Image-based activity pattern segmentation using longitudinal data of the German Mobility Panel

Sascha von Behren [a,*], Tim Hilgert [b], Sophia Kirchner [a], Bastian Chlond [a], Peter Vortisch [a]

[a] Institute for Transport Studies, Karlsruhe Institute of Technology (KIT), Kaiserstrasse 12, 76131 Karlsruhe, Germany
[b] INOVAPLAN GmbH, Degenfeldstrasse 3, 76131 Karlsruhe, Germany

ARTICLE INFO

ABSTRACT

In this paper, we present an approach to segment people based on a visualization of the longitudinal week activity data from the German Mobility Panel. In order to perform segmentations, different clustering methods are commonly used. Most of the approaches require comprehensive prior knowledge about the input data, e.g., condensing information to cluster-forming variables. As this may influence the method itself, we used images with a high degree of freedom. These images show week activity schedules of people, including all trips and activities with their purposes, modes as well as their duration or their temporal position within the week. Thus, we answer the question whether using only this type of image data as input will produce reasonable clustering results as well. For the clustering, we extracted the images from an existing tool, processed them for the method and finally used them again to select the final cluster solution based on the visual impression of cluster assignments. Our results are meaningful as we identified seven activity patterns (clusters) using this visual validation. The approach is confirmed by the data-based analysis of the cluster solution showing also interpretable key figures for all patterns. Thus, we show an approach taking into account many aspects of travel behavior as an input to clustering, while ensuring the interpretability of solutions. Usually, key figures from the data are used for validation, but this practice may obscure some aspects of the longitudinal data, which are visible when looking on the images as validation.

1. Introduction

Segmentation aims to allocate people into homogeneous groups showing a certain behavior or attitude due to common influencing factors. Thus, we generally assume that people of the same group can be influenced in the same way by external interventions. This approach has spread from market research to other disciplines and has been established in transportation research for a long time now. With different mobility concepts and services, providers can address the mobility needs of people in the segments. Such segmentations can provide various insights, e.g., into new technologies such as user characteristics of electric mobility.

A key challenge in segmentation relates to the input variables used for allocation. They significantly determine the formation of homogenous groups, often known as clusters. Existing literature on travel behavior research shows a wide range of approaches with different dimensions of input variables for segmentation (e.g., Schlich, 2004; Wittwer, 2014). These cluster-forming input variables are crucial in

this context. Most previous studies on mobility have selected a set of variables for later clustering through preliminary analysis (e.g., principal component analysis) and targeted selection. Targeted implies that the selection of variables is also strongly dependent on the desired target segments of the respective study. For example, if one intends to study multimodality segments, many variables for transport mode choice are selected from the reported behavior (e.g.,Oostendorp et al., 2019). In addition, the complexity increases with the number of variables in the clustering process and the curse of dimensionality arises. This also affects the following analysis and interpretation of the clusters.

In the existing literature, an approach to represent complex longitudinal mobility of people with many different aspects (e.g. trips, activities, distance, time) in a simple way in order to use it for clustering is missing. Such an approach may take into account many aspects of travel behavior as an input to clustering, while ensuring the interpretability of solutions. Wittwer (2014) emphasizes that a cluster solution is only useful if the cluster solution can be meaningfully

---

interpreted on the basis of the underlying input variables. Therefore the interpretability of the resulting cluster is a crucial aspect to be considered in the clustering procedure. In addition, most segmentation studies do not use panel data (i.e. both repeated and longitudinal). This results in a lack of intrapersonal comparisons at different points in time, both in terms of the day-to-day-variability as well as between years. Consequently, it is not possible to verify in such studies how stable the clustering solution and thus the behavior of an individual is. To close this gap, we formulated four main objectives for our study:

- Generating behaviorally homogeneous groups from longitudinal data over one week by considering many aspects of travel behavior (e.g. time, activities, trips, mode).
- Using an approach to simplify the presentation of travel behavior of individuals over one week with a high degree of freedom as input for segmentation of activity patterns.
- Providing an additional visual aid for interpreting cluster solutions.
- Evaluating stability of the cluster allocation over time.

To achieve these objectives, we used data from the longitudinal travel survey German Mobility Panel (MOP). Our clustering approach relies on images visualizing all activity and trip information of a whole week, taken from the plausibility checking tool "Graphical diagnosis of individual travel behavior" (GraDiV) which is especially designed for the MOP. The tool visualizes the out-of-home activity and travel behavior of people over a whole week from their reported trip diaries at a glance in so called GraDiV images. Persons with similar behavior within a week should be identified with their images. GraDiV images offer a suitable basis for the segmentation of behavior, as they show the necessary information over the period of one week. For our study, we used 9,062 travel diary images of 5,807 people from the MOP from the years 2016 to 2019. Due to the panel approach and thus repeated participations in consecutive years, different travel diaries of one person appear. In order to identify different groups, we first prepared and unified the GraDiV images. Based on the prepared images we applied an unsupervised machine learning procedure for clustering. Using this numerical method for segmentation, we objectified the subjective intuitive segmentation of the images by simply looking at them. As a strength of the approach, we can validate the resulting cluster solutions based on the images and the data. At the same time, the use of images with multiple aspects of mobility allows a high degree of freedom for the clustering process and avoids a selection of input variables. As a result, we obtain segments of similar images reflecting the individual activity behavior and mobility of people in the segment. Each cluster represents an activity pattern of a group of people.

The paper is structured as follows: First, we give an overview of existing segmentation approaches in the literature. Second, we describe the data used and the GraDiV tool. Third, we explain the methodology of our analysis. This includes the pre-processing as well as the clustering. Fourth, we evaluate the clustering result through the GraDiV images and the data in the received segments. Finally, we draw a conclusion, discuss the limits of our approach and refer to further work.

2. Segmentation approaches in transportation research

In general, two approaches are widely used for segmentation in transportation research: *a priori* and *post hoc* segmentation (Anable, 2005). A range of methods are summarized in the respective directions. The *a priori* segmentation is characterized by a detailed definition of the expected groups and their characteristics based on comprehensive prior knowledge and assumptions. Subsequently, the persons are assigned to the corresponding groups based on the characteristics in a first step. *Post hoc* segmentation uses statistical methods instead of comprehensive prior knowledge and assumptions to determine the groups and their characteristics from the available empirical data. From a scientific perspective, *post hoc* segmentation gives the researcher the advantage that the data itself forms the basis for segmentation. Nevertheless, the *post hoc* method requires the selection of input variables and thus prior knowledge for the procedure as well. The literature shows a large number of various studies with different approaches and target segments. All of them have in common that they attempt to choose an optimal selection of variables for the clustering process. We can divide *post hoc* segmentation in transportation research into three main categories relating to their content: external (geographical, socio-demographic), behavioral and psychological factors. The research presented in our study is a small selection from the existing literature showing only the most relevant aspects.

Segmentation based on *external factors* follows the assumption that people in similar life situations and with similar restrictions exhibit similar behavior. Salomon and Ben-Akiva (1983) used a k-means clustering to identify life-style groups for travel demand models. For their segmentation, they included household structure, income and education level. Hildebrand (2003) applied a *post hoc* segmentation based on socio-demographic characteristics in order to subsequently investigate the travel behavior of the determined groups. Further exemplary studies in this field are by Schöppe & Förschner (1983); Heuwinkel (1981); Kunert (1994).

Behavioral approaches play an important role in transportation research. People are surveyed to analyze trips or activities in an observed time period. For the segmentation, revealed behavior is captured through travel diaries or specific questionnaires. A travel diary includes many aspects of mobility. It provides information about trips such as purpose, start time, end time, distance and duration. In addition to trip information, activities and their duration are also covered indirectly in the diaries by the trip purposes and the time position of the trips. As a result, trips and activities can be included in the segmentation process. In particular, longitudinally oriented travel diaries over one week or more provide detailed information about individual travel behavior over a longer time period. This makes such surveys well suited for segmenting people, since a greater variance can be taken into account. In general, the aim of such behavior-oriented segmentation is to identify behaviorally homogeneous groups. Fundamental work in behavioral segmentation was done by Pas (1980); Hanson and Huff (1986); Schmiedel (1984). More recent approaches can be found in, Lipps (2001); Maat and Arentze (2003); Berger (2004); Schlich (2004); Wittwer (2014); Oostendorp et al. (2019); Niklas et al. (2020a); (2020b;). The challenge in the above-mentioned studies also lies in the selection of variables and the associated condensation of information for the segmentation process. The work of Schlich (2004) serves as an example for many other studies. Schlich chose 14 behavior-oriented variables from the longitudinal data of the *Mobidrive* survey for the consecutive segmentation process. The selection of input variables includes type of activity, time of activity (e.g., share of trips at the weekend), trip information (duration and distance), means of transport and frequency of activities. In other studies, information is pre-compressed through available techniques to reduce complexity (e.g., principal component analysis). This reduces variance and information is consequently lost in the clustering process. Before clustering, Wittwer (2014) selected 25 variables on travel behavior and condensed them into eight components. As a result, he received six clusters about young adults and their essential (i.e. necessary) mobility. Both studies also used a large set of data to describe and interpret the clustering solutions. However, no visual support for the interpretation of the clusters was given (e.g., visualization of the homogeneous behavior in the resulting clusters). Other studies focus only on the reported activities. Ectors et al. (2016) used a one-day travel diary survey from Flanders, Belgium to segment similar activity schedules. It was a completely data-driven approach to reveal the basic structure of individuals' schedules for one day, i.e. the skeleton schedule sequence. Allahviranloo et al. (2014) divided daily activities into

ten-minute intervals to segment activity into patterns carrying information about activity types, duration, schedule and travel distance of one-day travel diary data from the California Household Travel Survey. The objective was to forecast activity patterns. A different approach is from Zhao et al. (2015). They examined the day-to-day variability based on smartphone data of people from Singapore over an average period of ten days. They divided the daily activity patterns into 5-minute slots and converted each user day into a vector of 288 elements. On this basis they clustered the activity patterns of workdays. As a result, they received five workday patterns. For their analysis they used a visualization of the different days. The intrapersonal comparison with regards to day-to-day variability shows that 76% of the people end up in at least three clusters. These results are combined with data from a cross-sectional household travel survey (HTS) of employees over only one typical weekday. 93% of the people from the HTS correspond to one cluster. The authors therefore recommend to record more than one day of the person to capture intrapersonal variability. However, the authors did not examine week patterns, their focus was on the comparison of day patterns.

To complete the overview, we also present the *psychological* approaches briefly. In *post hoc* segmentation with statistical methods, attitudes or norms of individuals must be converted into concrete numerical values. For this purpose, psychological item sets on Likert scales are used in the survey. This was applied, for example, in studies by Hunecke et al. (2010); Anable (2005); Prillwitz and Barr (2011); Collum and Daigle (2015). Anable (2005) performed a cluster analysis in order to segment people in terms of their potential for change in their mode choice. The 17 psychological factors as input variables for the cluster analysis were determined by asking 105 psychological items on a 5-point Likert scale. There are also *hybrid* approaches in the literature where psychological data was mixed with behavioral data for clustering (e.g., Magdolen et al., 2019; von Behren et al., 2018). Magdolen et al. (2019) have used 6 psychological factors and 4 behavioral variables (trips per day, car share, long-distance trips and share of mandatory activities) to identify urban mobility types in the segmentation process. The disadvantage of psychological segmentation is the lack of availability of public data. In national household travel surveys (NHTS), the focus is on travel behavior. Only a few NHTS such as the Netherlands Mobility Panel (Hoogendoorn-Lanser et al., 2015) cover psychological factors in subsamples, otherwise the respondent burden increases.

The approaches presented show quite well the role of the selection and pre-compaction of input variables for segmentation in transportation research. Previous studies have taken the approach of compressing a large amount of different information into a few variables and thus decreasing the degree of freedom by pre-compacting the information. Since the cluster analysis procedure leaves a lot of influence to the users, they can easily be accused of manipulating the data to achieve the desired results (Götz et al., 1998). A further challenge is the interpretability of clustering solutions. Statistical methods provide indications such as homogeneity within the clusters and the differences between the clusters. Resulting in several possible clustering solutions, the main focus is usually on the interpretability of the solution. For this purpose, the evaluation through the data of the cluster-forming and cluster-describing variables remains. Statistical parameters such as mean values, standard deviations or distributions in the clusters are often used for evaluation and interpretation. A suitable visual support to assess interpretability of the behavior is missing.

Using GraDiV images as visualizations of activity behavior and travelling can capture multiple aspects of behavior and still simplify the interpretation through looking on images. Götz et al. (1998) stresses that the accusation of excessive influence or data manipulation in the segmentation of travel behavior must be countered only by transparency of the selection and interpretation process. To this aim, the GraDiV images can be used to create maximum transparency in the process, especially for the interpretation. To our knowledge, no

approach is available in the existing literature that uses the representations of a weekly activity schedules as images to segment travel behavior. The use of images makes intra- and intergroup comparisons easier to interpret.

Furthermore, the representation of all out-of-home activities and trips as an image with a defined format, independently from the number of episodes, allows for comparison or segmentation without any preliminary data analysis and without a definition of indicators for a clustering. Therefore, images with the defined format according to a week's time line do not need any preprocessing in terms of input variables, which then enables an unprejudiced segmentation of activity plans.

3. Longitudinal data for clustering

Our approach of clustering weekly activity schedules is based on longitudinal data of the German Mobility Panel. Following, we give deeper information about the study itself and the visualization tool GraDiV used to create the week activity schedule images.

3.1. German Mobility Panel

The German Mobility Panel (MOP) is a national household travel survey that has been conducted each year since 1994. It is carried out on behalf of and funded by the German Federal Ministry of Transport and Digital Infrastructure. The market research firm KANTAR is responsible for the field work (i.e., recruitment and data collection) and the Institute for Transport Studies of the Karlsruhe Institute of Technology (KIT) is in charge of the survey's design and scientific supervision. The data collection of the MOP takes place in autumn every year and the survey weeks are meant to not contain any school or bank holidays ("everyday travel"). The participants are asked to fill in a trip diary for one week. The diary provides information about all trips during this week (distances, means of transport, trip purposes and departure and arrival times). Participants also indicate whether each day was typical or non-typical, for example, whether they were ill or on holiday. Furthermore, sociodemographic information about the participants and the availability of cars, bicycles and transit passes is surveyed.

The overall sample size is 1,500–1,800 households with 2,600–3,100 persons (aged ten years and older) reporting each year. The MOP is designed as a rotating panel meaning that the participants are asked to report their travel behavior for up to three consecutive years. Every year a new cohort of first year reporters replaces a portion of the sample that retires. Besides these planned replacements also unplanned dropouts occur.

3.2. Graphical diagnosis of individual travel behavior (GraDiV)

The images used for the work in this paper are part of a plausibility tool called GraDiV. It has been developed by the Institute for Transport Studies of the KIT as part of the scientific supervision of the MOP (see previous section). Using this tool is part of the quality check process of the MOP. GraDiV is used for individual plausibility checks of each participant of the survey (e.g., check for completeness, identification of incompleted trip chains etc.). For this purpose, the tool prepares the reported data in two ways: as a list in alpha-numerical form and as an image.

The list contains all trips and corresponding activities during the week in chronological order. All raw information reported from the participant such as start and end time of the trip or purpose of the activity etc. are displayed, one row per reported trip. At the same time, the same data is formatted as image. As shown in the example of such an image (see Fig. 1), it displays the week activity schedule (one row for each day) at a glance, i.e. all trips (bars shifted up) and activities
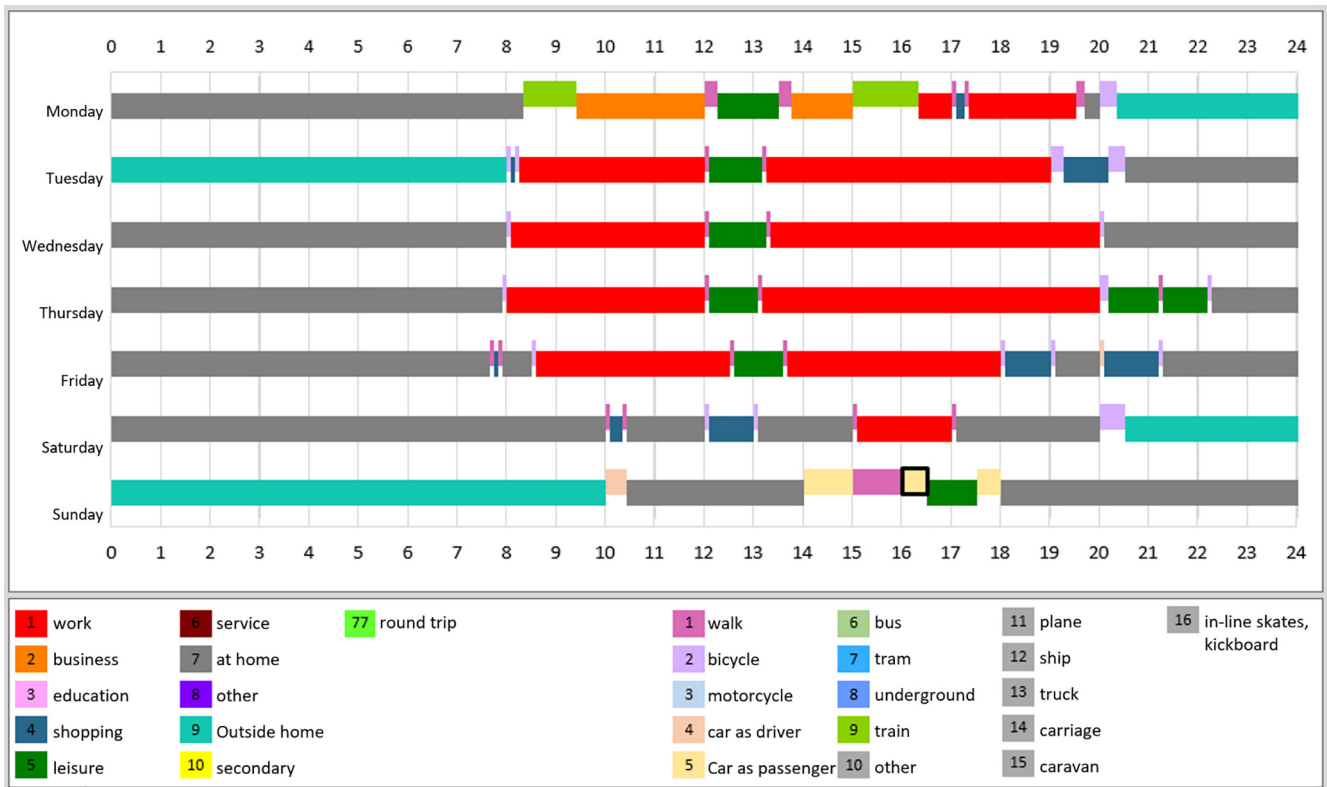
Fig. 1. Visualization of a week activity schedule as GraDiV image.

(bars shifted down) in chronological order for each day. While checking the data, the image is formatted at run-time, thus any adapted or corrected values in the chronological list will be directly reflected in an adapted image. Having a run-time visual "implementation" of the whole week activity schedule is probably the most powerful and important feature of GraDiV and a great help while checking for irregularities. After a short training phase, the images are directly and completely understandable as well as interpretable at a glance. Instead of checking the plausibility and completeness of information of all trips in a list with mainly alpha-numerical information in sequence, the visualization allows a fast and complete understanding as well as the identification of implausible or obviously wrong data. Using the chronological list and the image together, irregularities and minor faults can easily be identified and corrected. On the one side several types of mistakes are directly marked in the list (e.g., typing errors of departure or arrival times or implausible speeds) while others are noticeable only by looking at the images (e.g., uncompleted trip chains not returning home again) but are hard to detect directly in a list as they appear as a visual impression. The staff of the institute is instructed to check for certain cases, e.g. identification of circular trips such as walking the dog or identification and completion of missing values based on data of other reported days. Additionally, to simplify the process, participants of the same household can be visualized at the same time, allowing the identification of joint trips and activities.

Although the tool is mainly used for plausibility checks of the reported data, we can use the visualization of the week activity schedules for extended analysis as well. For this paper, we made some adaptions to the tool to be able to export the images of the final week activity schedules for further processing.

4. Methodology

In the following section we will go through the methodology of our *post hoc* approach. As shown in the following flowchart (see Fig. 2), we performed different steps in order to cluster the week activity schedule images. We first exported images from the GraDiV tool and then ran through several steps of the pre-processing. We then applied different types of cluster analyses to the data. The cluster analysis itself contained only the image data and no further descriptive information, i.e., no other prior knowledge. By means of the cluster analysis we received solutions which we afterwards validated by comparisons of the different images of all week activity schedules belonging to one of the clusters (see Fig. 4). Thus, our result clusters are based on both computed solutions that are also visually coherent.

4.1. Pre-processing GraDiV images

One central task for image processing and clustering consists in the condensation of the information contained in the images: These are made for human capabilities and human perception and include redundant information (e.g., in terms of size by reasons of perceptibility or as legends and scales such as the weekday and the time axis):
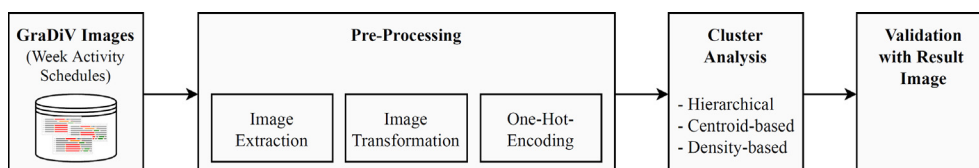

Fig. 2. Flow chart of the applied methodology.

This redundant information is not necessary for the computer, i.e. irrelevant information must be dropped, the remaining information must be filtered. A total of 9,062 GraDiV images from the MOP serve as the information basis, covering the data from a period of three years between 2016 and 2019. Within the *image extraction*, we first performed minor changes in the GraDiV tool for better usability in the following steps. Subsequently we exported the images and excluded all those of persons who showed irregularities during the reporting week, e.g. due to illness or vacation. This left 7,362 GraDiV images in total that we used for the analysis.

Another challenge when processing objects that contain a lot of information is the issue of dimensionality: A crucial aspect for comparison and clustering of images consists in the comparison of each pixel with the same position in other images. Applied to the image of activity behavior of a full week this means that all the pixels of every individual at a defined position will be compared to those at the same position of the other images. This results in a large dimensionality as every pixel can be different from others and dealt with as a single variable. Thus, to achieve a reasonable problem description for computing we needed to reduce the dimensionality. This took place in the *image transformation*, which consisted of several steps:

- First, only the relevant parts of the image were selected. I.e. we dropped all irrelevant information as being identical (e.g. the white areas outside the activity-bars as well as the legends within the images). This resulted in a "relevant size of 239 pixels in height and 606 pixels in length" per individual and week. In this image size each pixel represents approximately 2.4 min.
- Since the objective of the procedures is to find similarities between activity schedules of one week, only a condensed version of the GraDiV image of each day is necessary. Applying this to the image means that one row of pixels (1×606 pixels) is able to represent one day. For a presentation of the whole week activity schedule, which contains the full information, an image size per individual is 7 (days) $\times 606$ (time segments per day).
- Colored images store each pixel three times to represent the primary colors red, green and blue (cf. Beyerer et al., 2012). This enlarges the dimensionality and thus also the storage space as well as the computing efforts by reasons of the colors by factor 3. To ease the clustering there is the need to reduce the dimensionality, i.e. to simplify the problem's size without losing information. Therefore, we computed an average value for each pixel to merge the RGB color space to a single value and normalized it to a value range between 0 and 1.
- The pixel-values were first stored in a 7×606 matrix and then transformed into a vector (sequentially) with a total length of $4,242 (7 \times 606)$, which still representing the image. However, this still results in a high dimensionality. To reduce this and consequently the computing efforts, we decided that approximately 5-minute intervals would be sufficient to represent the activity schedules of complete weeks without losing too much information. We therefore removed every second value ($=$ time segment) to obtain an interval which represents approximately five minutes, finally resulting in $2,121$ dimensions per image of the weekly behavior. The reduction to a 5-minute interval and the representation of the image in one vector is similar to Zhao et al. (2015).

By combining the transformed image vectors of all $7,362$ images this finally resulted in a $7,362 \times 2,121$ matrix. Hereby each row represented the length of a vector of one transformed GraDiV image into the condensed time segments ($=$ dimensionality) and each picture column represented the number of individuals.

Still, the computed average values per pixel color may lead to misleading results in the subsequent clustering when using them as numerical numbers. In a similarity analysis in clustering, some of the average values of the pixel colors are closer to each other than others.

Thus, certain activities are considered more similar than others due to the different numerical distances caused by the pixel colors. This has a strong influence on the clustering solution. However, we decided not to include any prior knowledge and assumptions into the analysis. An assumption could be that, for example, red colored activities (work) are more similar to orange colored activities (business) as to blue colored activities (shopping)). Since we do not want to use a predefined similarity of activities, we treat all activities equally. To avoid the described bias, the so-called *one-hot encoding* (cf. Harris and Harris, 2013) was used. It represents the calculated color values in binary form so that the similarity between two values can either be the same or different but has no specific numeric distance. Therefore, a dimension was introduced for each possible color value a pixel receives. As there exist 14 different colors each pixel is represented by 14 new dimensions. Since each dimension represents one color value, it can take either the value 0 (pixel has not color value X) or 1 (pixel has color value X). The encoding procedure resulted in a fourteen times larger matrix, having $7,362 \times 29,694$ dimensions $(2,121 \times 14 = 29,694)$. Based on this preprocessed image data we were able to continue with the application of the different clustering algorithms.

4.2. Clustering GraDiV images

Clustering is an unsupervised learning method, which is able to explore valuable insights out of data. In general, those algorithms can find clusters where objects within the same cluster are similar to each other but differ from objects of other clusters. Available clustering algorithms can be divided into hierarchical, partitional or density-based clustering. To explore the data set and determining a reasonable number of clusters, we used agglomerative hierarchical clustering algorithms. Agglomerative clustering starts with each object as a single-element cluster. It then forms bigger clusters by merging clusters that are close by distance until finally resulting in one large cluster. The merging can be visualized in a dendrogram, illustrating the arrangement of the clusters as a tree. For this data set, we received the best results with Ward's method, an agglomerative approach, which tends to form clusters of similar size.

The dendrogram (see Fig. 3) shows the merging of the clusters and helps to identify a useful number of clusters in the data. The height hereby represents the distance between the clustered objects. The height of the specific dendrogram node can be thought of as the distance value between the right and left sub-branch clusters. The greater the height, the greater the distance between the merged clusters. By cutting the tree, we can receive different result clusters. There are several possibilities to define the optimal cut. Hereby one should consider the height and the amount of the clusters created by this cut. In this case, a cut at height $h = 500, 300$ or 125 results in $k = 3, 4$ or 7 clusters. The cut with $k = 7$ is illustrated in color in the dendrogram, forming one small, two medium-sized and four large cluster.

As hierarchical clustering does not necessarily define the optimal outcome, we also applied partitional and density-based algorithms in order to minimize the distance between the clustered objects. The partitional clustering method k-means is not able to select a suitable number of clusters for a dataset but it can minimize the intra-cluster variance. Therefore, we used the number of the possible clusters k of the Ward's method and performed the k-means algorithm afterwards. Due to its iterative improvement of the clustering solution it generates an output with lower distances between the clustered objects and therefore optimizes the output of Ward's method. However, this method is not outlier resistant. As it forces to include every object within a cluster this can influence the quality of the resulting clusters. Density-based methods like DBSCAN can handle outliers but couldn't generate overall meaningful results on this specific data set for different reasons: either the algorithms categorized too many images as outliers and generated only small clusters with the rest of the images, or it
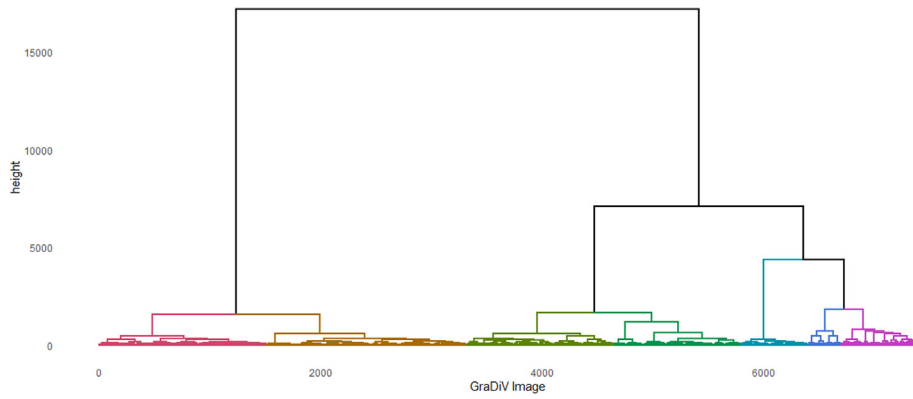
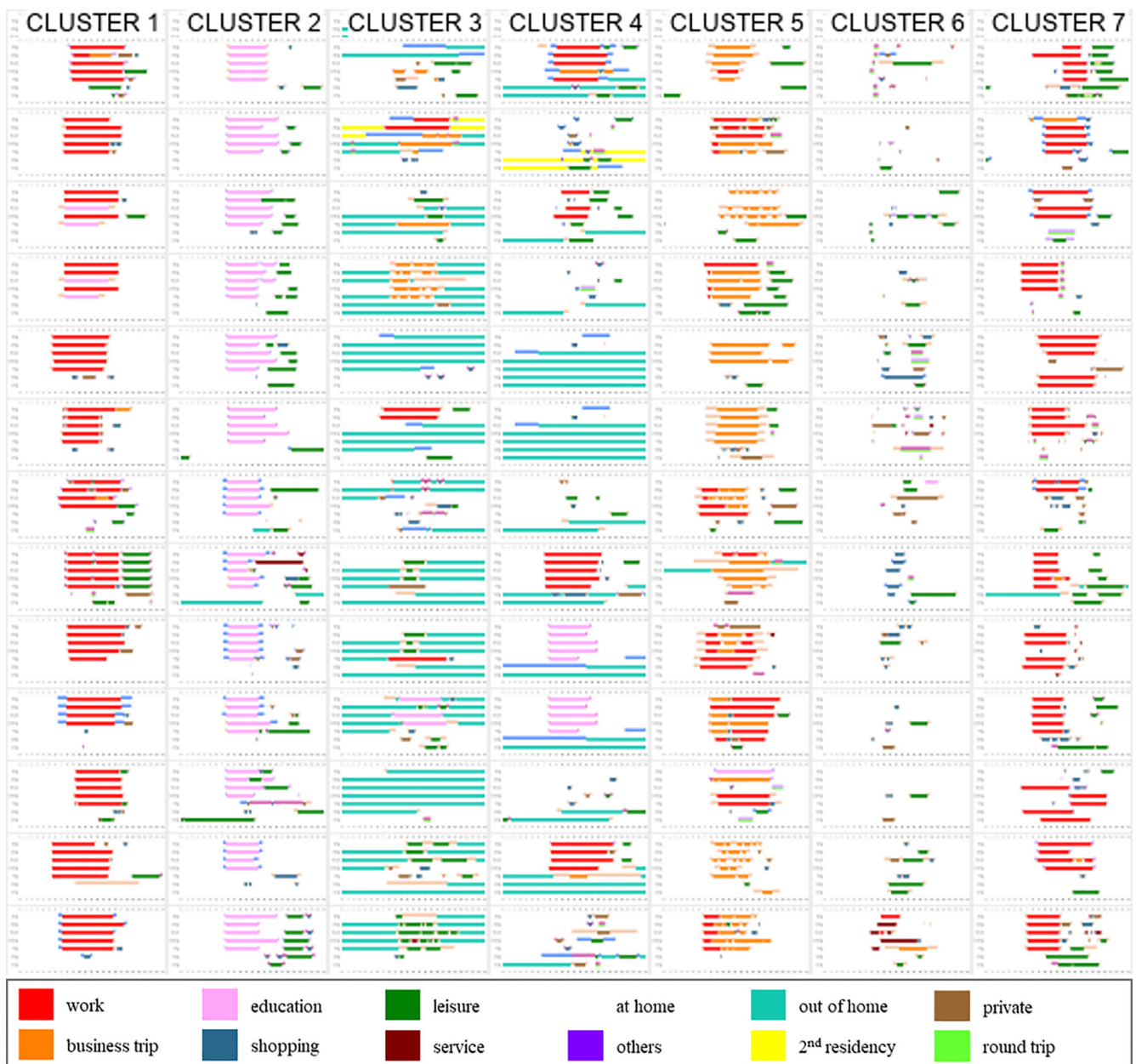Fig. 3. Dendrogram with Ward's method and split with k = 7.

Fig. 4. Cutout from result image with k-means k = 7.

resulted in one large cluster. It is also known that the DBSCAN has extreme difficulties in high-dimensional spaces (Xianting and Pan, 2016). And we have considered many dimensions through the images. Therefore, we continued working with the hierarchical and partitional clustering outputs.

5. Results

After applying the different clustering algorithms, the obtained results had to be validated. By generating a result image from the reasonable solutions, we were able to select the solution where the GraDiV images of each cluster are the most similar to each other. After choosing the best result image we evaluated the associated data to obtain a more detailed description of the clusters. Since the observation period covers three years, the same people may appear several times. Therefore, we then analyzed the intrapersonal stability of the clusters to evaluate the regularity of their daily life.

5.1. Visual validation of the final clustering solution

A big drawback of working with clustering algorithms is the validation of the obtained results. Normally one must choose the best segmentation by only using different error measures or plots that can be used for the applied methods. Since we worked with images, we were able to work with an additional validation: a visual validation. In order to visually validate the outcome of the applied clustering algorithms we have created a result image that displays all 7,362 images of the corresponding clusters underneath each other. For each outcome we received from the computed solutions, namely the hierarchical agglomerative Ward's method and the k-means algorithm, one result image was generated. Based on those images we were able to choose the one, that showed the clearest and most homogeneous segmentation. By evaluating the result images, we obtained the overall clearest clustering result with the k-means output of seven clusters. An extract from this result image is displayed in Fig. 4. Using the corresponding legend, the colors can be assigned to the activities. As we used the images of the week activity schedules for clustering, activities are usually dominating the images more than trips, as they are taking more time on average (cf. Fig. 1). Hence, we identified activity patterns with similarities in terms of color and length of activities rather than by mode choice.

Fig. 4 shows the advantage of our image-based method. In addition to the data-based evaluation, we are now able to visually assess and verify the clustering results. We can visually investigate how well the individual activity pattern images in the different segments fit together. For example, cluster 1 and 7 show a high dominance of the red color and therefore have a high work share during workdays. However, cluster 1 shows a higher interpersonal stability in the weekly activity schedules than cluster 7, where we can find considerable differences in the length of work and a higher interpersonal variance in the temporal positions caused by part-time work. In addition, people from cluster 7 undertake more leisure activities (green color) in the afternoons. In contrast, cluster 2 shows an even purple coloration with a high similarity for school or university. Cluster 3 and 4 are dominated by long out-of-home activities. In cluster 3, these overnight stays tend to take place during the week and in cluster 4 at the weekend. In cluster 3 it becomes apparent that patterns with different activities during the week fall into a common cluster, which is dominated by activities at the weekend. Fig. 4 is only a cutout. Theoretically it is possible to examine each GraDiV image in the cluster and eliminate any visual outliers. We have not eliminated any outliers in our work at this step.

As with other clustering methods, we can consult the cluster-describing variables for the validation and interpretation of the solution.

5.2. Analyzing key figures of the final clustering solution

In addition, using the underlying data is helpful to the visual analysis to receive more knowledge about the clusters. The evaluation in Table 1 contains the average values of sociodemographic and mobility-based key figures of each identified cluster. Values that differ significantly from the average value were marked in bold. The evaluation of the average values of the individual clusters indicate characteristics that are present in certain clusters. The coefficient of variation per attribute shows the differences of the cluster averages from each other.

To further support the interpretation, we created heat maps from the GraDiV images of each cluster to visualize the received activity patterns (see Fig. 5). As the time component is the determining dimension of the cluster result, it is interesting to see to what extent the activities in the clusters are distributed over the week. Fig. 5 illustrate the average out-of-home-activity time of the clusters over the week. In the trip diary and respectively in the image, we can only capture activities which are performed out of home. Thus the heat-maps show, based on the images in the specific segments, how long people perform similar activities (e.g., work, shopping) during the day. The more often out-of-home-activities take place at the same time, the darker the coloration is at this point. The coloration helps to understand the clusters and their time patterns independently from the type of activity. Cluster 2 has a high stability on weekdays in the morning compared to the other clusters. People from cluster 7 are more active in the morning than after noon. In cluster 7, it becomes obvious that a slight accumulation of activity can be observed at lunchtime and in the evening.

Additionally, the heat-maps can be used to analyze the day-to-day variability in the clusters. In Fig. 6, the differences in day-to-day variability are examined exemplarily on the two similar clusters 1 and 5. Similar coloration is an indication for stability between days. It is noticeable that cluster 1 shows a high stability in the morning (1). Furthermore, the first three days (Monday-Wednesday) are almost identical without variability (1, 2). But Thursday (3) and Friday (4) are very different in the evening. The activities on Thursday end late (3) and Fridays quite early (4). Cluster 5 shows a distinctly higher heterogeneity over the week (5, 6). The persons have less stability in everyday life – at least when comparing day-to-day-stability. This makes the differences between the two clusters clearer and shows the possibilities that exist with a visual representation of the longitudinal travel behavior of people.

Combining the information given in Table 1 and in Fig. 5, we can draw interesting conclusions for the identified clusters. In the following, we describe the seven activity patterns in more detail.

The *Employed middle-class* in cluster 1 includes people who work mostly full-time (87%). Slightly more men are contained in the cluster. The heat map shows a highly stable week (deep red coloration) in terms of working hours. In addition, there are few out-of-home activities in the evening and at the weekend. Work activities account for the highest proportion of out-of-home time (37%). With an average of 535 min of out-of-home activities per day, they have the third highest value among the clusters. In addition, car use dominates among the means of transport used (69%). Compared to the second cluster, *employed middle-class* people have longer activities in general, which can be seen in the heat-map and by the low ratio of travel time to out-of-home time.

Cluster 2 contains *Pupils & students*, but most of them are pupils, as can be seen from the age distribution. Cluster 2 also differs strongly to the heat maps of other clusters. The starting times of the activity in the morning have a high interpersonal stability within the cluster. In the afternoon further activities take place, but not as stable as in the morning. *Pupils and students* spend most of their out-of-home time with education (44%) and leisure activities (30%). As most of them do not yet have a driving license, the use of PT is much higher than in other clusters. For the known mobility figures such as distances, kilometers or

Table 1
Quantitative evaluation of the resulting clusters (of n = 7,362 weekly activity schedules).

Cluster #	1	2	3	4	5	6	7	Ø	CV
Cluster name	Employed middle-class	Pupils & students	Out-of-home long-distance traveler	Weekend-actives	Daily performers in job & life	Pensioners	Part-time mothers		
% Size (of n = 7,362 weekly activity schedules)	21%	9%	3%	7%	4%	42%	15%		
Sociodemographics									
Sex [1]	1.39	1.48	1.38	1.50	1.28	1.52	1.65	1.49	0.08
Age	47.70	**16.18**	45.18	50.27	49.69	**63.05**	49.52	51.86	0.28
% 10–17	0%	**74%**	6%	4%	0%	0%	0%	12%	2.25
% 18–25	4%	21%	15%	8%	3%	2%	4%	8%	0.89
% 26–35	12%	2%	11%	11%	11%	5%	9%	9%	0.42
% 36–50	36%	1%	22%	20%	33%	10%	34%	22%	0.59
% 51–60	40%	0%	26%	25%	38%	17%	38%	26%	0.55
% 61–70	8%	0%	13%	19%	12%	**32%**	13%	14%	0.72
% 70+	0%	0%	7%	13%	3%	**34%**	2%	8%	1.42
HH-size	2.49	3.77	2.63	2.39	2.67	2.12	2.61	2.70	0.21
Children within HH [2]	0.28	**0.61**	0.27	0.20	0.37	0.13	**0.35**	0.25	0.60
Income	7.36	7.73	7.50	7.28	7.60	**6.22**	7.15	6.89	0.07
Occupation									
% Full-time	**87%**	1%	55%	43%	**79%**	11%	44%	37%	0.87
% Part-time	10%	1%	8%	18%	15%	13%	**48%**	16%	0.91
% Education	2%	**95%**	18%	11%	1%	3%	3%	12%	2.92
% Others	2%	3%	19%	28%	5%	**73%**	5%	35%	0.74
Home Office [3]	0.14	0.00	**0.24**	0.16	**0.30**	0.09	0.18	0.12	0.80
Driving License	0.97	0.22	0.89	0.92	0.97	0.90	0.95	0.87	0.31
Car-availability [4]	**1.24**	2.25	1.36	1.36	**1.23**	1.41	1.31	1.38	0.26
Mobile Restricted	0.03	0.01	0.06	0.07	0.04	**0.18**	0.05	0.10	0.56
Geographic area [5]	2.08	2.15	2.00	**1.81**	2.09	2.00	2.03	2.02	0.05
Mobility behavior									
Km per day	47.6	30.0	**111.4**	88.6	91.2	27.8	40.8	43.1	0.78
Trips per day	3.4	3.1	3.1	3.4	**4.4**	3.1	3.8	3.3	0.15
Min per day	82.4	73.5	**138.1**	122.1	130.4	69.8	81.3	82.4	0.36
Min Out of Home per day	535	449	877	684	528	**206**	424	392	0.54
Ratio min/day travel time	**15%**	16%	16%	18%	25%	34%	19%	21%	0,32
Modal Split [6]									
% MIT	**69%**	39%	51%	57%	**69%**	60%	64%	61%	0.18
% PT	9%	**27%**	13%	10%	5%	6%	8%	9%	0.81
% NMIT	23%	34%	34%	33%	24%	34%	28%	30%	0.17
Activities									
% Work	**37%**	1%	10%	15%	16%	3%	27%	16%	0.81
% Business trips	3%	1%	10%	5%	**36%**	3%	4%	5%	2.32
% Education	0%	**44%**	3%	3%	0%	1%	0%	4%	4.01
% Shopping	19%	9%	14%	17%	14%	**30%**	20%	22%	0.30
% Leisure	18%	30%	22%	23%	15%	24%	20%	22%	0.23
% Service	7%	3%	3%	6%	7%	10%	**12%**	8%	0.40
% Outside home/2nd Residency	1%	1%	**25%**	14%	1%	0%	1%	2%	4.01
% Others	14%	11%	13%	16%	11%	30%	16%	20%	0.33

[1] Sex: 1 = *male* and 2 = *female;* [2]Children within HH: Number of children under 10 within the HH; [3]Home Office: 0 = *no use of Home Office* and 1 = *occasional to frequent use;* [4]Car-availability: 1 = *regularly* to 3 = *never;* [5]Geographic area: 1 = *more than 100.000 citizens* to 5 = *under 5.000 citizens;* [6]Modal split and activities proportional to the number of routes per day.

mobility time per day, the cluster is, on average, almost identical to cluster 6. Only the out-of-home activity time is longer. When looking at the heat map, the instability of cluster 6 is more noticeable.

Out-of-home long-distance travelers in cluster 3 consist of people that are usually out of home during the week, even overnight. Although individuals of this cluster do not cover more trips than the average, they comparatively cover longer distances (average 111 km per day). Looking at sociodemographic characteristics, they represent all age groups and half of them work full-time. The proportion of commuting to work (10%) is just as high as for business purposes (10%). In addition, people from this cluster work more from home (higher share of home office) and live mostly in urban areas. This also confirms the assumption that city dwellers tend to travel further (cf. Magdolen et al., 2020b).

People in cluster 4 are *Weekend-actives* because they were out of home on the weekend during the reported week, especially with an overnight stay from Saturday to Sunday. As in the previous cluster,

the distance covered is higher than in other clusters. As the commonalities during the week are lower and the persons are sociodemographic averages, the question arises whether this weekend activity is a constant component in their everyday life. Fig. 4 also shows a quite inhomogeneous cluster during the workdays. This cluster would be difficult to interpret without visual support.

The *Daily performers in job & life* in cluster 5 are mainly people who work as full-time employees. They have some work activities (16%), but even more than twice as many work-related activities (36%). These people have plenty of business trips nearby and generally have the most trips per day on average (4.4). Therefore, many self-employed people can be assumed to be part of this cluster. At the weekend, they are rather inactive, similar to cluster 1 and 2. As in cluster 1, the *Daily performers in job & life* also have an increased car availability in the household and use of MIT. Duration and distance per day are almost double the average. In terms of age, they settle in middle adulthood and have a male share of over 70%.
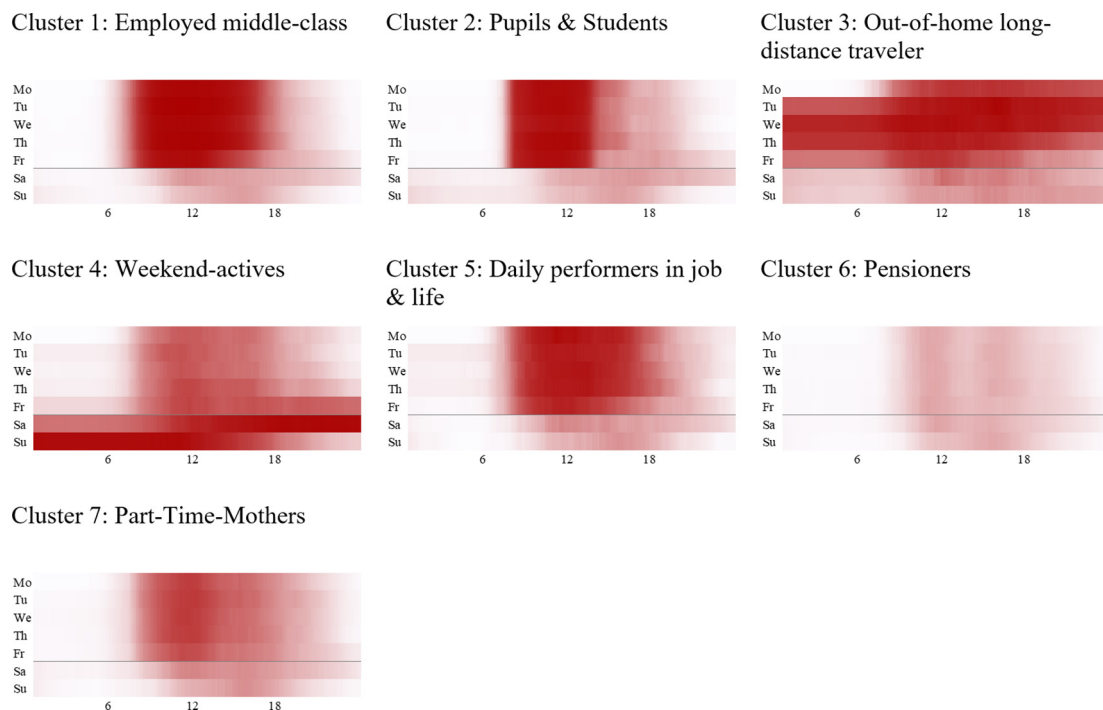
Cluster 1: Employed middle-class

Cluster 2: Pupils & Students

Cluster 3: Out-of-home long-distance traveler

Cluster 4: Weekend-actives

Cluster 5: Daily performers in job & life

Cluster 6: Pensioners

Cluster 7: Part-Time-Mothers

Fig. 5. Heat-Maps of activities per cluster.

The GraDiV images of the *Pensioners* in cluster 6 have sparse week activity schedules (see Fig. 4). This is also reflected in the short time they spend away from home and in mobility (206 min on average). The activities usually take place during the day (see Fig. 5). Each of the activities is short. The heat map shows no overlapping heats but only a low level of activity spread over the day. The number of trips per day is only slightly below average at 3.1. The cluster is dominated by pensioners. The commonality in mobility is therefore primarily that *Pensioners* rarely have longer out-of-home activities and spend much time at home. Their share of shopping activities is the highest (17%). They go out more often for small shopping trips in order to have social contacts. They are the only cluster with no large discrep-

ancy in out-of-home activities between weekdays and weekends (see Fig. 5).

People from the cluster *Part-time mothers* in cluster 7 are characterized by the high number of children within the household. They have many service activities to chauffeur their children (12%). More than 60% of the cluster are female, are in middle adulthood and almost half of them work part-time (48%). Although they do not travel longer than average, they make the second most trips per day (3.8). They have similarities to cluster 5 (e.g. number of children), but the people of cluster 5 mainly work full-time (79%). People from cluster 7 have a lot of leisure activities in the afternoon (see Fig. 4). We assume that they take care of their children in the household during this time. In cluster 5 there are more men and cluster 7 more women. Both clusters combine work with family life. In the following we use the cluster name as a synonym for the persons from the respective cluster.
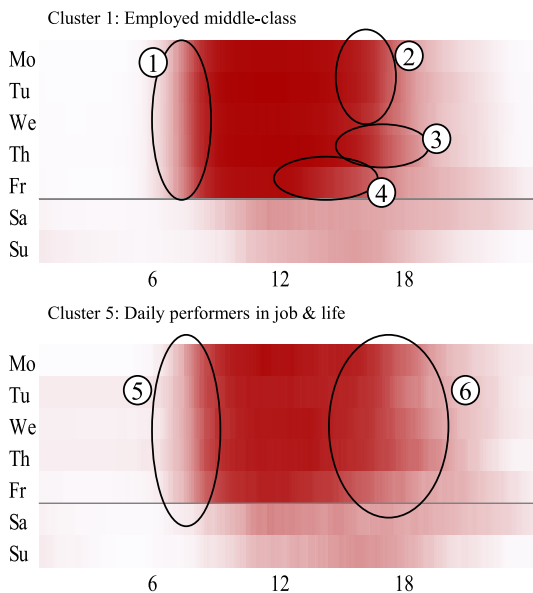
5.3. Intrapersonal cluster stability

In Table 2, we used the panel characteristics of the available data. The same individuals had been assigned to the clusters by their out-of-home activity behaviors in two consecutive years independently, in order to identify behavioral changes (Chlond and Eisenmann, 2018), which is reflected by an intrapersonal cluster change. For reasons of selective participation and unplanned dropouts (panel-attrition) between years, we again show only the unweighted results of the sub-sample of individuals who participated twice, i.e. the same person reported a weekly activity schedule two years consecutively. Table 2 illustrates the level of stability of cluster participation between any two years, overall 2,540 transitions. It should be read and interpreted as follows: the rows indicate the transitions between the clusters. For each of the clusters the relative proportion of transitions to other clusters as well as the stable share (remaining in the cluster) is given, therefore the row percentage in total is 100%. The stronger the stability, the more it is highlighted in color. The absolute share of clusters is given in the last column. It has to be mentioned that we have to expect a certain share of transitions due to normal changes in life situations during the life cycle, e.g. a student finishes his studies and starts

Cluster 1: Employed middle-class

Cluster 5: Daily performers in job & life

Fig. 6. Comparison of day-to-day variability in the reported week.

Table 2
Intrapersonal cluster stability (considering two consecutive years).

From Cluster (year 1)	To Cluster (year 2)							Total	Share of Cluster
	1	**2**	**3**	**4**	**5**	**6**	**7**		
1 Employed middle-class	59%	1%	3%	8%	3%	11%	14%	100%	**20%**
2 Pupils & Students	4%	78%	6%	1%	0%	9%	1%	100%	**7%**
3 Out-of-home long-distance traveler	13%	16%	28%	6%	0%	28%	9%	100%	**4%**
4 Weekend-actives	19%	8%	4%	20%	5%	38%	8%	100%	**8%**
5 Daily performers in job & life	24%	0%	4%	15%	33%	13%	11%	100%	**3%**
6 Pensioners	3%	0%	3%	7%	1%	81%	5%	100%	**44%**
7 Part-Time-Mothers	13%	1%	1%	10%	2%	22%	51%	100%	**14%**

working in following year. Those changes of life situations were not considered here and must therefore be taken into account when interpreting the outcomes. Hilgert et al. (2018) have studied the influence of life events on changing travel behavior between two consecutive years.

In the main diagonal the share is mostly the highest, i.e. the individuals have been assigned to the same clusters twice. By looking at the transitions, cluster 2 and 6 are the most stable clusters, having a stability of around 80%. We conclude that *Pupils & students* as well as *Pensioners* remain largely consistent with their reported everyday behavior in the following year. The high level of stability among *Pensioners* is also due to the low activity level in this cluster. It is interesting to note how some *Pupils & students* change to *Pensioners* cluster between the years. In some weeks, *Pupils & students* may have a low level of out-of-home activity, as attending university is not always compulsory in Germany.

Cluster 1 and 7 are still stable, but lose some people to each other or to the retired pensioners in cluster 6. The *Part-time mothers* hereby either switch back to a full-time employment, as childcare is not needed anymore, or they are less active in the reported week. The same accounts for the *Employed middle-class*. Either they cut back on working to a part-time job or also retire. As mentioned above, a further research of the underlying changes in life situations could be helpful to explain these transitions in more detail.

In contrast to the relatively stable clusters, we have highly unstable clusters as well. The most instable clusters are the *Out-of-home long distance travelers*, the *Weekend actives* as well as the *Daily performers in job & life*. The instability among the *Out-of-home long distance travelers* and *Weekend actives* exists because only a small part of these people continuously makes these long-distance trips. Many of them belong to other clusters, such as the *Pensioners* or the *Employed middle-class* and therefore cannot constantly be assigned to cluster 3 and 4. The *Out-of-home long-distance travelers* are also instable. Only 28% remain in the same cluster in the following year. As the same percentage that sticks to the cluster is found within the *Pensioners*, the year afterwards, it is likely to say that those persons were on vacation in the first year or visited some family. The same assumption applies to the transition to the clusters *Working middle-class* and *Pupils & students*. Especially with the *Weekend-actives*, everyday life during the week is very diverse and therefore makes one question whether activities at the weekend are exceptional. Only a relatively low share of the population performs short holidays and stays away from home at the weekends regularly. This cluster also describes the randomness of the reported week. The absolute low share of stable *Weekend actives* (20%) can be interpreted as those in the population who are long-distance commuters with several residences – regardless of being a student or belonging to the working population. Results show a high exchange of the weekend travelers to all other groups. The exchanges are less likely to those who are bound in daily activities (e.g. part-time mothers). This shows

a relevant limitation of our approach. The *Weekend-actives* are composed of persons from different clusters, which differ in their sociodemographic characteristics, but are bundled by their similarities in behavior in a random week. Therefore, *Weekend-actives* cannot be considered as a constant cluster of individuals. In contrast to cluster 3 and 4, the *Daily performers in job & life* cluster is not characterized by its out-of-home activities, but by its large proportion of business trips during the week. A third of its individuals (33%) reports constant business activities during the week. A large part of the cluster is migrating to the *Employed middle-class*. These persons reported in the next year a normal working week without business trips.

6. Evaluation of methodical approach

For this paper, we decided to take a different methodological approach for the segmentation of activity patterns. The data basis we used in the analysis consisted of transformed GraDiV images of the reported activity week schedules of the German Mobility Panel (MOP). By using this kind of data as raw information, we did not compress any data, we did not define any input variables and therefore we did not determine the input before the actual analysis. This allowed us to give the clustering algorithms more freedom in finding clusters without having to determine much in advance. In the results, we obtained activity patterns (clusters), which are similar in their week activity schedules and time use. They have a similar daily life regarding their out-of-home activities and the time these similar activities take place. The mode of transport is hereby less dominating for the cluster affiliation. Therefore, we have received activities and not mobility patterns. A valuable benefit of the images was the visual validation. Instead of only relying on the algorithm outputs or the corresponding error measures, we had another powerful possibility to validate the outcome by visually analyzing the result image. This not only increased the quality of the resulting clusters but also gave us the ability to visually describe the clusters and link the images of a cluster to the underlying data.

Compared to other segmentations within transportation research we did not use input variables that were based on prior knowledge or specific assumptions for the target segments. We used an image, which represents the out-of-home activity behavior and the combined travel of people over a week in a simplified way, while at the same time incorporating many aspects of their mobility. Therefore, we were able to additionally identify people who have a more atypical daily life such as the *Out-of-home long-distance traveler* or the *Weekend-actives*. Only by using the visual approach we were able to identify a common ground in the out-of-home time that could not be detected in prior publications.

Using the images as data basis also brought up some issues. By transforming the GraDiV images to computable data and especially by distinguishing the data by colors, we obtained a high-dimensional

data set. When processing high-dimensional data, the applied algorithms tend to become more inefficient as the dimensions increase and generate more unstable models (cf. Bellman, 2015). Another issue was the handling of outliers. As the k-means algorithm is not able to handle outliers, a way of identifying outliers is to either use methods that are able to handle outliers or to use the result image. For the first possibility, we identified outliers by using the hierarchical single-linkage and the DBSCAN algorithm. After leaving out the detected outliers and running the k-means again the results were visibly poorer. Another approach was then to select the outliers after generating the results. By scanning the result image, we could manually select each image that does not visually match the rest of the images. Since we would have to handle over 7,000 images, this process is not suitable.

7. Conclusion

Clustering people based on their behavior is not a new approach. In our literature review, we explained many existing approaches, which all require the selection of input variables as an important preliminary step. In our study, we have chosen a different approach. For this, we used GraDiV images of week activity schedules from the plausibility tool of the German Mobility Panel (MOP). These images were the input for our clustering. As a result, we received seven activity patterns (clusters), which are dominated by the activities performed and their respective time slots. The difference becomes visible in the heat maps of the respective clusters. Our approach has provided an image-based activity pattern segmentation based on longitudinal data. The approach worked well and provided helpful insights in identifying activity patterns. Another advantage of our approach was the evaluation and interpretation of potential clustering solutions. Instead of looking at key figures of the individual clusters only, we used the GraDiV images of each cluster. This visualization is very powerful, as differences between the weekly activity schedules become easily visible and mobility pattern can be presented through heat maps.

Altogether, this approach provides the chance to improve our understanding of activity participation and time use. On the one hand, merely clustering individuals by their socio-economic and demographic characteristics is not sufficient, as we also see similar behavior between different sociodemographic groups. On the other hand, a clustering based on typical key figures of travel behavior (trip rates, mileage) hides certain aspects, which can be uncovered by the approach as shown (e.g., long-distance travel). As a result, we received clusters with similar behavior in terms of out-of-home activities over the reported week. Especially the visualization of the two clusters *Weekend-actives* and *Out-of-home long-distance traveler* shows the randomness of a reported week in the travel diary. This behavior is not typical for most of them and does not occur every week. This was also shown by a comparison between the years in section 5.3. Some clusters are significantly more unstable than others. It is also clear, however, that longitudinal surveys are important for mapping intrapersonal variance during the week. The presented analysis also extended the approach of Zhao et al. (2015) and Schlich (2004), who considered day-to-day variability.

In summary, this work and the presented approach provides various contributions for research and practice.

- First, the visual approach helps decision makers to better understand how diverse mobility is and how much standard segmentation approaches generalize mobility. Segmentations are an important analysis tool for mobility concepts. In our approach, GraDiV images provide a detailed longitudinal section with temporal position, types of activities and their duration. On the other hand, the complex patterns are represented in a simplified way in images. This is an elementary advantage for the interpretation.

- Second, by using the predefined images this approach shows in a very transparent way that the results are not caused by data manipulation or excessive influence of the user. Only the transformation for the clustering process requires intervention. Contrary to other approaches, no indicators were calculated in advance for the clustering (cf. Oostendorp et al., 2019; Schlich, 2004; von Behren et al., 2018; Wittwer, 2014).

- Third, the results can also be used for big-data applications. The detailed information of the trip purpose respectively the activity is difficult to determine with passive data. The different clusters can be used especially for the interpretation of activity patterns in big-data information.

However, the results also show that further research is necessary. An examination of long-distance travel and its influence on the analysis of longitudinal surveys is crucial. Other quantitative segmentation approaches do not show this particularity as clearly as the GraDiV images did. However, it has a strong influence on relevant indicators such as kilometers per day as input key figures in the clustering. Additionally, another approach with the GraDiV images could be to consider weekdays only, similar to Zhao et al. (2015). With this approach, we could lay our emphasis more on everyday travel. To achieve this, we must also exclude long-distance travel during the week (see *Out-of-home long-distance traveler*, cluster 3). To remove these trips from the images, the approach of Magdolen et al. (2020a) could serve for the identification of non-routine trips during the reported week. In addition, the number of 7,362 GraDiV images is still limited in order to make general conclusions about the activity patterns of the German population. For this reason it is important to repeat the clustering with images from future MOP surveys. For further improvement, it is also relevant to consider mode use in more detail. Currently, activities strongly dominate the GraDiV image. This would require a relative increase in the visualization for the proportion of the trips. As a result, we would obtain more mobility types instead of activity types. A further innovation could be the use of image recognition programs, which are also used for face recognition. This application would even better consider the overall picture of the weekly activity schedules. Last but not least, it would also be possible to use a supervised procedure, such as random forest, in order to increase the influence of expert knowledge of the images. For this purpose, the GraDiV images could be pre-sorted into groups and the individual images could be labeled to train an allocation model. The remaining images are then allocated by means of the model. A similar approach has already been applied to the classification of spatial types by Niklas et al. (2020c). The disadvantage of this approach would again be the high influence of the user.

CRediT authorship contribution statement

Sascha von Behren: Conceptualization, Methodology, Investigation, Validation, Data curation, Formal analysis, Software, Writing - original draft, Writing - review & editing. **Tim Hilgert:** Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Sophia Kirchner:** Methodology, Software, Writing - original draft. **Bastian Chlond:** Conceptualization, Validation, Writing - original draft, Project administration. **Peter Vortisch:** Validation, Supervision, Project administration, Resources.

References

Allahviranloo, M., Regue, R., Recker, W., 2014. Pattern Clustering and Activity Inference, in: TRB 93rd Annual Meeting Compendium of Papers, Washington, D.C.

Anable, J., 2005. 'Complacent Car Addicts' or 'Aspiring Environmentalists'? Identifying travel behaviour segments using attitude theory. Transp. Policy 12, 65–78. https://doi.org/10.1016/j.tranpol.2004.11.004.

Bellman, R. (Ed.), 2015. Adaptive control processes: A guided tour. Princeton University Press, Princeton, N.J.

Berger, M., 2004. Typologiebildung und Erklärung des Aktivitäten-(Verkehrs-) verhaltens – ein Multimethodenansatz unter Verwendung der Optimal Matching Technik. Dissertation..

Beyerer, J., Puente León, F., Frese, C., 2012. Automatische Sichtprüfung: Grundlagen. Methoden und Praxis der Bildgewinnung und Bildauswertung, Springer, Berlin, Heidelberg, p. 940.

Chlond, B., Eisenmann, C., 2018. Workshop Synthesis: Behavioral changes in travel – challenges and implications for their identification and measurement. Transp. Res. Procedia 32, 563–572. https://doi.org/10.1016/j.trpro.2018.10.022.

Collum, K.K., Daigle, J.J., 2015. Combining attitude theory and segmentation analysis to understand travel mode choice at a national park. Journal of Outdoor Recreation and Tourism 9, 17–25. https://doi.org/10.1016/j.jort.2015.03.003.

Ectors, W., Kochan, B., Knapen, L., Janssens, D., Bellemans, T., 2016. A Generic Data-driven Sequential Clustering Algorithm Determining Activity Skeletons. Procedia Comput. Sci. 83, 34–41. https://doi.org/10.1016/j.procs.2016.04.096.

Götz, K., Jahn, T., Schultz, I., 1998. Mobilitätsstile. Ein sozial-ökologischer Untersuchungsansatz ; Arbeitsbericht ; Subprojekt 1: Mobilitätsleitbilder und Verkehrsverhalten. Öko-Inst, Freiburg Breisgau, 344 S.

Hanson, S., Huff, J., 1986. Classification issues in the analysis of complex travel behavior. Transportation 13 (3), 271–293. https://doi.org/10.1007/BF00148620.

Harris, D.M., Harris, S.L., 2013. Digital design and computer architecture. Morgan Kaufmann, Waltham, MA.

Heuwinkel, D., 1981. Aktionsräumliche Analysen und Bewertung von Wohngebieten: Ein verhaltensorientiertes Verfahren, entwickelt am Beispiel von Berlin (West). Zugl.: Hannover, Univ., Diss., 1979 u.d.T.: Heuwinkel, Dirk: Vergleichende Analyse räumlichen Verhaltens, Instrument zur Bewertung von Wohngebieten. Christians, Hamburg, p. 227.

Hildebrand, E.D., 2003. Dimensions in elderly travel behaviour: A simplified activity-based model using lifestyle clusters. Transportation 30 (3), 285–306. https://doi.org/10.1023/A:1023949330747.

Hilgert, T., von Behren, S., Eisenmann, C., Vortisch, P., 2018. Are Activity Patterns Stable or Variable? Analysis of Three-Year Panel Data. Transp. Res. Rec. 2672 (47), 46–56. https://doi.org/10.1177/0361198118773557.

Hoogendoorn-Lanser, S., Schaap, N.T.W., OldeKalter, M.-J., 2015. The Netherlands Mobility Panel: An Innovative Design Approach for Web-based Longitudinal Travel Data Collection. Transp. Res. Procedia 11, 311–329. https://doi.org/10.1016/j.trpro.2015.12.027.

Hunecke, M., Haustein, S., Böhler, S., Grischkat, S., 2010. Attitude-Based Target Groups to Reduce the Ecological Impact of Daily Mobility Behavior. Environment and Behavior 42 (1), 3–43. https://doi.org/10.1177/0013916508319587.

Kunert, U., 1994. Weekly mobility of life cycle groups. Transportation 21 (3), 271–288. https://doi.org/10.1007/BF01099214.

Lipps, O., 2001. Modellierung der individuellen Verhaltensvariationen bei der Verkehrsentstehung. Dissertation. Karlsruhe, 147 pp.

Maat, C. Arentze, T.A., 2003. Variation of activity patterns with features of the spatial context., in: TRB 82nd Annual Meeting. TRB 82nd Annual Meeting, Washington, D. C.

Magdolen, M., Ecke, L., Hilgert, T., Chlond, B., Vortisch, P., 2020a. Identification of Non-Routine Tours in Everyday Travel Behavior, in: 99th Transportation Research Board Annual Meeting, Washington D.C., January 12 - 16, 2020.

Magdolen, M., von Behren, S., Chlond, B., Hunecke, M., Vortisch, P., 2019. Combining attitudes and travel behavior - A comparison of urban mobility types identified in Shanghai, Berlin and San Francisco, in: TRB 98th Annual Meeting Compendium of Papers. TRB 98th Annual Meeting Compendium of Papers, Washington, D.C.

Magdolen, M., von Behren, S., Chlond, B., Vortisch, P., 2020b. Traveling Long-Distance with Bad Conscience? Discrepancies Between Everyday and Long-Distance Travel of Urbanites, in: TRB 99th Annual Meeting, Washington D.C.

Niklas, U., von Behren, S., Chlond, B., Vortisch, P., 2020a. Electric Factor—A Comparison of Car Usage Profiles of Electric and Conventional Vehicles by a Probabilistic Approach. WEVJ 11 (2), 36. https://doi.org/10.3390/wevj11020036.

Niklas, U., von Behren, S., Eisenmann, C., Chlond, B., Vortisch, P., 2020b. Premium factor – Analyzing usage of premium cars compared to conventional cars. Research in Transportation Business & Management 100456. https://doi.org/10.1016/j.rtbm.2020.100456.

Niklas, U., von Behren, S., Soylu, T., Kopp, J., Chlond, B., Vortisch, P., 2020c. Spatial Factor—Using a Random Forest Classification Model to Measure an Internationally Comparable Urbanity Index. Urban Science 4 (3), 36. https://doi.org/10.3390/urbansci4030036.

Oostendorp, R., Nieland, S., Gebhardt, L., 2019. Developing a user typology considering unimodal and intermodal mobility behavior: a cluster analysis approach using survey data. Eur. Transp. Res. Rev. 11 (1). https://doi.org/10.1186/s12544-019-0369-1.

Pas, E.I., 1980. Towards the understanding of urban travel behavior trough the classification of daily urban/activity patterns. Dissertation. Evanston.

Prillwitz, J., Barr, S., 2011. Moving towards sustainability? Mobility styles, attitudes and individual travel behaviour. J. Transp. Geogr. 19, 1590–1600. https://doi.org/10.1016/j.jtrangeo.2011.06.011.

Salomon, I., Ben-Akiva, M., 1983. The Use of the Life-Style Concept in Travel Demand Models. Environ Plan A 15 (5), 623–638. https://doi.org/10.1068/a150623.

Schlich, R., 2004. Verhaltenshomogene Gruppen in Längsschnitterhebungen. Dissertation. Zürich, 162 pp.

Schmiedel, R., 1984. Bestimmung verhaltensähnlicher Personenkreise für die Verkehrsplanung. Dissertation. Karlsruhe.

Schöppe, E., Förschner, G., 1983. Demografisch-soziologische Personengruppen und ihre Anwendung in der Verkehrsplanung. Die Straße 24, 353–357.

von Behren, S., Minster, C., Magdolen, M., Chlond, B., Hunecke, M., Vortisch, P., 2018. Bringing travel behavior and attitudes together: An integrated survey approach for clustering urban mobility types, in: TRB 97th Annual Meeting Compendium of Papers, Washington, D.C.

Wittwer, R., 2014. Zwangsmobilität und Verkehrsmittelorientierung junger Erwachsener: Eine Typologisierung. Habilitation, Dresden.

Xianting, Q., Pan, W., 2016. A Density-Based Clustering Algorithm for High-Dimensional Data with Feature Selection, in: International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII). International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), Wuhan, China. 03.12.2016. IEEE, pp. 114–118.

Zhao, F., Pereira, F.C., Ball, R., Kim, Y., Han, Y., Zegras, C., Ben-Akiva, M., 2015. Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore. Transp. Res. Rec. 2494 (1), 45–56. https://doi.org/10.3141/2494-06.