Montero et al.
European Transport Research Review (2023) 15:44
https://doi.org/10.1186/s12544-023-00621-1

European Transport Research Review

REVIEW

Open Access

The role of life course and gender in mobility patterns: a spatiotemporal sequence analysis in Barcelona

Lídia Montero[1*], Lucía Mejía-Dorantes[2] and Jaume Barceló[1]

Abstract

Citizens participate in various activities to fulfill their needs, advance their socio-economic status, and enhance their well-being through social and health-related engagements. However, activity participation is influenced by many factors in the built environment, such as the spatial and temporal dissemination of activities, which therefore necessitate travel to overcome physical distances. Moreover, individual attributes such as gender, daily schedules, and other socio-economic characteristics also influence mobility patterns. In this paper, we aim to investigate these factors in the specific context of the Barcelona Metropolitan Area using three different samples of residents from annual mobility surveys conducted between 2018 and 2020. To this end, we employ a sequence analysis method that examines the entire trajectory of an individual's daily activities and travel, considering the number, order, and duration of activities. In this way, we analyse in detail how various individual characteristics and the built environment influence the fragmentation of activities. Our study yields multiple results. First, we find that even in a transport-oriented city, the fragmentation of activities is shaped by gender, especially after age 30, when major changes occur in an individual's life course, in particular caring responsibilities and family status. Second, we observe that the educational level and year of the sample also play a central role in shaping mobility patterns. Finally, our paper makes a methodological contribution by defining sequence distances, after projecting the original space onto the factorial one defined by the Multiple Correspondence Analysis. This study shows that mobility policies should not focus solely on transport aspects, but also consider the built environment, dwelling location, gender, equity, and individual lifetime characteristics in an integrated manner.

Keywords Travel behavior, Fragmentation, Sequence analysis, Gender, Equity, Classification analysis, Life course, Activity participation

1 Introduction

The spatial and temporal distribution of activities in the urban environment determines how people arrive at places safely and on time. However, our access to different services in the built environment is affected by available transport systems, topography, and weather. Furthermore, citizens' mobility is determined by certain intrinsic characteristics, such as socioeconomic level, profession, gender, disabilities, family status, religion, ethnicity, and other factors.

Research has consistently shown that lifestyles differ notably by gender. For example, women are overrepresented in certain professions such as healthcare or education, and their family status influences their income, contracts, work schedules, and conditions. Furthermore, the burden of unpaid work and family responsibilities, [35], also known as care work, affects

*Correspondence:
Lídia Montero
lidia.montero@upc.edu
[1] Universitat Politècnica de Catalunya (UPC), Carrer Jordi Girona 1-3, 08034 Barcelona, Spain
[2] Karlsruhe, Germany

their participation in economic activities and thus influences their mobility choices, thereby leading to gendered mobility patterns [15, 19]. As a result, women tend to have more complex commuting patterns than men, characterised by shorter, chained trips [16, 20], and they often prefer certain transport modes [17, 32].

In the years before the COVID-19 pandemic, it was widely accepted that emerging information and communication technologies (ICTs) would transform the way people move, such as their travel to work [6], which would in turn lead to an expected increase in travel demand, especially in the developed world [12–14]. Because the traditional schedules and sequences of daily activities were thus anticipated to become more fluid and flexible as people switched between multiple activities throughout their days, it was predicted that transport demand would increase as many activities became untethered from specific times and specific places.

Since the COVID-19 outbreak, we have seen a remarkable increase in the use of ICTs for daily activities, which has indeed influenced our travel behaviour. Despite this trend, many activities still require physical presence and cannot be replaced by virtual services, such as school attendance, doctor visits, and various activities that require an escort. While working from home has become a viable option for some professions with better wages and job conditions [26], it remains relatively uncommon. Certain other activities are preferably carried out in person, such as those that foster community engagement and provide physical and mental health benefits, thus contributing to people's well-being. In all these cases, individuals rely on available transportation options.

The current study aims to understand spatiotemporal travel behaviours in the Metropolitan Area of Barcelona while taking into account the abovementioned issues and ultimately shedding light on policies that would increase sustainable practices. We take a sequence analysis approach that relies on three recent and consecutive mobility surveys in the Metropolitan Region of Barcelona (BMR). As far as we know, ours is the first attempt to apply this technique in the Spanish context, in a transport-oriented development (TOD) urban environment using a longitudinal dataset. This paper is divided into eight sections and organized as follows. After this Sect. 1 introduction, we present the research objectives in Sect. 2, followed by the literature review and methodological framework in Sect. 3. We next introduce the case study in Sect. 4, describe the data in Sect. 5, explain the application of our methodology to the case study in Sect. 6, and present the results in Sect. 7. Finally, Sect. 8 presents the discussion, main conclusions, and implications of this study.

2 Research objectives

The objectives of this study are multifaceted:

1. Firstly, we aim to enhance our understanding of the spatiotemporal patterns and travel behavior in the Barcelona Metropolitan Area (BMA), both before and after the pandemic outbreak. To the best of our knowledge, a sequence analysis of daily travel patterns has never been previously carried out in Barcelona.
2. Secondly, we endeavour to quantify activity participation in Barcelona by employing various fragmentation indicators, which will be discussed extensively in Sect. 3. Barcelona, being a major European city with a dense multimodal public transport system, serves as a relevant case study.
3. Lastly, we aim to analyse gendered mobility patterns and explore the significant role played by individual characteristics, as well as those derived from the built environment, in shaping mobility.

In this study, we present an alternative approach to handling a large dataset. Rather than aggregating the dataset into arbitrary time intervals that could induce bias in the study, we maintain a detailed minute-by-minute dataset and subsequently utilise a multiple correspondence analysis to address dimensionality issues effectively.

3 Literature review and methodological framework

3.1 Sequence analysis (SA)

Sequence analysis has emerged in the field of social sciences as a valuable tool for comprehending the occurrence of social events in a structured manner. Andrew Abbott is widely recognized as a trailblazer in sequence analysis, having developed the fundamental concepts and methodology that extend beyond how historians order events and how quantitative analysis deals with sequence in social processes. His contributions are evident in his diverse range of publications elaborating on the evolution of these concepts and methodologies [1–5, 38]. According to Abbott and DeViney [3, p. 428], "Social reality happens in sequences of actions with constraining or enabling structures [...]. It is a matter of particular social actors, in particular social places, at particular social times".

Numerous authors within the sociology field have used activity sequence analysis to conduct studies on life course events such as marriage, childbearing, and employment [22, 29]. These studies all understand that a sequence is a structured list of events or actions performed in a specific order. Additionally, Abbott and Forrest [4] describe how a sequence dataset can encompass two different patterns: one involving activities that occur only once, and another involving activities that occur multiple times within a particular sequence.

This approach has also been recently applied to the analysis of mobility behaviour. Bhat and Pinjari [9] noted that analysing travel behavior crucially involves examining sequences of activities, the daily transitions between them, and the amount of time spent on each activity.

The analysis of activity sequences plays a critical role in formulating econometric models integrated into activity-based daily simulations of household activity-travel patterns for large-scale travel demand analysis, as observed by many authors [8, 11, 43, 44].

State sequences can offer valuable insights by highlighting differences or similarities among groups. Studer and Ritschard [48] identify the following intertwined characteristics that are crucial for analysing and comparing activity sequences:

- Experienced states: These refer to each alternative activity in the sequence, such as being at home, work, school, or travelling by car, public transport, or other. State sequences can provide essential information that sheds light on group differences or similarities.
- Distribution: This refers to the total time allocated to each state within a sequence.
- Timing: This is the specific moment in time when each state appears within the sequence.
- Duration: This pertains to the length of time spent in each successive state.
- Sequencing: This refers to the specific order in which distinct successive states occur. A sequence represents an ordered string of activities spanning a specific period.

To identify similar sequences and quantify the mismatch between them, a rule for sequence comparison is needed. This rule enables the measurement of dissimilarity, with the number of operations required to align two sequences representing a distance or dissimilarity score. The distance between two sequences is the minimum combination of operations necessary to transform one sequence into the other [5]. A matrix of dissimilarity scores is generated as the output of an algorithm that performs these operations on all the sequences. Studer and Ritschard [48] present a thorough overview of various similarity measures that are commonly found in the literature for comparing sequences.

3.2 Fragmentation

As previously explained, a sequence refers to a series of time points during which a subject can move from one discrete "state" to another. Individuals with numerous states in their daily schedules are considered to have fragmented schedules. Fragmentation characterises how activities are reorganized spatiotemporally, with activities being subdivided into smaller components performed at different times and/or locations. Various authors such as Alexander et al. [6], Couclelis [13, 14], Hubers et al. [33], and McBride et al. [40] have defined travel and activity fragmentation as the sequence of many short trips occurring within a person's daily schedule. The concept of activity fragmentation, which forms the basis of this study, is inspired by Couclelis [13], who describes fragmentation as "a process whereby a certain activity is divided into several smaller pieces, which are performed at different times and/or locations" (page 11). Temporal fragmentation relates to the different times at which activities are carried out, while spatial fragmentation pertains to the locations where activities are performed. Together with other activities and movements that occurr in a larger frametime, they collectively form a string of activities with varying durations and purposes. The complexity of this activity string can vary depending on certain extrinsic and intrinsic characteristics of both the individual and the urban environment.

The classification of activity-travel fragmentation segments into clusters offers the opportunity to understand different groups, although it may not explain how and why individuals engage in activity-travel fragmentation, as noted by Su et al. [49]. However, it does allow us to explore the relationship between the segments and socioeconomic characteristics. For example, research conducted by McBride et al. [39] and Su et al. [49] in California found that individuals aged 25–54 with children tended to have the most fragmented schedules. These studies also revealed significant differences among people with different income levels, with poverty inhibiting activity variety and ethnicity/nativity playing a role. Their analyses also showed that specific age groups tend to have very long sequences of short activities and trips. People living in urban and suburban environments tend to have more fragmented schedules, most likely due to the combination of short and long activities within their daily routines. A more in-depth analysis suggests that social exclusion can occur in two ways: Individuals either stay at home with limited access to opportunities, indicating immobility, or they become extremely active, dedicating little time to personal matters and prioritizing the needs of others [30, 31, 40].

Furthermore, researchers have observed gender-differentiated patterns in segmentation analyses. McBride et al. [39, 40] conducted a study in California and concluded that gender roles need to be further investigated, due to the relationship between time allocation for activities and activity-travel. In Europe, Leszczyc and Timmermans [38] analysed Dutch activity diaries and concluded that gender and age play important roles in determining transitions between different activity types. Burchell et al. [11] analysed the gender differences in the segmentation of workplace patterns using the 2015 European Working Conditions Survey [25], which captured information before the pandemic. Their analysis reveals clear differences between genders, such as a higher likelihood for women to work at their employer's offices compared to men. Additionally, von Behren et al. [51] analysed individual pattern segmentation using image-based clustering analysis in combination with the German Mobility Panel [10]. The authors identified two clusters characterised by households with children, with one cluster predominantly consisting of women and part-time workers.

3.2.1 Fragmentation indicators: entropy, turbulence, complexity, and travel time ratio

To understand the complex behaviour of an individual's activity sequences, various indicators have been developed. Ritschard [46] provides a comprehensive review of indicators that characterize state sequences, although only a subset of them is applicable to travel behaviour analyses. In this field, the most relevant and commonly used indicators are *entropy, turbulence, complexity,* and *travel time ratio* [21, 28, 39]**.** These indicators allow for the analysis and measurement of activity durations and transition rates from one activity to the other, thus providing insights into the diversity and complexity of sequences. An activity sequence is a list of states indicating for each minute in a day the activity/state that applies.

The index Entropy provides a measure of variety in daily schedules and represents the proportion of total time spent in each state McBride et al. [40]. However, it does not take into account the number of state transitions [28]:

$$h(x) = h(\pi_1 \ldots \pi_S) = -\sum_{i=1}^{S} \pi_i log(\pi_i) \tag{1}$$

Here, π_i is the proportion of occurrences of the ith state in the considered sequence. S is the number of potential states and $\boldsymbol{x}$ is the sequence of daily activities defined from minute to minute. Function *log()* refers to natural logarithm.

The entropy indicator is calculated based on the proportion of minutes allocated to each state during a day. This measure disregards the number of state changes and the specific ordering of states in the sequence. If a person experiences no state changes during the entire day, their entropy value would be 0. Conversely, visiting several states increases their entropy value. The potential range of values depends on the number of states, with the maximum value achieved when the sequence evenly distributes time among all states. Therefore, a normalized entropy score is commonly used, dividing the entropy by the maximum entropy value, thus obtaining a range of 0–1.

To address the limitations of not considering the ordering and number of state changes, the following indicators have been developed.

The *turbulence index* was proposed by Elzinga and Liefbroer [21] for measuring the number of state recurrences and the variability of durations of daily activities. It is based on sequence permanence and employs two components: the number of distinct subsequences that can be derived from the distinct state sequence,and the variance of consecutive time points spent in a distinct state. For a given sequence $\boldsymbol{x}$, the formula for $T(\boldsymbol{x})$ turbulence is [39]:

$$T(x) = log_2\left(\phi(x)\frac{s_{max}^2 + 1}{s^2 + 1}\right) \tag{2}$$

In the formula: $\phi(\boldsymbol{x})$ is the number of distinct subsequences that can be extracted from the distinct state sequence, considering time precedence. s^2 is the variance for the state duration. s_{max}^2 is the maximum variance, based on the sequence duration, and it is computed as $s_{max}^2 = (n-1)\left(1-\bar{t}\right)^2$, where n-1 is the number of transitions in the sequence and $\bar{t}$ is the sequence duration divided by the number of distinct states in the sequence.

The *complexity* index, proposed as a more sensitive indicator of entropy, was initially developed by Elzinga and Liefbroer [21] and later presented in the work of Gabadinho et al. [27]. This index is a normalized score [1] based on the entropy, and it takes into account both the order of successive states, measured by transitions, and the distribution of different states. It is defined as follows:

$$C(\boldsymbol{x}) = \sqrt{\left(\frac{nt(\boldsymbol{x})}{(l(\boldsymbol{x})-1)}\frac{h(\boldsymbol{x})}{h_{max}}\right)} \tag{3}$$

where $nt(\boldsymbol{x})$ is the number of distinct transitions within a sequence, $l(\boldsymbol{x})$ is the length of the sequence, $h(\boldsymbol{x})$ is the entropy indicator, and h_{max} is the maximum entropy in the sample. This indicator will have a value between 0 and 1, with zero corresponding to zero entropy and no transitions (e.g., staying at a single place for the entire day of observation).

The travel time ratio (TTR), initially defined in the research by Dijst and Vidakovic [18] and recently modified by McBride et al. [39], serves as a concise indicator that represents the trade-offs that people make between travel time and activity time. In this study, TTR is calculated as the total time spent on daily activities (T_t) divided by the sum of the total time at home (T_s) plus the total time on daily activities (T_t). Thus, TTR ranges from 0.5 (no trips made) to 1.0 (the entire day spent away from home).

$$\tau = \frac{T_t}{T_t + T_s} \tag{4}$$

The following sections will describe the application of these indicators to our case study, highlighting their role and importance.

4 Case study

We apply this theory in the Barcelona Metropolitan Area (BMA), which is depicted with a red line in Fig. 1. The Barcelona Metropolitan Region (BMR) consists of 164 municipalities (light grey zone in the map, blue perimeter) and has a population of over 5 million inhabitants. Within this region, the BMA (red perimeter) encompasses 36 municipalities with a population of 3.2 million inhabitants. The BMA has a well-developed public transportation network, with more than 200 bus lines, 4,000 stops, 10 metro lines, 15 railway lines, and two tramway lines. It serves more than 9 million trips every day. The urban and metropolitan bus network provides services throughout the metropolitan area, while the metro network covers the Barcelona municipality and the closest municipalities within the BMA. The commuter rail connects the most populated municipalities in the BMR. The Primary Crown of the BMA, depicted with a green line in Fig. 1, includes the 18 most populated municipalities and is partially covered by the two tram networks. The main data used in this research is derived from detailed mobility surveys conducted every year in the BMR. For further information on the BMA and BMR, please refer to Mejía Dorantes et al. [41] and the AMB [7].

According to EU statistics [24], before the pandemic in 2019, only 5.3% of the labour force in the EU typically worked from home. In Spain, this share was approximately 4.8%. This source observes that following the pandemic outbreak in 2020, this percentage doubled. A recent survey by the National Statistics Institute in Spain [36] reveals that the proportion of people engaging in partial or full teleworking in Spain and Catalonia was 17.6% and 23.4%, respectively. However, as Fana et al. [26] observe, people who work from home (WfH) tend to be high-skilled workers with better wages and contract conditions.

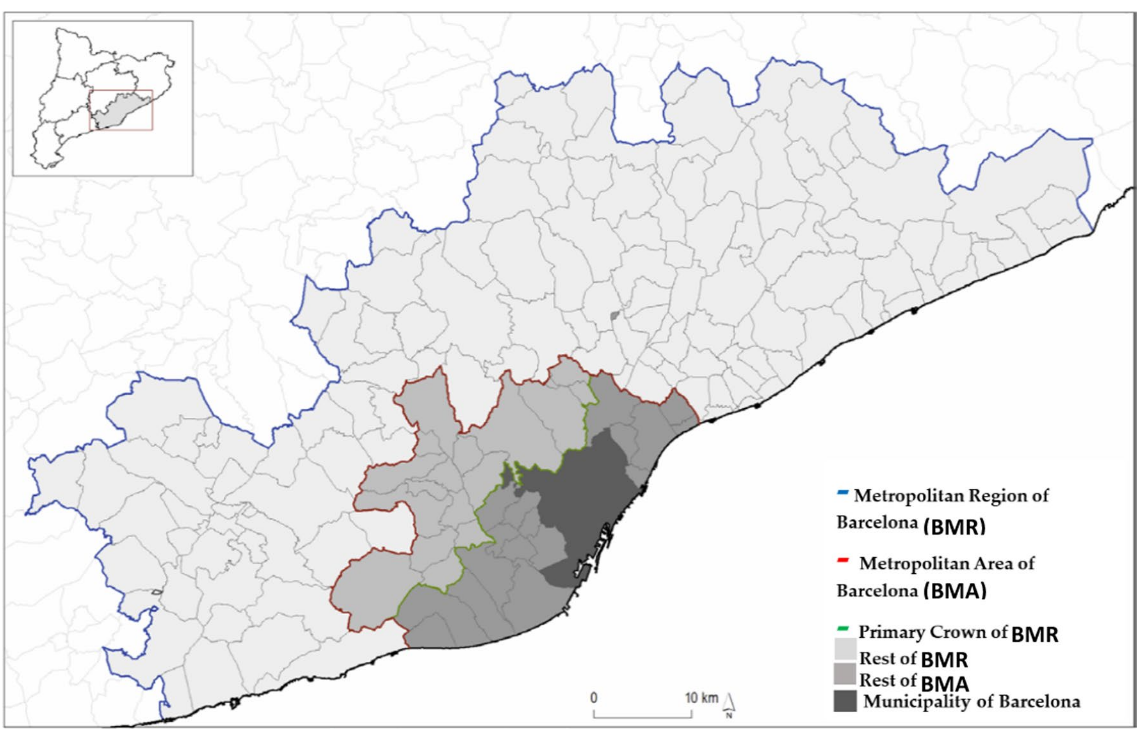

Fig. 1 BMR study area: Transportation analysis zones. BMA subarea in green

5 Data description

The working days' mobility surveys (EMEF) used in this research cover the period from 2018 to 2020, and they are traditional mobility surveys that analyse the mobility of residents in BMR who are 16 years of age and older. The spatial granularity is at the municipality level, and Barcelona is divided into ten districts, resulting in a total of 296 macrozones. Of these, only 45 macrozones are located in the BMA, which is our main study area (as indicated by the green perimeter in Fig. 1).

The survey design and field data collection for the EMEF surveys are supervised by local authorities. A brief description of the process can be found on the official website of the [42]. Each survey consists of two waves: the first wave conducted in April and May; and the second one in October and November. The EMEF 2020 survey was conducted only during the autumn of 2020 when certain mobility restrictions and the prevalence of online activities persisted due to the COVID-19 pandemic. Therefore, special attention should be given to the EMEF 2020 survey, since activity patterns may differ compared to the pre-COVID-19 situation.

The sample size is approximately 10,000 units, selected by stratification using residential area, gender, and age group. The data collected for each journey includes trips for the previous day: origin and destination (macrozones), the purpose of the trip, mode of transportation (with a highly detailed list of possibilities), travel start time, duration (in minutes), vehicle use, parking use, and more. The sample units represent individual residents rather than households. The Barcelona datasets include only working days and no data on household composition. To avoid sample bias, professional drivers are excluded from the analysis, as their travel behaviour is not representative of the population sample. The final sample sizes for each year are as follows: 9,930 for 2018, 9,934 for 2019, and 10,024 for 2020. The total number of trips recorded in the BMA is 36,368 for 2018, 37,463 for 2019, and 30,591 for 2020 BMA. After filtering out cases with missing information, such as residential areas, the final number of residents included in the total sample is 26,860.

We also make use of demographic and land use data. This information is defined with the same spatial granularity as the traffic analysis zones (TAZ-EMEF) used in the surveys. This data includes the population segmented by gender (male and female only), age group (5 groups), education level, educational institutions, services, land use, residential morphology, average income, and number of stops in the public transportation network.

6 Methodology for the case study analysis

We aim to understand how mobility patterns in the BMA are influenced not only by individual characteristics such as gender, educational attainment, origin, and life-course but also by the pandemic outbreak. To achieve this, we employ a time series dataset and, in line with previous studies and theoretical approaches, we apply sequence analysis to statistically analyse the fragmentation of activities.

Furthermore, we analyse the normalized entropy, turbulence, complexity, and TTR for trip makers. Using a minute-by-minute time series, we capture every minute of the day associated with a specific state for each person in the study.

The workflow of the procedure is as follows (see Fig. 2 for a summary):

1. First, we analyse the activities and modal choices, defining ten activities, including home (H), work (W), casual (C, infrequently visited places), other (O, frequently visited places outside of work), school (S, students only), escorting (A), and the four travel categories: active (TW, walking and cycling), public transport (TP), travelling by car (TC), and other travel categories (TM). The final analysis considers an

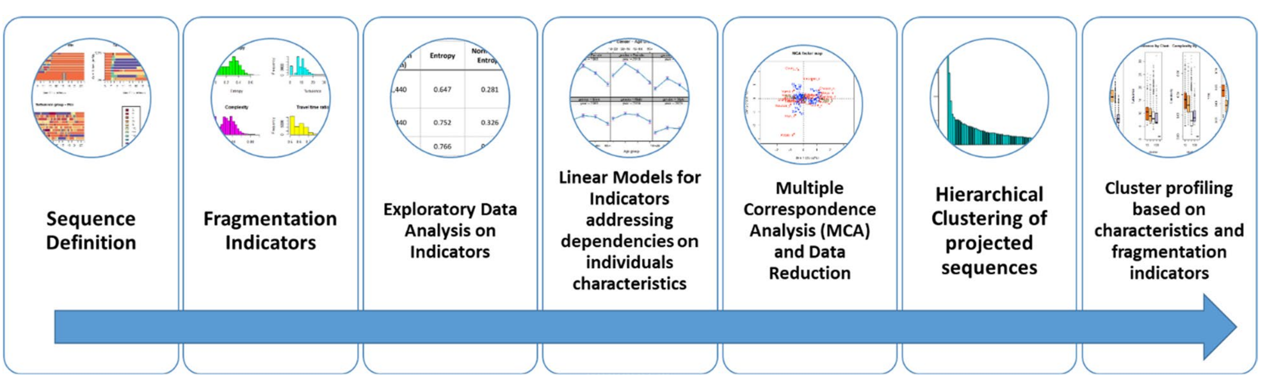

Fig. 2 Methodological workflow

alphabetic list of 10 activities and transport choices. Both trip makers and non-trip makers are included in a sequence of 1440 min or more (for example, those returning home after midnight), with each minute classified according to an activity category.

2. Next, we analyze entropy, turbulence, complexity, and the travel time ratio. Since we consider ten activities, the maximum entropy is 2.3 and we thus use the normalized entropy. The statistical analysis of these indicators is carried out with the TraMineR package in R [28, 45], which has been widely used in the social sciences for the analysis of biographical longitudinal data.
3. The previously defined fragmentation indicators are used as inputs for the different linear models, focusing on the subsample of trip makers from 2018 to 2020. The goal is to examine the dependence and behaviours of these trip-makers in terms of their individual characteristics and information in the built environment. Our models incorporate individual characteristics like gender, age group, activity, disability, working-from-home possibility, day of the week, and the zone (TAZ-EMEF) in the Barcelona area. The following descriptive variables are also included: the percentage of residential areas with dwellings larger than 100 m^2, and the number of buses, metros, trams, and train stops in each zone.
4. We next compare all sequences by computing pairwise mean fragmentation indicators for each year, using Tukey's honestly significant difference test (Tukey's HSD) [50] to evaluate variations in fragmentation indicators across years.
5. To address the dimensionality issue of this database, we employ a dimension reduction technique. As discussed in the literature review, various alternatives are available. Instead of selecting a subsample like McBride et al. [40] or aggregating the data into longer timeframes, we chose to reduce the dimensionality. This alternative is recommended for handling such a large dataset, which in our case comprises daily sequences from 5 to 24 h. Our database thus remains highly detailed, as we maintain relevant information rather than reduce the timeframe, which would lead to lost information, such as when an activity duration does not fit in a specific timeframe. A more methodologically suitable approach thus maintains the level of detail and later employs data analytics techniques to evaluate the database. Our analysis uses a $26{,}860 \times 1{,}140$ matrix containing activities from the selected alphabetic list (10 options), where each column represents a categorical factor with 10 possible levels (activities).
6. For clustering analysis, the similarity/dissimilarity matrix has millions of cells ($26{,}860 \times 26{,}860 = 721{,}459{,}600$) that contain the dissimilarity scores for the sequences of each person in the working sample. The method seqdist() from the TraMineR package in R allows us to calculate the dissimilarity matrix based on several metrics [48]. However, due to large memory requirements, it cannot be applied. For this reason, we must use multiple correspondence analysis (MCA) instead of principal component analysis (PCA) to detect underlying structures in the dataset, given the large memory requirements of 2.3 Gb.
   After the data reduction, we proceed with a clustering technique to group sequences of activities with similar dissimilarity scores, which are obtained from the sequence comparison after projection. The fragmentation indicators are used to interpret the clusters. The final number of clusters is determined by using optimized criteria for balancing within-group similarity and between-group dissimilarity. Specifically, we apply hierarchical clustering (HC) [34] to the daily travel patterns contained in the minute activity matrix, where each minute is based on distances between daily travel sequences. Each cluster comprises points that are more similar to each other than to the points in other groups. The hierarchical clustering method in the FactoMineR package can reduce the computational burden by starting the agglomerative process on a heuristic partition that represents 10% of the original length. The hierarchical agglomerative tree is then cut using a balanced combination of commonly used techniques, such as the between sum of squares to total sum of squares, gap statistic, and silhouette methods.

7 Results

The next sections present an overview of the different analyses conducted in this study.

7.1 Analysis of activities and modal choice

Our analysis encompasses a broad range of activities based on the places individuals visit throughout the day. The final analysis considers an alphabetic list of 10 activities and transport choices. Both trip makers and non-trip makers are included in the analysis, along with a sequence of 1440 min or more, which accounts for individuals returning home after midnight. Each minute is classified according to the corresponding activity or transport choice indicated by the assigned alphabetic letters. Figure 8 shows the distribution of daily activity patterns across the years. It is obvious that the state activity pattern distribution for 2020 shows an overrepresentation

Table 1 Percentages of activity states and transition modes distribution across main factors such as gender, age group, year, and residential area (See Fig. 1, Barcelona city, Primary Crown and Rest of the Barcelona Metropolitan region). From 5:00 to 24:00 h using the EMEF 2018 to 2020

Gender	**Activity states and transition modes**									
	A Escort (%)	**C Casual (%)**	**H Home (%)**	**O Other (%)**	**S School (%)**	**TC Private Tr (%)**	**TM Other Tr.mode (%)**	**TP Public Tr. (%)**	**TW Active modes (%)**	**W Work (%)**
Male	1.0	1.9	62.6	4.9	2.5	3.0	0.2	1.4	2.2	20.2
Female	1.3	2.3	68.1	4.6	2.4	2.0	0.1	1.8	2.2	15.2
Age group	**Activity states and transition modes**									
	A (%)	**C (%)**	**H (%)**	**O (%)**	**S (%)**	**TC (%)**	**TM (%)**	**TP (%)**	**TW (%)**	**W (%)**
16–29	0.3	2.3	60.1	5.4	12.2	1.9	0.1	3.2	1.6	12.9
30–44	1.7	1.6	58.2	3.8	0.6	3.2	0.2	1.5	1.9	27.4
45–64	1.1	2.0	62.8	4.4	0.3	2.9	0.1	1.4	2.1	22.9
65 +	1.2	2.9	82.9	6.1	0.2	1.4	0.1	0.9	3.5	0.9
Year	**Activity states and transition modes**									
	A (%)	**C (%)**	**H (%)**	**O (%)**	**S (%)**	**TC (%)**	**TM (%)**	**TP (%)**	**TW (%)**	**W (%)**
2018	1.0	2.7	61.7	5.1	2.8	2.9	0.0	2.0	2.1	19.6
2019	1.5	2.1	60.8	5.9	3.0	2.8	0.1	1.9	2.4	19.5
2020	0.9	1.6	73.5	3.5	1.5	1.8	0.2	0.9	2.2	13.8
Residential area	**Activity states and transition modes**									
	A (%)	**C (%)**	**H (%)**	**O (%)**	**S (%)**	**TC (%)**	**TM (%)**	**TP (%)**	**TW (%)**	**W (%)**
Barcelona	1.2	2.7	61.9	5.8	2.8	1.5	0.1	2.8	2.9	18.4
Primary	1.4	2.4	63.8	4.6	2.5	2.2	0.1	2.0	2.5	18.4
Rest	1.3	2.0	63.3	4.5	2.8	3.1	0.1	1.4	2.2	19.1

of entire-day home activities, likely influenced by the COVID-19 pandemic and related restrictions.

The percentages of activity distribution across the day for men and women and by year are shown in Table 1. Work activities outside the home are greater for men than for women, while school activity is similar for both genders. In 2020, home activity increases by almost 13 points compared to previous years, reaching 73.5%, while school and work activities outside the home were clearly reduced. The percentage of private transport use is substantially influenced by the residential area (analysed here according to the crowns, as described in Sects. 4 and 5). Non-central crowns show a higher incidence of private transport activity (3.4%) compared to Barcelona City (1.5%), while public transport usage is higher in the central crown, as Barcelona City shows an incidence of 2.8%. External crowns fall within 1.2–1.4% incidence. Similarly, active modes are preferred in the central part of Barcelona city. Gendered mobility patterns also emerge. Women tend to prefer public transport, while men are more inclined to use private transportation. Female individuals escort others more often and spend more time at home. Men have more trips related to work compared to women. Furthermore, car use increases in the 30–44 age group, which, again, aligns with life-course patterns.

7.2 Fragmentation analysis

Table 2 shows three examples of trip makers along with their daily activity sequences and durations. The first unit's daily activity sequence is as follows: H-TW-O-TW-H-TW-S-TW-H-TP-O-TP-H. After midnight, this person allocates their time according to the previous activities' sequence in the following manner for each episode (in minutes): 600-1-29-1-314-15-120-5-55-30-90-30-150. In this case, the total duration is 1,440 min. The values of the fragmentation indicators for this example are: 0.647 for entropy, 0.281 for normalized entropy, 14.421 for turbulence, 0.041 for complexity, and 0.563 for TTR. This illustrates the rationale for maintaining a minute-by-minute level of detail instead of grouping activities into arbitrary time intervals. While grouping would reduce the dimensionality of the dataset, it would also result in the loss of relevant information.

Table 2 Daily-travel pattern example for 4 units of the working sample (maximum duration of dataset 1965 min)

Unit	Daily activity sequence	Used time per episode (min)	Total duration (min)	Entropy	Normalized entropy	Turbulence	Complexity	TTR (travel time ratio)
1	H-TW-O-TW-H-TW-S-TW-H-TP-O-TP-H*	600-1-29-1-314-15-120-5-55-30-90-30-150*	1,440	0.647	0.281	14.421	0.041	0.563
2	H-TW-S-TW-H-TW-S-TW-H*	600-30-240-30-120-30-210-30-150*	1,440	0.752	0.326	10.898	0.036	0.623
3	H-TP-W-TP-H*	465-40-590-40-305*	1,440	0.766	0.333	7.421	0.026	0.652
4	H-TC-S-TP-O-TW-O-TP-H	960-30-195-40-155-15-450-120-0	1,965	1.247	0.542	9.972	0.044	0.672

*For the calculations, Homestay extends from 1,440 to 1,965 min to account for the maximum day duration recorded in the dataset (Unit 4)

The range of values for normalized entropy are 0 to 1. Units 2 and 3 exhibit greater normalized entropy values (0.33) compared to Unit 1 (0.28). Despite the number of episodes, Unit 1 predominantly remains at home for most of the time (as reflected in the TTR for each unit). Unit 4 has the largest (normalized) entropy 0.542 and TTR (0.672) involving eight episodes and six different activities; it leaves home at 16 h (960 min from midnight) contributing to increase entropy.

It is worth noting that Units 2 and 3 have fewer changes in activities throughout the day, with only three different activities compared to Unit 1's five different activities (or Unit 4). Consequently, Unit 1 has a higher turbulence value than Units 2 and 3. The Complexity indicators for Units 2 and 3 are lower than for Unit 1 because they have fewer episodes (9 and 5, respectively, compared to Unit 1's 13 episodes). The complexity of Unit 4 combines large entropy and turbulence.

Higher values of turbulence and complexity serve as clear indicators of fragmentation, but a comprehensive understanding of fragmentation can also be obtained through the additional indicators, Entropy and Travel Time Ratio.

In general terms, these indicators provide an overall perspective on the extent of fragmentation experienced by individuals, and they exhibit a direct correlation, with correlation coefficients exceeding 0.6 (Pearson/Spearman). To gain further insights into the behaviour of fragmentation indexes, a further analysis was conducted using nonparametric statistical methods. The Kruskal–Wallis test was employed to assess the homogeneity of means across groups (defined by factors such as gender, age group, and year, among others). Additionally, the Tukey multiple comparison HSD test was utilised to perform pairwise comparisons between means within the same groups.

The Kruskal–Wallis test confirms the presence of non-homogeneous normalized entropy and TTR across years (p-value = 0), while homogeneity of variances cannot be rejected. Tukey multiple comparisons of means confirm that the entropy mean in 2020 is lower than those in 2018 and 2019 at a 95% confidence level. When considering only the subset of trip makers, the same conclusions hold for the previous years, as entropy is reduced by 18.57% in 2020. The Tukey multiple comparisons of means confirm that the TTR mean in 2020 is lower than those in 2018 and 2019 at a 95% confidence level. When considering only the subset of trip makers, the same conclusion holds, and TTR in 2020 is reduced by 5.5% compared to 2018 and 2019.

7.3 Linear models

In the following section, we provide detailed information on the most relevant results regarding the various fragmentation indicators.

7.3.1 Linear model turbulence

To gain a deeper understanding of *the behaviour of the turbulence* and its relationship to individual characteristics, a linear model was employed on their logarithms using the *subsample of trip makers* from the 2018–2020 EMEF dataset. These individual characteristics include gender, age group, activity, disability status, WfH possibility (for the year 2020 only, due to the lack of information in previous years), and day of the week. Additionally, descriptive variables related to TAZ-zone were included, such as the percentage of households over 100m^2 and the number of bus, metro, tram, and train stops. However, the following variables were found to be non-significant: day of the week, dwellings larger than 100m2, number of public transport stops, and professional activity. The last case is because professional activity provides no additional information once the educational level is included in the model.

The marginal effect of the variables having significant net effects is shown in Fig. 3. The results show that the *fragmentation* of female out-of-home activities is significantly higher than for men in the 30–44 age group.

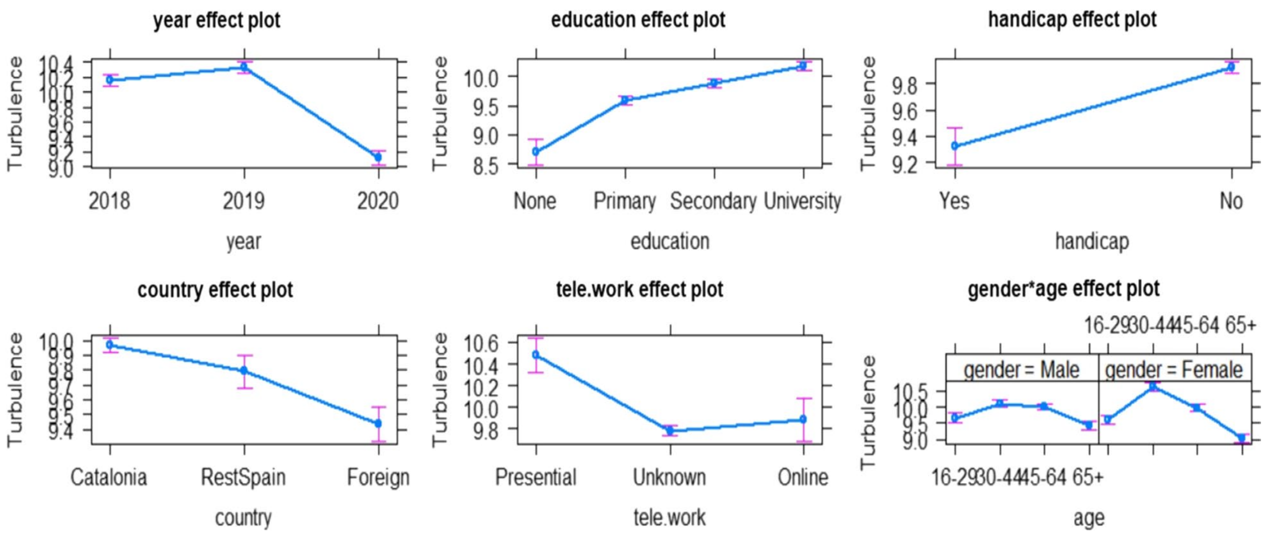

Fig. 3 Turbulence transformation marginal effect for 2018–2020 EMEF (only trip makers)

A notable impact of the year 2020 on the logarithm of turbulence is observed after controlling for other factors. Trip makers originating from other places in Spain or abroad appear to have lower *turbulence* than local trip makers, potentially due to having less social interaction.

Figure 4 shows the logarithm of the **turbulence** according to gender, age group, and year, after controlling for education, birthplace, disability status, and telework possibilities in the model. Women in the 30–44 age group exhibit the highest turbulence across all the examined years. While the turbulence for men remains relatively stable in the 16–64 age groups in 2018 and 2019, the youngest groups for both men and women experienced a substantial decrease in fragmentation in 2020. Generally,

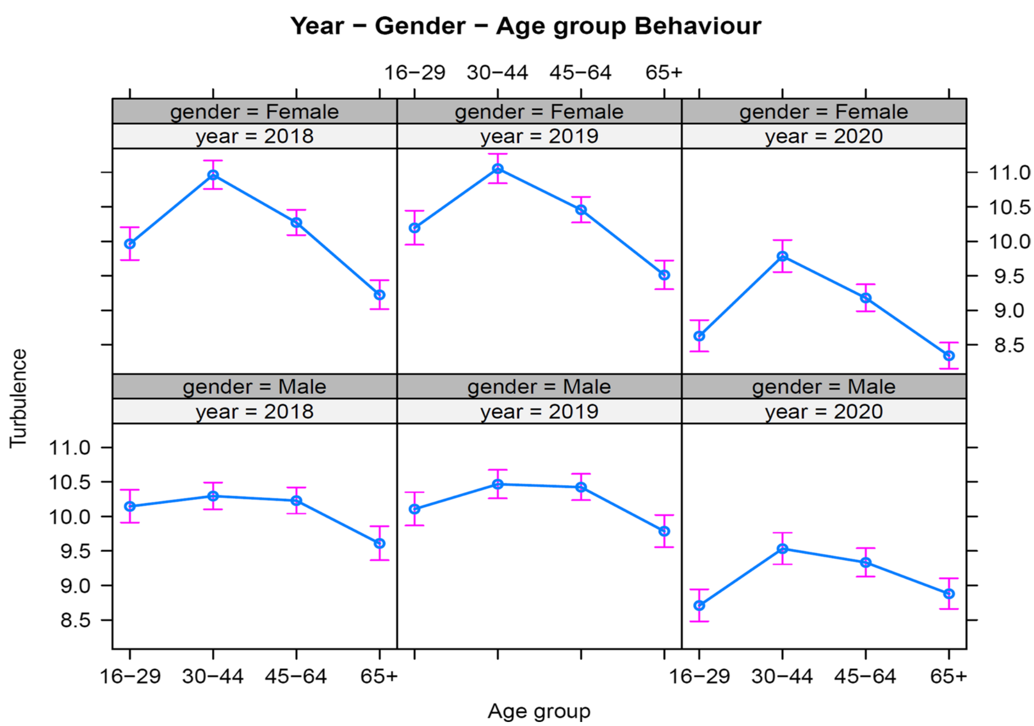

Fig. 4 Marginal effect of turbulence transformation, considering year, gender, and age-group interaction in the entire 2018–2020 EMEF dataset (limited to trip makers) after controlling for education, origin country, disability status, and telework effects

turbulence levels were reduced in 2020 after the COVID-19 outbreak. The higher turbulence among women confirms that gender plays a role in the complexity of daily activities, particularly regarding women's lifetime situation, which involves various caregiving activities. Women above the age of 65 appear to decrease their activities more than men.

7.3.2 Linear models of entropy and TTR

The analysis reveals that male entropy surpasses female entropy, reaching its highest value within the youngest age group (see Fig. 5). Essentially, irrespective of gender, individuals under 30 years of age exhibit a more balanced time allocation across various activities during the years 2018 and 2019. However, as individuals progress through different stages of life, this equilibrium diminishes, particularly for women. Furthermore, the normalized entropy for the year 2020 illustrates the influence of the COVID-19 pandemic on activity patterns compared to previous years, regardless of gender. It indicates a reduction in the number of activities with unequal time distribution, signifying a departure from established patterns.

Figure 5 also shows the TTR. Similar to the previous indicator, individuals below the age of 30 tend to spend a greater portion of their time outside of their homes, which gradually decreases as they grow older. This trend aligns with the increased parental responsibilities typically experienced with age. However, this decline is particularly pronounced among women, reflecting the patterns observed in household and family care participation, as reported by Eurostat [23] and other sources. Furthermore, even though mandatory lockdowns were no longer in place in Spain during the time of the survey in 2020, individuals continued to minimize their travel and preferred to stay at home as much as possible.

7.4 Cluster analysis

An agglomerative hierarchical clustering method (HCPC method in FactoMineR library for R) was employed

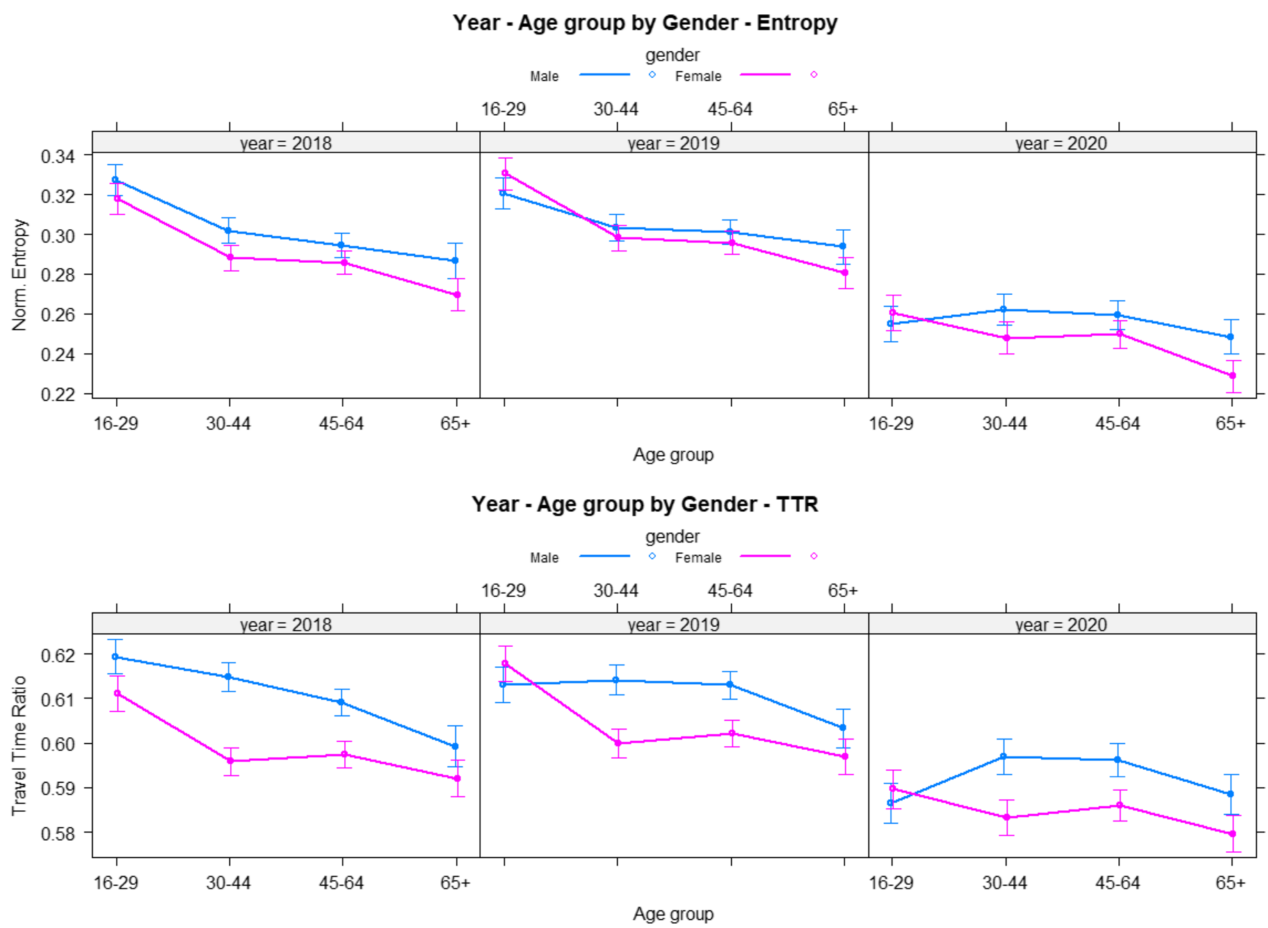

Fig. 5 Marginal effects of normalized entropy (upper panel) and travel time ratio (TTR, lower panel), by gender, age-group, and year of interaction (2018–2020). Data source: EMEF subsample of trip-makers after controlling for year, education, activity, country of origin, disability, and telework effects

following dimensionality reduction through multiple correspondence analysis to mitigate biases resulting from information loss. This analytical procedure ensures the utilisation of all available information (minute-to-minute activities), and it was designed to avoid distortions during dimensionality reduction. The reduction is based on the Kaiser criteria [37], with the retained factorial axes accounting for 93.47% of the data variability, thus ensuring representativeness. These axes were used to project the sequences into the new space, achieving a 95% dimension reduction. The 100 clusters chosen for analysis capture 43% of the total data variability. The cluster outcomes have been incorporated into the dataset containing descriptive characteristics of the samples. It is worth noting that some clusters have a relatively small number of samples.

The five largest clusters exhibit distinct distributions of entropy, turbulence, complexity, and travel time ratio, highlighting the divergent behaviors of individuals within each cluster regarding these indicators. Figure 6 shows the various statistical values for fragmentation indicators, along with their outliers, emphasising their relevance in characterising the fragmentation within each cluster. These clusters account for 32% of the sample, with a median size of 89 units, and 90% of the clusters comprising less than 373 units. The fifth cluster consists of non-trip makers, who appear at the bottom of each panel as a reflection of their 0 values for entropy, turbulence, and complexity, while their TTR is 0.5. The presence of numerous clusters with a small number of observations can be attributed to the fine granularity of minute-based activity sequences taking all sequences in the sample, which allows for highlighting differences in activity patterns. Homogenous fragmentation means across clusters (inter-cluster variability) have been rejected using the Kruskal–Wallis test (p value = 0) for all four indicators.

To provide an illustrative example, we use categorical variables to examine one of the largest clusters, number 13, which consists of 1482 observations (in the 90% percentile of cluster size). This cluster primarily comprises car drivers who occasionally use public transport. They are working men in the age groups of 30–44 and 45–64, have higher education levels, and originate from Catalonia. This cluster is over-represented by sample units (trip makers) from 2018 and 2019. Profiling of clusters based on characteristics and fragmentation indicators can be addressed for any cluster, obtaining similar results based on statistical significance.

An in-depth analysis was carried out to examine two specific aspects:

- Clusters significantly overrepresented by male samples compared to those overrepresented by female samples.
- Clusters significantly overrepresented by samples from the year 2020 compared to those from 2018 and 2019.

For instance, clusters 19, 22, and 62 exhibit a significantly higher percentage of women (70%). Figure 7 presents the activity pattern distribution for each of these clusters. Women in these clusters allocate a significant portion of their daily activities (from 15 to 21 h) to escorting activities (A) and staying at home (H). Conversely,

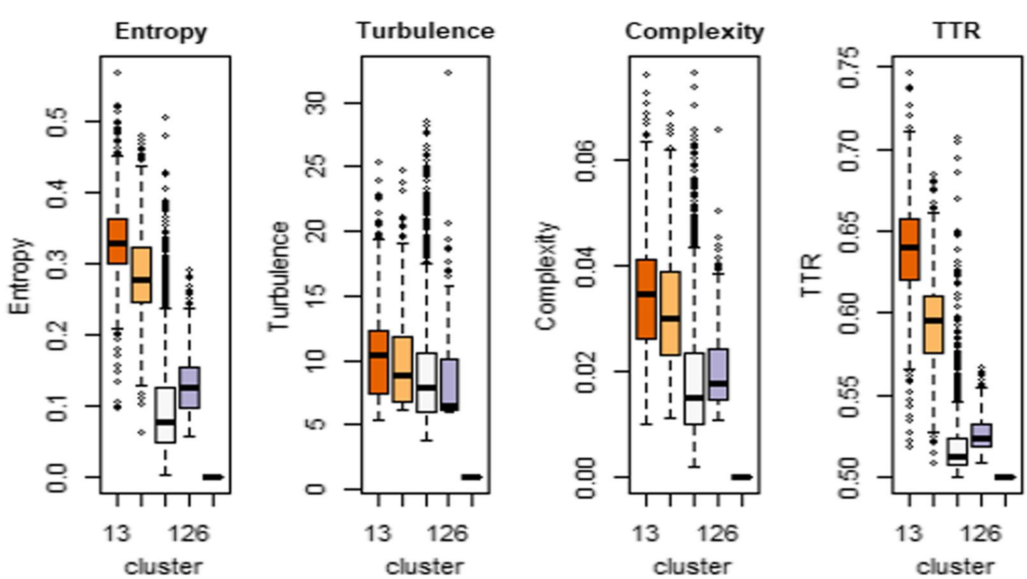

Fig. 6 Entropy, turbulence, complexity, and TTR for largest clusters. EMEF 2018 to 2020. The rightmost cluster collects all non-traveller units

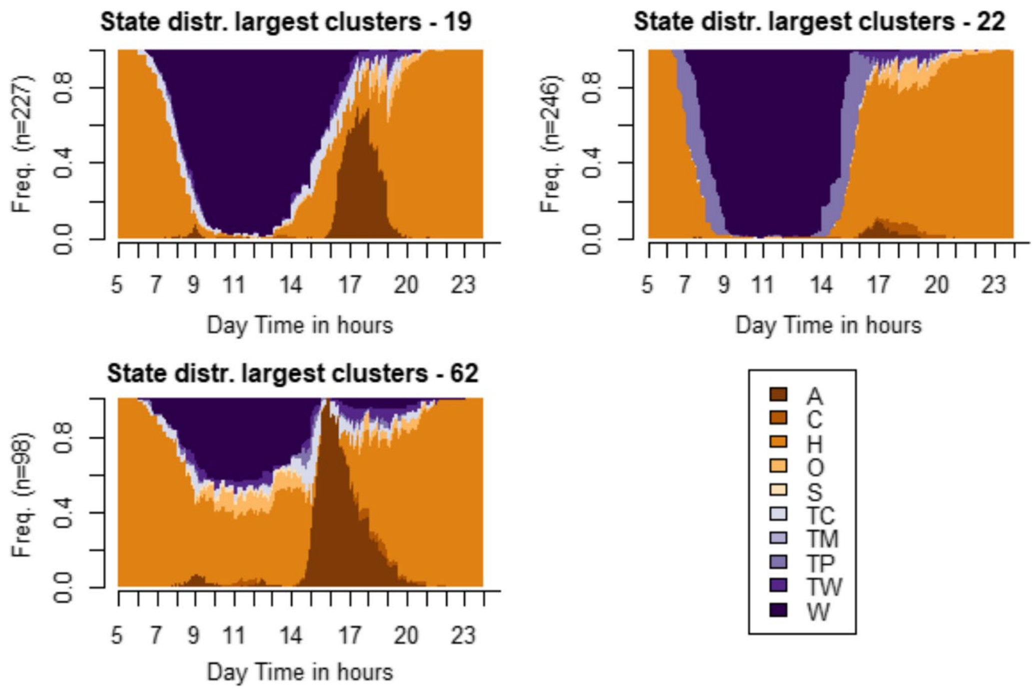

Fig. 7 Classification of daily activity patterns in EMEF sample, 2018–2020. The three clusters that are most over-represented by women

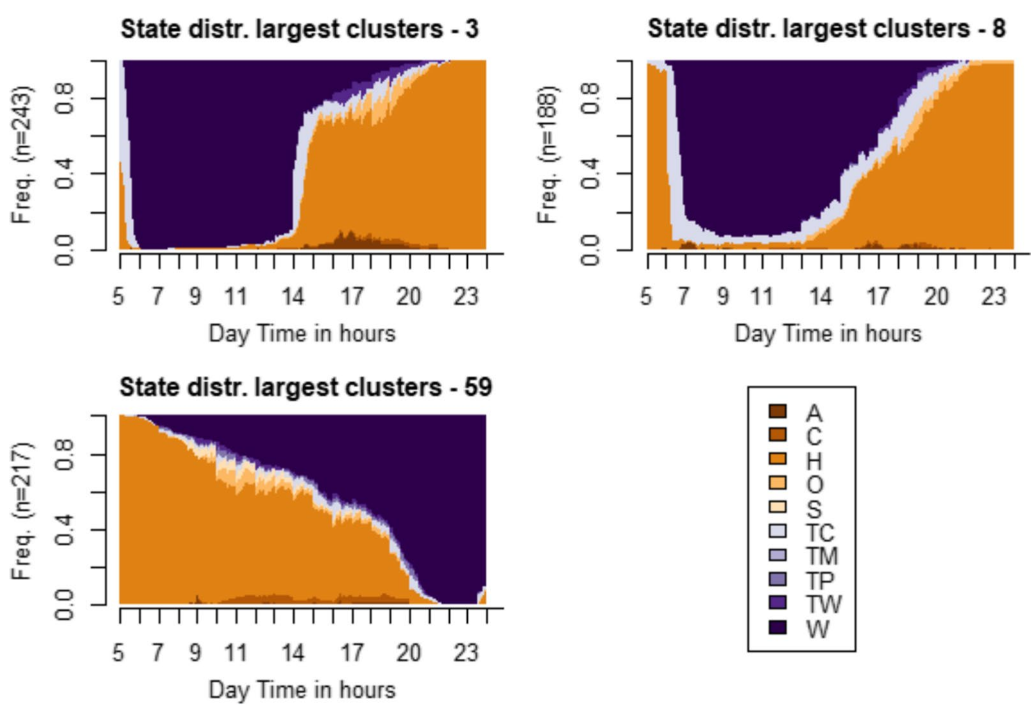

Fig. 8 Classification of daily activity patterns in EMEF sample, 2018 to 2020. The three clusters that are most over-represented with men

working activities (W) are particularly prominent from 8 a.m. to 3 p.m., and transportation activities peak during lunchtime.

In contrast, clusters 3, 8, and 59 exhibit a significantly higher percentage of men. The activity pattern distributions for these clusters are shown in Fig. 8. These clusters exhibit a high incidence of work (W). The relevance of private transport (TC) as the modal choice is evident. Furthermore, clusters 3 and 8 show a distinct pattern of staying at home (H) during the late afternoon. Notably, cluster 59 contains men who work night shifts. Compared to the female-dominated clusters, the incidence of escorting activities (A) differs notably in these male-dominated clusters.

Table 1 in Sect. 7.1 shows how escorting activity has a greater incidence in females, than in males and this is consistent with the fact that clusters over-represented by women show a remarkable presence of escorting activity in the state distribution of activity sequences (Fig. 7). Escorting activity is more common in the age group 30–44 years (Table 1) and turbulence linear model shows a marginal effect on turbulence for the 30–44 age group and females (Fig. 4). In Table 1, the residential area effect in the incidence of the transition modes shows that the built environment plays a relevant role in terms of the use of the private vs public transport and that it is possible because of the coverage of multimodal transport system in the study area. In Sect. 7.4, we address an analysis in greater detail based on classifying activity sequences, an extensive analysis of all obtained clusters is not addressed in this work. We examine some of the clusters using state distribution graphics and profiling tools on characteristics and fragmentation indicators, significant information is obtained that summarizes the critical features of each cluster.

Data analytics techniques, clustering analysis, MCA, PCA, and others used in this paper on a rich multivariate dataset, offer many different alternatives for leading to the analysis and understanding of the features and properties of the studied samples, we are aware that we have conducted only partial studies that represent specific choices aimed at servicing the main objectives.

7.5 Insights on the built environment and fragmentation indicators

The results reported in Table 1 make evident that travel patterns also depend on other factors, the last sub-table reveals that the zone of residence, the city of Barcelona, the Metropolitan Area (BMA), or the rest of the Metropolitan Region (BMR) has a strong in the way travelers use the available transport modes but, a deeper analysis to get a better understanding of these travel patterns requires resorting to the information provided by other data sources as those related to socioeconomic aspects of the area object of study, the spatial characteristics of the area, like accessibility to public transport or the built environment, in addition to the fragmentation indicators used so far. To achieve that objective let's consider which are the available variables in our case study. On one hand, there are aggregated variables accessible at the Traffic Analysis Zone (TAZ) level, as those corresponding to the socioeconomic and built environment attributes characterizing each TAZ. While, on the other hand, fragmentation variables are disaggregated, available at the individual level for each member of the sample.

Concerning the aggregated variables, for the study area, and its zones, depicted in Fig. 9, there is an additional data set containing land use, demographic, and other socioeconomic data at the TAZ level contained in the TAZ-EMO, which were defined in the Commuters' Mobility survey in 2001 [47].

However, the zones used in the EMEF surveys, called from now on TAZ-EMEF, whose data have been utilized in the fragmentation analysis of previous sections of the research reported in this paper, are not the same as the TAZ-EMO containing the new additional data, therefore the use of these data requires a:

- Zonal aggregation: TAZ-EMO and TAZ-EMEF zonings had different origins and were set up with different objectives. For BMR TAZ-EMO consists of 582 ones and TAZ-EMEF of 296, therefore there is a need to aggregate the 582 BMR TAZ-EMO zones into 173 TAZ-EMEF zones in the BMR area, Table 3.
- Data orchestration is needed to account for the three EMEF sources (2018, 2019, and 2020) because they were delivered independently and the recorded fields are not the same. The orchestration of EMEF datasets selects common subsets of fields and reorders them properly. Although EMEF data allow access to specific periods of the day the orchestrations have not considered them and have been conducted for total daily trips.
- Land use variables are made available for research by local authorities (Autoritat del Transport Metropolità, ATM) and they can be transferred (aggregated) from the TAZ-EMO zoning system for the Barcelona Metropolitan Region (BMR) to the TAZ-EMEF zoning system. These variables are m2 and m2 of ceiling for residential use, shopping use, industrial use, services, etc., m2 of ceiling for residential areas classified as the traditional city center, and, according to the terminology used in Barcelona for the typology of the different districts classified into four groups: old city, urban block (Eixample-like), isolated blocks and individual houses.

7.5.1 Land use and built environment variables combined to yearly fragmentation indicators

These variables describe the socioeconomic characteristics of the population and other information explicitly

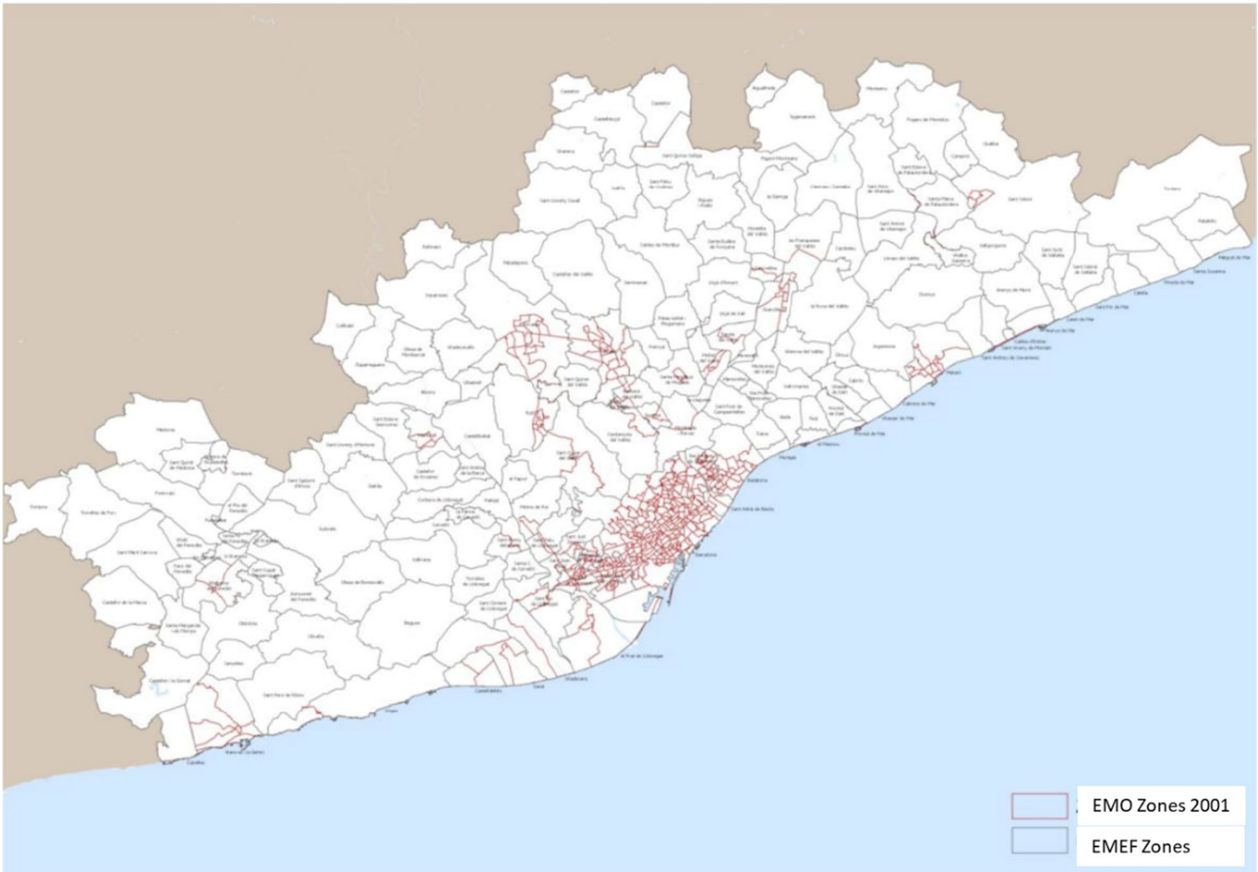

Fig. 9 TAZ-EMEF and TAZ-EMO Zoning of BMR. Source: [47]

Table 3 Comparison of the number of zones EMEF and EMO Zoning Systems

CROWN	TAZ-EMEF	TAZ-EMO
BCN	10	198
Rest of primary crown (ETM)	17	151
Rest of BMA	18	33
Rest of BMR	128	210
Rest of Barcelona province	134	–

or implicitly related to the built environment according to TAZ-EMO zoning system granularity:

- Densitypop: Residential Density defined by the ratio: TAZ Population / TAZ Area Size.
- Densityjob: Employment Density defined by m2 ceiling for activities (commerce, cultural, industrial, warehouses, entertainment, public buildings, shops, offices) divided into total m2 ceiling any activity).
- Diversity: defined after the measure of Land Use Diversity in terms of the Land Use Mix (LUM) indicator, whose value for TAZi is given by:

$$LUM_i = -\left\{ \frac{\sum_{j=1}^{n} P_{ij} lnP_{ij}}{ln(n)} \right\} \tag{5}$$

where P_{ij} is the proportion of land use of type j in TAZi, and n is the number of different types of land uses, i.e. residential, services, retail, other). Named diversity.
- Blocktype: Design (Blocks Typology) for each TAZ we have m2 of residential land use split into percentages of residential block types (old city, urban block, isolated block, and houses).
- Accesstpu: Accessibility to public transport facilities (i.e. distance to transit). Defined as the access time at origin to public transport.
- Access: Destinations' accessibility defined as the median access time walking at origin and travel time to any destination. Estimated by skim matrices from a traffic assignment analysis.

- Access.D12. Defined as the median access time walking at origin and travel time (any mode) to the center in Barcelona (Districts 1 and 2).

Additional complementary variables, which are not usually directly available, but can be derived from the sources provided by planners:

- Scholar places: Number of scholar places in each TAZ.
- Services: Number of services or services' typologies).
- Av_rent: Average per capita income.

The former variables have been aggregated into the TAZ-EMEF zoning system and fragmentation variables for each year (2018–2020) calculated at the individual level have been also aggregated at the TAZ-EMEF level taking into account weighted means according to the expansion factor given at each individual unit. After a first exploratory data analysis, we decided to combine fragmentation indicators for 2018–2019 (average) in four variables and 2020 fragmentation variables stand as defined:

- Entropy 2018–2019 and 2020 Average normalized entropy of each TAZ – EMEF and years
- Turbulence 2018–2019 and 2020 Average turbulence of each TAZ – EMEF and years
- Complexity 2018–2019 and 2020 Average complexity of each TAZ – EMEF and years
- TTR 2018–2019 and 2020 Average travel time ratio of each TAZ – EMEF and years.

The dataset has as many rows as TAZ-EMEF zones in the BMR and as many columns as the variables indicated above. The first step of the new analysis consisted of conducting a Principal Component Analysis (PCA), taking the TAZ-EMEF as units, since all the variables, defined in the previous section, characterize the attributes of each TAZ are quantitative. Fragmentation indicators were taken as active variables in the calculation of the factorial axes in the PCA. Figure 10 (left) presents the projection of the variables in the factorial plane of the first two main components. This analysis has been complemented by a hierarchical clustering based on the principal components defined by PCA analysis based on population and employment density, average per capita income, etc. as socioeconomic indicators, residential structure, accessibility, diversity, etc., as indirect indicators of built environment, and fragmentation indicators. Figure 10 (right) presents the projection in those of the third and fourth main principal components.

The first two principal components explain 45% of the variance and the first main component can be interpreted as a size component, in terms of population, accessibility variables, and services, and the second although has a diluted role, is mainly determined by income, public transport accessibility, and 2018–2019 fragmentation variables. Given the relatively low

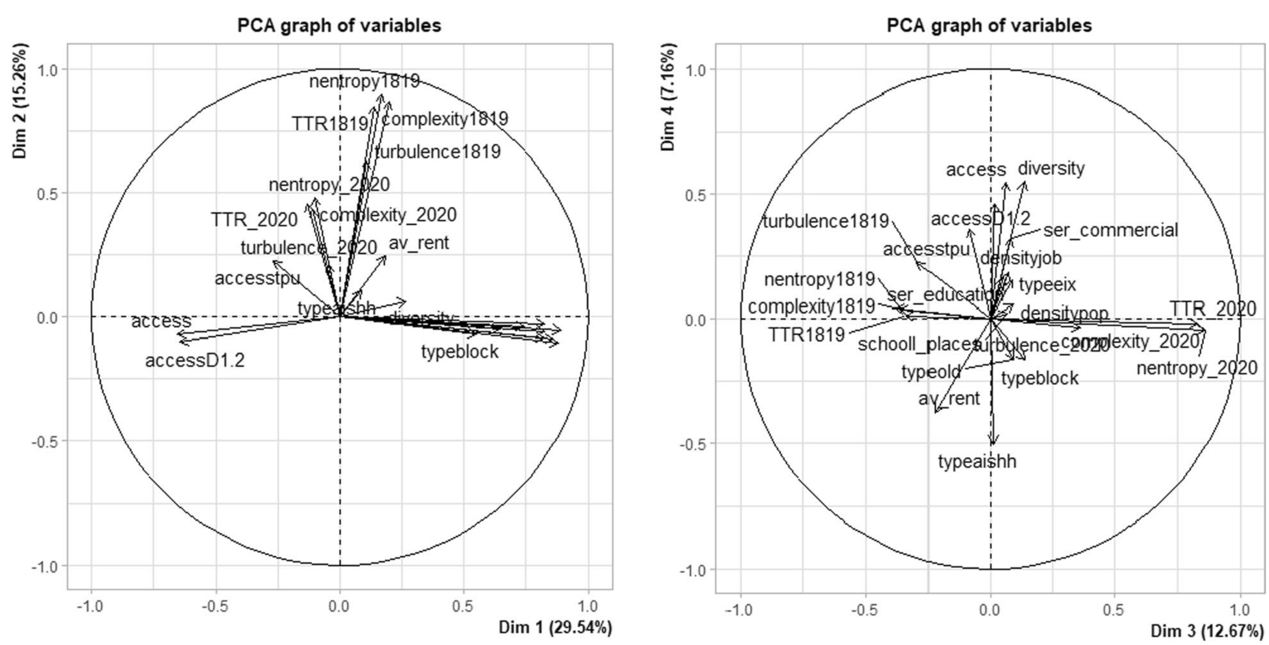

Fig. 10 Projection of all variables in the plane of the first two principal components (left) and 3rd-4th principal (right)

percentage of variance explained, we wondered what role the next two components, the third and the fourth, could play.

The third and fourth components, which represent a significant explanation of the variance (20%), are fundamentally determined by the fragmentation indicators and, what is relevant, those of accessibility highly correlated to diversity. Thus, for example, fragmentation variables in 2020 as opposed to 2018–2019 ones at the same time.

The hierarchical clustering, in the space of the first six principal components, captures the main combined information of all variables (75% of data variability). The number of selected clusters is 9 (plus missing data cluster) because goodness of fit statistics show no remarkable gain by increasing this number and it captures 65% of intercluster data variability. Cluster grouping based on TAZ-EMEF zone-related data can be seen in Fig. 11. The first factorial axis quantifies service availability on the positive side and good accessibility to Barcelona municipality. Cluster 9 (at the right of axis 1 corresponds to the historic city of Barcelona (many services for tourists and residential areas with a high percentage of immigration); it is a multivariate outlier. Cluster 8 (13 zones) corresponds to the rest of the districts of Barcelona's municipality and closest municipalities (L'Hospitalet, Badalona, etc.) containing many services and commerces for tourists and residents of the whole BMR, high job density). Increasing values of fragmentation indicators dominate the second principal axis and correspond to clusters 3 (22 zones) and 5 (10 zones) aligned to increasing values of accessibility by public transport assigned to BMR municipalities. Cluster 7 (25 zones) collects the primary crown (BMA), excluding Barcelona city and BMR municipalities well-connected to Barcelona city. Clusters 1 (16 zones) and 2 (3 zones) show low values of fragmentation indicators and belong to small municipalities in the BMR with a relatively weak relation to Barcelona city. Cluster 6 (19 zones) contains municipalities with high income and low diversity (Sant Cugat, Sant Joan Despí). Finally, cluster 4 (49 zones) consists of BMR municipalities (excluding BMA).

Hierarchical clustering has been translated into geographical terms, assigning TAZ colors according to the cluster to which they belong provides the map of Fig. 12. Missing data corresponds to municipalities not appearing in the EMEF surveys (thus, no fragmentation indicators can be computed). It is remarkable the exceptions to the general pattern described above:

- A green TAZ-EMEF zone (cluster 7) inside cluster 8 (Barcelona City, Santa Coloma, and Badalona). It is Sant Adrià del Besòs a very low-income residential

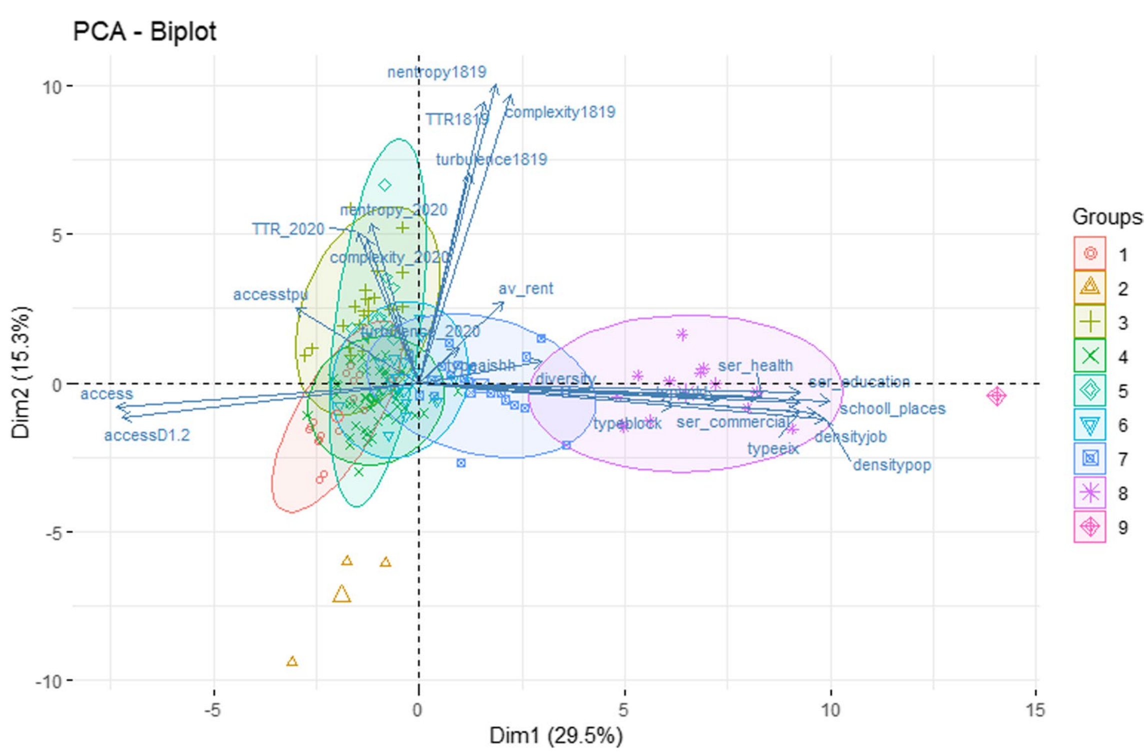

Fig. 11 Hierarchical clustering of TAZ-EMEF zones based on build environment and fragmentation indicators

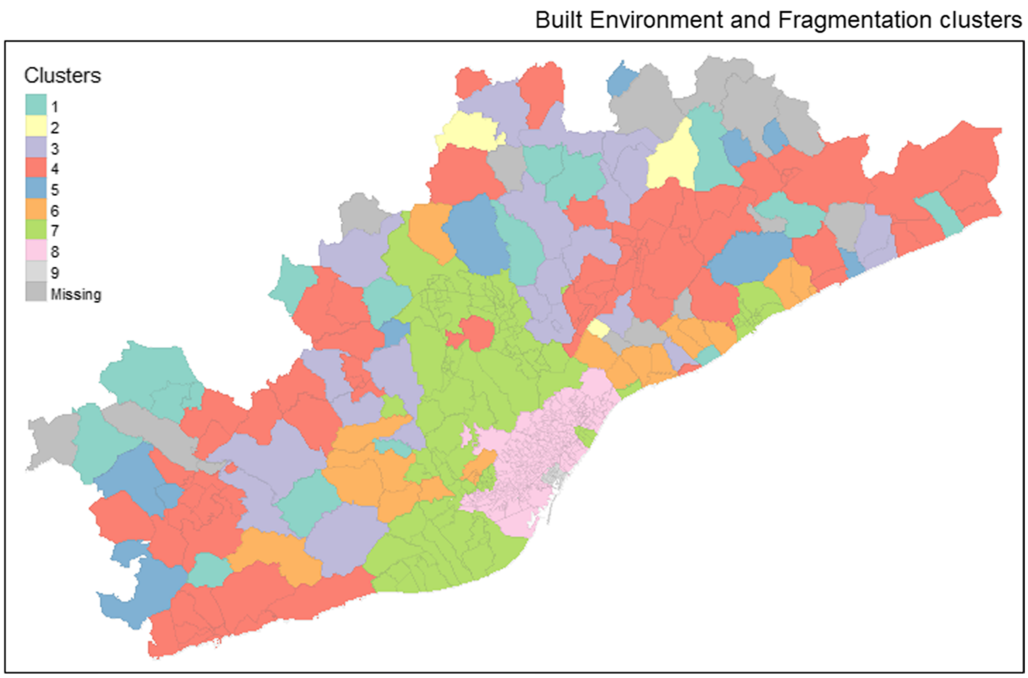

Fig. 12 Map of clusters with fragmentation variables

area with poor communication to the city. Montgat on the north coast also belongs to this cluster.

- Cluster 6 (orange) contains one zone in the northwest of Cluster 8 (pink) corresponding to Sant Just Desvern (high income and isolated houses in a large area).
- Cluster 4 (red) contains three zones in the middle of Cluster 7. It corresponds to Sabadell (the capital of the county) and two adjacent municipalities with high diversity and access time over the mean.

As a conclusion, the average income per capita is directly related to fragmentation indicators in 2018–2019. Access time by public transport and average income per capita are variables directly related to fragmentation indicators. The south coast of Barcelona is grouped into the same cluster, spatial continuity can be seen in a large proportion of the BMR.

8 Discussion and conclusions

Addressing fragmentation indicators and their potential connection to modal preferences in one of the ten largest European regions is notably relevant. This paper focuses on conducting an activity analysis across days in the BMA. Our research draws inspiration from various studies, such as Abbott and Tsay [5] and McBride et al. [40], among others. However, it should be noted that our study is conducted in a completely distinct context, the Metropolitan Area of Barcelona, which is characterised by unique spatial and temporal behaviour of transport demand, as evidenced by the underlying socioeconomic characteristics, transport-oriented development (TOD), urban structure, and variety of activity alternatives and services.

The study of daily activity sequences involves examining a person's entire daily trajectory of activities, accounting for specific locations and associated trips. This analysis considers various factors, including the number of activities and trips, their sequential order, and their durations.

Although comparing our results with those of other studies would provide valuable insights, direct comparisons are unfortunately limited due to differences in the available datasets. While Barcelona's dataset is larger than others, it is based on individual-level data rather than household-level data, which imposes certain constraints on the exploratory data analysis and multivariate analysis involving fragmentation indicators. Nevertheless, the Barcelona dataset offers a wealth of information specifically related to travel details, encompassing various modes of transportation (public, private, and others), which allows us to define a more detailed alphabetic list consisting of 10 activities and transitions between them. This represents a novel contribution. Furthermore, the dataset covers data collected over a minimum period of one month for three consecutive years, capturing year-to-year variations specifically for working days. These characteristics are particularly pertinent in addressing gender-related issues.

The analysis conducted in this study reveals significant differences in activity distributions according to gender, residential area, and year. The examination of the marginal effect of age demonstrates that fragmentation indicators among young individuals are similar regardless of gender. However, significant gender disparities in mobility patterns emerge after the age of 30. The fragmentation of activities is influenced by individual life courses, with gender playing a significant role, even in a TOD city like Barcelona. The life course perspective proves crucial in understanding the dynamics of how, when, and with whom people move, and this is particularly relevant for women.

Our study also examines the relationship between spatial distribution and transport mode choices. For example, we observe a notable impact of residential areas on the percentage of private transport usage. The outer areas of Barcelona exhibit a higher prevalence of private transport usage compared to the central part of the city, whereas public transport and active modes are more prominent. Furthermore, women tend to prefer public transport, while men more frequently opt for private transport. The year 2020 shows significant deviations from previous years, with home activities gaining considerable importance and a noticeable decrease in the use of public transport compared to previous years. If we are to understand the ongoing evolution of mobility behaviors, it is necessary to analyse post-2020 mobility patterns.

In the existing literature, the most common approaches conduct an a priori dimensionality reduction by defining short time intervals. However, as shown in this paper, this method can lead to substantial information loss and potential biases. To mitigate these risks, we have adopted a different approach in this study. In computing the fragmentation indicators, we preserved the full dimensionality of the minute-by-minute time-based data. Later, we employed a dimensionality reduction method based on the data analytic technique MCA, which is well-suited for reducing dimensions while maintaining control over information loss. This new dataset allowed us to conduct a more detailed clustering analysis that revealed the differences between clusters overrepresented by women and by men.

Our analyses of daily activity sequences reveal distinct segments of citizens, each with specific needs that cannot be disregarded by urban planners. Our proposed approach, which involves segmenting based on daily sequence analysis of travel and activity patterns, serves as an accurate indicator for transport planning. This method enables the formulation of targeted policies for each segment, taking into account their representation within the population to quantify policy coverage.

The addition of complementary aggregated information at the TAZ level, on demographic, socioeconomic, variables, on characteristics of the built environment, supplemented with the fragmentation indicators aggregated at the TAZ level, enabled an additional analysis of their spatial impacts.

The identified relationships between people's mobility, the individual life course, gender equity, and the built environment indicate the need for further detailed analysis. Such analysis is crucial not only for understanding these relationships but also for translating this understanding into appropriate modelling approaches for effective transport and city planning.

Finally, future research should prioritize a detailed analysis of mobility behaviour, utilizing new datasets and additional information that was not available for this study. A comprehensive examination of all clusters and daily mobility patterns should be undertaken to gain deeper insights. Undoubtedly, an analysis of the urban environment at a finer spatial scale is necessary.

Acknowledgements
The authors would like to express their sincere gratitude to the Autoritat del Transport Metropolità (ATM) for generously sharing the datasets used in this study. Their contribution to our research is greatly appreciated.

Author contributions
Conceptualization, L.M., L.M.D. and J.B.; formal analysis, L.M., L.M.D. and J.B.; funding acquisition, L.M. and J.B.; methodology, L.M., L.M.D. and J.B.; software, L.M.; supervision, J.B.; writing original draft, L.M.; writing—review and editing, L.M.D. and J.B. All authors have read and agreed to the published version of the manuscript.

Funding
This research was funded by PID2020-112967GB-C31 Spanish R+D Programs and by Secretaria d'Universitats-i-Recerca-Generalitat de Catalunya- 2021 SGR 01252 Information Modeling and Processing.

Availability of data and materials
The authors declare that they have obtained permission to use the data from the ATM authorities (Autoritat del Transport Metropolità) for this research. The availability of data should be coordinated directly with the ATM.

Declarations

Competing interests
The authors declare no conflict of interest.

Received: 21 October 2022 Accepted: 14 November 2023
Published online: 27 November 2023

References

1. Abbott, A. (1983). Sequences of social events: concepts and methods for the analysis of order in social processes. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 16*(4), 129–147. https://doi.org/10.1080/01615440.1983.10594107
2. Abbott, A. (1984). Event sequence and event duration: colligation and measurement. *Historical Methods: A Journal of Quantitative and*

Interdisciplinary History, 17(4), 192–204. https://doi.org/10.1080/01615440.1984.10594134
3. Abbott, A., & DeViney, S. (1992). The welfare state as transnational event: evidence from sequences of policy adoption. *Social Science History, 16*(2), 245. https://doi.org/10.2307/1171289
4. Abbott, A., & Forrest, J. (1986). Optimal matching methods for historical sequences. *Journal of Interdisciplinary History, 16*(3), 471. https://doi.org/10.2307/204500
5. Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: review and prospect. *Sociological Methods & Research, 29*(1), 3–33. https://doi.org/10.1177/0049124100029001001
6. Alexander, B., Ettema, D., & Dijst, M. (2010). Fragmentation of work activity as a multi-dimensional construct and its association with ICT, employment and sociodemographic characteristics. *Journal of Transport Geography, 18*(1), 55–64. https://doi.org/10.1016/j.jtrangeo.2009.05.010
7. AMB. (2023). Metropolitan Area of Barcelona. Website. https://www.ambmobilitat.cat/Principales/MediosTransporte.aspx?idioma=3
8. Bhat, C. R., Konstadinos-Goulias, G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., Hu, H.-H., Bhat, C. R., Sidharthan, Á. R., Sidharthan, R., Goulias, K. G., Pendyala, R. M., Paleti, R., Brinckerhoff, P., Plaza, P., Schmitt, L., & Hu, H.-H. (2013). A household-level activity pattern generation model with an application for Southern California. *Transportation, 40*(5), 1063–1086. https://doi.org/10.1007/S11116-013-9452-Y
9. Bhat, C. R., & Pinjari, A. R. (2007). Duration modeling. In: *Handbook of transport modelling* (pp. 105–131). Emerald Group Publishing. https://doi.org/10.1108/9780857245670-006
10. BMDV. (2020). Mobility in Germany. https://www.bmvi.de/EN/Services/Statistics/Mobility-in-Germany/mobility-in-germany.html
11. Burchell, B., Reuschke, D., & Zhang, M. (2020). Spatial and temporal segmenting of urban workplaces: the gendering of multi-locational working. *Urban Studies, 58*(11), 2207–2232. https://doi.org/10.1177/0042098020903248
12. Couclelis, H. (2000). From sustainable transportation to sustainable accessibility: Can we avoid a new (pp. 341–356). https://doi.org/10.1007/978-3-662-04027-0_20
13. Couclelis, H. (2003). Housing and the new geography of accessibility in the information age. *Open House International, 28*(4), 7–13.
14. Couclelis, H. (2006). Pizza over the Internet: E-commerce, the fragmentation of activity and the tyranny of the region. *Entrepreneurship & Regional Development, 16*(1), 41–54. https://doi.org/10.1080/0898562042000205027
15. Crane, R. (2007). Is there a quiet revolution in women's travel? Revisiting the gender gap in commuting. *Journal of the American Planning Association, 73*(3), 298–316. https://doi.org/10.1080/01944360708977979
16. Cresswell, T. (2008). Gendering mobility: Insights into the construction of spatial concepts. In T. Uteng & T. Cresswell (Eds.), *Gendered Mobilities* (pp. 97–112). Routledge. https://doi.org/10.4324/9781315584201-12
17. Cubells, J., Marquet, O., & Miralles-Guasch, C. (2020). Gender and age differences in metropolitan car use. Recent gender gap trends in private transport. *Sustainability, 12*(18), 5. https://doi.org/10.3390/SU12187286
18. Dijst, M., & Vidakovic, V. (2000). Travel time ratio: The key factor of spatial reach. *Transportation, 27*(2), 179–199. https://doi.org/10.1023/A:1005293330869/METRICS
19. EIGE. (2019). Browse Gender Statistics | Gender Statistics Database | European Institute for Gender Equality. What Lies behind the Gender Pay Gap? https://eige.europa.eu/gender-statistics/dgs/data-talks/what-lies-behind-gender-pay-gap
20. EIGE. (2022). European Institute for Gender Equality. https://eige.europa.eu/
21. Elzinga, C. H., & Liefbroer, A. C. (2007). De-standardization of family-life trajectories of young adults: A cross-national comparison using sequence analysis. *European Journal of Population, 23*(3–4), 225–250. https://doi.org/10.1007/S10680-007-9133-7/FIGURES/2
22. Elzinga, C. H., & Studer, M. (2019). Normalization of distance and similarity in sequence analysis. *Sociological Methods & Research, 48*(4), 877–904. https://doi.org/10.1177/0049124119867849
23. Eurostat. (2020). How do women and men use their time - statistics - Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=How_do_women_and_men_use_their_time_-_statistics&oldid=463738#Overview https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Participation_time_per_day_in_household_and_f
24. EUROSTAT. (2022). Working from home in the EU. Products Eurostat News. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20180620-1
25. EWCS. (2015). Sixth European Working Conditions Survey: 2015. Eurofound. https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys/sixth-european-working-conditions-survey-2015
26. Fana, M., Tolan, S., Torrejón Pérez, S., Urzi Brancati, M. C., & Fernández-Macías, E. (2020). The COVID confinement measures and EU labour markets. *JCR Technical Reports*. https://doi.org/10.2760/079230
27. Gabadinho, A., Ritschard, G., Studer, M., & Müller, N. (2010). Indice de complexité pour le tri et la comparaison de séquences catégorielles - Editions RNTI. *Revue Des Nouvelles Technologies de l'Information*, *Extraction*, 61–66. https://editions-rnti.fr/?inprocid=1001267
28. Gabadinho, A., Ritschard, G., Studer, M., & Müller, N. (2011). *Mining sequence data in R with the TraMineR package: A user s guide* (1.8; p. 129). https://docplayer.net/7853982-Mining-sequence-data-in-r-with-the-traminer-package-a-user-s-guide-1.html
29. Giele, J., & Elder, G. (1998). Methods of life course research: qualitative and quantitative approaches. In *Methods of life course research: Qualitative and quantitative approaches*. SAGE Publications, Inc. https://doi.org/10.4135/9781483348919
30. Goulias, K. (2020). *An Analysis of Accessibility, Social Interaction, and Activity-Travel Fragmentation in California in METRANS | Research Projects*. https://www.metrans.org/research/an-analysis-of-accessibility-social-interaction-and-activity-travel-fragmentation-in-california
31. Goulias, K. G., McBride, E. C., & Su, R. (2020). Life cycle stages, daily contacts, and activity-travel time allocation for the benefit of self and others. *Mobility and Travel Behaviour Across the Life Course*. https://doi.org/10.4337/9781789907810.00023
32. Hanson, S. (2010). Gender and mobility: New approaches for informing sustainability. *Gender, Place and Culture, 17*(1), 5–23. https://doi.org/10.1080/09663690903498225
33. Hubers, C., Schwanen, T., & Dijst, M. (2008). ICT and temporal fragmentation of activities: An analytical framework and initial empirical findings. *Tijdschrift Voor Economische En Sociale Geografie, 99*(5), 528–546. https://doi.org/10.1111/j.1467-9663.2008.00490.x
34. Husson, F., Lê, S., & Pages, J. (2010). *Exploratory multivariate analysis by example using R Analysis*. CRC Press.
35. Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *The American Psychologist, 74*(2), 171–193. https://doi.org/10.1037/AMP0000307
36. INE. (2021). *Posibilidad de teletrabajar y teletrabajo efectivo, de forma total o parcial, por características demográficas.* Instituto Nacional de Estadística. https://datos.gob.es/es/catalogo/ea0010587-posibilidad-de-teletrabajar-y-teletrabajo-efectivo-de-forma-total-o-parcial-por-caracteristicas-demograficas-identificador-api-t25-p450-base_2011-a2021-l0-04055-px
37. Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141–151.
38. Leszczyc, P. T. L. P., & Timmermans, H. (2002). Unconditional and conditional competing risk models of activity duration and activity sequencing decisions: An empirical comparison. *Journal of Geographical Systems, 4*(2), 157–170. https://doi.org/10.1007/S101090200083
39. McBride, E., Davis, A., & Goulias, K. (2019). Fragmentation in daily schedule of activities using activity sequences. *Transportation Research Record, 2673*(4), 844–854. https://doi.org/10.1177/0361198119837501
40. McBride, E., Davis, A., & Goulias, K. (2020). Exploration of statewide fragmentation of activity and travel and a taxonomy of daily time use patterns using sequence analysis in California. *Transportation Research Record, 2674*(12), 38–51. https://doi.org/10.1177/0361198120946011
41. Mejía-Dorantes, L., Montero, L., & Barceló, J. (2021). Mobility trends before and after the pandemic outbreak: analyzing the metropolitan area of barcelona through the lens of equality and sustainability. *Sustainability*. https://doi.org/10.3390/su13147908
42. OMC. (2021). *Working day mobility survey (EMEF) of the ATM of the Barcelona area*. Website. https://omc.cat/en/w/surveys-emef
43. Paleti, R., Vovsha, P., Vyas, G., Anderson, R., & Giaimo, G. (2017). Activity sequencing, location, and formation of individual non-mandatory

tours: Application to the activity-based models for Columbus, Cincinnati, and Cleveland, OH. *Transportation, 44*(3), 615–640. https://doi.org/10.1007/S11116-015-9671-5/FIGURES/2
44. Rasouli, S., & Timmermans, H. (2014). Activity-based models of travel demand: Promises, progress and prospects. *International Journal of Urban Sciences, 18*(1), 31–60. https://doi.org/10.1080/12265934.2013.835118
45. R Development Core Team R. (2021). The R Project for Statistical Computing. The R Foundation: Ames, IA, USA.
46. Ritschard, G. (2021). Measuring the nature of individual sequences. *Sociological Methods & Research*. https://doi.org/10.1177/00491241211036156
47. Roselló, X. (2010). *Descripció de les enquestes de mobilitat a catalunya. Procediment de fusió de matrius de mobilitat per obtenir la matriu 2006 general. (In Catalan)*. Technical Note. https://doc.atm.cat/ca/_dir_notestecniques/2010-11-enquestes-a-catalunya.pdf
48. Studer, M., & Ritschard, G. (2016). What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures on JSTOR. *Journal of the Royal Statistical Society A, 179*(Part 2), 481–511.
49. Su, R., McBride, E. C., & Goulias, K. G. (2020). Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. *Transportation Research Part C: Emerging Technologies, 120*, 102796. https://doi.org/10.1016/J.TRC.2020.102796
50. Tukey, J. W., Brillinger, D. R., Cox, D. R., & Braun, H. I. (1984). The Collected Works of John W. Tukey. Wadsworth Advanced Books & Software: Belmont, CA, USA.
51. von Behren, S., Hilgert, T., Kirchner, S., Chlond, B., & Vortisch, P. (2020). Image-based activity pattern segmentation using longitudinal data of the German Mobility Panel. *Transportation Research Interdisciplinary Perspectives, 8*, 100264. https://doi.org/10.1016/j.trip.2020.100264

Publisher's Note