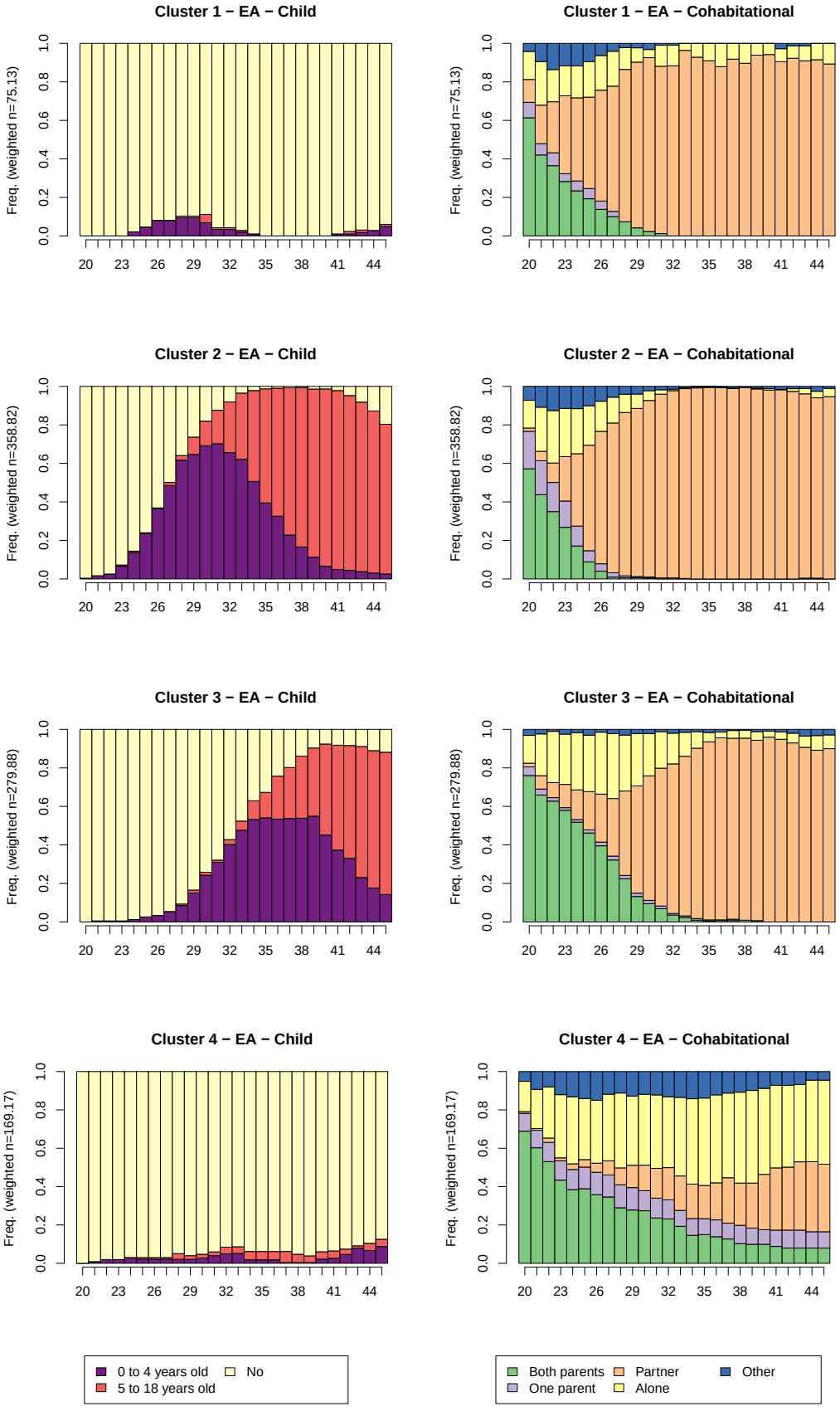

Figure 33: Chronograms of the child and cohabitational status typology in four groups obtained with EA for men.

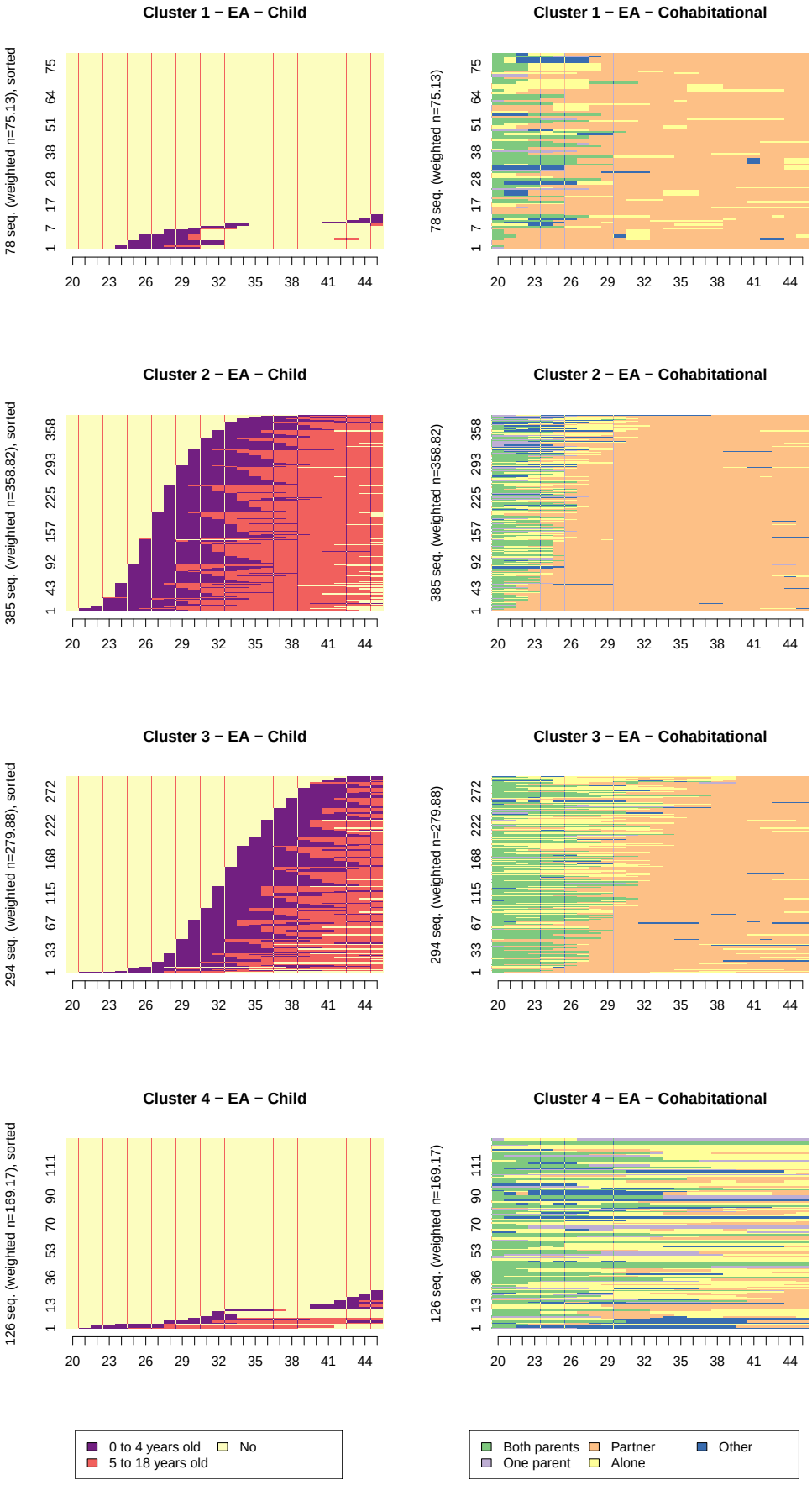

Figure 34: Index plots of the child and cohabitational status typology in four groups obtained with EA for men.

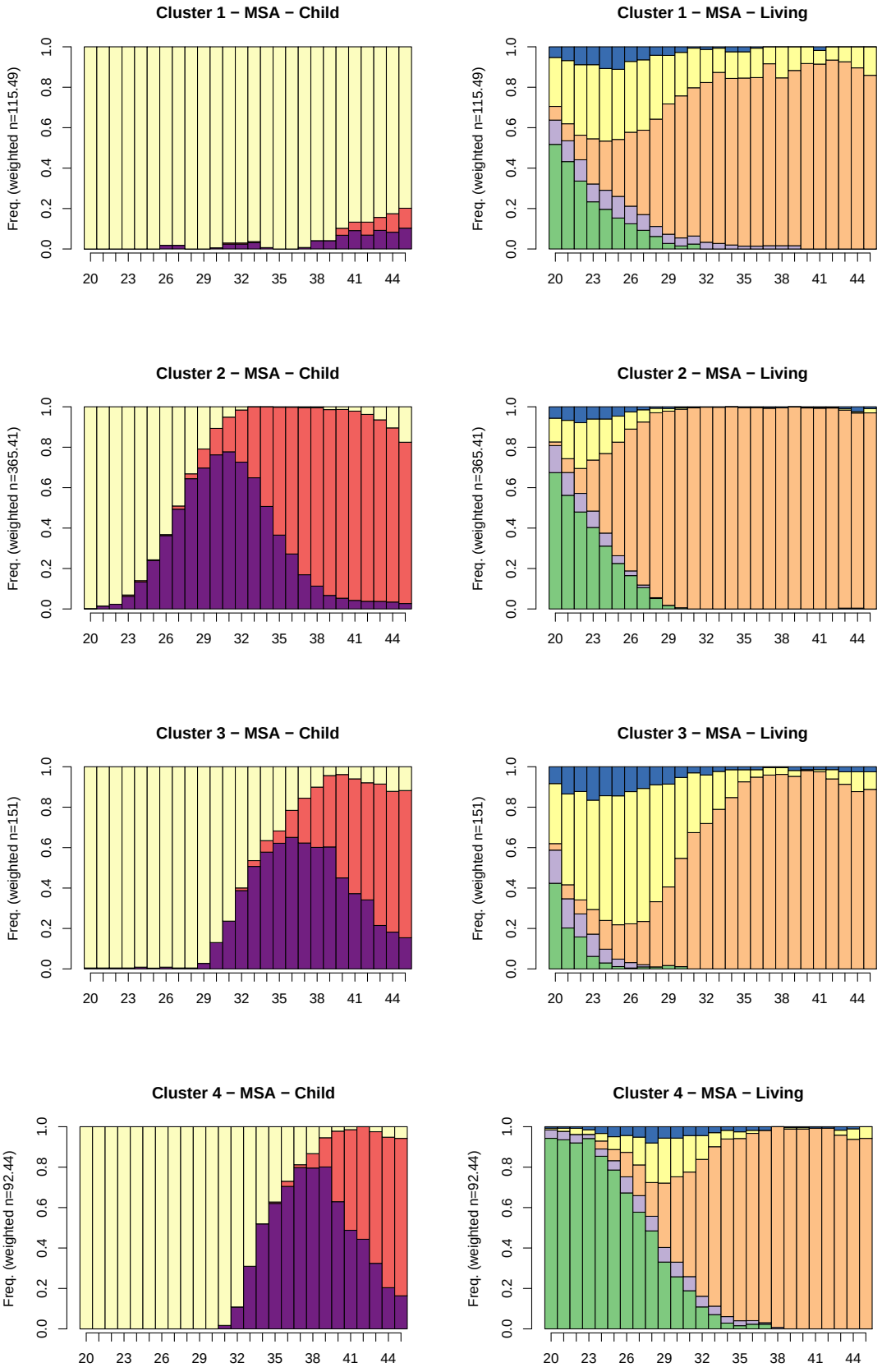

Figure 35: Chronograms of the child and cohabitational status typology in seven groups obtained with MSA for men.

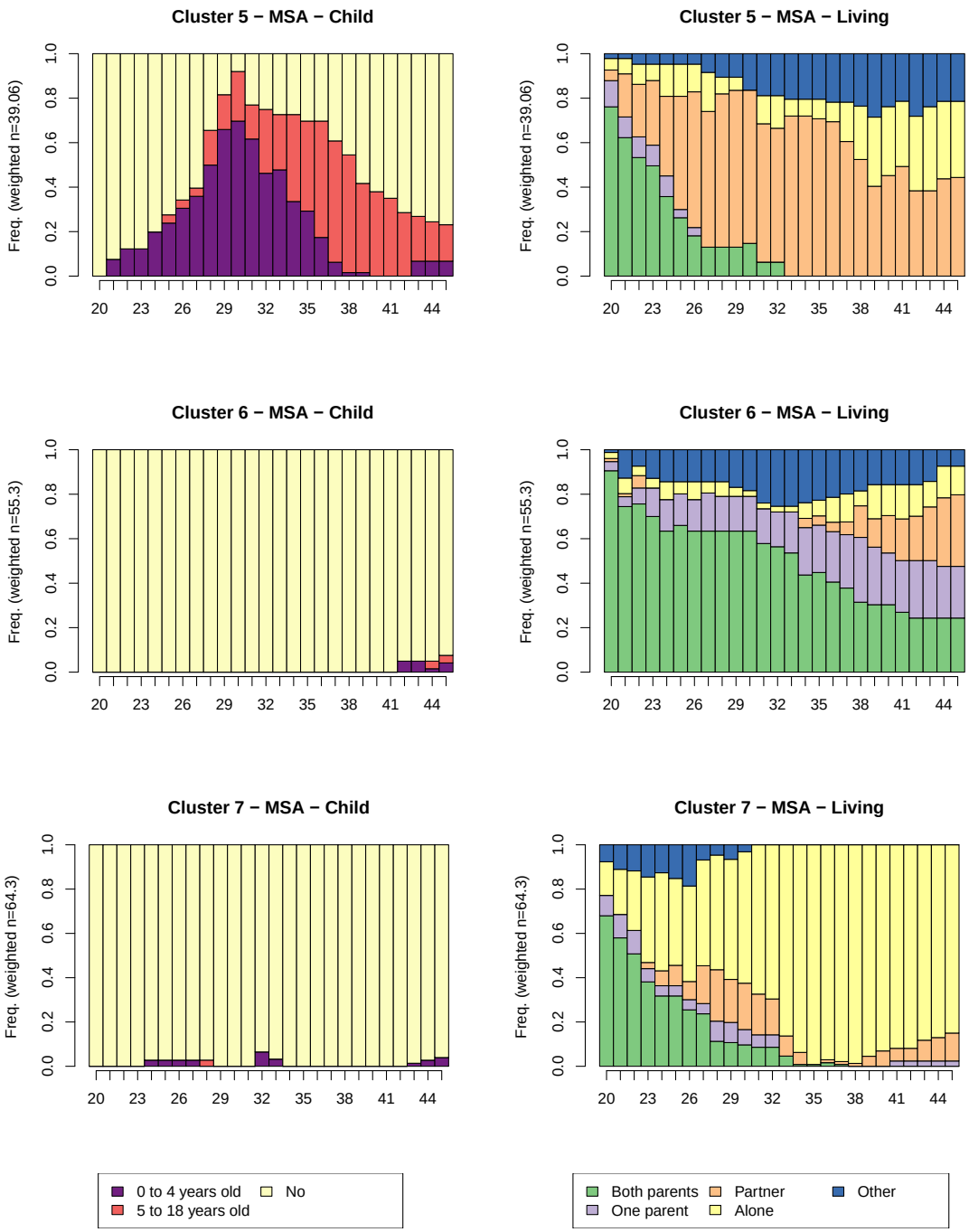

Figure 35: Chronograms of the child and cohabitational status typology in seven groups obtained with MSA for men (continued).

4.2.2 Women

For women, as for the full dataset, the health issues trajectories were disconnected from all other domains. Each pair of domains involving health issues had a Cronbach's α smaller than 0.04. The conclusion was different for the professional status domain. Although the cohabitational status and child trajectories, which had a Cronbach's α of 0.54, were still the most interrelated domains, the professional status domain was somehow linked to them. Taking the three domains together, we obtained a Cronbach's α of 0.51, while the pairs of professional status and, respectively, child and cohabitational status domains gave values of 0.39 and 0.26, respectively. Although these values are not high, they are larger than those obtained with the full dataset. Therefore, a joint typology of the cohabitational status, child, and professional status trajectories might be suitable, like expected beforehand, in addition to those typologies involving only two domains.

Child–Cohabitational status–Professional status At first sight, MSA seemed better at summarising the information issued from the three domains. We obtained correlations of 0.75, 0.63, and 0.7 between d_{MSA} and, respectively, d_{Child}, d_{Prof}, and d_{Cohab}, while the respective correlations for d_{EA} with the single domains were 0.48, 0.7, and 0.48.

As before, we determined which clusterings were meaningful using bootstrap validation. Figure 36 shows that the solutions in the two and six groups built by MSA were significant in terms of both ASWw and HC, with splitting into two groups clearly more significant than into six groups. For EA, the solution in three groups was significant. Other clusterings were also significant, but the number of groups was too important (e.g. nine or ten). Therefore, some were ill-defined, while others were too small to allow us to generalise (Table 7).

Since more than two domains were involved, we applied the bootstrap procedure to determine whether each individual domain was taken into account by these clusterings. Figure 37 shows that the three domains were taken into account by the two-group solution built by MSA since, for each individual domain, the ASWw and HC obtained with our data were respectively above and below the confidence intervals built under the hypothesis that the individual domain is disassociated from the others. However, there was a larger significance, both in terms of ASWw and in terms of HC, when the sequences of the child domain were randomly permuted compared with randomising the two other domains, probably because the child domain was more linked, according to the Cronbach's α value, to the other domains. Concerning the grouping in three groups built with EA, the cohabitational status

domain was clearly not taken into account (Figure 38). We cannot really draw any conclusions for the other two domains since only ASWw was significant. This suggests that EA had some difficulties combining the information issued from the different domains. This is explainable by the fact that combining the three domains produced an extended alphabet of 60 states, with a few rare categories. Some substitution costs violated the triangle inequality producing an inconsistent dissimilarity measure. It was, for example, less costly to first substitute the extended state *0 to 4 years old/Both parents/Education* with *5 to 18 years old/Other/Education* before substituting the latter with *0 to 4 years old/Other/Education* than to directly substitute *0 to 4 years old/Both parents/Education* with *0 to 4 years old/Other/Education*. Therefore, we have reached a limitation of the EA approach here.

The two-group solution built with MSA and presented in Figure 39 was characterised by one cluster of women having a child, mainly living with a partner, which were more prone to part-time working or non-working, and one cluster of women not having a child with a variety of living statuses and working mainly full-time. Both groups were relatively homogeneous since the ASWw values by group, 0.36 and 0.35, were balanced. The R^2 values by group were 0.82 for the child domain, 0.66 for the cohabitational status domain, and 0.47 for the professional status domain. These values confirmed that the clustering was more driven by the central domain according to Cronbach's α, namely, the child one.

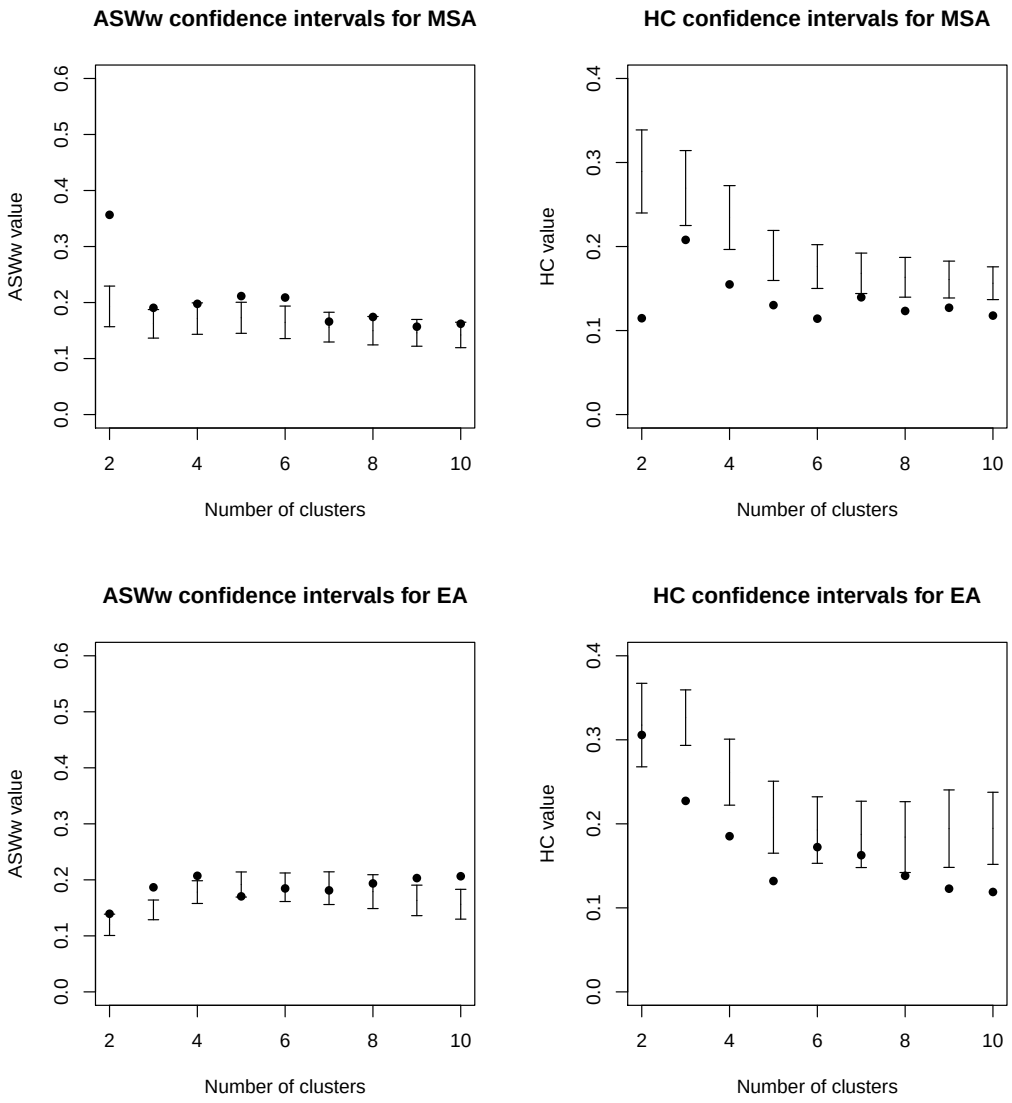

Figure 36: ASWw and HC values obtained by clustering the data, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that child, cohabitational and professional domains are disassociated.

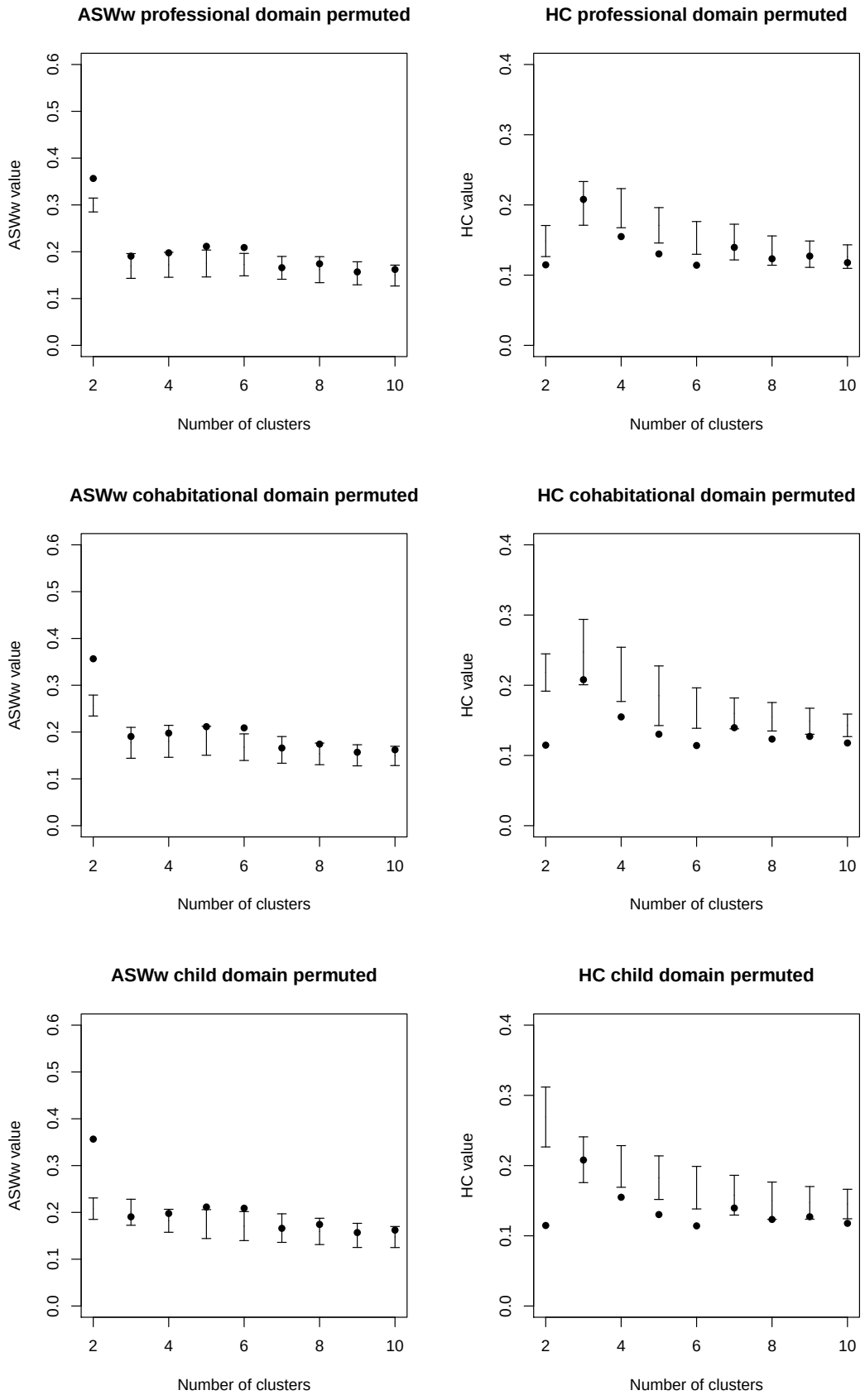

Figure 37: ASWw and HC values obtained by clustering the data with MSA, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that the professional (top), cohabitational (middle) or child domain (bottom) is disassociated from the others.

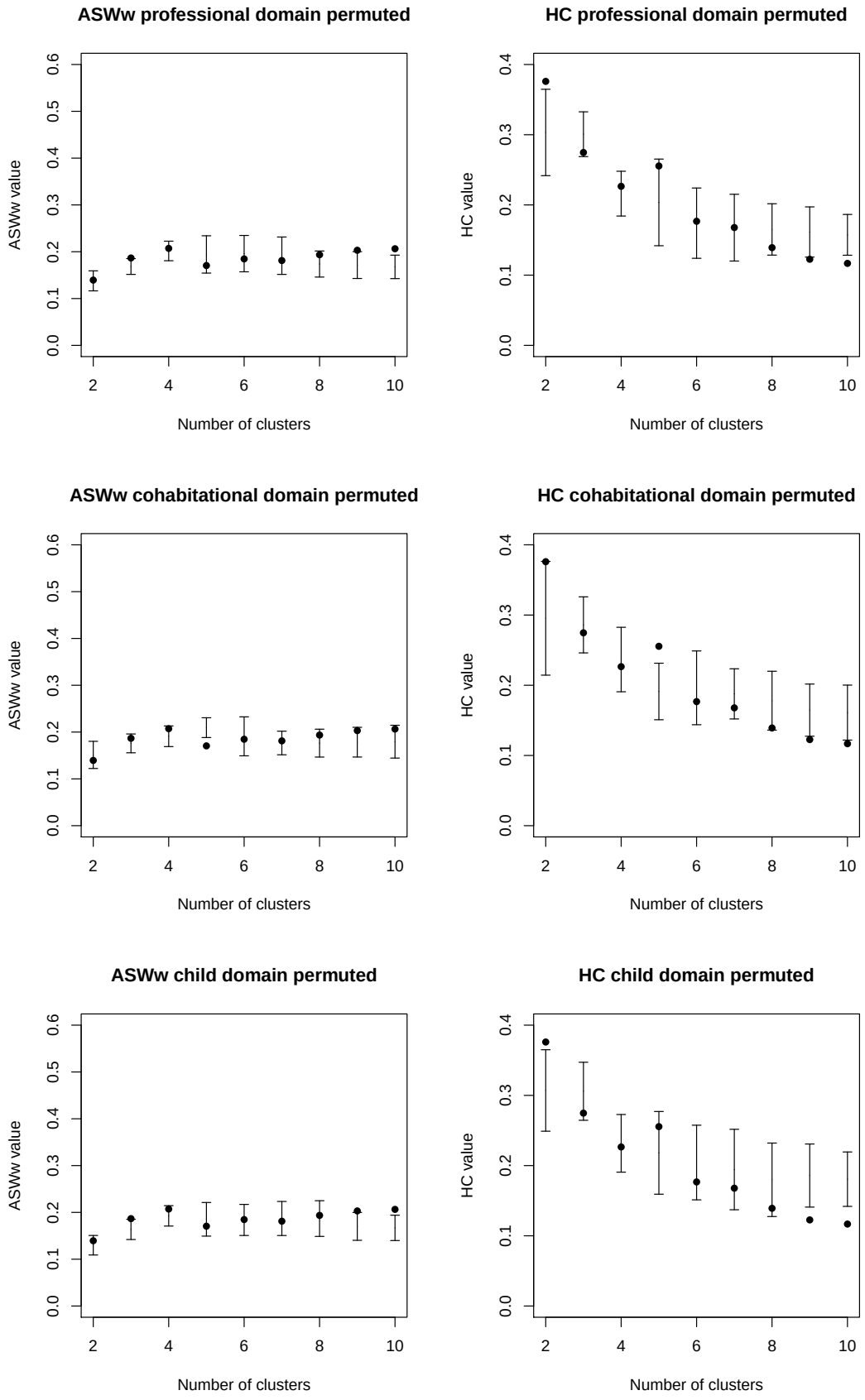

Figure 38: ASWw and HC values obtained by clustering the data with EA, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that the professional (top), cohabitational (middle) or child domain (bottom) is disassociated from the others.

Table 7: Summary of the results obtained by clustering the child, cohabitational and professional status channels with MSA and EA. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	Method	CQI		Group ASWw		Size clusters (%)		R^2 by channel		
		ASWw	HC	min	max	min	max	child	cohab	prof
2	MSA	**0.36**	**0.11**	0.35	0.36	23	77	0.82	0.66	0.47
	EA	**0.14**	0.38	0.01	0.47	26	74	0.51	0.6	0.57
3	MSA	**0.19**	**0.21**	0.01	0.35	23	45	0.82	0.67	0.62
	EA	**0.19**	**0.27**	0.01	0.42	26	43	0.63	0.68	0.68
4	MSA	0.2	**0.15**	0.01	0.28	20	32	0.84	0.67	0.69
	EA	**0.21**	0.23	0	0.42	10	33	0.69	0.71	0.69
5	MSA	**0.21**	**0.13**	0.14	0.26	9	32	0.88	0.68	0.72
	EA	0.17	0.26	-0.01	0.3	10	33	0.7	0.71	0.73
6	MSA	**0.21**	**0.11**	0.12	0.26	7	32	0.88	0.7	0.74
	EA	0.18	0.18	-0.12	0.3	10	26	0.75	0.73	0.74
7	MSA	0.17	**0.14**	0.08	0.29	7	25	0.9	0.7	0.76
	EA	0.18	0.17	-0.14	0.44	6	26	0.75	0.77	0.75
8	MSA	0.17	**0.12**	0.07	0.29	4	21	0.9	0.73	0.76
	EA	0.19	0.14	-0.11	0.44	5	26	0.78	0.77	0.76
9	MSA	0.16	**0.13**	0.03	0.26	45	18	0.9	0.73	0.79
	EA	**0.2**	**0.12**	-0.08	0.44	4	26	0.81	0.78	0.77
10	MSA	0.16	**0.12**	0.07	0.25	4	18	0.9	0.74	0.81
	EA	**0.21**	**0.12**	-0.1	0.4	4	26	0.81	0.78	0.78

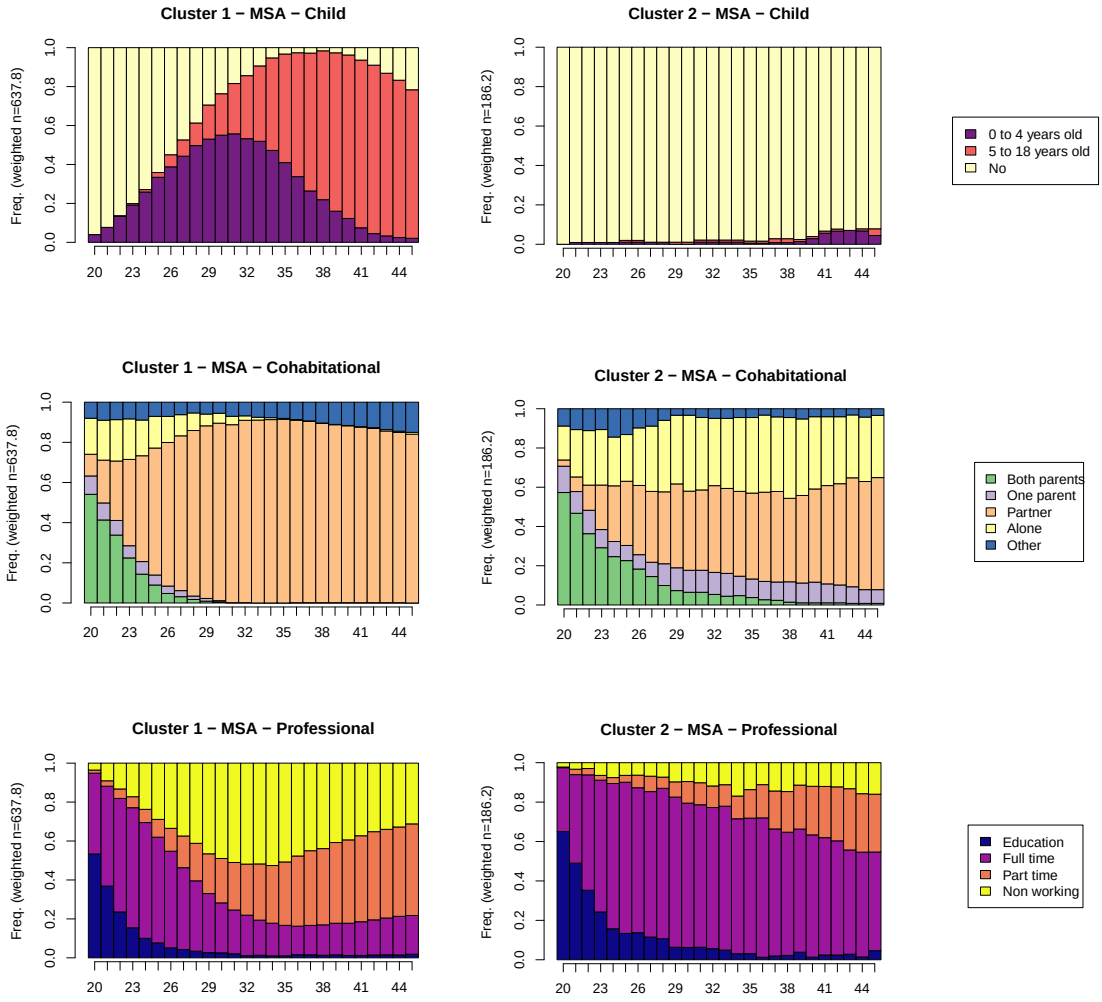

Figure 39: Chronograms of the child, cohabitational and professional status typology in two groups obtained with MSA for women.

Cohabitational status–Professional status Among the three related domains (child, cohabitational status, and professional status), the cohabitational status–professional status pair had the lowest Cronbach's α (0.26). According to the correlations, EA summarised the professional status domain well (correlation of 0.80), probably at the expense of the cohabitational status domain (0.53). On the contrary, both correlations were higher than 0.7 for MSA, with a slightly higher value for the cohabitational status domain (0.76 vs 0.73).

Regarding the clusterings, only separation into two groups was significant for MSA in terms of both HC and ASWw (Figure 40). The first cluster (Figure 41) was composed of women working part-time or not working any longer in their forties who mainly lived with a partner. The second cluster contained women that worked full-time in their forties. The women in this cluster had a variety of cohabitational statuses. The first group, which had an ASWw of 0.32, was more homogeneous than the second group, which had an ASWw of 0.24. The R^2 was 0.63 for the cohabitational status domain and only 0.57 for the professional status domain. Therefore, this clustering was slightly more driven by cohabitational status (Table 8).

For EA, only the three-group solution was significant in terms of both HC and ASWw. The women were mainly split according to their professional status (part-time vs non-working vs full-time) (Figure 42). The second group was the most homogeneous (ASWw of 0.49). The first group was still relatively well defined (ASWw of 0.28), while the last group was ill-defined (0.09). According to the R^2 values (0.73 for the professional status domain and 0.69 for the cohabitational status domain), the clustering was also slightly more driven by the professional status.

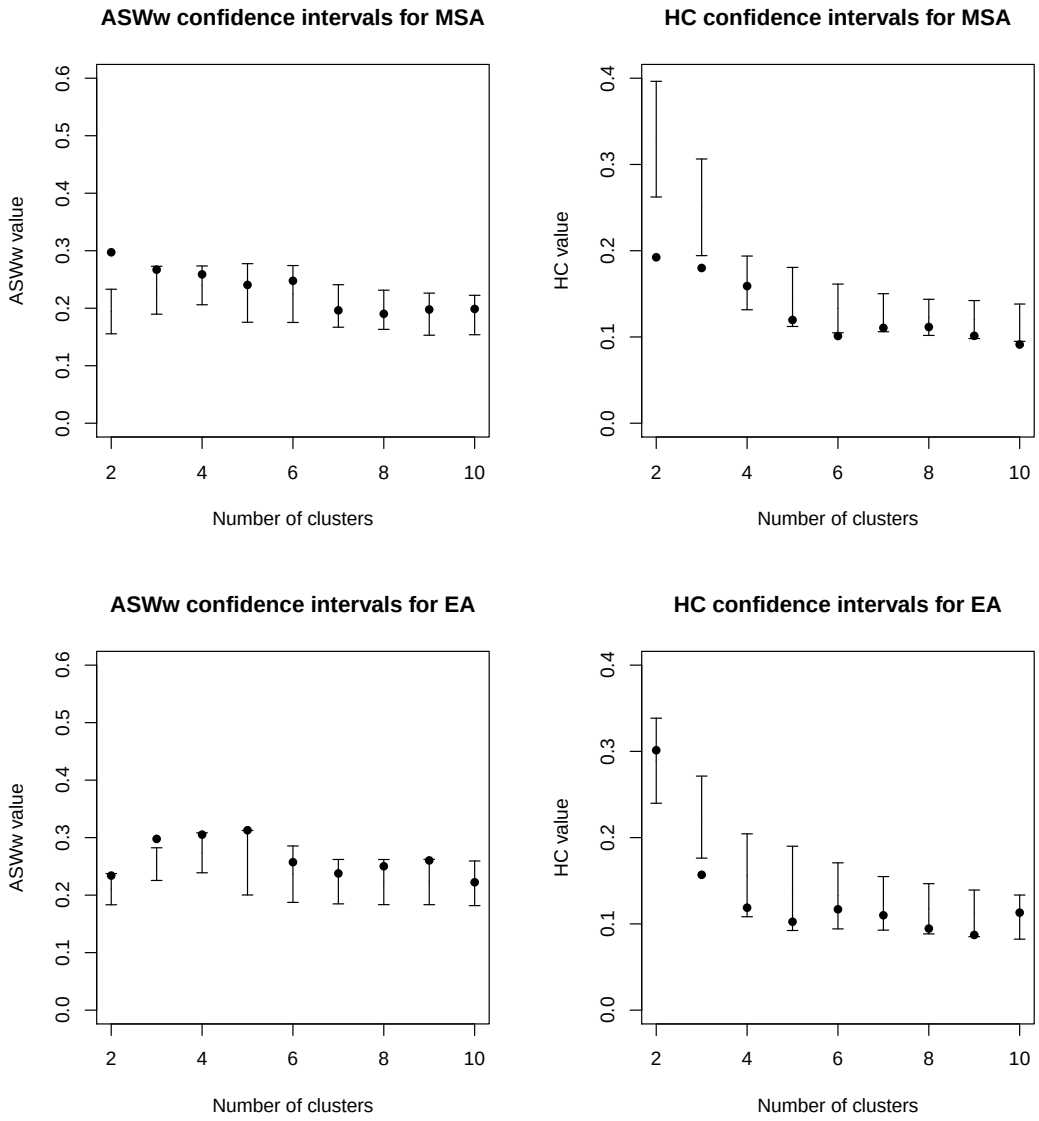

Figure 40: ASWw and HC values obtained by clustering the data, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that cohabitational and professional domains are disassociated.

Table 8: Summary of the results obtained by clustering the cohabitational and professional status channels for women with MSA and EA. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	Method	CQI		Group ASWw		Size clusters (%)		R^2 by channel	
		ASWw	HC	min	max	min	max	cohab	prof
2	MSA	**0.3**	**0.19**	0.24	0.32	27	73	0.63	0.57
	EA	0.23	0.3	0.09	0.52	32	68	0.6	0.61
3	MSA	0.27	**0.18**	0.16	0.45	27	41	0.64	0.76
	EA	**0.3**	**0.16**	0.09	0.49	32	36	0.69	0.73
4	MSA	0.26	0.16	0.14	0.43	13	41	0.7	0.76
	EA	0.31	0.12	-0.05	0.49	12	36	0.74	0.75
5	MSA	0.24	0.12	-0.1	0.38	13	32	0.75	0.78
	EA	**0.31**	0.1	-0.11	0.47	6	36	0.77	0.75
6	MSA	0.25	**0.1**	0.01	0.38	7	32	0.77	0.78
	EA	0.26	0.12	-0.11	0.44	6	32	0.77	0.78
7	MSA	0.2	0.11	-0.04	0.26	7	32	0.77	0.82
	EA	0.24	0.11	-0.12	0.42	6	32	0.77	0.81
8	MSA	0.19	0.11	-0.05	0.3	7	25	0.77	0.84
	EA	0.25	0.09	-0.08	0.42	4	32	0.78	0.82
9	MSA	0.2	0.1	-0.05	0.35	5	25	0.8	0.84
	EA	0.26	0.09	-0.05	0.42	3	32	0.8	0.83
10	MSA	0.2	**0.09**	-0.12	0.44	5	25	0.8	0.85
	EA	0.22	0.11	-0.05	0.42	3	21	0.8	0.85

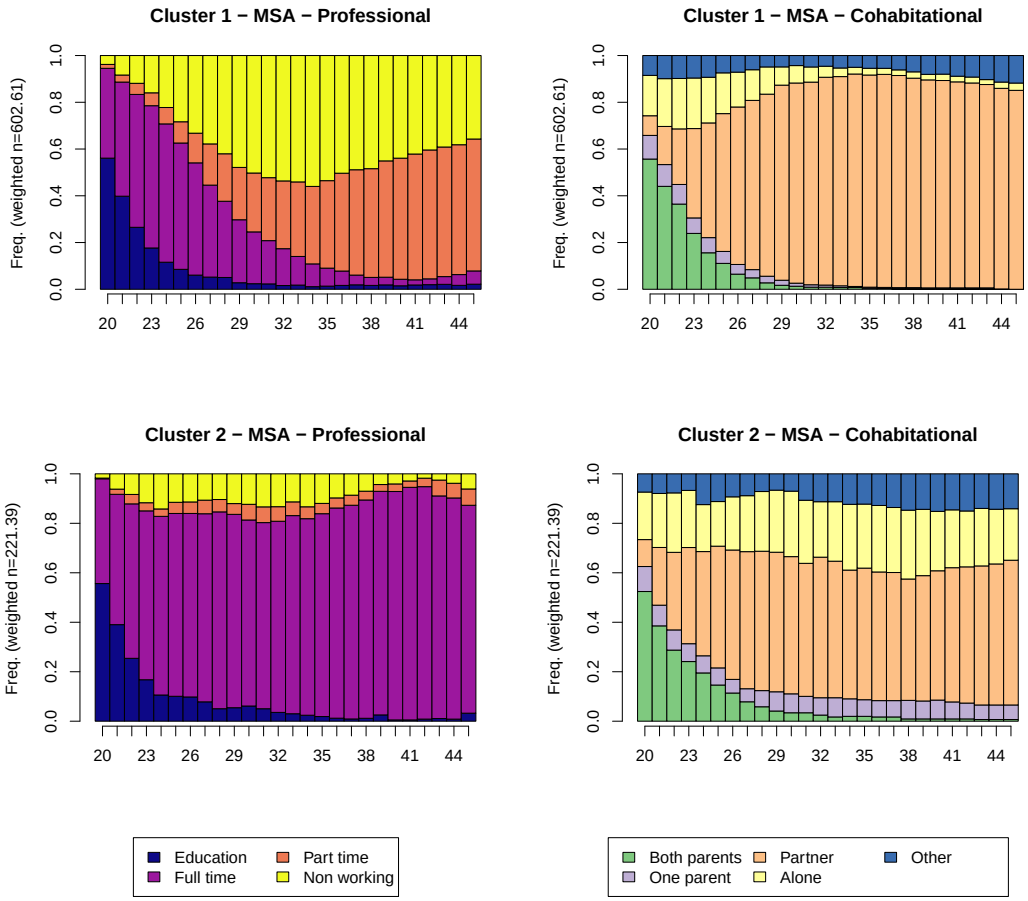

Figure 41: Chronograms of the cohabitational and professional status typology in two groups obtained with MSA for women.

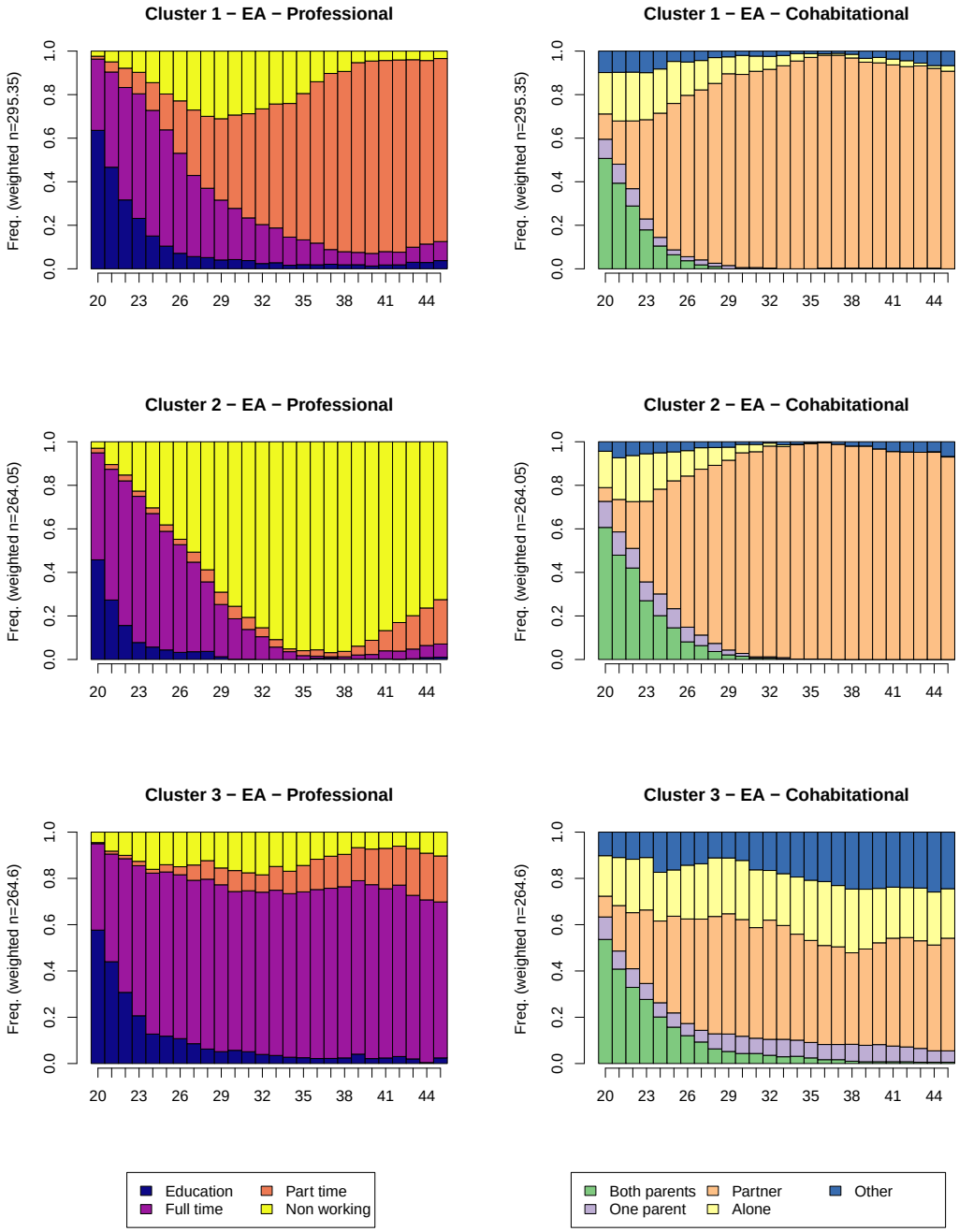

Figure 42: Chronograms of the cohabitational and professional status typology in three groups obtained with EA for women.

Child–Professional status Since the Cronbach's α obtained from these two domains was 0.39, it was worth extracting a joint typology based on the child and professional status domains for women. According to the correlations, MSA summarised the pairwise dissimilarities from each individual domain almost equivalently (0.78 for child and 0.77 for professional status), while EA placed higher importance on the professional status domain (0.81) than the child domain (0.55).

Separation into two clusters was significant both in terms of ASWw and in terms of HC for MSA, while only ASWw was significant for EA. Thus, both approaches provided significant clusterings for the separation into three and four groups (Figure 43). Clusterings into seven and eight groups were also significant for EA but not as much as for the two others. The two-group solution of MSA mainly separated women based on whether they had a child (Figure 44), while the separation made by EA involved one cluster of women having a child and not working and a cluster aggregating women not having a child and women working and having a child (Figure 45). The groups were more homogeneous for MSA since the ASWw values by cluster were 0.33 and 0.55 compared with 0.43 and 0.12 for EA. Then, separation into three groups (Figures 46, 47, 48, and 49) and four groups (Figures 50, 51, 52, and 53) obtained by both approaches were relatively similar (85% and 95% agreement, respectively). However, the three-group solution built by MSA was less balanced in terms of ASWw by group. The value of the second cluster was only 0.1, while none of the values were smaller than 0.2 for EA. The solution was more driven by the child domain for MSA, while the clustering made by EA explained an almost equivalent part of the variance in pairwise dissimilarities for each of the individual domains (Table 9). For both approaches, the ASWw by group values were relatively balanced for the four-group solution. They were between 0.33 and 0.39 for EA, while the values for MSA were more heterogeneous, varying between 0.27 and 0.43. Both four-group solutions had almost identical R^2 values and were more driven by the child domain.

With EA, one would either choose the three-group solution because it had the larger relative distance to the confidence intervals in terms of both ASWw and HC and R^2 was almost equal for both channels or the four-group solution because ASWw by group was well-balanced and we observed an important increase in the R^2 of the child domain between the three- and four- group solutions. With MSA, the relative distance to the confidence intervals was also maximised for the clustering into three group. However, one of group had a small ASWw and R^2 was imbalanced. The four-group solutions was therefore preferable. This illustrated that it was not always possible to find a solution that optimize all the criteria.

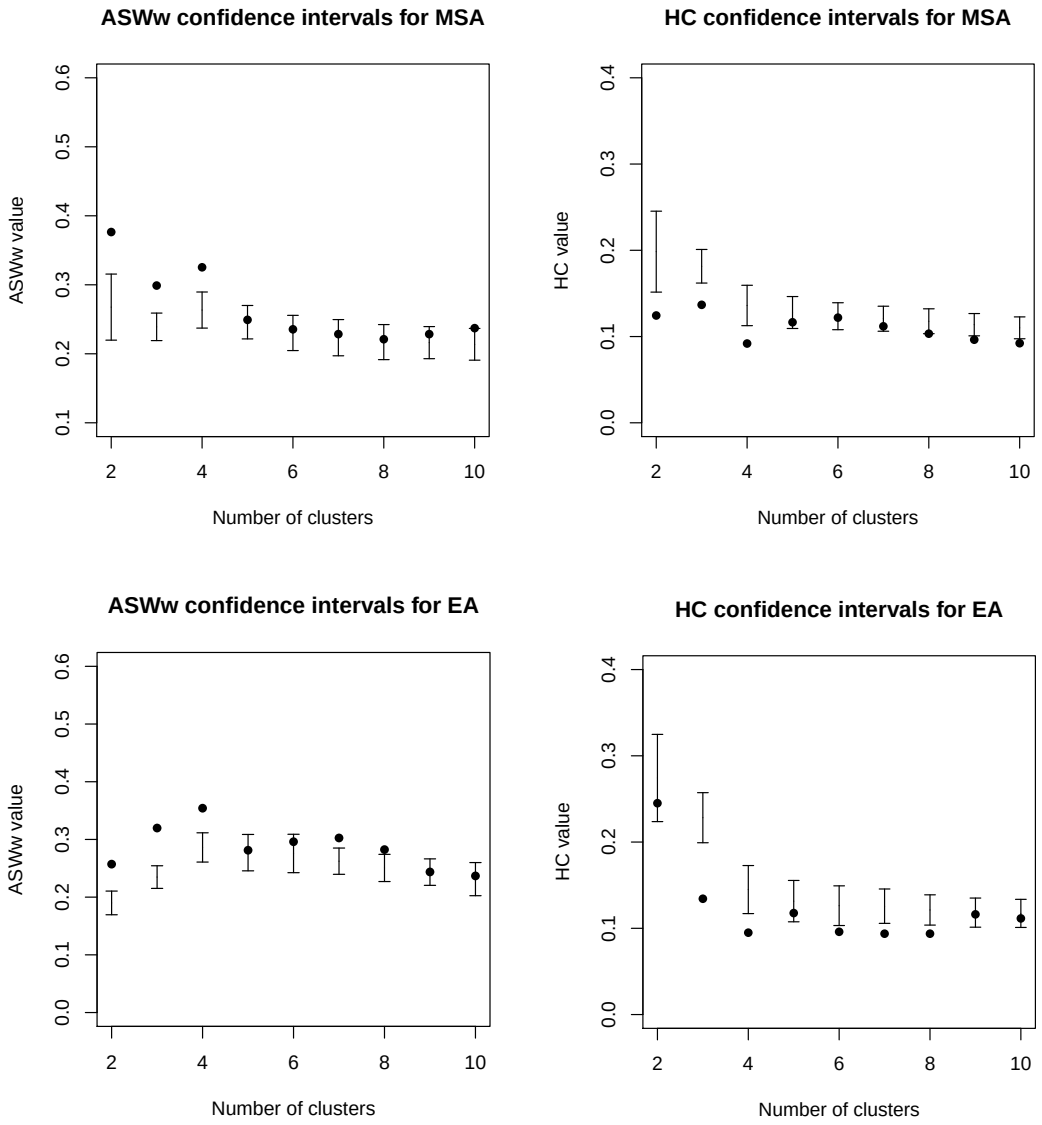

Figure 43: ASWw and HC values obtained by clustering the data, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that professional and child domains are disassociated.

Table 9: Summary of the results obtained by clustering the child and professional status channels for women with MSA and EA. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	Method	CQI		Group ASWw		Size clusters (%)		R^2 by channel	
		ASWw	HC	min	max	min	max	child	prof
2	MSA	**0.38**	**0.12**	0.33	0.55	23	77	0.82	0.47
	EA	**0.26**	0.25	0.12	0.43	42	58	0.57	0.61
3	MSA	**0.3**	**0.14**	0.1	0.51	23	40	0.83	0.64
	EA	**0.32**	**0.13**	0.24	0.36	24	42	0.7	0.74
4	MSA	**0.33**	**0.09**	0.27	0.43	12	40	0.83	0.74
	EA	**0.35**	**0.09**	0.33	0.39	11	42	0.83	0.75
5	MSA	0.25	0.12	0.19	0.41	12	26	0.86	0.75
	EA	0.28	0.12	0.18	0.39	11	30	0.83	0.78
6	MSA	0.24	0.12	0.12	0.41	7	24	0.86	0.78
	EA	0.3	**0.1**	0.17	0.58	5	30	0.83	0.82
7	MSA	0.23	0.11	0.1	0.37	6	23	0.88	0.79
	EA	**0.3**	**0.09**	0.16	0.6	2	30	0.83	0.83
8	MSA	0.22	0.1	0.04	0.61	6	19	0.88	0.81
	EA	**0.28**	**0.09**	0.05	0.6	2	30	0.85	0.84
9	MSA	0.23	**0.1**	0.04	0.61	5	19	0.9	0.82
	EA	0.24	0.12	-0.01	0.6	2	17	0.85	0.85
10	MSA	0.24	**0.09**	0.1	0.58	3	19	0.9	0.83
	EA	0.24	0.11	-0.01	0.6	2	17	0.87	0.85

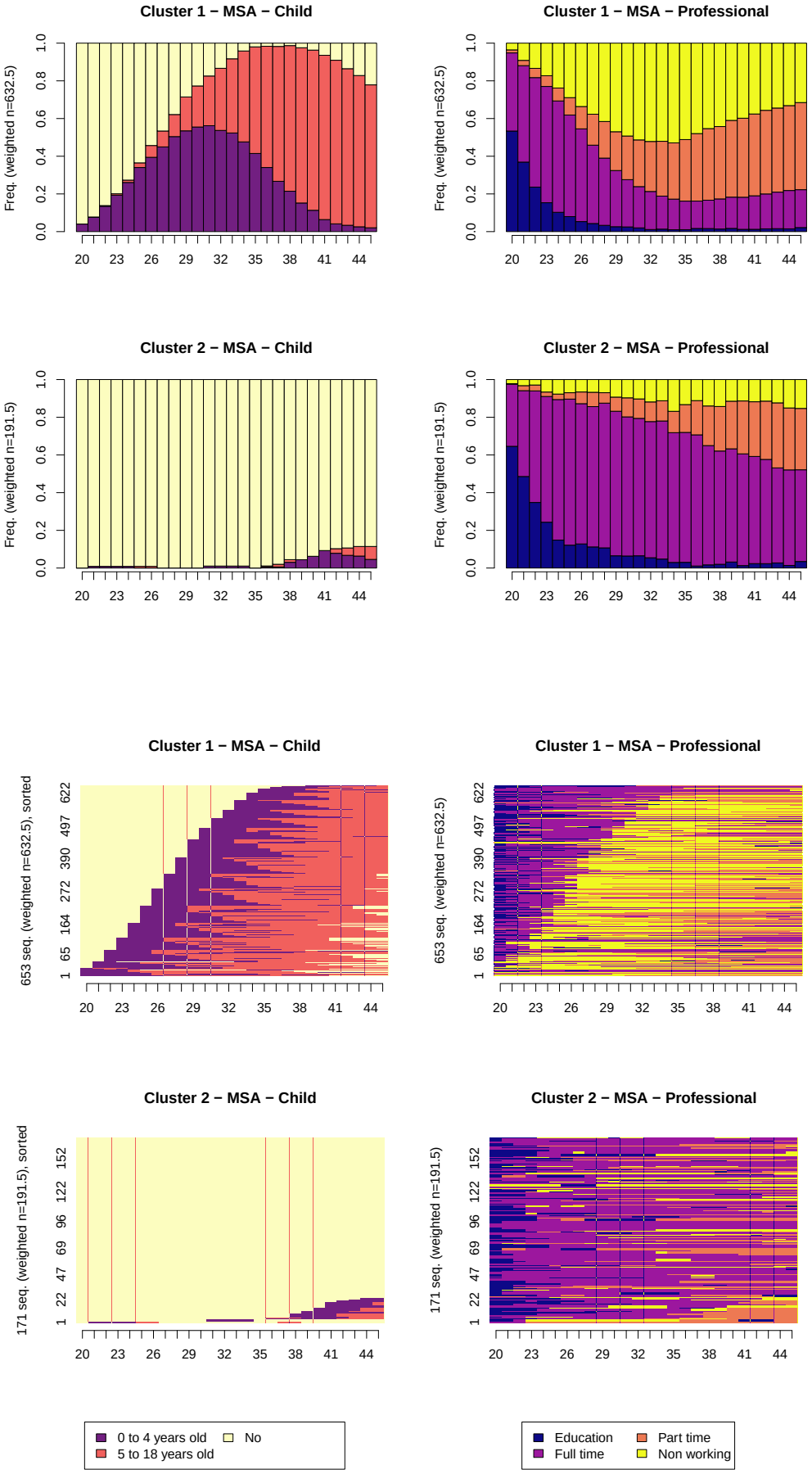

Figure 44: Chronograms (top) and index plots (bottom) of the child and professional status typology in two groups obtained with MSA for women.

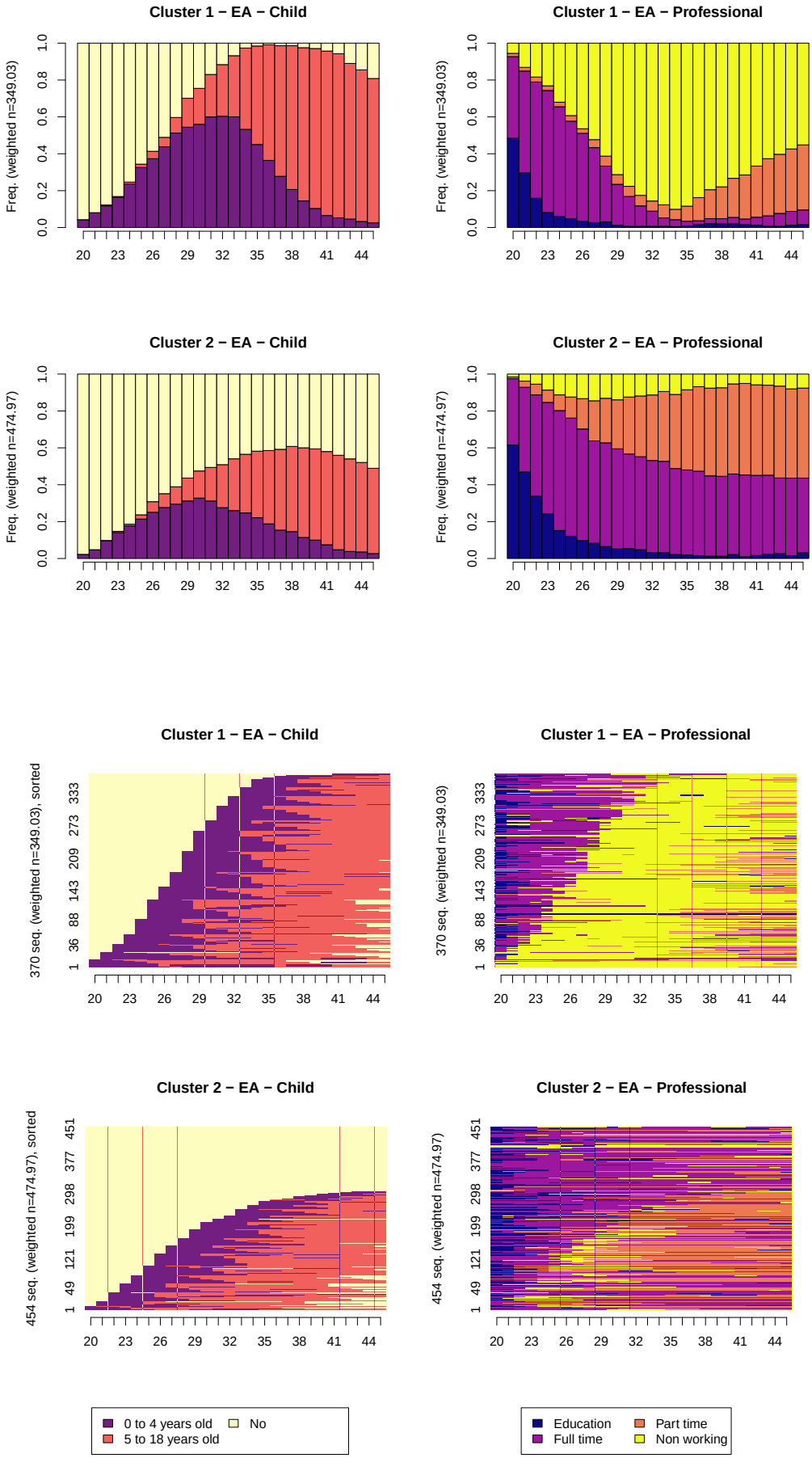

Figure 45: Chronograms (top) and index plots(bottom) of the child and professional status typology in two groups obtained with EA for women.

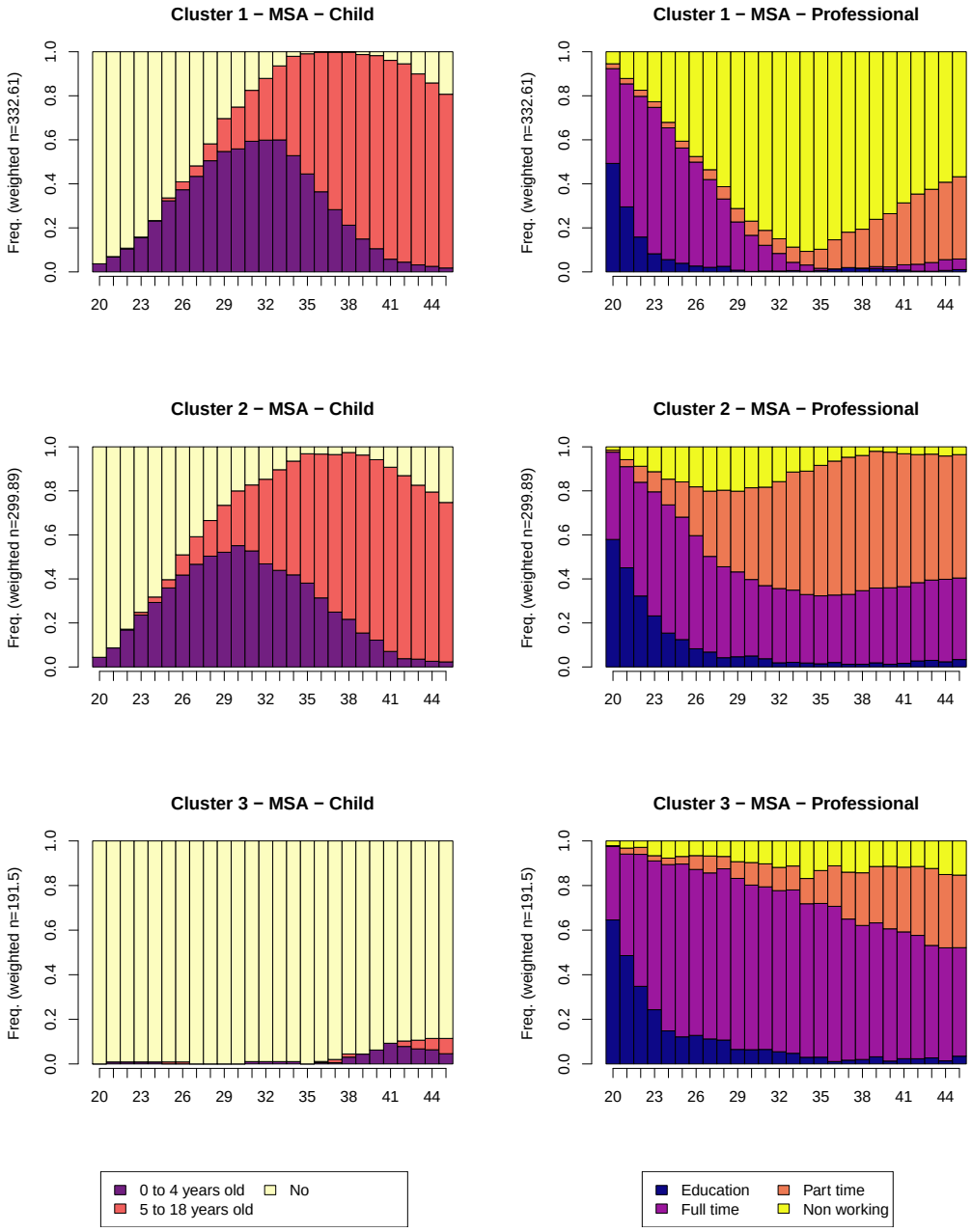

Figure 46: Chronograms of the child and professional status typology in three groups obtained with MSA for women.

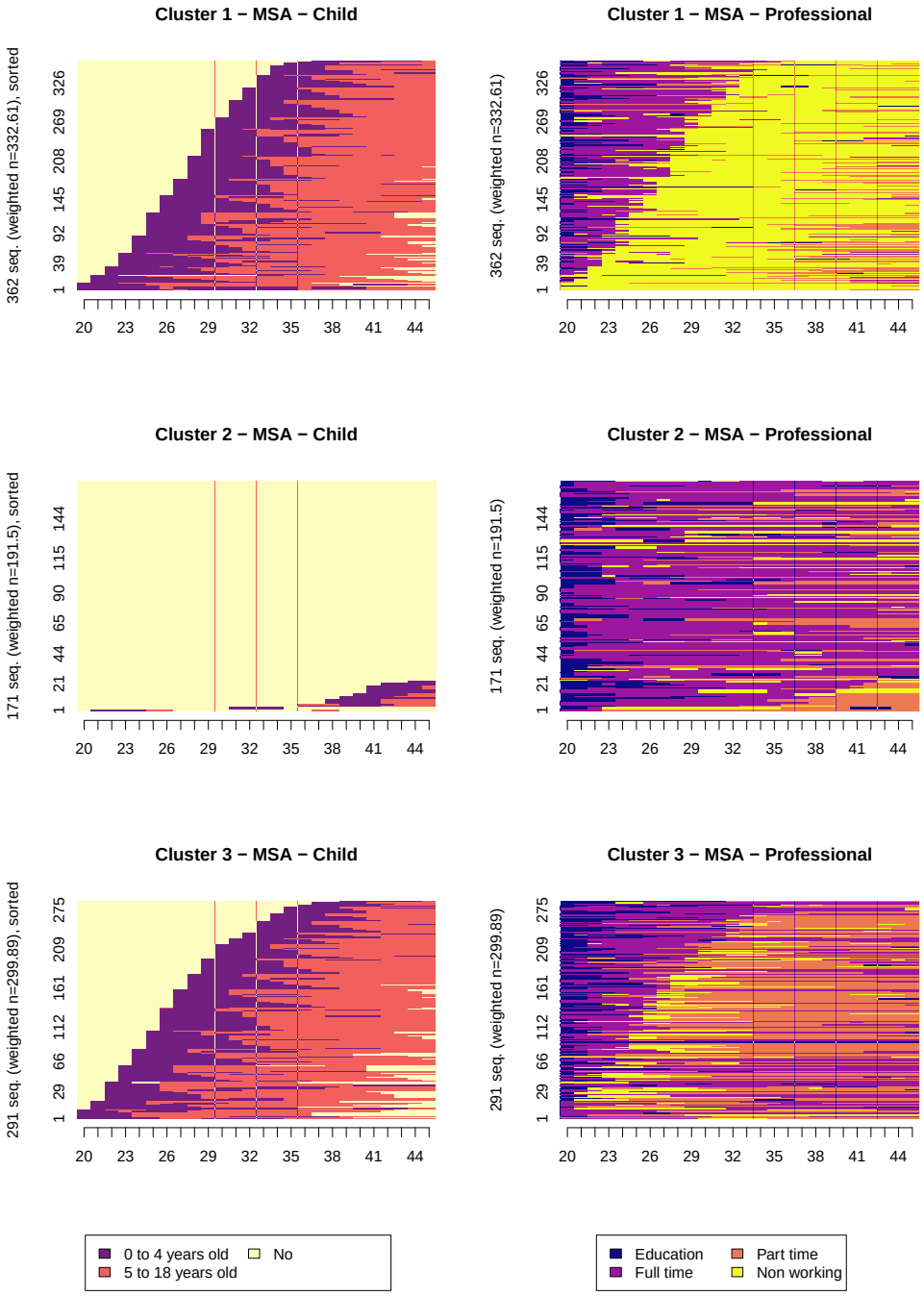

Figure 47: Index plots of the child and professional status typology in three groups obtained with MSA for women.

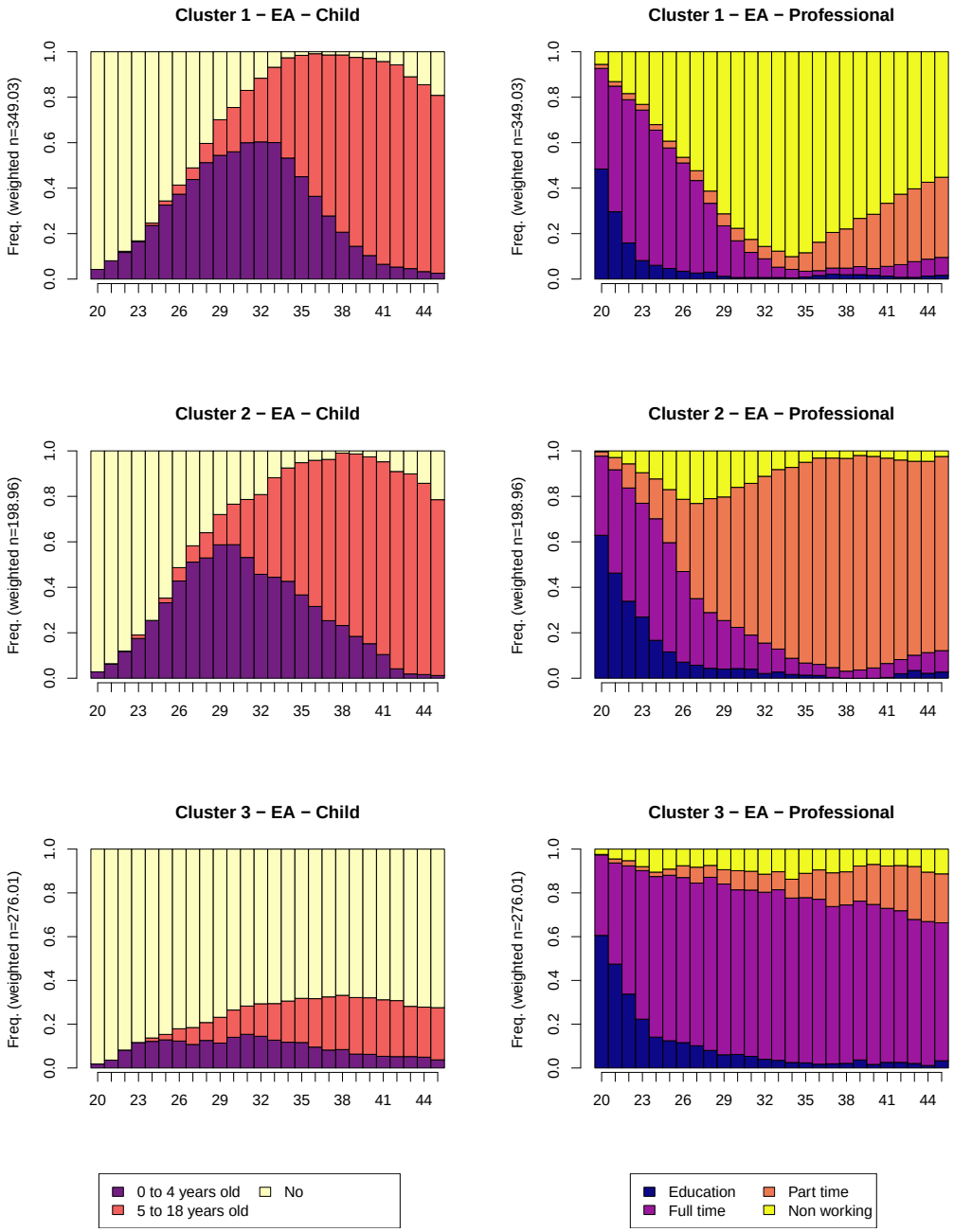

Figure 48: Chronograms of the child and professional status typology in three groups obtained with EA for women.

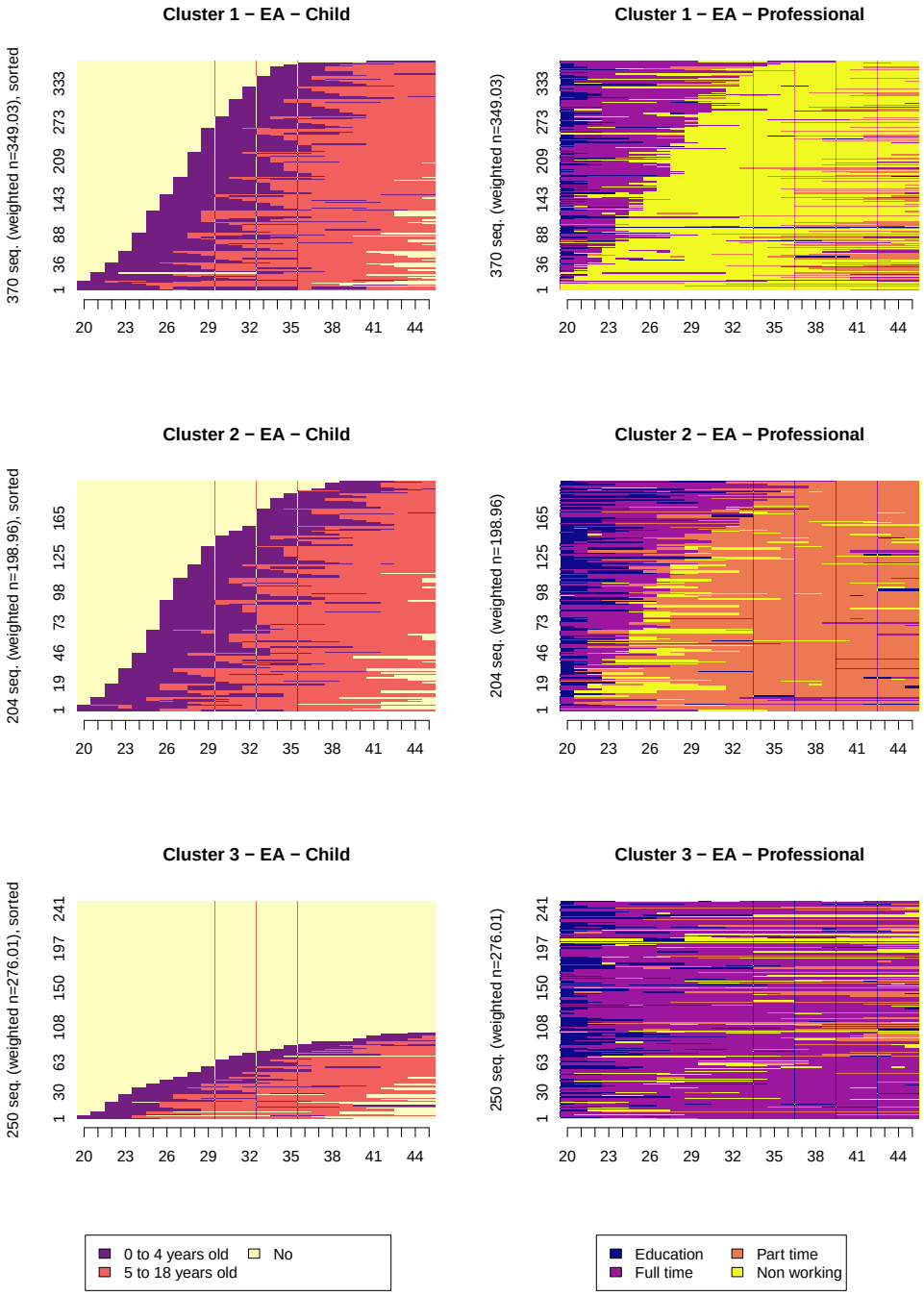

Figure 49: Index plots of the child and professional status typology in three groups obtained with EA for women.

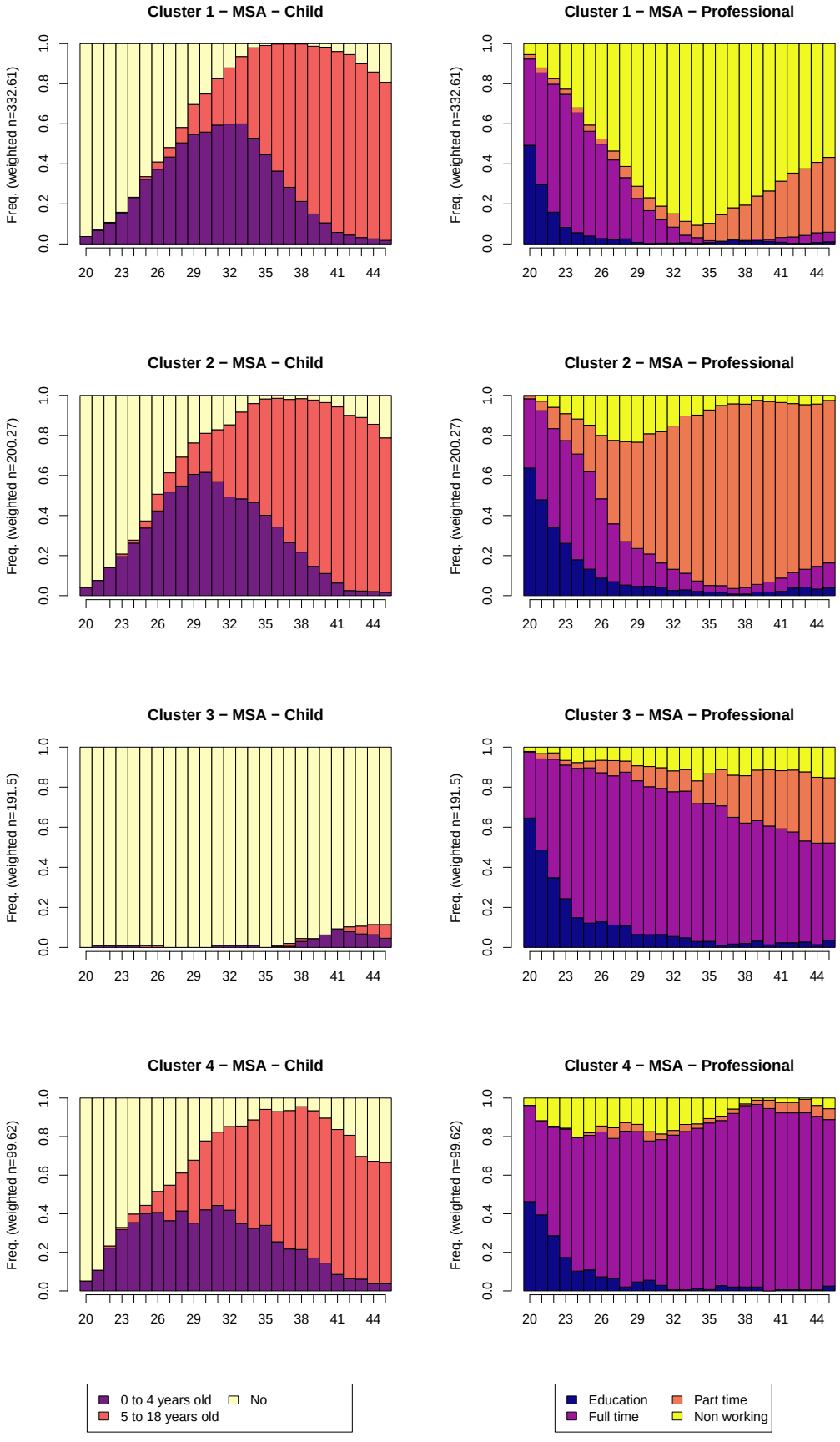

Figure 50: Chronograms of the child and professional status typology in four groups obtained with MSA for women.

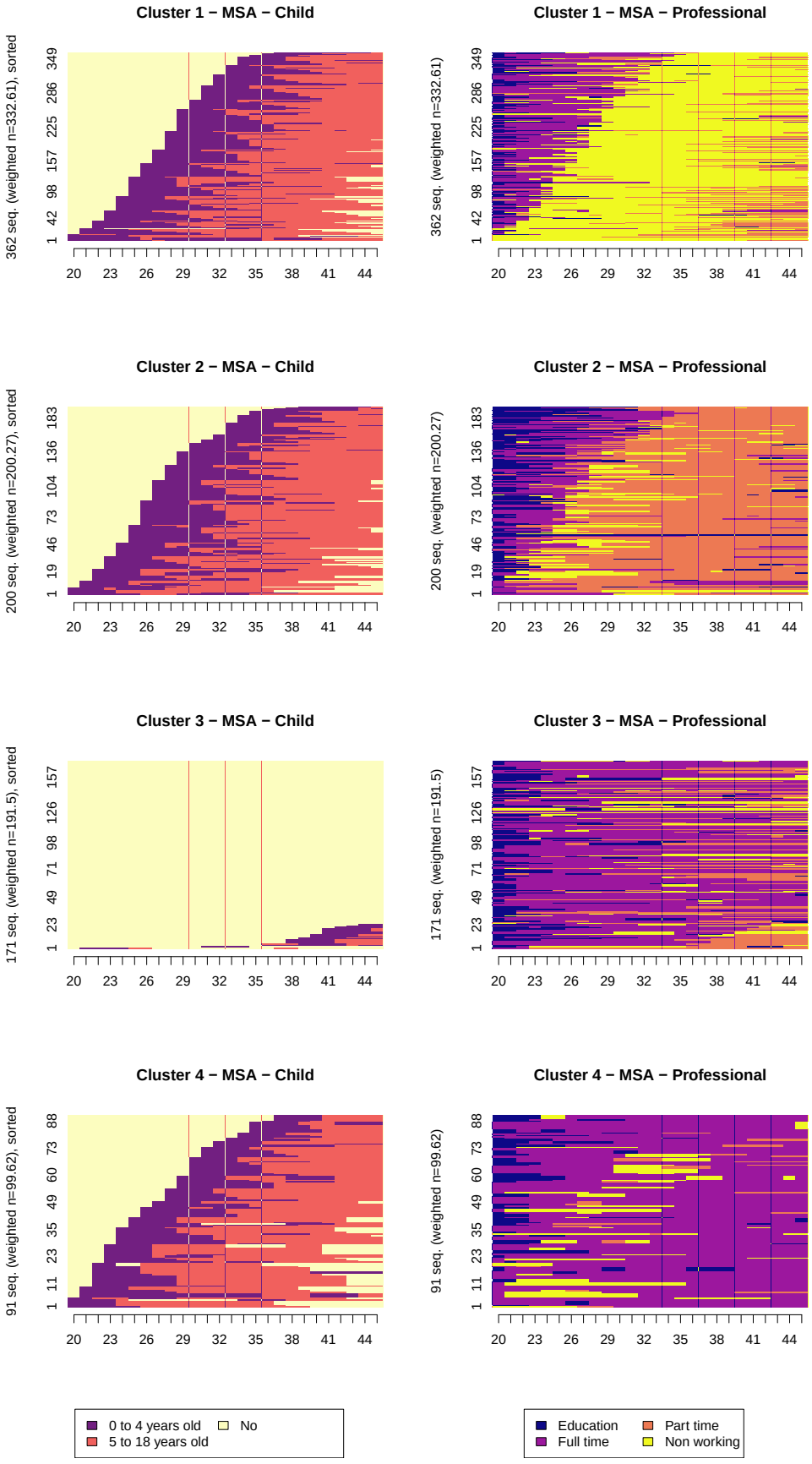

Figure 51: Index plots of the child and professional status typology in four groups obtained with MSA for women.

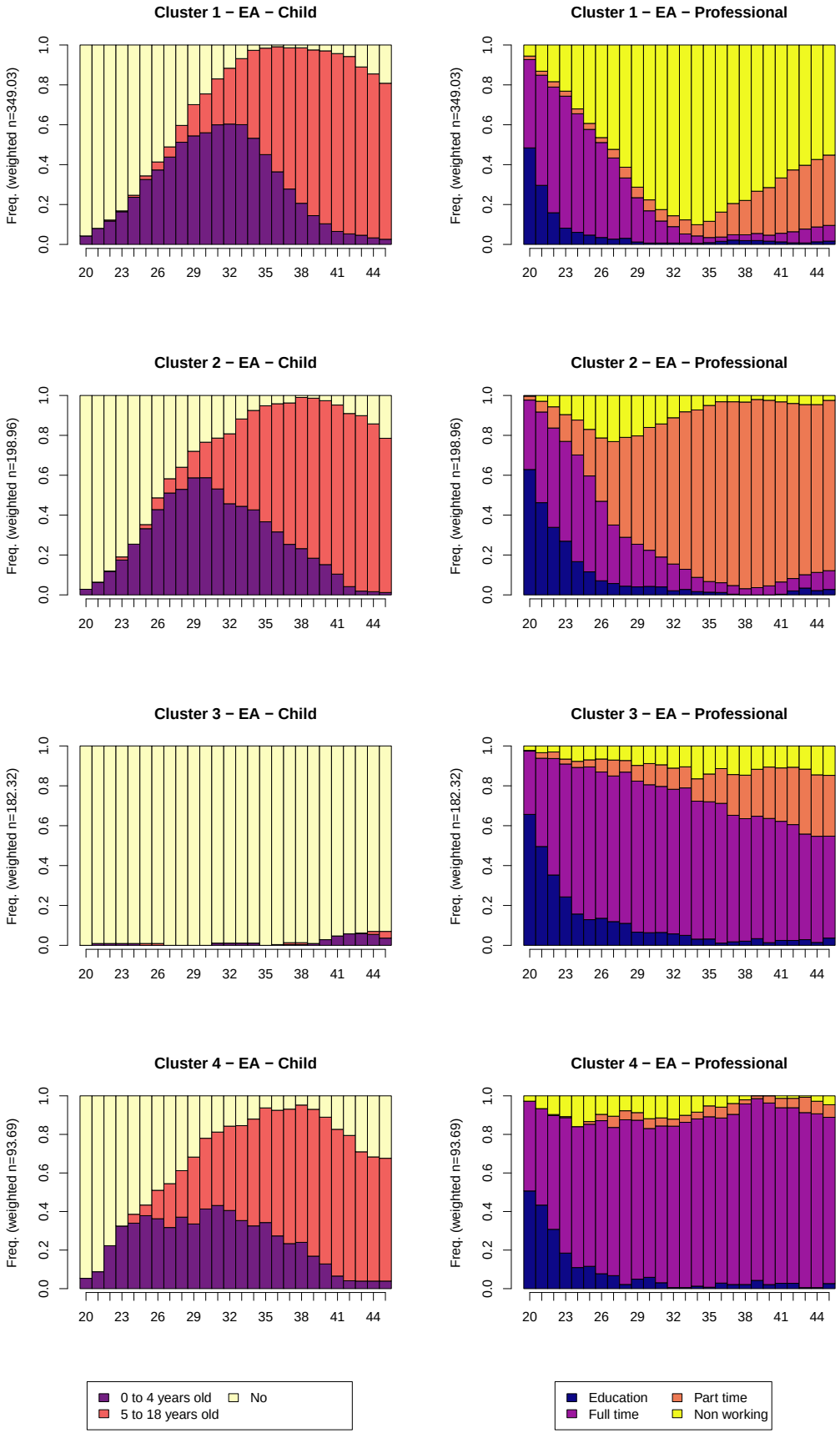

Figure 52: Chronograms of the child and professional status typology in four groups obtained with EA for women.

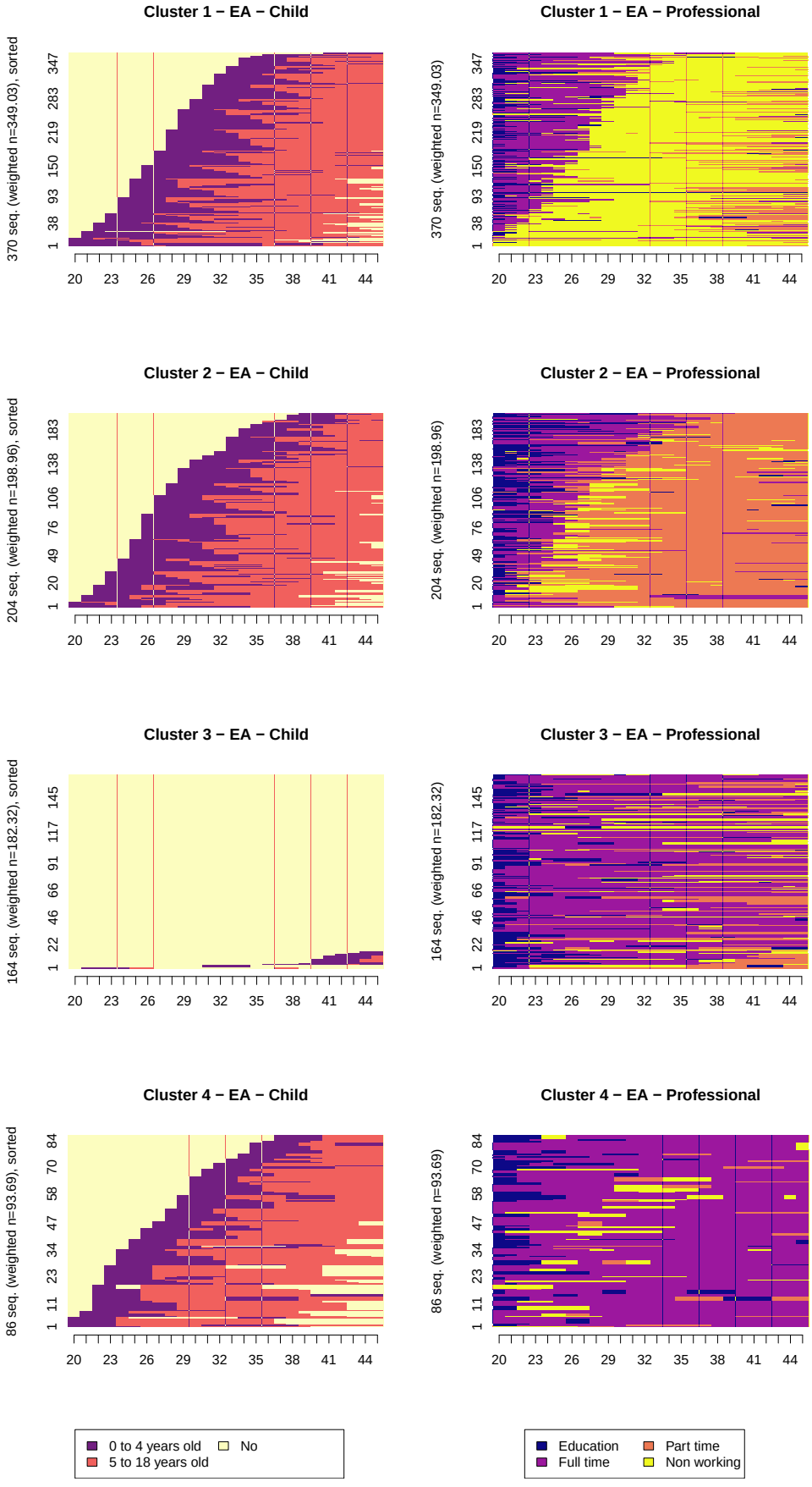

Figure 53: Index plots of the child and professional status typology in four groups obtained with EA for women.

Child–Cohabitational status This pair of domains was, as for the full dataset, the most interrelated. According to the correlations between dissimilarities, both domains were equivalently taken into account but the values were slightly better for MSA (both correlations equal to 0.82 for MSA and 0.76 for EA).

Only the two-group solution was significant for MSA both in terms of ASWw and in terms of HC, while, for EA, the clusterings from two to four groups were significant. Other groupings were therefore not able to summarize accordingly the link between the channels into account. The two-group solution built with MSA split women according to whether they had a child (Figure 55), while under the EA approach, women with other cohabitational statuses and a child were associated with childless women (Figure 56). The clusters were more homogeneous for MSA. Indeed, the ASWw by group values were 0.5 and 0.46, respectively, while they were 0.52 and 0.17 for EA. The clustering issued from MSA was more driven by the child domain, while EA produced more balanced R^2 values (Table 10). The three-cluster solution built by EA involved one cluster of women having a child and living with a partner, one cluster of women not having a child and living either with a partner or alone, and one residual cluster of women having a child with other cohabitational statuses and women not having a child and living with one of their parents (Figures 57 and 58). This latter group was ill-defined since the ASWw by group value was 0.06. To build the four-cluster solution, the group of women not having a child was mainly split according to cohabitational status (partner vs alone) (Figures 59 and 60). However, the last group was still ill-defined and the groups were relatively small, giving poor generalisation. The solution provided by MSA had better defined clusters since ASWw by group was more balanced, while for EA the R^2 by group was more balanced.

Summary To summarise, in the case of women, three domains were linked, namely, the child, cohabitational status, and professional status domains. These three domains were first considered simultaneously. Owing to the number of states in the alphabet, EA was unable to extract a typology. On the contrary, the clustering in two groups built with MSA, which mainly separated women according to whether they had a child in their household at some point, was significant and took into account each individual domain. The EA and MSA approaches provided relatively similar clusterings for the pairs of cohabitational status–professional status and child–professional status domains. For the child and cohabitational status domains, as for the full dataset, the clustering in two groups was significant for MSA and split women mainly according to whether they had a child. Concerning EA, the

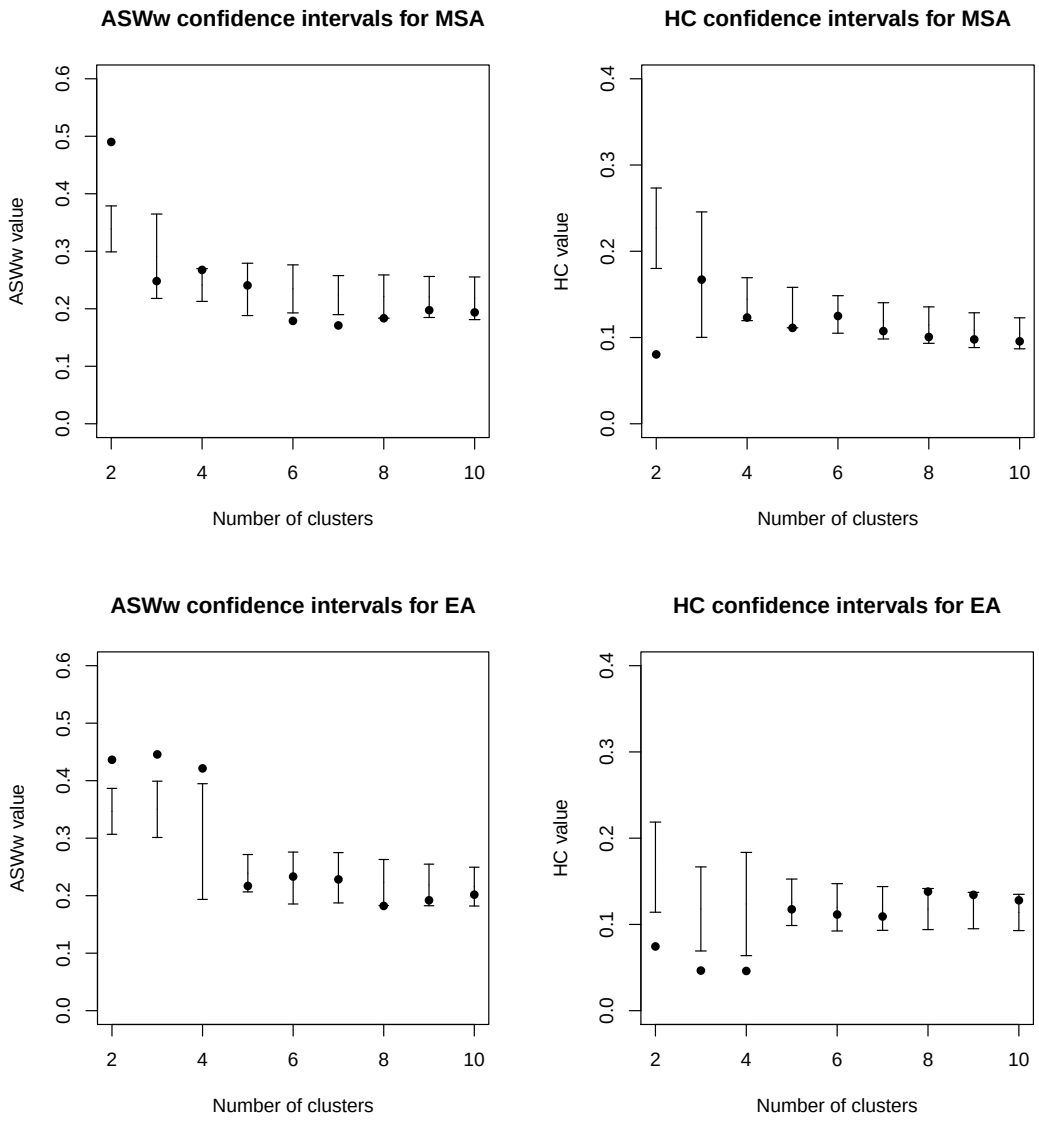

Figure 54: ASWw and HC values obtained by clustering the data, represented by the dots, together with bootstrap confidence intervals built under the hypothesis that the child and cohabitational status domains are disassociated.

clusterings between two and four groups were significant but some clusters were ill-defined or too small, inducing poor generalisation, and the groupings were difficult to interpret overall. Therefore, only the separation into two groups built by MSA was suitable in this case.

Table 10: Summary of the results obtained by clustering the child and cohabitational status channels for women with MSA and EA. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	Method	CQI		Group ASWw		Size clusters (%)		R^2 by channel	
		ASWw	HC	min	max	min	max	child	cohab
2	MSA	**0.49**	**0.08**	0.46	0.5	21	79	0.82	0.66
	EA	**0.44**	**0.07**	0.17	0.52	29	71	0.73	0.69
3	MSA	0.25	0.17	0.18	0.39	21	43	0.89	0.67
	EA	**0.45**	**0.05**	0.06	0.5	11	71	0.78	0.73
4	MSA	0.27	0.12	0.1	0.39	10	36	0.89	0.74
	EA	**0.42**	**0.05**	0.03	0.65	6	71	0.78	0.78
5	MSA	0.24	**0.11**	0.1	0.47	9	36	0.89	0.78
	EA	0.22	0.12	0.01	0.65	6	50	0.84	0.79
6	MSA	0.18	0.12	0.05	0.47	9	36	0.9	0.78
	EA	0.23	0.11	0.1	0.64	3	50	0.86	0.81
7	MSA	0.17	0.11	0.05	0.42	5	31	0.92	0.78
	EA	0.23	0.11	0.13	0.64	3	50	0.86	0.83
8	MSA	0.18	0.1	0.05	0.4	4	31	0.92	0.81
	EA	0.18	0.14	0.02	0.64	3	34	0.89	0.83
9	MSA	0.2	0.1	0.03	0.4	4	21	0.92	0.83
	EA	0.19	0.13	0.01	0.64	1	34	0.89	0.84
10	MSA	0.19	0.1	0.03	0.57	3	21	0.92	0.85
	EA	0.2	0.13	0	0.64	1	34	0.9	0.85

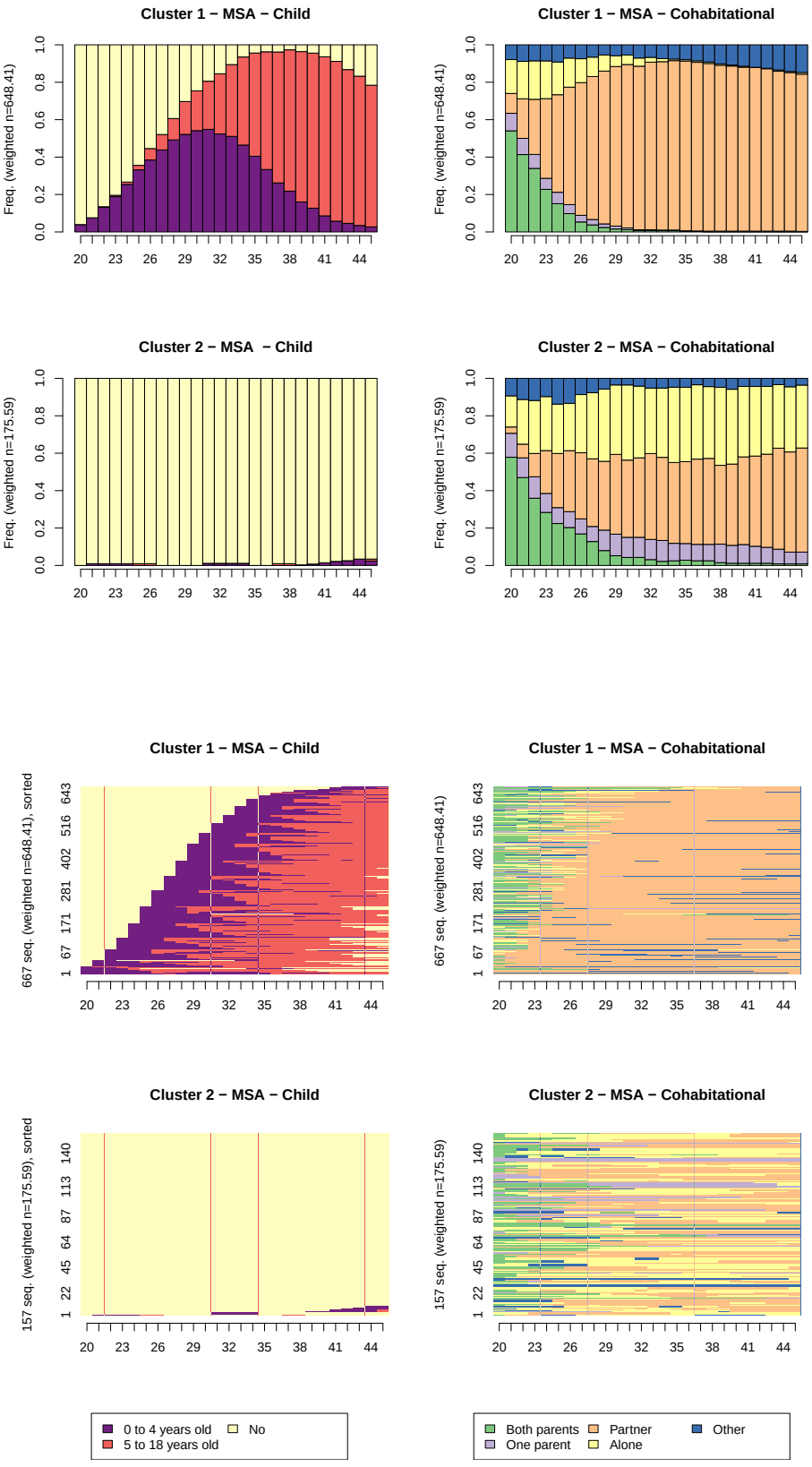

Figure 55: Chronograms (top) and index plots (bottom) of the child and cohabitational status typology in two groups obtained with MSA for women.

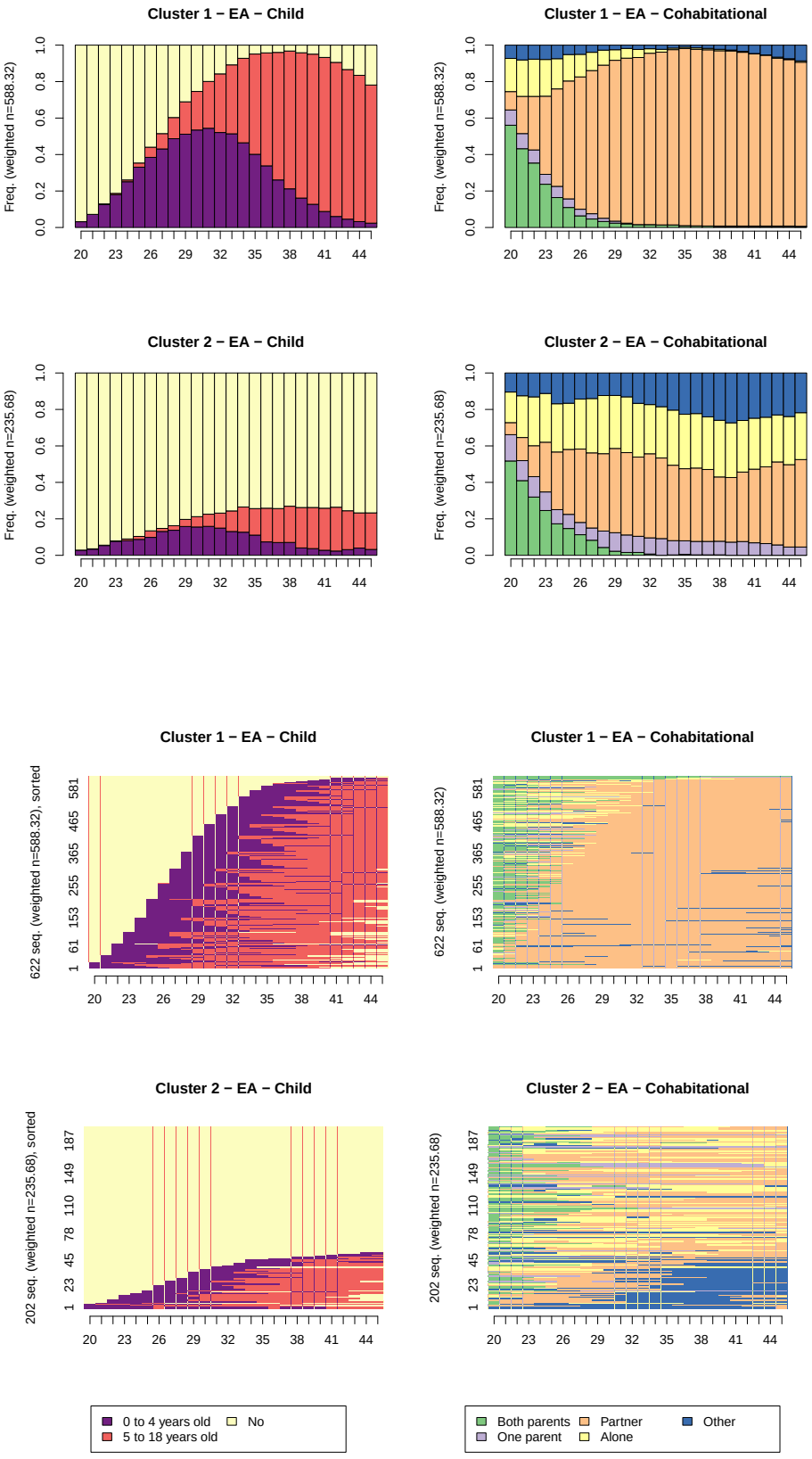

Figure 56: Chronograms (top) and index plots (bottom) of the child and cohabitational status typology in two groups obtained with EA for women.

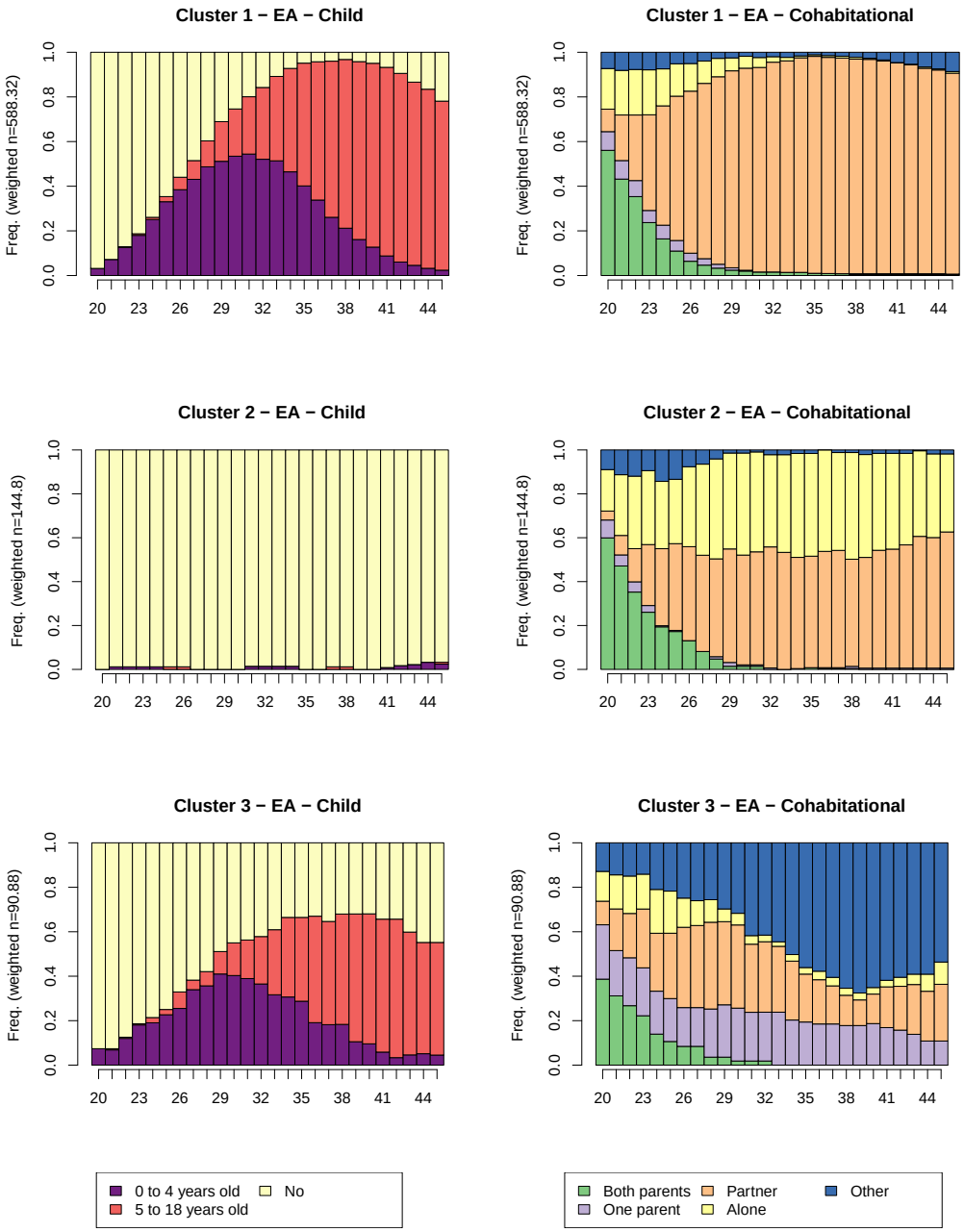

Figure 57: Chronograms of the child and cohabitational status typology in three groups obtained with EA for women.

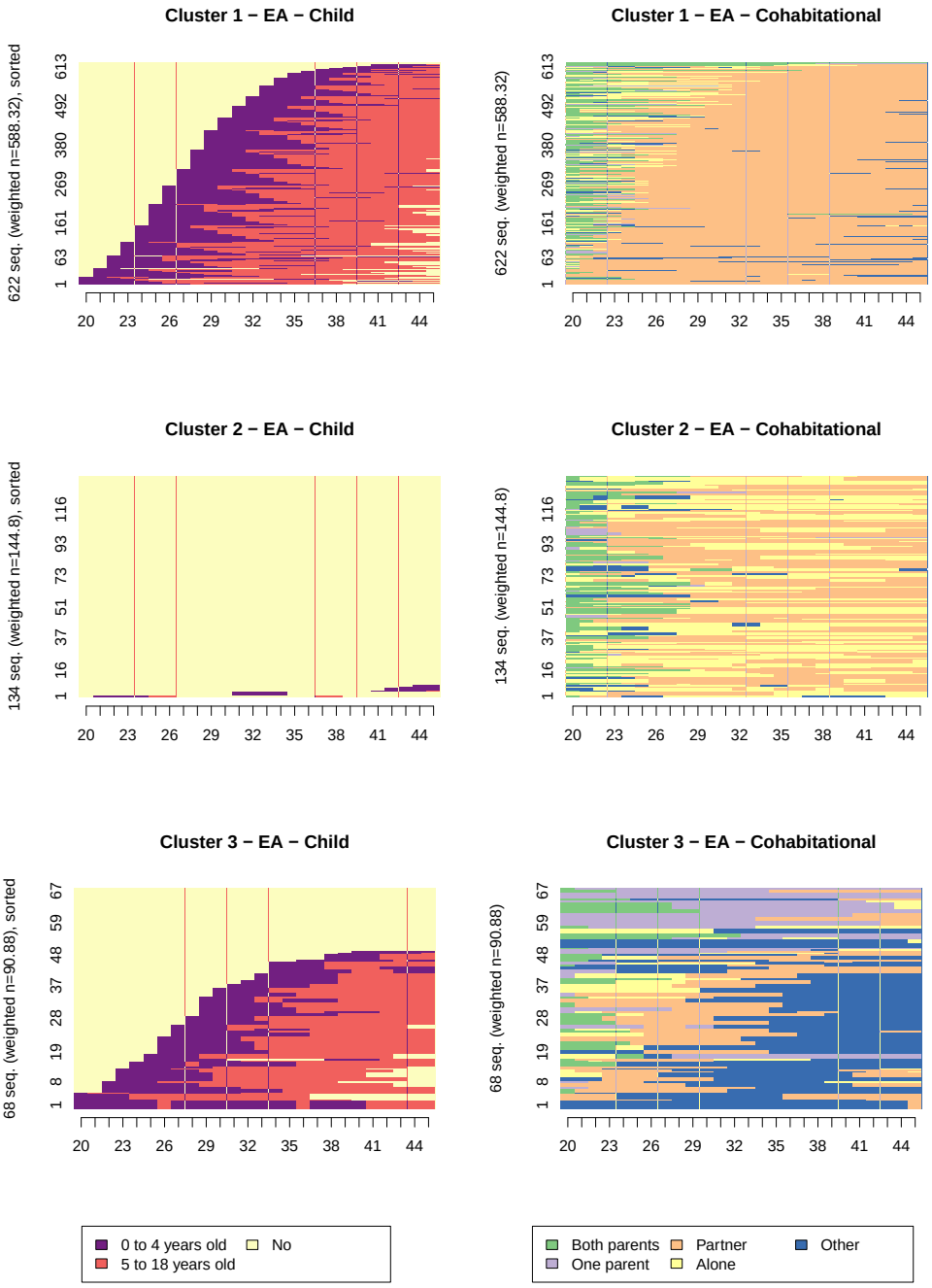

Figure 58: Index plots of the child and cohabitational status typology in three groups obtained with EA for women.

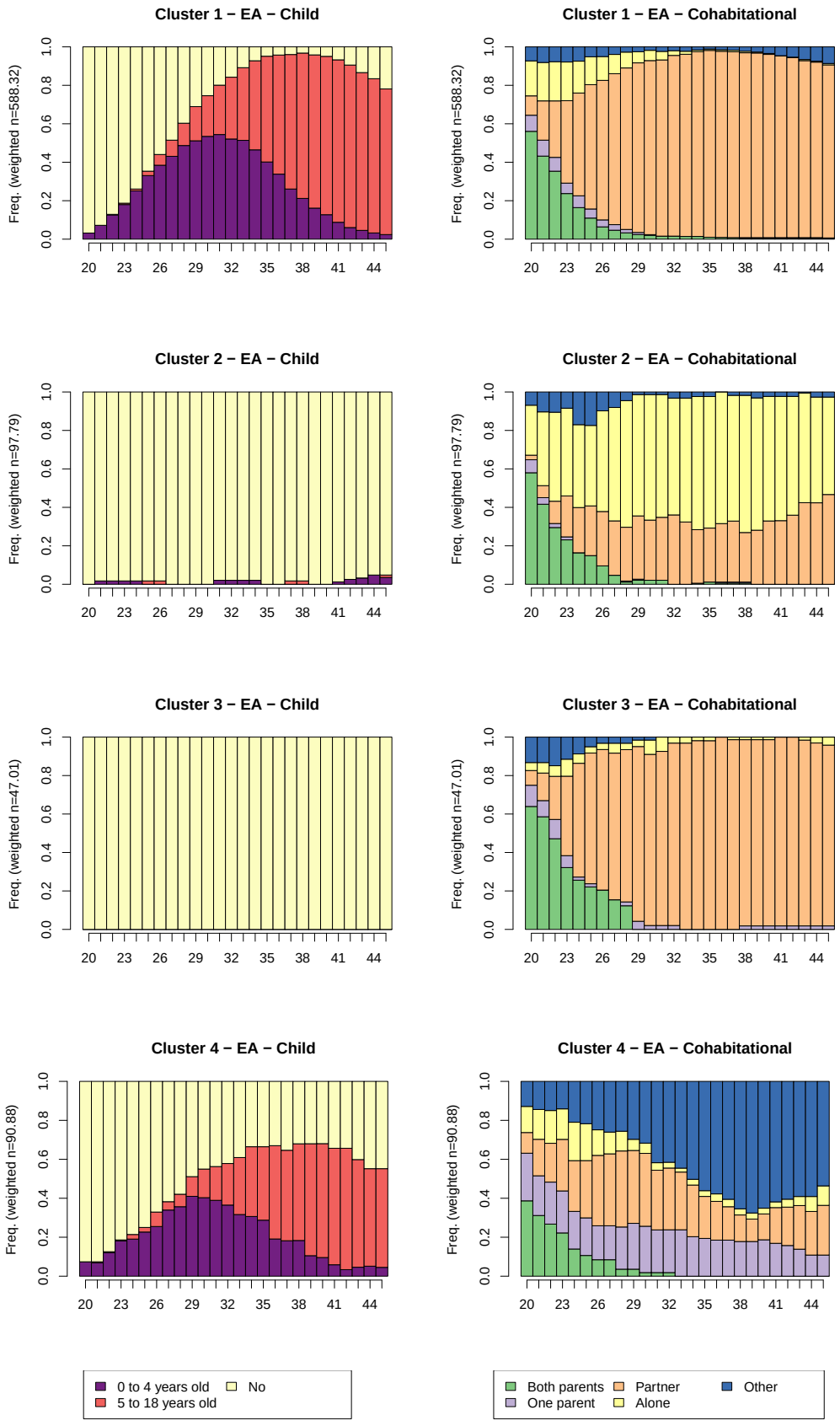

Figure 59: Chronograms of the child and cohabitational status typology in four groups obtained with EA for women.

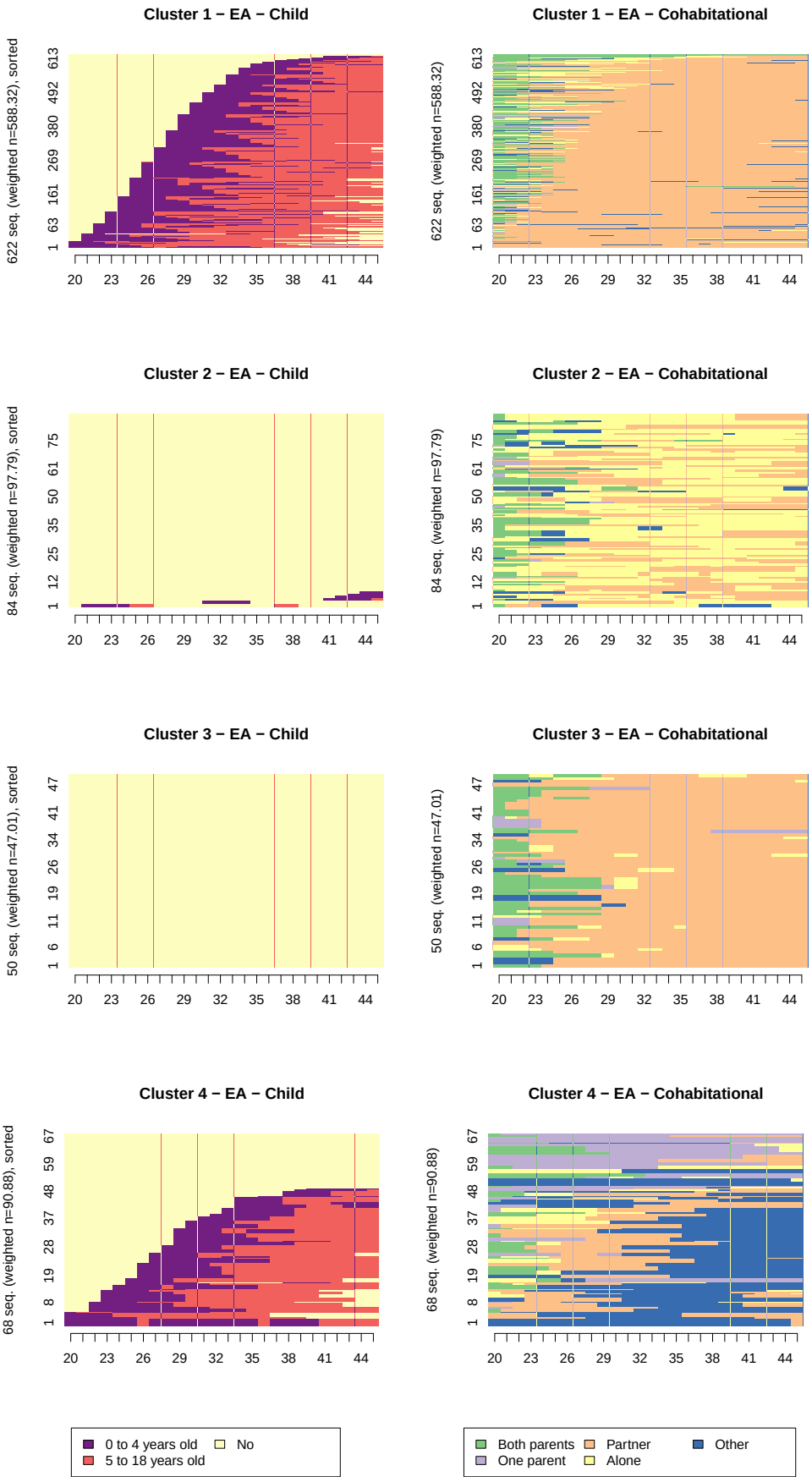

Figure 60: Index plots of the child and cohabitational status typology in four groups obtained with EA for women.

4.3 Alternatives for the computation of pairwise dissimilarities

As explained before, it is of interest to explore the behaviour of the EA and MSA approaches under different computations of pairwise dissimilarities. Owing to the principles of the MSA algorithm, we were constrained to use optimal matching algorithms. As pointed out by Studer and Ritschard (2016), the sensitivity of optimal matching to the difference in timing can be controllable by modifying the ratio between the substitution and indel costs. In the extreme case with high indel costs, only substitution costs were used. With this algorithm, called the Hamming distance (Hamming, 1950), the sensitivity to timing is at its maximum. On the contrary, the sensitivity to duration is at its maximum with the Levenstein distance, where only substitutions are available. The full dataset was used here.

4.3.1 Levenstein distance

With the Levenstein distance, only insertions and deletions are used. Therefore, EA and MSA give identical results. Indeed, with MSA, the synchronisation of the sequences is a crucial hypothesis. States are inserted or deleted simultaneously in all the channels, which mimics EA behaviour. To illustrate this, consider the example in Table 11. We have two domains, where the states are coded as A, B and C, D, respectively. Theoretically, the extended alphabet can thus be composed of four super-states (AC, BC, AD, BD); however, in this example, the combination BC never appears, meaning that the extended alphabet comprises only the remaining three super-states.

Denote by c the indel cost, which is the cost for either the insertion or the deletion of a state. In the EA representation, to transform the second sequence into the first one, state AD is inserted in the third position. The dissimilarity between the two sequences is therefore c, the cost of one insertion. For MSA, we also add an indel in the third position of the second sequence: state A is inserted in the first domain and state D is inserted in the second domain. Both operations have a cost of c; however, since MSA considered the average cost per domain, we finally have $(c + c)/2 = c$, which is the same cost as under the EA approach. Hence, both approaches behave identically when no substitution between states is considered.

The value of Cronbach's α depends on how the dissimilarities are computed on the individual domains. We therefore determined the interrelation between domains when the pairwise dissimilarities are computed with the Levenstein distance. The results of the Cronbach's α were similar to those obtained with the standard optimal

Table 11: Example with two multichannel sequences (seq 1 and seq 2) on two domains, and their equivalent representation with an extended alphabet (ext 1 and ext 2).

	MSA representation							EA representation						
seq 1	domain 1	A	A	A	B	B	B	ext 1	AC	AC	AD	BD	BD	BD
	domain 2	C	C	D	D	D	D							
seq 2	domain 1	A	A	B	B	B		ext 2	AC	AC	BD	BD	BD	
	domain 2	C	C	D	D	D								

matching algorithm. Indeed, the child–cohabitational status pair provided a value of 0.53, while all the other pairs of domains gave values smaller than 0.1. Adding either the health issues or the professional status domain to the pair of child–cohabitational status domains decreased the Cronbach's α value substantially.

As with standard optimal matching, the two-group clustering obtained from the child and cohabitational status domains separated individuals mainly according to whether they had a child (Figure 62). The solution in three groups was composed of two groups of childless people differentiated by who they live with and a group of individuals having a child. The two first clusters were relatively homogeneous since their ASWw values were respectively 0.42 and 0.49, while the third one was more heterogeneous with an ASWw value of 0.15. For both solutions, R^2 was larger for the child domain (Table 12).

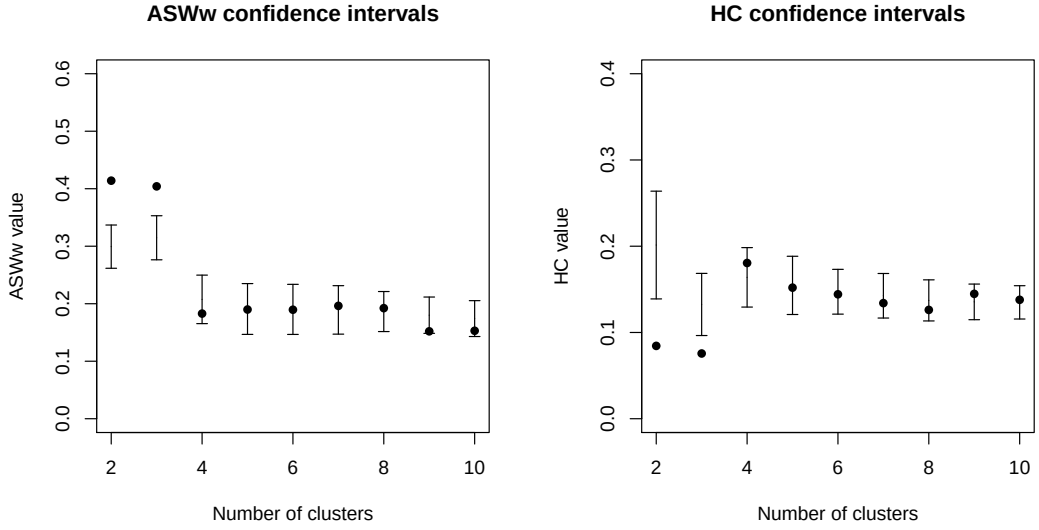

Figure 61: ASWw and HC values, represented by the dots, obtained by clustering the data when the Levenstein distance is used, together with bootstrap confidence intervals built under the hypothesis that child and cohabitational status domains are disassociated.

Table 12: Summary of the results obtained by clustering the child and cohabitational status channels with Levenstein distance. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	CQI		Group ASWw		Size clusters (%)		R^2 by channel	
	ASWw	HC	min	max	min	max	child	cohab
2	**0.41**	**0.08**	0.25	0.46	24	76	0.78	0.66
3	**0.4**	**0.08**	0.15	0.49	11	76	0.78	0.71
4	0.18	0.18	-0.01	0.48	11	42	0.82	0.73
5	0.19	0.15	0.05	0.47	5	42	0.83	0.77
6	0.19	0.14	0.11	0.43	5	42	0.84	0.79
7	0.2	0.13	-0.01	0.5	5	42	0.84	0.81
8	0.19	0.13	-0.01	0.5	5	32	0.85	0.82
9	0.15	0.14	-0.04	0.5	5	22	0.86	0.83
10	0.15	0.14	-0.08	0.49	5	22	0.86	0.84

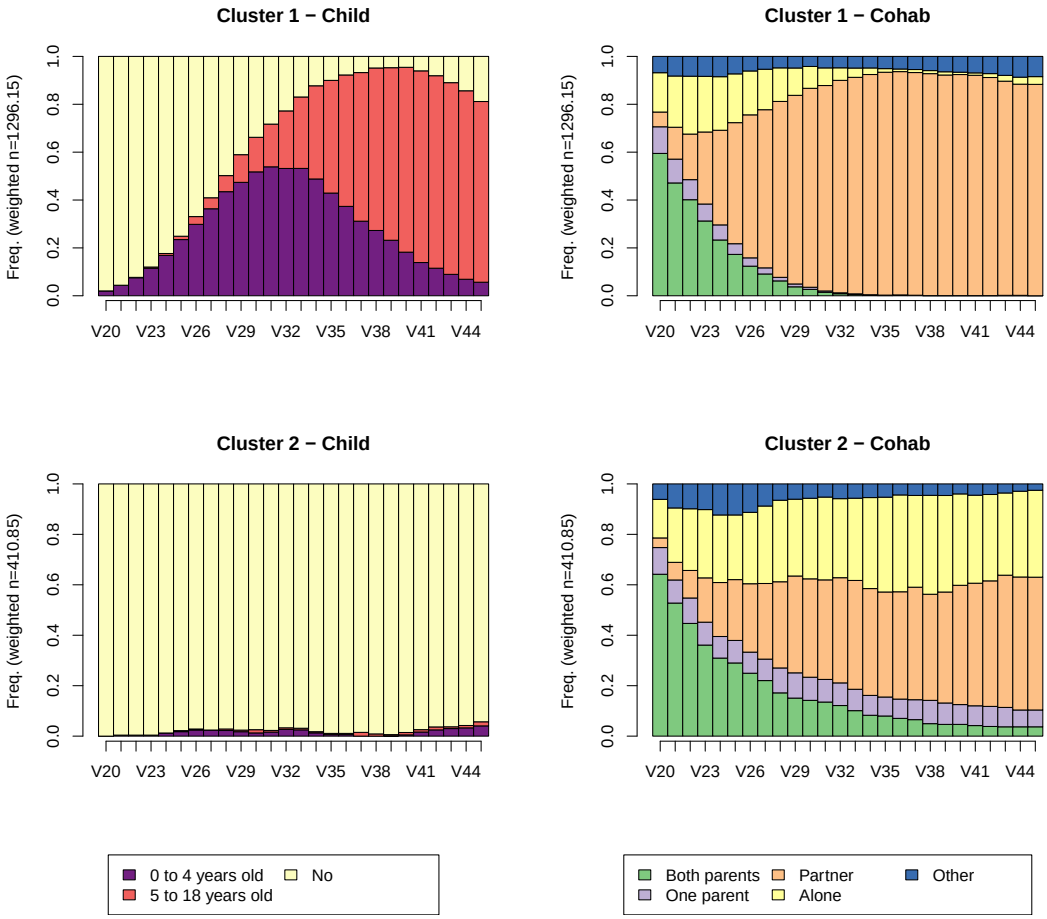

Figure 62: Chronograms of the child and cohabitational status typology in two groups obtained with Levenstein distance on the full dataset.

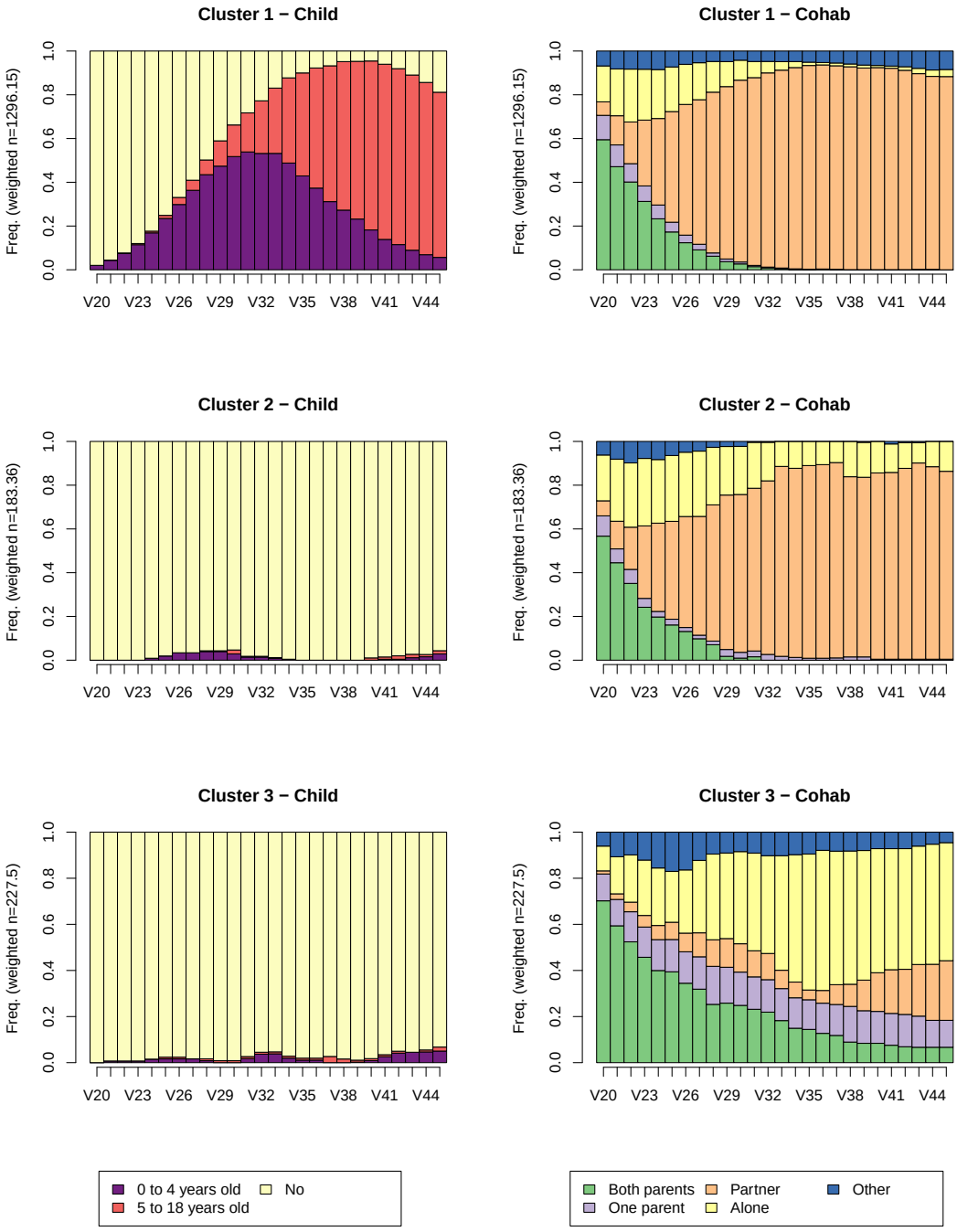

Figure 63: Chronograms of the child and cohabitational status typology in three groups obtained with Levenstein distance on the full dataset.

4.3.2 Hamming distance

When pairwise dissimilarities are computed with the Hamming distance, only substitution costs are used. In this case, the MSA and EA approaches give different results. For instance, returning to the example in Table 11 and considering a substitution cost between the states of c, the transformation of the second sequence into the first one in the EA representation involves two substitutions: substitute state BD in the third position by state AD and substitute the missing state in the last position by state BD. The total cost is then equal to $2c$. For the MSA representation, three substitutions are necessary: substitute state B in the third position of domain 1 by state A, substitute the missing data in the last position of domain 1 by state B, and substitute the missing data in the last position of domain 2 by state D. Therefore, the average cost by domain is $3c/2 = 1.5c$. In this example, the cost is lower under the MSA approach than under the EA approach.

Compared with the standard features of optimal matching, the Cronbach's α value obtained with the pair of child and cohabitational status domains was lower (0.39 vs 0.53). Only the clustering in two groups was significant for MSA and EA, both in terms of ASWw and in terms of HC (Figure 64). However, there was only a 77% agreement between the two clusterings. We found that MSA seemed to group individuals having a child at a late age with childless individuals (Figure 65), while EA grouped individuals not having a child, individuals with another cohabitational status, and individuals having a child after living alone (Figure 66). In both approaches, the first group was more homogeneous than the second (values of 0.43 and 0.2 for MSA and of 0.37 and 0.16 for EA). EA was more balanced in terms of R^2 by domain than MSA (Table 13).

4.3.3 Summary

The Levenstein and Hamming distances are two extreme cases of the standard optimal matching algorithm. By modifying the ratio between the substitution and indel costs, the focus of optimal matching is shifted between the duration spent in each state and the timing of the states. With the Levenstein distance, the MSA and EA approaches become identical since state insertions and deletions behave identically in both cases. The clusterings are, as expected, driven by the duration spent in each state. Regarding the Hamming distance, the results differ between the MSA and EA approaches markedly. In our experiments, timing is a central focus for MSA since the clusterings are mainly driven by the arrival of a new-born in the household. On the contrary, EA has more trouble accounting for the timing compound and the clusterings are more difficult to interpret.

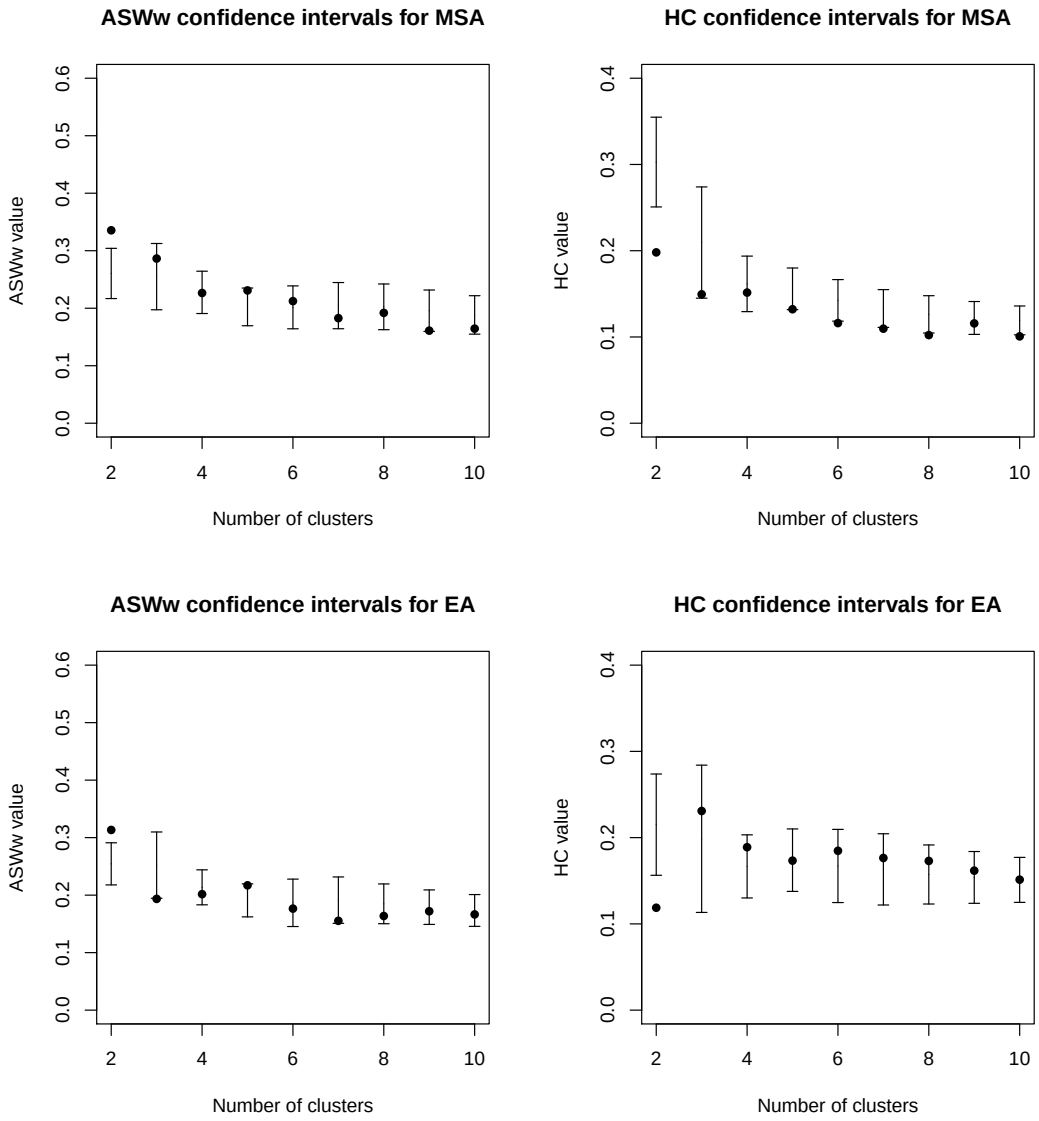

Figure 64: ASWw and HC values obtained by clustering the data when the Hamming distance is used, represented by the points, together with confidence intervals built under the hypothesis that child and cohabitational status domains are disassociated.

Table 13: Summary of the results obtained by clustering the child, cohabitational and professional status channels with MSA and EA when Hamming distance is used. The cluster quality indices (CQI) that are significant according to bootstrap validation are in bold. The minimum and maximum size of the weighted clusters are given in percent.

Clusters	Method	CQI		Group ASWw		Size clusters (%)		R^2 by channel	
		ASWw	HC	min	max	min	max	child	cohab
2	MSA	**0.34**	**0.2**	0.2	0.43	44	56	0.78	0.65
	EA	**0.31**	**0.12**	0.16	0.37	32	68	0.73	0.69
3	MSA	0.29	0.15	0.27	0.29	19	56	0.87	0.69
	EA	0.19	0.23	0.12	0.3	32	35	0.81	0.7
4	MSA	0.23	0.15	0.2	0.27	19	36	0.9	0.69
	EA	0.2	0.19	-0.01	0.34	14	35	0.82	0.74
5	MSA	0.23	0.13	0.09	0.29	4	32	0.9	0.72
	EA	0.22	0.17	0.12	0.31	5	35	0.88	0.77
6	MSA	0.21	**0.12**	0.06	0.41	4	32	0.9	0.76
	EA	0.18	0.18	0.08	0.31	5	35	0.89	0.77
7	MSA	0.18	**0.11**	0.06	0.41	4	32	0.9	0.77
	EA	0.16	0.18	0	0.28	5	26	0.9	0.78
8	MSA	0.19	**0.1**	0.08	0.39	4	32	0.91	0.78
	EA	0.16	0.17	0.03	0.28	5	24	0.9	0.8
9	MSA	0.16	0.12	0.05	0.39	4	20	0.91	0.79
	EA	0.17	0.16	0.03	0.3	4	24	0.9	0.81
10	MSA	0.16	**0.1**	0.03	0.36	4	20	0.92	0.8
	EA	0.17	0.15	0.03	0.38	4	24	0.9	0.83

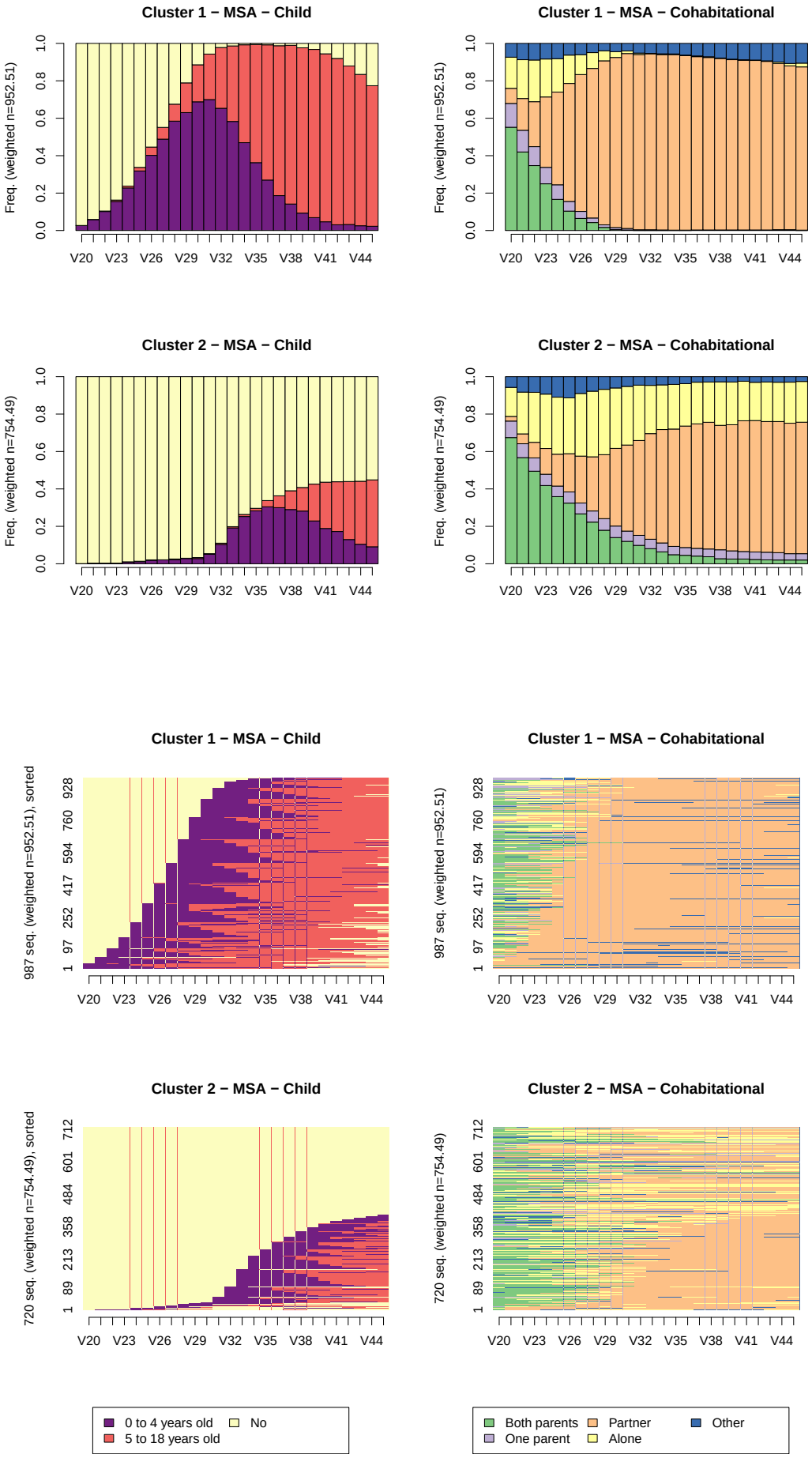

Figure 65: Chronograms (top) and index plots(bottom) of the child and cohabitational status typology in two groups obtained with MSA using the Hamming distance on the full dataset.

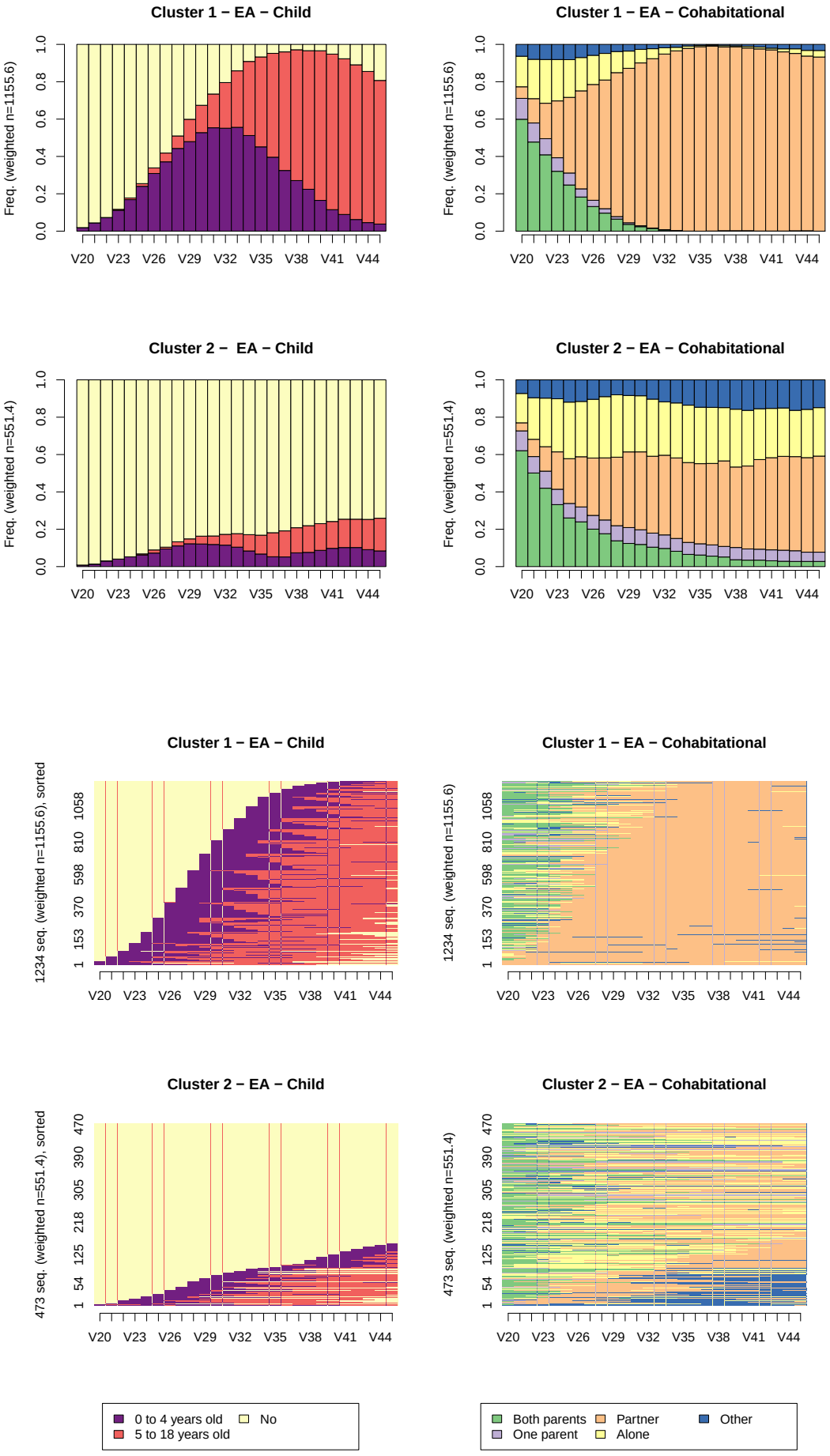

Figure 66: Chronograms (top) and index plots(bottom) of the child and cohabitational status typology in two groups obtained with EA using the Hamming distance on the full dataset.

5 Discussion

Life domains often influence each other and cannot be considered separately. In this context, EA and MSA are the two main strategies that allow us to combine sequences from different domains. To compare, in a real context, these two approaches, we used data from the Swiss Household Panel. Four domains, namely child, cohabitational status, professional status, and health issues, were considered. Child and cohabitational trajectories were expected to be highly interrelated, since they are often considered as two facets of a family trajectory. Family and professional domains are the typical application of a joint analysis, especially for women. On the other hand, no link was expected between health issues and any other domains. We first determined which domains were linked using Cronbach's α and PCA, which were adapted to this specific situation by Piccarreta (2017). Only the cohabitational status and child domains were linked for the full dataset according to these tools. The pairwise dissimilarities between sequences were computed on one side with MSA and on the other side with EA. The standard parameters of these algorithms were used following the literature. Then, a hierarchical clustering with Ward linkage was applied to extract a joint typology of the cohabitational status and child domains. A procedure involving the randomisation of the link between these domains was then applied to validate the typology. Moreover, we also examined the ASW computed by domain to determine the homogeneity of the individual clusters. Finally, R^2 values were used to assess the share of the discrepancy in the individual domains explained by each clustering.

The evidence in the literature suggests that trajectories differ between women and men, at least in the professional status domain, and our dataset confirmed this finding. Therefore, the research approach described herein for the overall sample was replicated separately by sex. Finally, we analysed the two extreme cases of optimal matching using the Levenstein and Hamming distances. In standard optimal matching, by controlling the ratio between the substitution and indel costs, one can shift the focus of optimal matching between the differences in the total duration spent in each state and the timing of states, meaning the age someone is in a specific state. By studying these two extreme cases, we obtained a better understanding of the relationship between timing and duration, on the one hand, and the MSA and EA approaches, on the other. Concerning the third type of sequence feature, namely sequencing, no clear differences were screened between MSA and EA. However, the chosen dataset may not be ideal to investigate differences in terms of sequencing.

To the best of our knowledge, this research is the first to systematically compare the MSA and EA approaches using a real dataset, and results show that neither of

the approaches is obviously superior. This is not surprising, since there is generally no absolute truth when examining a typology based on real data. The results provided by both approaches are relatively similar but differ significantly in some cases. When MSA and EA are used with their standard features, MSA seems to emphasise the timing of the occurrence of the states. Indeed, when the cohabitational status and child domains are considered simultaneously, the typologies extracted with MSA are often driven by the timing of the arrival of a new-born in the household, while EA focuses more on the time spent in each state overall. Moreover, when the Hamming distance, which tends to focus on timing, is used, the typologies built with MSA are driven by such timing, while the results provided by EA are more difficult to interpret. At the other extreme, using the Levenstein distance, the focus was clearly on the time spent in each state, and both approaches provide identical results in this case.

In practice, MSA is easier to use than EA and works in a broader range of situations. For example, with a relatively small dataset, EA is difficult to use when the extended alphabet is large (e.g. when the cohabitational status, child, and professional status domains were considered simultaneously for women). This could even induce violations of the metric properties, such as the triangle inequality. Since algorithms computing optimal matching distance assume that the metric properties are satisfied by the costs, extended alphabet with substitution costs derived from transition rates should not be used in this case. Moreover, although this was beyond the scope of the present research, individual domains can be weighted differently with MSA, allowing us to place different levels of importance on them. However, the drawback is that a clear strategy to determine the weights to be given to each domain is lacking, and there is a risk of influencing the analysis toward a specific result. EA can overcome this shortcoming by allowing the use of a larger range of dissimilarity measures, but only when optimal matching-like dissimilarities are available with MSA.

It was worth testing both approaches since they generally provide slightly different results. However, it led to a difficulty in identifying a final clustering. For that purpose, joint sequence analysis tools are useful. The bootstrap procedure developed by Studer (2019) allows us to determine whether a clustering takes into account the association between domains and it is usable both with two and with more than two domains simultaneously. The R^2 value by domain also provides the share of the variation in a domain explained by a clustering, allowing us to verify that no clustering is too influenced by any one domain in particular. Finally, the ASW computed for each group of a clustering measures the degree of internal cohesion.

However, in addition to all these statistical coefficients, it is worth remembering that the experience of the researcher and her/his knowledge of the data are essential to extract a useful typology and interpret it correctly.

The main limitation of this research lies in its empirical nature. Therefore, our conclusions could have been different with another dataset. However, most of our findings can be related to the existing literature. Moreover, given the complexity of the algorithms used here and the number of possible choices during their utilisation, a theoretical comparison does not seem possible. By selecting only complete sequences without any missing data, we based our analyses on a relatively simple dataset, which could also be considered as a limitation even though our goal was to better evidence the strict comparison between MSA and EA. Moreover, we restricted ourselves to the standard features of the algorithms, but a number of parameters could have been set differently, possibly providing different conclusions than those resulting from our analyses. First, with the exception of the Levenstein and Hamming distances, we restricted ourselves to standard optimal matching even though a large range of alternative dissimilarity measures are available. Then, we used a hierarchical clustering with Ward linkage, whereas many other clustering algorithms are available. We could have opted for a different linkage (e.g. a complete one) and partitioning around medoids would also have been possible. Finally, a broad range of cluster quality indices are available that have the ability to capture the additional characteristics of a clustering.

As mentioned before, when the dataset is small, the EA approach becomes difficult to use. In this case, some states are rare and therefore the substitution costs may not be well estimated with the transition rates. However, except for potential computational issues, a sufficiently large dataset should allow the use of a large extended alphabet. The link between the size of an extended alphabet and number of sequences necessary for its use could be interesting to examine. More broadly, this raises questions about the number of independent sequences necessary to use sequence analysis tools and thus the statistical power of sequence analysis.

Although we decided to keep only sequences without missing data to avoid an interaction between them and the object of our research, missing data are unavoidable in practice. It would thus be interesting to determine how missing data, and the procedures to deal with them, interact with the MSA and EA approaches. The two most commonly used strategies to deal with missing data in the case of sequences are to consider missing data as an additional state in the alphabet, or to impute missing data. With the first strategy, an extended alphabet could become even larger, especially if the missing data in each domain correspond are replaced

by a different additional state. For example, in the case of the child, cohabitational status, and professional status domains, the extended alphabet would have a size of 120 with missing data as an extra state in each individual domain, while it would be 60 with no missing data. Moreover, when states are missing in multiple channels, EA takes that into account since it is considered as a different state, while this does not affect MSA markedly since the missing data are substituted into each channel separately. Regarding an imputation strategy such as that proposed by Halpin (2017), the impact on the results provided by both approaches is less clear. Nevertheless, if multiple imputation is used and if the typology is identified on the basis of a large dataset combining all replications of the multiple imputation instead of working independently on each replication, the size of the extended alphabet could increase greatly. More research is clearly necessary on this point.

To summarise, we found that although the results are sometimes close, the MSA and EA approaches are still two distinct methods. In particular, they do not consider the timing of an event or time spent in a specific state in the same way, and this mean that they can be useful in different contexts. Although MSA is generally easier to use, and since it applies to more situations, EA can sometimes identify original typologies. Hence, it should also be considered when multiple domains are analysed simultaneously. It could also be of interest to combine the two approaches by building an extended alphabet from some domains and then using MSA to combine it with other domains. In this way, it could be possible to control for the risk of a too large extended alphabet.

Acknowledgements

This paper benefited from the support of the Swiss National Centre of Competence in Research LIVES - Overcoming vulnerability: Life course perspectives (NCCR LIVES), which is financed by the Swiss National Science Foundation (grant number: 51NF40-185901). The authors are grateful to the Swiss National Science Foundation for its financial assistance.

6 Bibliography

Abbott, A. and Forrest, J. (1986). Optimal matching methods for historical sequences. *The Journal of Interdisciplinary History*, 16(3):471–494.

Abbott, A. and Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological Methods & Research*, 29(1):3–33.

Aisenbrey, S. and Fasang, A. (2017). The interplay of work and family trajectories over the life course: Germany and the united states in comparison. *American Journal of Sociology*, 122(5):1448–1484.

Bernardi, L., Huinink, J., and Settersten, R. A. (2019). The life course cube: A tool for studying lives. *Advances in Life Course Research*, 41:100258. Theoretical and Methodological Frontiers in Life Course Research.

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

Halpin, B. (2017). SADI: Sequence Analysis Tools for Stata. *The Stata Journal*, 17(3):546–572.

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell system technical journal*, 29(2):147–160.

Hennig, C. and Liao, T. (2010). Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification.

Hennig, C. and Lin, C.-J. (2015). Flexible parametric bootstrap for testing homogeneity against clustering and assessing the number of clusters. *Statistics and Computing*, 25:821–833.

Hubert, L. J. and Levin, J. R. (1976). A general statistical framework for assessing categorical clustering in free recall. *Psychological bulletin*, 83(6):1072.

Jalovaara, M. and Fasang, A. (2017). From never partnered to serial cohabitors: Union trajectories to childlessness. *Demographic Research*, 36:1703–1720.

Kühr, J., Bühlmann, F., Gauthier, J.-A., Morselli, D., Mugnari, E., Bumbaru, A., Ryser, V.-A., Spini, D., Le Goff, J.-M., Dasoki, N., Roberts, C., Tillmann, R., Bernardi, L., and Brandle, K. (2013). Assessing the performance of the Swiss Panel LIVES Calendar: Evidence from a pilot study. Publisher: LIVES.

Lesnard, L. (2008). Off-scheduling within dual-earner couples: An unequal and negative externality for family time. *American Journal of Sociology*, 114(2):447–490.

Levy, R., Gauthier, J.-A., Widmer, E., et al. (2006). Entre contraintes institutionnelle et domestique: les parcours de vie masculins et féminins en suisse. *The Canadian Journal of Sociology*, 31(4):461–489.

Lorentzen, T., Bäckman, O., Ilmakunnas, I., and Kauppinen, T. (2019). Pathways to adulthood: Sequences in the school-to-work transition in Finland, Norway and Sweden. *Social Indicators Research*, 141(3):1285–1305.

Madero-Cabib, I. and Fasang, A. E. (2016). Gendered work–family life courses and financial well-being in retirement. *Advances in Life Course Research*, 27:43–60.

Malin, L. and Wise, R. (2018). Glass ceilings, glass escalators and revolving doors. In Ritschard, G. and Studer, M., editors, *Sequence Analysis and Related Approaches*, volume 10 of *Life Course Research and Social Policies*, pages 49–68. Springer, Cham.

Mattijssen, L. and Pavlopoulos, D. (2018). A multichannel typology of temporary employment careers in the netherlands: Identifying traps and stepping stones in terms of employment and income security. *Social Science Research*.

McMunn, A., Lacey, R., Worts, D., McDonough, P., Stafford, M., Booker, C., Kumari, M., and Sacker, A. (2015). De-standardization and gender convergence in work–family life courses in great britain: A multi-channel sequence analysis. *Advances in Life Course Research*, 26:60–75.

Müller, N. S., Sapin, M., Jacques-Antoine, G., 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.

Oris, M. and Ritschard, G. (2014). Sequence analysis and transition to adulthood: An exploration of the access to reproduction in nineteenth-century East Belgium.

In Blanchard, P., Bühlmann, F., and Gauthier, J.-A., editors, *Advances in sequence analysis: Theory, method, applications*, volume 2 of *Life Course Research and Social Policies*, pages 151–167. Springer, Cham.

Piccarreta, R. (2017). Joint sequence analysis: Association and clustering. *Sociological Methods & Research*, 46(2):252–287.

Piccarreta, R. and Billari, F. C. (2007). Clustering work and family trajectories by using a divisive algorithm. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(4):1061–1078.

Piccarreta, R. and Studer, M. (2019). Holistic analysis of the life course: Methodological challenges and new perspectives. *Advances in Life Course Research*, 41:100251.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rousseeuw, P. (1987). Rousseeuw, p.j.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. comput. appl. math. 20, 53-65. *Journal of Computational and Applied Mathematics*, 20:53–65.

Schwanitz, K. (2017). The transition to adulthood and pathways out of the parental home: A cross-national analysis. *Advances in Life Course Research*, 32:21–34.

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.

Stafford, M., Lacey, R., Murray, E., Carr, E., Fleischmann, M., Stansfeld, S., Xue, B., Zaninotto, P., Head, J., Kuh, D., and Mcmunn, A. (2018). Work–family life course patterns and work participation in later life. *European Journal of Ageing*, 16.

Studer, M. (2013). WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R. *LIVES Working Papers*, 24.

Studer, M. (2015). Comment: On the use of globally interdependent multiple sequence analysis. *Sociological Methodology*, 45(1):81–88.

Studer, M. (2018). Divisive property-based and fuzzy clustering for sequence analysis. In Ritschard, G. and Studer, M., editors, *Sequence Analysis and Related*

Approaches, volume 10 of *Life Course Research and Social Policies*, pages 223–239. Springer, Cham.

Studer, M. (2019). Validating sequence analysis typologies using bootstrapping.

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. S. (2011). Discrepancy analysis of state sequences. *Sociological Methods & Research*, 40(3):471–510.

Tillmann, R., Voorpostel, M., Kuhn, U., Lebert, F., Ryser, V.-A., Lipps, O., Wernli, B., and Antal, E. (2016). The swiss household panel study: Observing social change since 1999. *Longitudinal and Life Course Studies*, 7(1):64–78.

Wahrendorf, M., Marr, A., Antoni, M., Pesch, B., Jöckel, K.-H., Lunau, T., Moebus, S., Arendt, M., Brüning, T., Behrens, T., et al. (2018). Agreement of self-reported and administrative data on employment histories in a german cohort study: A sequence analysis. *European Journal of Population*, 35(2):329–346.

Widmer, E. D. and Ritschard, G. (2009). The de-standardization of the life course: Are men and women equal? *Advances in Life Course Research*, 14(1-2):28–39.