Clustering railway passenger demand patterns from large-scale origin–destination data

van der Knaap, Renate J.H.; de Bruyn, Menno; van Oort, Niels; Huisman, Dennis; Goverde, Rob M.P.

DOI
10.1016/j.jrtpm.2024.100452

Publication date
2024

Document Version
Proof

Published in
Journal of Rail Transport Planning and Management

ARTICLE IN PRESS

Journal of Rail Transport Planning & Management xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Rail Transport Planning & Management

journal homepage: www.elsevier.com/locate/jrtpm

FULL LENGTH ARTICLE

Clustering railway passenger demand patterns from large-scale origin–destination data

Renate J.H. van der Knaap [a,*], Menno de Bruyn [b], Niels van Oort [a], Dennis Huisman [b,c], Rob M.P. Goverde [a]

[a] Department of Transport and Planning, Delft University of Technology, P.O. Box 5048, Delft, 2600 GA, The Netherlands
[b] Netherlands Railways, P.O. Box 2025, Utrecht, 3500 HA, The Netherlands
[c] Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, Rotterdam, 3000 DR, The Netherlands

ARTICLE INFO

Keywords:
Passenger demand pattern
Railway
Origin–destination data
Clustering
Homogeneous period

ABSTRACT

Train passenger demand fluctuates throughout the day. In order to let train services, such as the line plan and timetable, match this fluctuating demand, insights are needed into how the demand is changing and for which periods the demand is relatively stable. Hierarchical clustering on both regular and normalized origin–destination (OD) data is used to determine for each workday continuous time-of-day periods in which the passenger demand is homogeneous. The periods found for each workday are subsequently used as input in a clustering algorithm to look for similarities and differences between workdays. The methods for finding homogeneous periods during the day and week are applied to a case study covering a large part of the railway network in the Netherlands. We find large differences between the periods based on regular OD matrices and those based on normalized OD matrices. The periods based on regular OD matrices are more compact in terms of passenger volumes and average kms travelled and therefore more suitable to use as input for designing a service plan. Comparison of different workdays shows that mainly the peak periods on Friday are far away from Monday to Thursday, and hence could benefit from an altered service plan.

1. Introduction

It is well known that the passenger demand in railway services fluctuates with the time of day. During the peak hours there is typically a high demand for transport, while in the middle of the day or in the evening the demand is much lower. The demand also fluctuates over the week: not all (week)days have the same demand. To illustrate these fluctuations, Fig. 1 shows the number of arriving (in red) and departing (in blue) passengers and the total number of passengers (in black) per time of day for a station in the Dutch railway network. The figure displays the arrival and departure data of two days: an average Tuesday in 2019 (denoted by solid lines) and an average Friday in 2019 (denoted by dashed lines). Fig. 1 shows the pattern of a typical commuter station with a lot of jobs in the vicinity: many people arriving during the morning peak period and leaving during the evening peak period and not so many people arriving and/or departing outside the peak hours. We can also observe in Fig. 1 that the demand is not the same for each day of the week. The demand at the peak periods on Friday is much lower than the demand on Tuesday, while the demand is similar to Tuesday outside the peak hours. Due to the COVID-19 pandemic, these differences between days are expected to increase. Several studies in the Netherlands show that more people will partly work from home after the pandemic, with a preference

* Corresponding author.
E-mail addresses: r.j.h.vanderknaap@tudelft.nl (R.J.H. van der Knaap), Menno.deBruyn@NS.nl (M. de Bruyn), N.vanOort@tudelft.nl (N. van Oort), Dennis.Huisman@NS.nl (D. Huisman), R.M.P.Goverde@tudelft.nl (R.M.P. Goverde).

https://doi.org/10.1016/j.jrtpm.2024.100452
Received 30 May 2023; Received in revised form 8 May 2024; Accepted 9 May 2024

Please cite this article as: Renate J.H. van der Knaap et al., *Journal of Rail Transport Planning & Management*, https://doi.org/10.1016/j.jrtpm.2024.100452

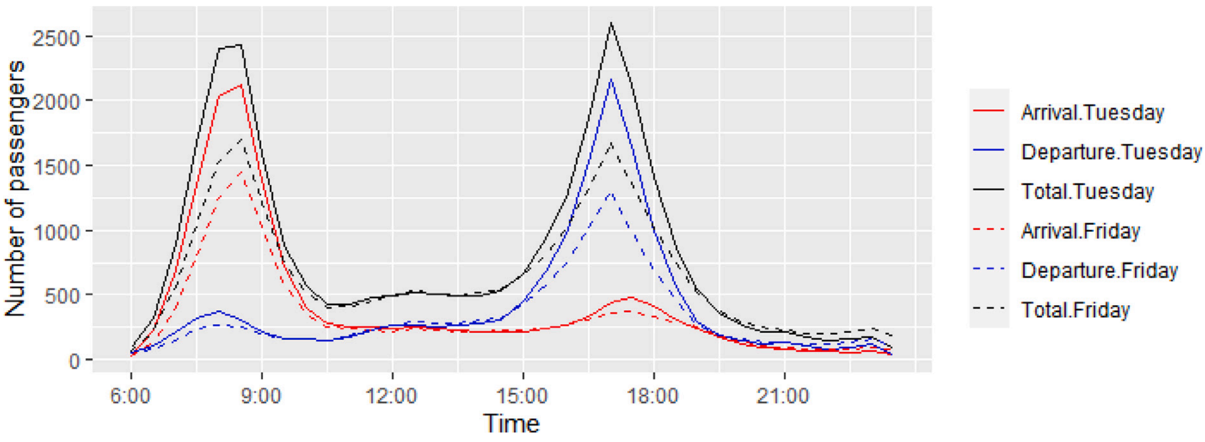


Fig. 1. The figure displays the number of arriving (red), departing (blue), and total (black) passengers at a station in the Dutch railway network. The data of two days are given: Tuesday (solid lines) and Friday (dashed lines).

for Friday and Wednesday (see e.g. Kennisinstituut voor Mobiliteitsbeleid et al., 2021; Van Hagen et al., 2021; Ton et al., 2022). Although this may cause the difference between peak and off-peak periods to decrease, it will increase the difference between days.

Nevertheless, many European countries including the Netherlands have fixed line systems and cyclic railway timetables. These timetables provide regular interval services throughout the day, including fixed departure times (e.g., 10 min past every hour) and sufficient time to transfer at stations where different services meet. Therefore, these timetables are easy to remember for passengers. Another benefit is that it is easier for people to travel outside the peak hours, since the service frequency is the same throughout the day. Wardman et al. (2004) and Johnson et al. (2006) confirm these aspects are beneficial for the customers, by showing that regularity of a timetable increases the passenger demand. One of the disadvantages however is the fact that the train lines and cyclic timetables are usually based on the peak hour demand, and are therefore not tailored for other periods outside the peak hours. This is both in terms of volume, and in terms of the structure of the demand (i.e., where passengers are coming from and/or going to). People who travel outside the peak hours are likely to have different travel purposes and hence destinations. This leads to heterogeneous travel demand patterns, which does not align well with the fixed cyclic service supply.

Better matching the train services (line plan and timetable) to the travel demand can have multiple benefits for both the passengers and the railway undertaking (RU). Passengers will have a better travel experience with faster trips and fewer transfers. Furthermore, the RU can save money on energy, rolling stock and personnel by only operating those trains for which there is demand. Lastly, when train services better match the demand, the train becomes a more attractive mode of transport. This would result in more people taking the train and higher ticket sale revenues for the RU.

Since both regular and irregular schedules have their benefits, it might be useful to have a schedule with both regular and irregular characteristics. One example of this could be to have multiple periods during a day, where there is a regular schedule within each period, and changes are made to the schedule when switching between periods to better match the demand in the next period. However, having multiple periods during a day also has some costs. For example, the RU who is used to having a regular schedule during the entire day would have to make several new schedules: one for each period. This is a challenging task since both line planning and timetabling problems are difficult to solve. Schöbel (2012) shows that even some special cases of line planning are NP hard. Furthermore, the Periodic Event Scheduling Problem, which is widely used for modelling cyclic railway timetabling is shown to be NP complete by Serafini and Ukovich (1989). Moreover, the literature review paper by Durán-Micco and Vansteenwegen (2022) about the line planning problem concludes that there exists a large gap between the problems addressed in the literature and the extremely complex problems that have to be solved in practice. In practice, the process of making a line plan and timetable is not fully automated yet, but still requires a large amount of manual labour. Therefore, there is a limit to the number of different schedules that the RU can create and hence operate in a day. Besides the cost for the RU, there is also a cost for the passenger, because they have to adjust to new schedules during the day. Therefore, the RU should only make changes to the schedule if significant improvements for the passenger and/or the RU can be realized.

Due to the reasons mentioned above, first a good insight is needed in how the demand patterns change throughout the day and week, before changes are made to the schedule. To aid the RU in finding periods for which a different schedule might be interesting, this paper proposes a data-driven method to derive periods during the day and week in which the railway demand is homogeneous. We define periods that are homogeneous in demand as periods in which the travel patterns in terms of origin–destination flows are more or less the same. The idea behind this is that in periods of homogeneous demand and operational conditions, the public transport schedules can also be the same in terms of e.g., routes operated, stops, and frequencies. By determining the periods with homogeneous demand during the day and week, we can therefore put a maximum on the amount of periods an RU needs to consider for providing a different schedule to.

Since the emergence of Automatic Fare Collection (AFC) systems, collecting data about passenger trips has become much easier for the public transport agencies and researchers. Several papers analyse such data to describe periods of homogeneous demand or passenger groups with similar demand characteristics. Data mining techniques (like k-means, DBSCAN, and hierarchical clustering) are often used to find homogeneity in AFC data. The first ones to do this are Agard et al. (2006), who analyse user behaviours and determine different market segments among bus users. Other papers that use data mining techniques to get a better understanding of public transport passengers include (Ma et al., 2013; Kieu et al., 2015; El Mahrsi et al., 2017; Deschaintres et al., 2019). They suggest that these insights can be used by public transport operators to provide better pricing options and/or information to passengers. Furthermore, data mining techniques are also used on AFC data to investigate the change in passenger behaviour over time (Briand et al., 2017), the effect of long-term service disruptions (Eltved et al., 2021), and the effect of COVID-19 on public transport usage (Mützel and Scheiner, 2022; Henrion et al., 2023).

Besides the aforementioned goals, there are multiple papers that want to determine periods with homogeneous operational performance and/or demand, with the aim to improve the bus schedule. For example, Muller and Furth (2001) use the data from Trip Time Analysers on operational performance of buses, to determine periods in which the trip time is homogeneous. The trip time is defined as the time spent from the first to the last stop on the bus route. These homogeneous periods with corresponding trip times can then be used to improve the bus schedule, in order to minimize early and late arrivals. Lu and Reddy (2012) use hierarchical cluster analysis on the total hourly bus trip volumes to determine whether different days should have different bus schedules. Mahmoudzadeh and Wang (2020) also investigate the effects of days on travel patterns for a university campus bus shuttle service. Clustering based on the aggregate demand and aggregate delay is used to group days with similar travel patterns together and to find homogeneous periods of demand during the day. The knowledge gained is used for improving the shuttle schedule. Ji et al. (2011) take a different approach. They use hierarchical clustering on probability flow matrices to find contiguous periods during a normal workday in which the travel patterns are homogeneous. Probability flow matrices are OD matrices in which every cell is divided by the total number of trips in the matrix. Hence, the cells in a probability flow matrix sum to 1 and the value of a cell denotes the probability that a passenger travels from the origin to the destination corresponding to that cell. Ji et al. (2011) conclude that clustering probability flow matrices provides a better result than clustering that only looks at the total trip volume. However, De Bruyn and Mestrum (2021), who look at OD matrices of train demand, argue that both the volume and the structure of the demand are important when determining the similarity of demand in different periods. They introduce two measures that denote the similarity of two OD matrices in structure and volume, respectively.

Although there are several papers on finding homogeneous demand patterns in bus services, the literature on determining periods with homogeneous train demand is limited. To the best of our knowledge, the paper by De Bruyn and Mestrum (2021) is the only paper about finding periods with homogeneous demand patterns in the rail sector. Therefore, testing the performance of the current methods on train demand data is an important addition to the academic literature. Furthermore, we believe that both the structure and the volume of the demand are relevant when constructing the line plan and the timetable of a railway service. Although the demand structure plays an important role in constructing the train lines, the volume is also a crucial aspect in determining the frequencies of the lines. Therefore, it seems reasonable to take both the structure and the volume of demand into account when determining periods of homogeneous demand, as De Bruyn and Mestrum (2021) propose. However, the metrics that they propose are not directly suitable for constructing periods with homogeneous demand and the literature on bus demand only considers either the trip volumes or the structure of the demand (in the form of probability flow matrices). Therefore, we propose a new method using hierarchical clustering that takes into account both changes in structure and volume.

To summarize, the main contributions of this paper are:

1. Providing a new method for finding periods of homogeneous demand, where both the structure and the volume of the demand are taken into account;
2. Applying the methods for finding periods of homogeneous demand based on the structure of demand and based on the structure and the volume of demand on a case study with railway passenger demand.

The remainder of this paper is organized as follows. In Section 2 the methodology of the analysis is described. Next, Section 3 describes the case study data and Section 4 provides the results of applying the methods on this case. We conclude the paper with some conclusions and recommendations for future research in Section 5.

2. Methodology

This section provides the methods used to derive homogeneous periods in railway travel demand during the day and week. First we describe the method employed to find homogeneous periods in demand during the day in Section 2.1. Next, in Section 2.2 we describe how this method is adapted to look for homogeneity during the week.

2.1. Homogeneous periods during the day

We propose to use clustering of OD matrices to determine periods of homogeneous demand. Clustering methods can be partitioned in two main groups: hierarchical approaches and partitional approaches (Jain et al., 1999). Partitional approaches (like k-means) produce only one partition of the data, while hierarchical methods produce a nested series of partitions. Since beforehand the optimal number of periods to divide a day in is not known, a hierarchical clustering approach seems to be the most fitting. Hierarchical clustering approaches can be agglomerative or divisive. In agglomerative approaches, all objects that need to be clustered start in their own cluster. In each iteration of the clustering algorithm, the two clusters which are closest to each other are merged, until only a single cluster is left. With divisive clustering all objects start in a single cluster and in each step of

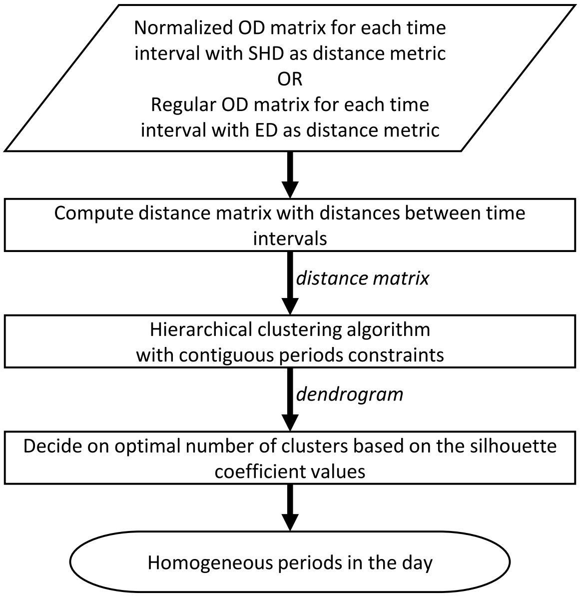


Fig. 2. Flowchart describing the proposed process for finding homogeneous periods during the day.

the algorithm a cluster is split into two parts. We refer the interested reader to Han et al. (2011) for an overview of data mining techniques including clustering.

Another important decision in hierarchical clustering algorithms is how the distance between clusters is defined, which is also known as the linkage measure. There are three widely used linkage measures: single linkage, complete linkage and average linkage. Each cluster contains one or more objects (which in this case are OD matrices). To determine the distance between two clusters, each of the linkage measures looks at the distance between each pair of objects such that the two objects are not from the same cluster. The linkage types differ in the way they use these distances to determine the distance between two clusters. In single linkage, the distance between two clusters is defined as the smallest distance between any pair of objects. When single linkage is used, this is sometimes referred to as nearest-neighbour clustering. Contrarily, in complete linkage, the distance between two clusters is defined as the largest distance between any pair of objects. Farthest-neighbour clustering is sometimes used to refer to clustering using this measure. Lastly, average linkage defines the distance between two clusters as the average distance between all object pairs.

In this paper, we will use an agglomerative hierarchical clustering approach using complete linkage. This clustering method is already successfully applied in the work of Ji et al. (2011) to find passenger demand patterns in bus OD data. Complete linkage is chosen as it produces more compact clusters and since it has been found to be more effective across many applications (Jain et al., 1999). Furthermore, just like Ji et al. (2011), we will search for contiguous periods. The aim of this paper is to find periods of homogeneous demand, for which a service plan (which includes a line plan and timetable) can be designed. Since switching between service plans is likely to be difficult, we add a constraint that periods must be contiguous, to reduce the potential number of switches during the day. Fig. 2 provides a flowchart that the describes the process that is used to find homogeneous periods during the day. All the terms in this figure will be explained in the remainder of this section.

In short, the hierarchical contiguous clustering algorithm works as follows. Suppose we want to cluster N objects, so in our case N OD matrices. The first step is to calculate an $N \times N$ distance matrix containing the distances between each pair of OD matrices. The distance between two OD matrices can also be seen as the degree of dissimilarity: if the distance is small, then the matrices are similar, and if the distance is large, the matrices are dissimilar. Once the distance matrix is calculated, the algorithm starts with N number of clusters (each containing one object). Then in each iteration of the algorithm, we determine which pairs of clusters are contiguous and find which of these pairs have the smallest distance to each other. This pair is then merged into a new cluster and the distance matrix is updated. This process is repeated until all clusters are combined into a single cluster. The result of the clustering algorithm can be visualized in a dendrogram. For a full description of the clustering algorithm, we refer to Ji et al. (2011).

In this paper, we use two different methods to determine the distance matrix required as input for the clustering algorithm: one method using normalized OD matrices and another method using regular OD matrices. These methods are described in the next two paragraphs, respectively.

The first method using normalized OD matrices is based on the work of Ji et al. (2011) and McCord et al. (2012). Normalized OD matrices are also known as probability flow matrices. A regular OD matrix can be converted to a normalized OD matrix by dividing each cell of the OD matrix by the total sum of the matrix' cells. Let A_t^d be an OD matrix corresponding to day d and time t. Then the normalized OD matrix $\bar{A}_t^d$ can be calculated as:

$$\bar{A}_t^d(i,j) = \frac{A_t^d(i,j)}{\sum_{i=1}^{n} \sum_{j=1}^{n} A_t^d(i,j)}. \tag{1}$$

Here $A_t^d(i,j)$ ($\bar{A}_t^d(i,j)$) denotes the value in the cell of matrix A_t^d ($\bar{A}_t^d$) corresponding to the ith row and the jth column. Since each cell in an OD matrix denotes the number of passengers travelling from an origin to a destination, the value in each cell must be greater or equal to zero. The transformation to normalized OD matrices by Eq. (1) preserves this nonnegativity property. Furthermore, the total sum of all cells in a normalized OD matrix is equal to 1. Therefore, a normalized OD matrix can be seen as a discrete probability distribution. Hence, a similarity measure for probability distributions can be used to determine the distance between two normalized OD matrices. In this paper the squared Hellinger distance measure (SHD) is used. SHD is chosen as distance measure, since it is a suitable measure for comparing probability distributions (Le Cam and Yang, 2000), and it has been used before to compare normalized OD matrices (Ji et al., 2011; McCord et al., 2012). The squared Hellinger distance between two $n \times n$ normalized OD matrices $\bar{A}_t^d$ and $\bar{A}_{t'}^d$, denoted by $\mathrm{H}(\bar{A}_t^d, \bar{A}_{t'}^d)$, is defined as:

$$\mathrm{H}(\bar{A}_t^d, \bar{A}_{t'}^d) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sqrt{\bar{A}_t^d(i,j)} - \sqrt{\bar{A}_{t'}^d(i,j)} \right)^2. \tag{2}$$

When the distance matrix is based on normalized OD matrices, only the structure of the demand is taken into account. The volume is cancelled out completely, since the OD matrices are divided by their total volume to create the normalized OD matrices. However, as mentioned before the passenger volumes also play an important role in determining what service is required. Therefore, we propose a new method to determine the distance matrix based on the original OD matrices, since these matrices include both the structure and the volume of the demand. As regular OD matrices usually do not meet the requirements of a probability density function, the squared Hellinger distance cannot be used in this case. Instead, we propose to use the well-known Euclidean distance (ED) as the distance metric. Given two non-normalized $n \times n$ OD matrices A_t^d and $A_{t'}^d$, the ED between the two matrices, denoted by $\mathrm{E}(A_t^d, A_{t'}^d)$ is calculated as:

$$\mathrm{E}(A_t^d, A_{t'}^d) = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \left(A_t^d(i,j) - A_{t'}^d(i,j) \right)^2}. \tag{3}$$

The hierarchical clustering algorithm produces in each iteration a clustering result. Hence, if N objects are being clustered, the algorithm provides N different ways to cluster these objects (including all objects form a single cluster and each object forms its own cluster). To determine which of these results is the best, the silhouette coefficient is used. The silhouette coefficient was introduced by Rousseeuw (1987) as a measure for the quality of a clustering result. The result with the highest measured quality can then be selected as the final result.

We use the following example to explain the calculation of the silhouette coefficient. Let o be an object (e.g., an OD matrix) for which we calculate the silhouette coefficient and let O be the set of all objects. In the clustering result we want to assess, the objects from O are divided over k clusters $C_1, \ldots, C_k$. Furthermore, let object o belong to cluster C_i ($o \in C_i$) with $1 \le i \le k$ and let $|C_i|$ denote the number of objects in cluster C_i. The equation given by Rousseeuw (1987) to calculate the silhouette coefficient of object o (denoted by $s(o)$) is

$$s(o) = \frac{b(o) - a(o)}{\max\{a(o), b(o)\}}. \tag{4}$$

The $a(o)$ in Eq. (4) denotes the average distance from o to other objects within cluster C_i and hence is a measure of the cluster's compactness. Let $o, o' \in O$ and let $\mathrm{dist}(o, o')$ denote the distance between objects o and o'. Then $a(o)$ is calculated as:

$$a(o) = \frac{\sum_{o' \in C_i, o \neq o'} \mathrm{dist}(o, o')}{|C_i| - 1}. \tag{5}$$

The distance is measured using either the SHD or the ED, depending on which input is used: normalized OD matrices or regular OD matrices. On the other hand $b(o)$ in Eq. (4) denotes the average distance from o to the objects in the nearest other cluster. This serves as a measure of closeness from object $o \in C_i$ to the $k-1$ other clusters and is calculated as:

$$b(o) = \min_{C_j : 1 \le j \le k, j \neq i} \left\{ \frac{\sum_{o' \in C_j} \mathrm{dist}(o, o')}{|C_j|} \right\}. \tag{6}$$

The silhouette coefficient can take any value between −1 and 1. High values are desirable, since this denotes that an object is very close to other objects in its cluster and very far from objects in other clusters. On the other hand are negative values undesirable, since this denotes that an object is closer to the objects in another cluster than to the objects in its own cluster. The silhouette coefficient for an entire clustering result can be determined by computing the silhouette coefficient for each object and taking the average of these values. Let $|O|$ denote the number of objects that are clustered, then the silhouette coefficient for an entire clustering result (denoted by SC) is defined as Eq. (7):

$$\mathrm{SC} = \frac{1}{|O|} \sum_{o \in O} s(o). \tag{7}$$

As higher values of the silhouette coefficient are more desirable, one way to choose the appropriate number of clusters is by selecting the result which has the highest average silhouette coefficient over all objects (Rousseeuw, 1987).

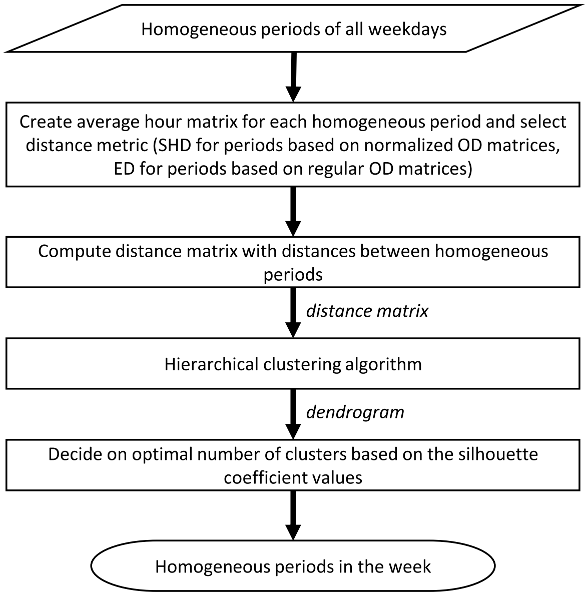


Fig. 3. Flowchart describing the proposed process for finding homogeneous periods during the week.

2.2. *Homogeneous periods during the week*

With the methods described in Section 2.1, we can determine for each day separately a certain number of contiguous homogeneous periods (based on the values of the silhouette coefficient). However, (parts of) certain days of the week might be very similar in terms of demand patterns. For example, the morning peak period on Tuesday might be similar to the morning peak on Thursday, because many people work at the office on those days. Therefore, in this section we describe the method used to compare the clusters found with the method described in Section 2.1. In Fig. 3, a flowchart is given that displays the proposed process for finding homogeneous periods during the week. The remainder of this section will describe this process as well.

The clustering algorithm used for determining homogeneous periods during the week is the same as the one described in Section 2.1, with one exception: the constraint to find only contiguous periods is dropped. Hence, in each iteration of the clustering algorithm any pair of clusters is eligible to be merged into a new cluster, not only the clusters that are contiguous in time. At this stage, we are looking for homogeneity during the week to determine if for example the service plan for a morning period on Tuesday can also be used for a morning period on Thursday. This cannot be found if the contiguity constraint would be maintained, therefore it is dropped.

The input for the algorithm is also slightly different compared to determining periods of homogeneous demand during the day. To keep the size of the clustering instance to an acceptable level, we construct an OD matrix for each of the determined periods. Each period consists of one or more subperiods which have their own OD matrices. The OD matrices of these subperiods are used as input in Section 2.1. An OD matrix for a larger period is established by adding up the OD matrices of all its subperiods and subsequently divide each cell by the number of subperiods. This division is necessary to be able to compare periods consisting of a different number of subperiods more fairly.

The distance metrics used to create the distance matrix are the same as before. For the cluster results based on regular OD matrices, a distance matrix is created using the Euclidean distance. For the cluster results based on normalized OD matrices, the periods' OD matrices are transformed to normalized OD matrices using Eq. (1) and a distance matrix is created using the squared Hellinger distance.

3. Case study data

The hierarchical clustering methods described in Section 2 require multiple time-dependent OD matrices as input. The time-span that is covered in an OD matrix can vary, but should be sufficiently small to get the desired level of detail, e.g., per hour or per half hour. Public transport operators that use smart card automatic fare collection systems, usually possess the data to create such time-dependent OD matrices (Pelletier et al., 2010), since smart cards often register time and location of check-in and/or check-out. This is also the case at the Dutch railway undertaking NS, which provides the data for the case study in this paper. NS is the principal passenger railway undertaking in the Netherlands and serves 253 Dutch stations. Every working day in 2019, NS facilitated over 1.3 million passenger journeys (Nederlandse Spoorwegen, 2020). In 2019, NS operated essentially a cyclic timetable with a cycle time of 30 min. Adjusting the train services to the changing demand is mainly done by varying the length of the trains. Besides that, extra peak train lines are operated at a small part of the network and in the evening hours (after 22:00) the frequency of the service

Table 1
Sample of fictitious OD data.

Date (d)	Start time of 30 min period (t)	Check-in station (i)	Check-out station (j)	Number of trips ($A_t^d(i,j)$)
01/01/2019	00:00	1	2	1
18/04/2019	15:30	34	196	53
31/12/2019	23:30	253	252	0

Table 2
Data selection process.

	Number of days removed in step	Number of remaining days
		365
No weekend days	104	261
No school holidays	86	175
No public holidays	4	171
No (un)scheduled major service withdrawals	17	**154**

is gradually reduced. Given the great regularity of the schedule and the limited amount, size and frequency of the peak lines, the schedule is not expected to hinder the interpretation of the clustering result.

At NS, the main way in which the travel fare is calculated is via a check-in and check-out system using a smart card. Data recorded by this smart card includes the origin and destination station of the trip, and the check-in and check-out times. For a full description of the data recorded by the smart card, we refer to Van Oort et al. (2015). Since the check-in and check-out times and locations are recorded, the data can be easily translated into a time-dependent OD matrix. Besides using a (disposable) smart card, people can also order a ticket online. With these tickets, the origin, destination and travel date of the trip are known, but the exact travel time is not. The e-tickets are allocated over the day based on the distribution of trips made with similar smart-card products. In principle, ticket prices at NS are solely based on the distance travelled and not on time of day or the location in the network. However, to encourage people to travel outside the peak hours, NS has subscriptions that offer a discount on the trip price in the off-peak hours for a monthly fee. To get this discounted price, passengers need to check in outside of the defined peak hours: 6:30–9:00 and 16:00–18:30.

The data that is available to this research is the OD data in 2019, aggregated to half hour periods. A fictitious sample of the provided OD data is given in Table 1. The dataset contains for each half hour of each of the 365 days in 2019 how many trips were made between every possible OD pair. A possible OD pair is any combination of the 253 stations served by NS, which gives $253 * 252 = 63{,}756$ possible OD pairs. The check-in time determines in which half hour a trip is recorded. Hence, from the second row in Table 1 we conclude that on April 18, 2019, there are 53 trips from station 34 to station 196 that started between 15:30 and 15:59. The data from Table 1 can be used to fill OD matrices. As we have OD data for every day of the 365 days in 2019 and for every of the 48 half hours of those days, we can create $365 * 48 = 17{,}520$ OD matrices. For example, the data in the 2nd row of Table 1 belongs to OD matrix $A_{15:30}^{18.04.2019}$, and $A_{15:30}^{18.04.2019}(34,196) = 53$. The data of 2019 is used to exclude any effects of the COVID-19 pandemic. Furthermore, as we are interested in the passenger travel patterns during a normal workday, several days are excluded from the dataset, such as weekend days, school holidays, public holidays, and several (un)scheduled withdrawals from service as reported in NS' Annual Report of 2019 (Nederlandse Spoorwegen, 2020). Table 2 provides an overview of the number of days removed and the number of remaining days after each type of non-normal day is removed from the data. After removing all these outlier dates, there are 154 days left. Let D denote the set of all dates left in the dataset. Since we also want to compare the demand patterns of different workdays, we split up set D into five sets based on the day of the week. Let these subsets of D be denoted by $D^{\mathrm{Mo}}, D^{\mathrm{Tu}}, D^{\mathrm{We}}, D^{\mathrm{Th}}, D^{\mathrm{Fr}}$, for Monday until Friday, respectively. As we have the data of 154 days, distributed over five workdays, we have data of approximately 31 days per workday. Furthermore, we only focus on the period between 06:00 and 23:59, since this is the period in which most of the journeys occur. Let T denote the set of all half hour periods between 06:00 and 23:59.

As we are interested in the travel patterns at ordinary workdays, we create a median OD matrix, by taking for each workday and for each half hour the median number of trips between each OD pair. So for every half hour period $t \in T$ and for every workday $x \in \{\mathrm{Mo}, \mathrm{Tu}, \mathrm{We}, \mathrm{Th}, \mathrm{Fr}\}$ we create an OD matrix. The entry of the OD matrix for day x for period t for the trip between origin station i and destination station j is calculated as:

$$A_t^x(i,j) = \mathrm{med}_{d \in D^x}\{A_t^d(i,j)\}, \tag{8}$$

where med$\{\cdot\}$ denotes the calculation of the median value. The median is chosen instead of the average, to reduce the effect of outliers. With 253 stations in the network, there are many OD pairs that are used only sporadically. By creating median OD matrices, instead of average OD matrices, the number of trips between these origins and destinations will be equal to zero instead of a very small positive number. As usually these OD trips are not made, the (median) value of zero seems the best choice. After an OD matrix is created for each day $x \in \{\mathrm{Mo}, \mathrm{Tu}, \mathrm{We}, \mathrm{Th}, \mathrm{Fr}\}$ and each time period $t \in T$, the normalized OD matrices $\bar{A}_t^x$ can be created using Eq. (1).

Table 3
Distances of the morning matrix to the regular afternoon/evening matrix (column 2), the transposed afternoon/evening matrix (column 3), and matrices of other days (column 4).

Day	Distance (SHD) of morning matrix to		
	Afternoon/evening matrix	Transposed afternoon/evening matrix	Closest other matrix (closest matrices)
Monday	0.2517	0.0238	0.0067 (Mo.A-Th.A)
Tuesday	0.2190	0.0144	0.0079 (Tu.A-Mo.A)
Wednesday	0.2319	0.0399	0.0060 (We.A-Th.A)
Thursday	0.2316	0.0237	0.0060 (Th.A-We.A)
Friday	0.1903	0.0358	0.0123 (Fr.M-Th.M)

Note that although median OD matrices are used in this case study, the method presented works for any type of OD matrix. For example, if an operator feels that the data of one particular week is representative enough for the demand, then just the realized OD matrices of that week can be used as input. Alternatively, if an operator wants to know what demand patterns can be found in the forecasted demand, then an OD matrix with forecasted demand can be used as input. In this case study, we want to find demand patterns during regular workdays in 2019. Therefore, a median OD is selected as the best way to represent this regular demand.

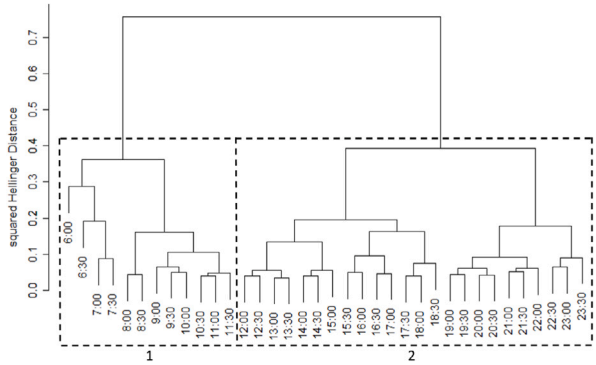
4. Results

This section presents the results of applying the methods described in Section 2 to the data described in Section 3 in order to find periods that are homogeneous in demand. The homogeneous demand periods within the different workdays are presented in Section 4.1. Next, Section 4.2 presents how homogeneous the demand is during the week. Lastly, in Section 4.3 the differences between the identified periods are illustrated by looking at some demand characteristics in the different periods.

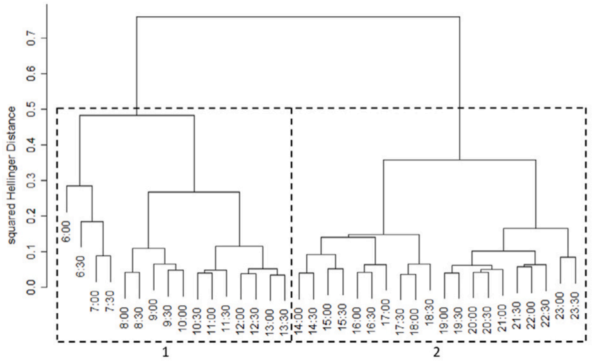
4.1. Homogeneous periods during the day

The results of the hierarchical clustering method can be visualized in a dendrogram plot, which shows the grouping of the (normalized) OD matrices. The dendrograms of the normalized and regular OD matrices of the five median workdays in 2019 are given in Figs. 4 and 5, respectively. The dendrograms should be read as follows. The horizontal axis displays the time, where for example 6:00 denotes the OD matrix containing all trips that start between 6:00 and 6:29. All matrices start in separate clusters (as can be seen at the bottom of the dendrogram) and in every step two matrices are clustered together until all matrices are in a single cluster (as can be seen at the top of the dendrogram). The merging of two clusters is denoted in the dendrogram by a horizontal line connecting the two clusters. Note that when looking at homogeneous periods during the day, only two clusters that are adjacent in time are allowed to be combined into a new cluster. The vertical axis denotes the distance between two clusters at the moment they are combined. For each day, the cluster result with the highest silhouette coefficient is chosen as the result with the optimal number of clusters. This result is visualized in the dendrogram using dashed lines and each of these clusters is numbered for easier reference in the text. For example, for the Monday (see Fig. 4(a)) the result with two clusters received the highest silhouette coefficient (namely 0.54), which resulted in the clusters 6:00 until 11:59 and 12:00 until 23:59.

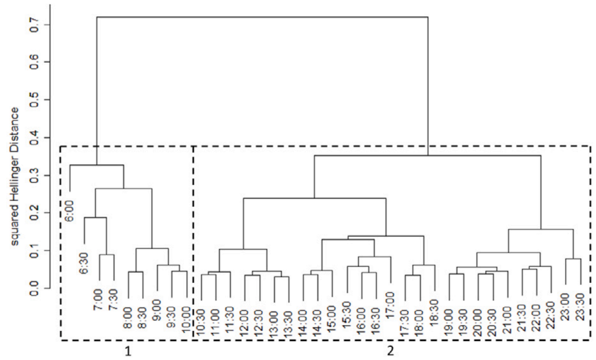
When looking at the clustering results of the normalized OD matrices (see Fig. 4), the results with the highest silhouette coefficient are the results that divide each workday into two parts: the morning hours on the one side (cluster 1) and the afternoon and evening hours on the other side (cluster 2). The transition from morning to afternoon is usually at 12:00, but is later on Tuesday (at 14:00) and earlier on Wednesday (at 10:30). A possible explanation for this division between morning and afternoon/evening is that people usually go somewhere in the morning and return to where they came from in the afternoon or evening of the same day, which is also clearly visible in Fig. 1. To check the effect of this phenomenon on the clustering result, we compare the distance between the morning OD matrix and the afternoon/evening OD matrix to the distance between the morning OD matrix and the transpose of the afternoon/evening OD matrix. The results for this analysis are given in Table 3. The columns from this table show from left to right: the day, the squared Hellinger distance (SHD) between the morning and the afternoon/evening matrix, the SHD between the morning and the transposed afternoon/evening matrix, and the shortest distance from one of the day's matrices to one of the other days' matrices. This last column also denotes between brackets which matrices are the closest together. Here the first two letters denote the day (e.g., Mo stands for Monday) and the last letter denotes whether it is the morning matrix (.M) or the afternoon/evening matrix (.A). So for example for Monday, the distance between the morning and the afternoon/evening matrix is 0.2517, while the distance from the morning to the transposed afternoon/evening matrix is 0.0238. Furthermore, the shortest distance from one of the Monday matrices to another day's matrix is 0.0067, which is the distance between the Monday afternoon/evening matrix (Mo.A) and the Thursday afternoon/evening matrix (Th.A). The results show that the morning matrix is between 5 and 15 times closer to the transposed matrix than to the normal OD matrix of the afternoon/evening. Hence, it seems reasonable that the cluster result at least partly reflects this travel phenomenon. However, it is also clear that the "back and forth"-effect is not the only effect happening during the day. For each day the distance to another day's morning or afternoon/evening matrix is still a lot smaller than to its own transposed afternoon/evening matrix.



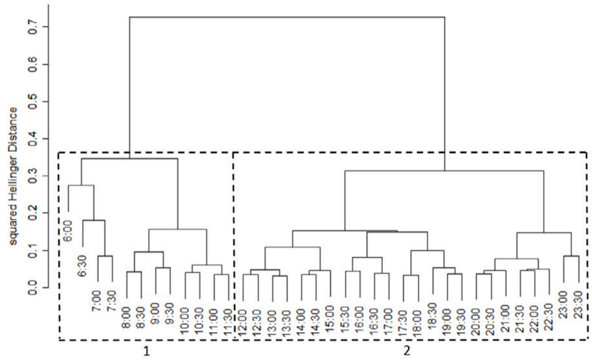
(a) Clustering results Monday - SC = 0.54.



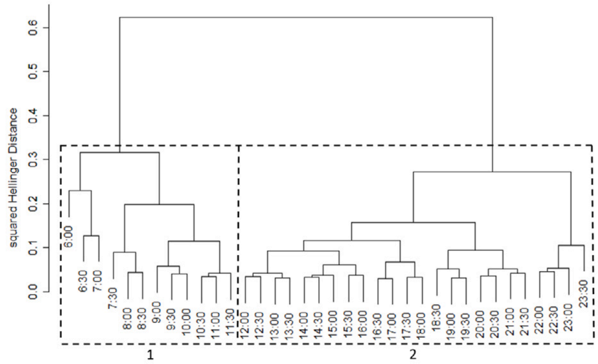
(b) Clustering results Tuesday - SC = 0.46.



(c) Clustering results Wednesday - SC = 0.55.



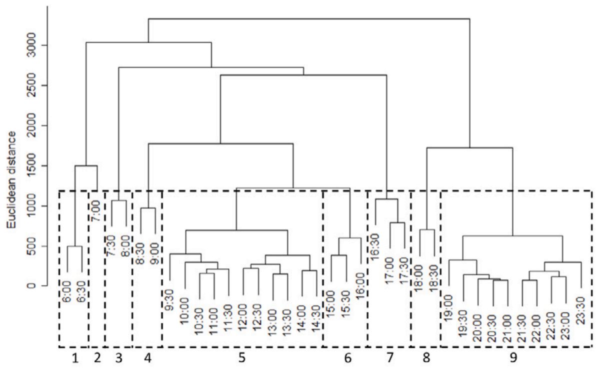
(d) Clustering results Thursday - SC = 0.56.



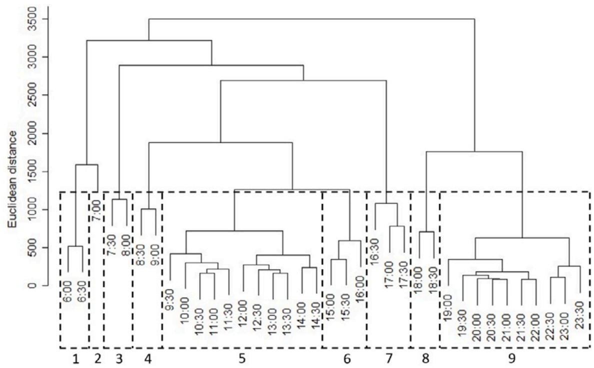
(e) Clustering results Friday - SC = 0.53.

Fig. 4. Clustering results using squared Hellinger distance on normalized OD matrices. Dashed lines denote the optimal number of clusters, according to the silhouette coefficient (SC).

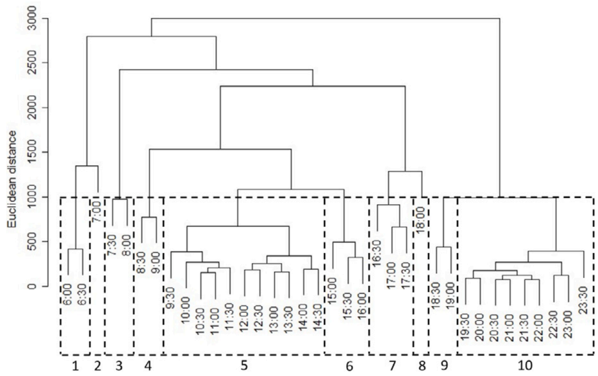
When we take the volume into account by looking at the regular OD matrices (see Fig. 5), the morning versus afternoon/evening pattern does not emerge. The recommended number of clusters (based on the silhouette coefficient) is also much higher: 10 clusters for the Wednesday, and 9 clusters for the other workdays. In the results, we see larger clusters during the day (from 9:30 until approximately 15:00) and in the evening (from 19:00/19:30 until 00:00). These clusters can be characterized as the midday off-peak (cluster 5 in Fig. 5) and the evening period (cluster 10 on Wednesday and cluster 9 on other days), respectively. Around and during the peak hours we see many small clusters with between one and three matrices per cluster. On each day, the peak hours are divided into three periods, clusters 2, 3, and 4 for the morning peak and clusters 6, 7, and 8 for the afternoon peak. It is likely that the middle of these periods is the hyper peak (clusters 3 & 7), while the other periods can be seen as the shoulders of the peak. If we compare the start times of the different clusters with Fig. 1, we see that the transition between clusters coincides with large changes in the passenger volume at the station. When comparing the clustering results of the different workdays, many similarities can be seen. For example, the first four periods of each workday have the same start and end time. Moreover, on Monday, Tuesday, and Friday, all the periods have all the same start and end times. Another noteworthy characteristic about the times, is that the peak



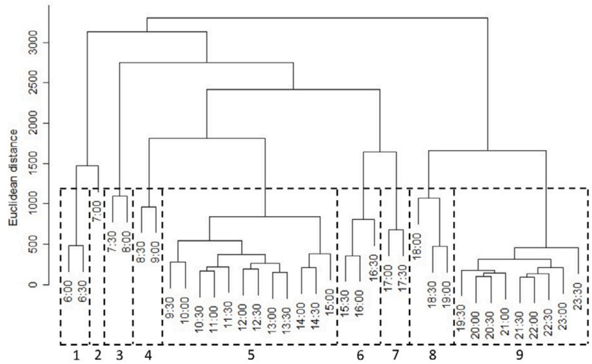
(a) Clustering results Monday - SC = 0.41.



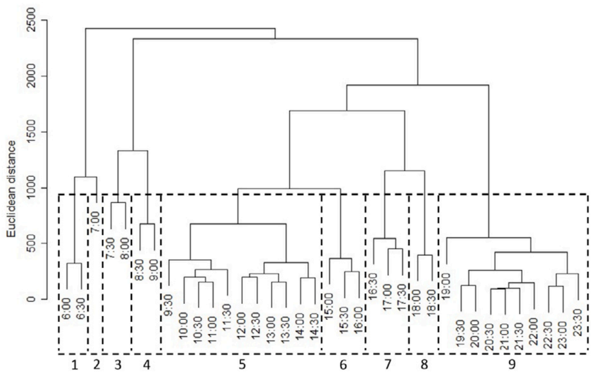
(b) Clustering results Tuesday - SC = 0.41.



(c) Clustering results Wednesday - SC = 0.38.



(d) Clustering results Thursday - SC = 0.39.



(e) Clustering results Friday - SC = 0.39.

Fig. 5. Clustering results using Euclidean distance on regular OD matrices. Dashed lines denote the optimal number of clusters, according to the silhouette coefficient (SC).

hours according to the clustering result do not coincide completely with the peak hours that NS uses for pricing. For example, the morning peak hours end at 9:00 according to NS, while the midday off-peak demand only starts at 9:30 according to the clustering result. A similar result is seen at the afternoon peak. The peak hours used for pricing are 16:00–18:30, while the clustering result indicate that the peak hours start already at 15:00 (15:30 on Thursday) and end on most days at 19:00 (18:30 on Wednesday, 19:30 on Thursday). This result is not surprising: the pricing strategy is used to encourage people to travel outside the busiest periods. So, this result shows that the pricing strategy is working to some extent. In the next section, we check whether these periods are not only similar in start and end time, but also in terms of demand during those periods.

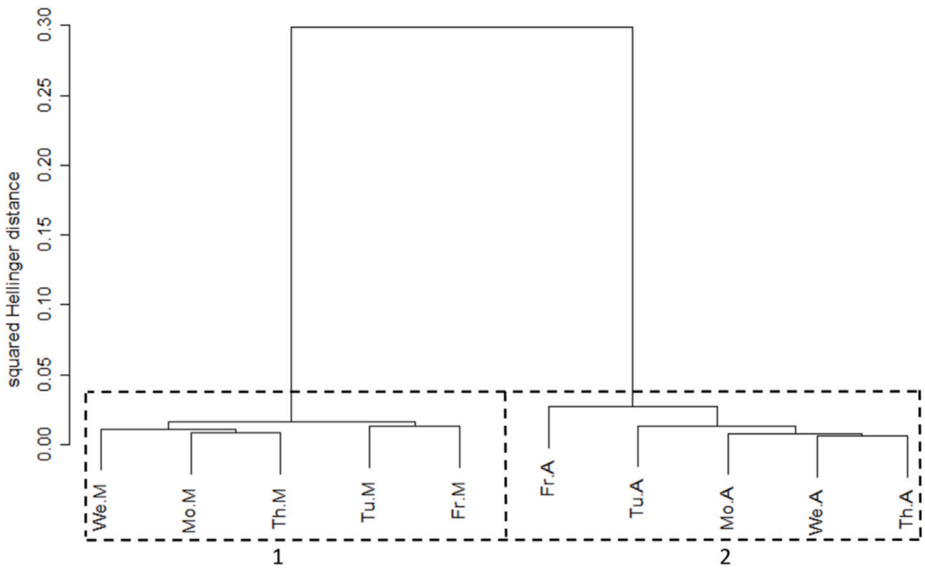


Fig. 6. Clustering results for homogeneity during the week using squared Hellinger distance on normalized OD matrices. Dashed lines denote the optimal number of clusters, according to the silhouette coefficient (SC = 0.94).

4.2. *Homogeneous periods during the week*

In Section 4.1, for each workday a number of homogeneous periods was determined in two different ways: based on the structure of the demand (using normalized OD matrices) and based on the structure and volume of the demand (using regular OD matrices). In this section, the demands in these periods are compared, to see if similarities exist across days of the work week.

First the periods based on the structure of demand are compared. As described in Section 3, this is done by creating a new OD matrix for each of the periods, and using these as input for the clustering algorithm. The constraint that contiguous periods must be formed is not taken into account at this stage. The dendrogram produced by the clustering algorithm is shown in Fig. 6. In this figure, on the horizontal axis the different periods are depicted. In Section 4.1 we have found for each day two periods, which can be roughly viewed as a morning period and an afternoon/evening period. This is also reflected in the labelling on the horizontal axis: the first two letters denote the day (e.g., Mo stands for Monday) and the last letter denotes whether it is the morning matrix (.M) or the afternoon/evening matrix (.A). As before, the vertical axis displays the distance between two clusters at the moment when they are combined in a single clusters. This distance is measured by the squared Hellinger distance.

Looking at Fig. 6, there is a clear division between morning and afternoon/evening matrices. When looking at the silhouette coefficient, the optimal number of clusters is two, where the morning matrices are in cluster 1 and the afternoon/evening matrices are in cluster 2. The silhouette coefficient corresponding to this clustering result is 0.94, which is very high since the silhouette coefficient can take a value between −1 and 1. The division in morning and afternoon/evening periods is not surprising given the analysis of the "back and forth"-effect in Section 4.1. This analysis shows that morning matrices and afternoon/evening matrices of the same day are relatively far apart, especially when compared to the distances to morning or afternoon/evening matrices of other days. Furthermore, this division is also illustrated in Fig. 1: both on Tuesday and on Friday people travel towards the station mainly during the morning peak period, and travel back mainly during the afternoon peak period. Hence, it makes sense that a morning matrix is more similar to another morning's matrix than to an afternoon/evening matrix.

Next, the periods based on the structure and volume of the demand are compared. The dendrogram produced by the clustering algorithm is shown in Fig. 7. These periods are labelled based on their occurrence during the day, so Mo.3 denotes the third cluster on Monday. In Fig. 5 we can find that this is the cluster from 7:30 until 8:29.

Based on the silhouette coefficient values, we find that the optimal number of clusters is twelve. These clusters are denoted by dashed lines in Fig. 7. Note that the majority of these clusters only contains matrices of the same time of the day. For example, the third cluster in Fig. 7 contains the first period (.1) of each weekday. This homogeneity within the determined clusters implies that the demand develops similarly during the different weekdays.

Besides providing the optimal clusters, the shape of the dendrogram also gives information about the similarity of different clusters. When starting at the top of the dendrogram in Fig. 7 and going down, the first division that is made is between low volume periods and the high volume (peak) periods. On the left side of the dendrogram, we see the three periods with a relatively low volume. From left to right, these periods can be characterized as the midday off-peak period (.5; ±9:30–15:00), the evening off-peak period (.9 or .10; ±19:00–00:00), and the early morning (.1; ±6:00–7:00). Note that although these periods are in separate clusters, we can see from the height of the horizontal lines joining them that they are quite close to each other and quite far apart from the rest of the periods. On the right side of this first division in the dendrogram, we see the peak periods. When zooming in on

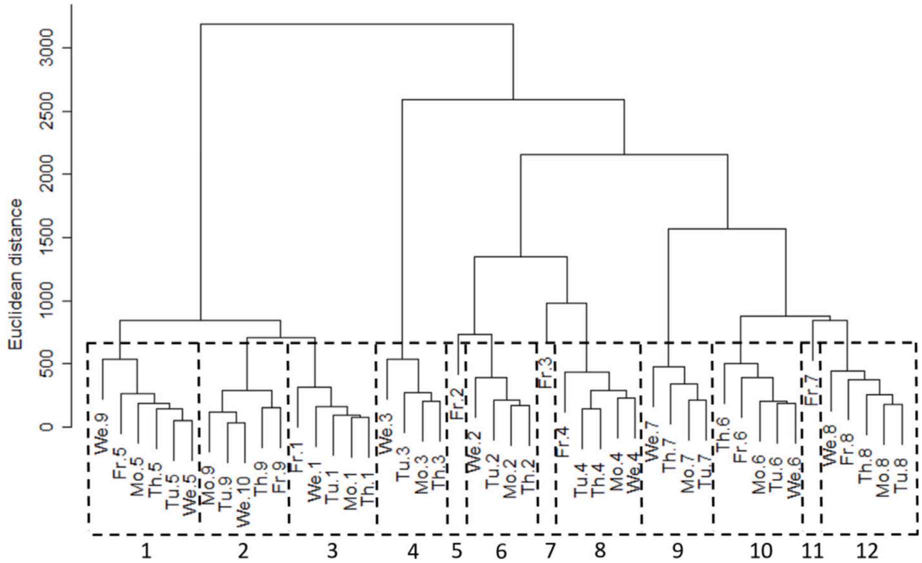


Fig. 7. Clustering results for homogeneity during the week using Euclidean distance on regular OD matrices. Dashed lines denote the optimal number of clusters, according to the silhouette coefficient (SC = 0.54).

the peak periods, we see a division into three larger groups: hyper morning peak (.3; 7:30–8:30), the rest of the morning peak (.2 and .4; 7:00–7:30 and 8:30–9:30) and the evening peak (.6, .7 and .8; ±15:00–19:00). Note that the shoulders of the peak periods (e.g., .2 & .4 or .6 & .8) are more similar to each other than to the main peak period (.3 or .7), but are not so similar that they end up in the same cluster.

Although we see many similarities in the demand across the different days, the peak demand on Friday does not follow the same pattern as the peak demand on other workdays. Several of Friday's peak periods ended up in its own cluster (Fr.2, Fr.3 and Fr.7 are in clusters 5, 7, and 11). Furthermore, Friday's evening hyper peak period (Fr.7) is not clustered with the rest of the .7 periods, but instead is more similar to the evening peak shoulders (.6 and .8). The same holds for Friday's morning hyper peak period (Fr.3) which is closer to the morning peak shoulders (.4 and .2) than to the morning hyper peak of the other workdays (.3). Hence, Friday does not really show a hyper peak demand pattern. This result is not surprising, since people who work part-time often do not work on Fridays and Friday is a popular day to work from home (even before the COVID-19 pandemic).

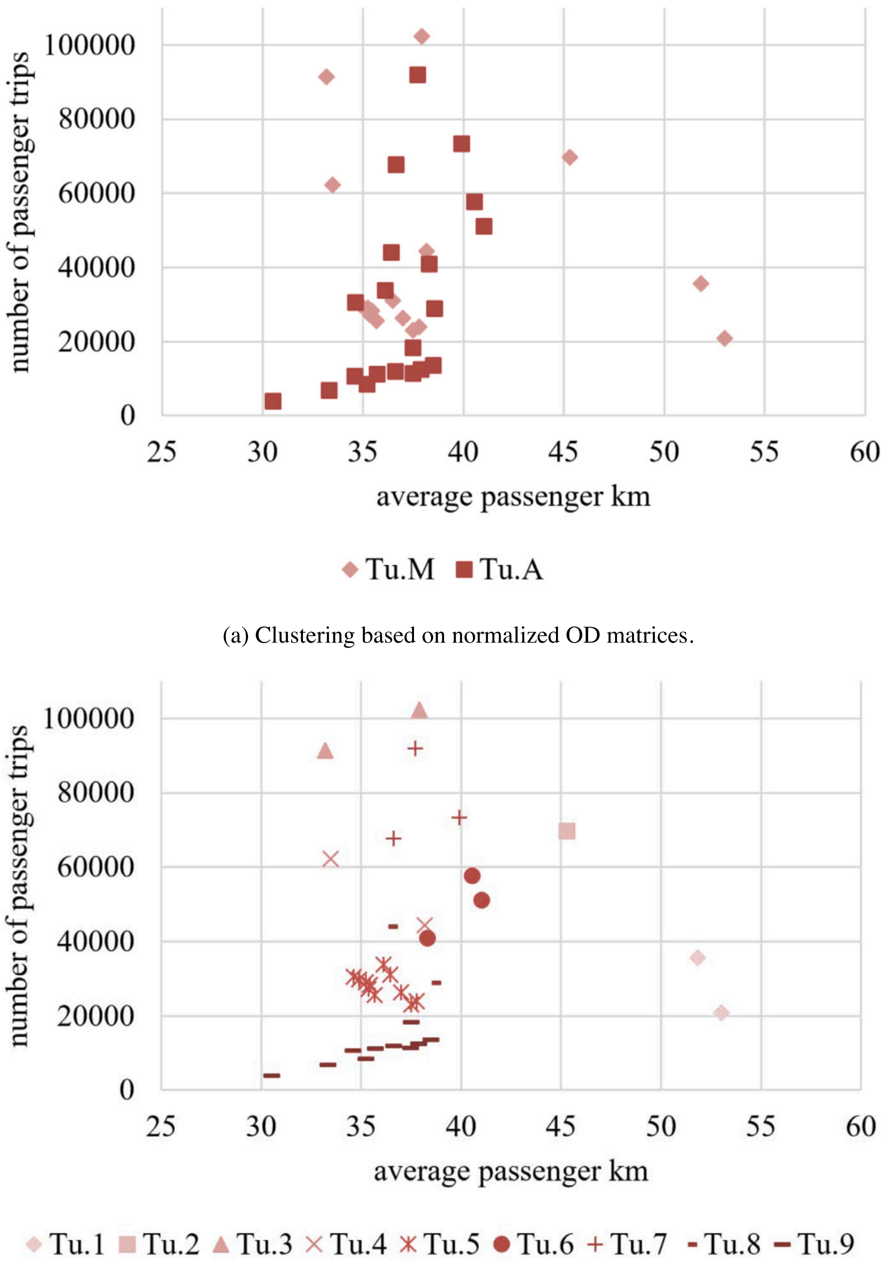
4.3. Demand characterization

In Sections 4.1 and 4.2, we showed which periods are homogeneous in demand. However, the cluster results on its own do not give a good understanding of how the demand is changing throughout the day. Therefore, in this section we will illustrate this by providing some demand characteristics for the demand on Tuesday. The Tuesday is chosen because it is one of the most popular days of the week to travel by train, together with Thursday. We chose Tuesday instead of Thursday, since when we look at the clustering results, Tuesday has more similarities with the other days than Thursday and hence is more representative.

Besides giving a better understanding of changes in the demand, looking at demand characteristics of the found periods can also validate the clustering results. Investigating the demand characteristics can validate the clustering results, since when the demand is the same in two different adjacent periods, then there is no valid reason to define multiple periods. Furthermore, the changes in demand characteristics can provide input to the service plan. If it is known how demand differs throughout the day, then this also provides information about how the provided service should change throughout the day to match this demand.

According to Balcombe et al. (2004), demand can be described by the total trip volume and the distance travelled, where the distance travelled is usually measured in passenger kilometres. Therefore, to show how the demand is changing throughout the day, for each half hour OD matrix of the Tuesday the total number of trips and the average number of passenger kilometres travelled in those trips is calculated. These results are then plotted in a scatter plot, which is given in Fig. 8. In this figure, each of the 36 half hour OD matrices is a data point, where the average passenger kilometres travelled is displayed on the horizontal axis, and the total number of passenger trips is displayed on the vertical axis. To also visualize how the demand changes between periods, a different icon and colour is used for each period. In the plot, the light colour is used for the first period of the day, and the dark colour is used for the last period of the day. As we have two types of cluster results (namely based on the structure of the demand and based on the structure and volume of the demand), we also have two scatter plots (Figs. 8(a) and 8(b), respectively).

The first thing that stands out from Fig. 8(a) is that periods Tu.M and Tu.A both show great variation in the demand characteristics. That periods Tu.M and Tu.A have a large variation in volume is not surprising. The clustering that found Tu.M and Tu.A is solely based on the structure of the demand, so the volume of demand is not taken into account at all. Therefore, it



(a) Clustering based on normalized OD matrices.

(b) Clustering based on regular OD matrices.

Fig. 8. Average passenger km and volume for each half hour between 6:00 and 23:59 on the Tuesday. The symbols indicate in which period a half hour is placed. The two subfigures correspond to the two types of clustering results: based on normalized OD matrices (a) and based on regular OD matrices (b). The colour of the symbol relates to the time of day, where the light colour corresponds to the first and the dark colour to the last period of the day.

would have been more surprising if this clustering would provide a result in which the volumes are also neatly divided into different clusters. However, especially period Tu.M also displays a large variation in average passenger km, which could be seen as a (rough) measure of the structure of the demand. Thus, it would most likely be quite difficult to find a service plan that satisfies the demand in each half hour of Tu.M.

Fig. 8(b) gives a much more detailed view of how the demand is changing throughout the day than Fig. 8(a). For example, we see that the early morning period Tu.1 (6:00–6:59), the midday off-peak period Tu.5 (9:30–14:59), and the evening period Tu.9 (19:00–23:59) are the periods with the lowest average volume. Furthermore, we see that during the morning peak (periods Tu.2, Tu.3, and Tu.4) and the evening peak (periods Tu.6, Tu.7, and Tu.8) the volume rises to a peak in the hyper peak periods Tu.3 and

Tu.7, before declining again. Two data points that really stand out, are the data points from the early morning period Tu.1. Within this period there is not only a relatively low volume, but also a high average passenger km. Most likely, these are people who live far away from their job, and hence have to leave earlier to arrive on time at their job. A similar trend cannot be seen at the end of the workday, likely because at that time there are also more short-distance travellers in the train. However, note that the average passenger km during the evening peak periods (Tu.6, Tu.7, Tu.8), are higher than the average passenger km during the majority of the morning peak periods (Tu.3 and Tu.4), which might be due to the passengers from periods Tu.1 and Tu.2 returning home.

When comparing Figs. 8(a) and 8(b), the clustering based on the structure and volume of the demand is more compact compared to the clustering based solely on the structure of the demand. When looking at the different volumes within clusters, in Fig. 8(a) the cluster with the largest range in volume is Tu.A. The difference in volumes between the half hour period with the highest volume and the period with the lowest volume is 88 056. When we look at Fig. 8(b), the difference in volume within the clusters is much smaller. The cluster with the highest difference is cluster Tu.7, which has a difference in volume within the cluster of 24 272. So the difference in volume within the clusters is much lower for the result based on regular OD matrices compared to the result based on normalized OD matrices. Similarly, if we look at the difference in average passenger kilometres within the different clusters, the difference is much lower in the results based on regular OD matrices. The highest range in average passenger km is 7.96, which is observed in cluster Tu.9. However, if we look at the result based only on the structure of the demand, we find a highest range of 19.82 within cluster Tu.M. Hence, the cluster results based on regular OD matrices are more compact in terms of volume and structure compared to the cluster results based on normalized OD matrices. Furthermore, if data points are close to each other in Fig. 8(b), we see that either these points are in the same period, or that these points are in periods that are not adjacent to each other and hence cannot be combined due to the contiguity constraint. In Fig. 8(a), this is not completely the case. For example, if we look at the points that are contained in Tu.5, we see that these points are split between Tu.M and Tu.A in Fig. 8(a). As the points in Tu.5 are very close together in terms of volume and average passenger km, it is questionable whether different service plans are needed to serve this demand. Therefore, the periods based on clustering regular OD matrices seem more suitable as input for creating demand-responsive railway schedules than the periods based on normalized OD matrices.

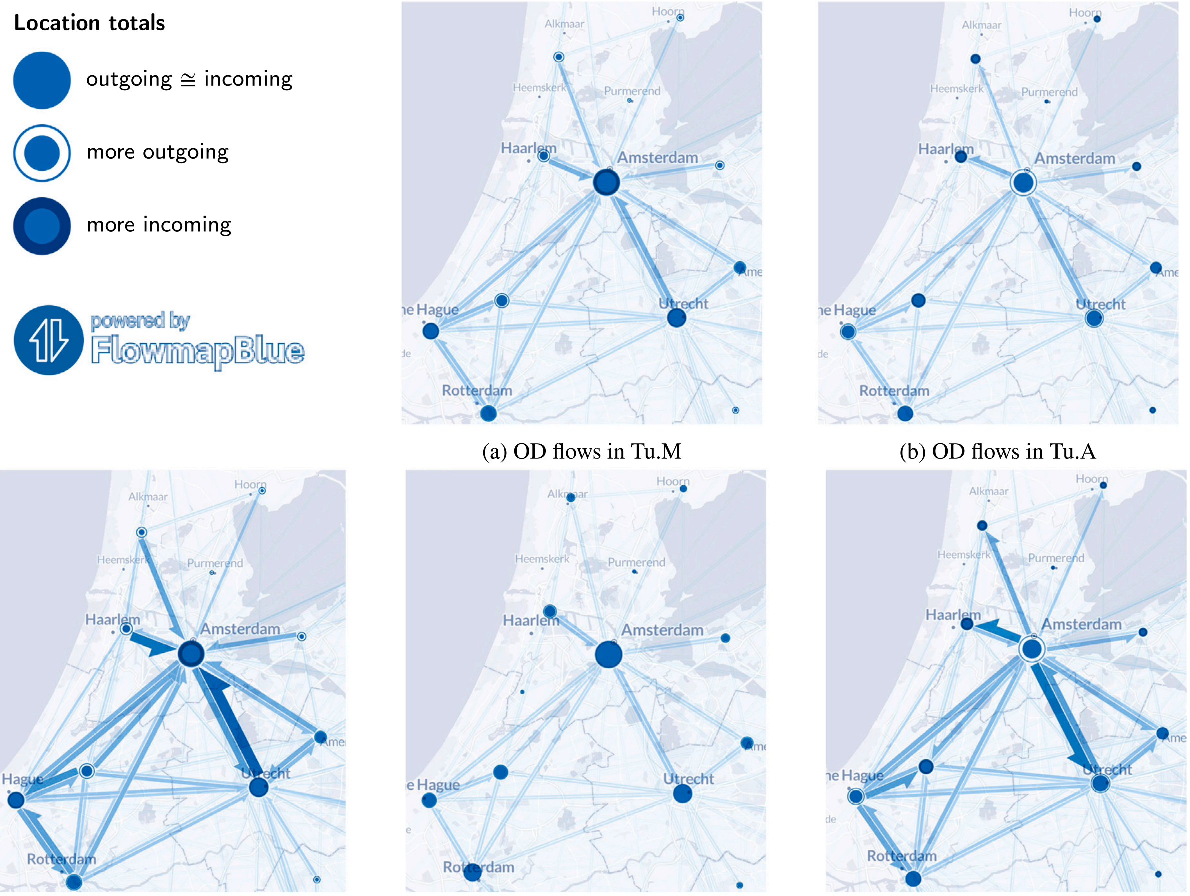
To strengthen this conclusion, we visualize the OD flows in the different clusters. These visualizations are provided in Fig. 9 and are made using the open-source software FlowmapBlue (Boyandin, 2019). To create these visualizations, an average one-hour OD matrix was created per cluster. Furthermore, to improve the readability of the figures, only a part of the network that includes the 4 largest cities of the Netherlands is displayed and nearby stations are grouped together. In this figure, the arcs display OD flows, where the colour and width of the arcs denote the volume of the OD flows. Furthermore, each group of stations is denoted by a circle on the map, where the size of the circle depends on the total volumes going to or departing from those stations. As shown by the legend in the top left corner, the colour of this circle also denotes whether the outgoing and incoming flows are balanced or if either the outgoing or incoming flow is higher.

Figs. 9(a) and 9(b) display the OD flows of the cluster result based on the normalized OD matrices, where a morning and afternoon/evening cluster were found. These figures give further confirmation of the "back-and-forth" effect discussed in Section 4.1. In the morning (Fig. 9(a)), the cities Amsterdam, Utrecht and The Hague all have more incoming flow than outgoing flow, while in the afternoon (Fig. 9(b)) this is reversed. Furthermore, the direction of the flows reverses when we compare the morning to the afternoon. For example, in Fig. 9(a) many people want to go from Haarlem to Amsterdam in the morning, while the other direction is not so popular. However, in the afternoon, the widest arrow points from Amsterdam towards Haarlem instead. We see similar a similar pattern in the flows between Alkmaar, Hoorn and Utrecht on the one side and Amsterdam on the other side.

From the clustering result based on regular OD matrices, we visualize three of the nine clusters: the morning hyper peak (Fig. 9(c)), the midday off-peak (Fig. 9(d)), and the afternoon hyper peak (Fig. 9(e)). When we compare the three figures, we see a lot of difference in volumes between the three clusters, where the hyper peaks have a much higher demand than the midday off-peak period. Furthermore, similar to Figs. 9(a) and 9(b), we see a "back-and-forth" effect in the two peak hours. However, in the midday off-peak period this effect is not visible. In Fig. 9(d), most station groups have an incoming flow that is approximately equal to the outflow, as displayed by the single-coloured circles. Furthermore, for each OD pair in the figure, the arcs have approximately the same width and colour for both directions. From this we conclude that the demand in the midday off-peak is not only lower in volume, but also much more spatially balanced than the demand in the peak hours. Moreover, the midday off-peak is quite a long period (from 9:30 to 14:59). Hence, when we want to determine input for creating demand-responsive railway schedules, it seems logical that the midday off-peak gets its own cluster instead of being divided between a morning and afternoon cluster, as in the result based on normalized OD matrices. Since these visualizations show that there is more variation in demand during the day than just the "back and forth"-effect, we conclude that in this case the clusters based on regular OD matrices are better suited to be used as input for creating demand-responsive railway schedules.

5. Conclusions and recommendations

In this paper, we investigated how homogeneity in railway demand can be discovered. This method could be useful for RUs that would like to create service plans (line plans and timetables) that better match the changing demand throughout the day and week, but still need a lot of manual labour to create their service plans. By determining the periods with homogeneous demand, the RU can reduce the amount of periods during the day that need to be considered for a different service plan. There have been several studies about homogeneous demand in bus services, but so far the academic literature for railway demand has been lacking. This paper contributes to the academic literature by evaluating if the method for determining periods of homogeneous demand in bus services also work for railway demand. Furthermore, this paper extends those methods to create homogeneous periods not only



(a) OD flows in Tu.M (b) OD flows in Tu.A

(c) OD flows in Tu.3 (d) OD flows in Tu.5 (e) OD flows in Tu.7

Fig. 9. Visualization of OD flows in different clusters. Figs. 9(a) and 9(b) show the OD flows from the clusters based on normalized OD matrices and Figs. 9(c), 9(d), and 9(e) show the OD flows from the morning hyper peak, midday off-peak, and afternoon hyper peak clusters based on regular OD matrices.

based on the demand structure, but also the demand volume, since volume is an important aspect when creating railway service plans.

The methods in this paper are demonstrated on a case study from the Dutch passenger railway operator NS. When applying the clustering method on the normalized OD matrices (containing the structure of demand), each workday was split into two periods with the switch between periods around 12:00. An explanation for this is that people usually travel somewhere in the morning, and travel back in the afternoon or evening. The clustering result therefore provides some evidence that a symmetric service plan, in which each train line is operated with the same frequency and stops in both directions, is not the optimal choice when strictly looking at the demand structure. Applying the clustering method on the regular OD matrices (containing both structure & volume of demand) results in these periods: morning peak, midday off-peak, afternoon peak, and evening. The peak periods are more diverse, and hence divided into smaller time intervals, than the off-peak periods. The results of this analysis help to determine the time at which switching between service plans makes sense. The switching times between periods are quite consistent throughout the week. Hence, switching between service plans can happen at the same times every day, which is favourable for the customers. When looking at demand characteristics such as volume and average passenger kilometres per period, the introduced clustering method based on the regular OD matrices provides periods that are more compact and hence more suitable for creating railway service plans than the clustering method from the bus literature, which was based on the normalized OD matrices.

The analysis for homogeneity in demand during the week shows that the days of the week are quite similar. However, when looking at the structure & volume of the demand, the peak demand on Friday forms an exception to this. These results suggest that the service plan for Friday's morning and evening peak hours could be different from the service plan for the other weekdays' peak hours. The differences between days might increase further due to the COVID-19 pandemic (e.g, because people will work from home more on Wednesdays and Fridays). If this is the case then potentially the demand on other days will also change enough to warrant a separate service plan.

The methods discussed in this paper can be used by public transport authorities and operators to get a better understanding of their demand. The method can be especially useful for operators that would like to make their schedule more demand-responsive, but do not have the ability and/or desire to create a fully flexible schedule. Firstly, the clustering results provide information about the maximum number of periods that need to be considered during the day/week when creating demand-responsive schedules. In time intervals in which the demand is homogeneous, the demand can be served effectively by the same service schedule. Since there is a cost to creating and switching between service plans, the number of periods with homogeneous demand gives us the maximum number of schedules that we need to consider. However, this does not mean that it is also optimal to use this number of schedules in practice. For example, if some of the identified periods cover a very short time period (e.g., 30 min), it might not be beneficial to create a separate schedule for this. Secondly, the clustering result provides information on when the switch should be made between different schedules. For example, in the presented case study the midday off-peak period starts at 9:30, while the pricing peak hour at NS ends at 9:00. Therefore, if the RU wants to create a separate schedule for the morning peak hours, this result shows that it is better not to let the end of this peak period coincide with the end of the pricing peak period, but instead postpone it with 30 min. As a final note, we want to point out that the train passenger demand is sensitive to the train schedule and the pricing. Therefore, when a RU makes changes to its schedule or pricing strategy, this will influence the demand, which might change the clustering result as well. One way to deal with this is by using price and travel time elasticities to forecast the new demand and redo the clustering with this forecasted demand. If these clusters are in line with the new schedule, or with the old schedule if the pricing strategy is changed, then we can be more confident that the schedule can serve the new demand well. If instead the new clusters are not in line with the new schedule, further adjustments might be necessary.

Since travel behavioural changes are expected due to the COVID-19 pandemic, it would be very interesting to repeat this study with post-COVID data to see if there are (significant) changes in the results. Furthermore, it would be interesting to test these methods on other datasets, to check the methods' validity. Our case study shows that creating homogeneous periods based on the regular OD matrices provides periods that are more compact and hence more suitable for creating service plans than the clustering based on normalized OD matrices. It would be interesting to see if this result also holds for different application contexts, like bus and metro. Another promising research direction would be to investigate whether adapting the service plan based on the results of such analysis can be beneficial for railway undertakings and its customers. Although the adapted service plan should fit the changing demand better, adapting the service plan also comes with some costs. For example, switching between service plans during the day can be difficult to plan or execute for railway undertakings. Furthermore, customers might find it difficult to get used to schedules that change throughout the day. Therefore, it should be investigated whether the benefits will outweigh these costs. Part of this investigation should deal with what indicators should be considered to go from the clustering result to the optimal number of schedules. The method presented in this paper only determines the maximum number of periods that need to be considered when adapting the line plans and timetables to the changing demand. However, there could be several reasons to further merge periods together, including that the length of certain periods is too short to create a different schedule for, or that the realized reduction in travel time with a different schedule is too small. Hence, further research is needed to determine how we can go from the maximum to the optimal number of periods.

CRediT authorship contribution statement

Renate J.H. van der Knaap: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft. **Menno de Bruyn:** Writing – review & editing, Supervision. **Niels van Oort:** Writing – review & editing, Supervision. **Dennis Huisman:** Writing – review & editing, Supervision. **Rob M.P. Goverde:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Co-author is Editor in Chief of the Journal of Rail Transport Planning and Management (JRTPM) - R.M.P. Goverde

Acknowledgement

This research is supported by the Netherlands Railways (NS), the largest passenger railway undertaking in the Netherlands.

References

Agard, B., Morency, C., Trépanier, M., 2006. Mining public transport user behavior from smart card data. IFAC Proc. Vol. (IFAC-PapersOnline).

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., White, P., 2004. The Demand for Public Transport: A Practical Guide. Transportation Research Laboratory, London.

Boyandin, I., 2019. FlowmapBlue. URL: https://www.flowmap.blue/. (Accessed 7 May 2024).

Briand, A.-S., Côme, E., Trépanier, M., Oukhellou, L., 2017. Analyzing year-to-year changes in public transport passenger behaviour using smart card data. Transp. Res. C 79, 274–289.

De Bruyn, M., Mestrum, D., 2021. Een studie naar een vraagafhankelijke dienstregeling voor NS. URL: https://www.cvs-congres.nl/e2/site/cvs/custom/site/upload/file/cvs_2021/sessie_a/a4/cvs_34_een_studie_naar_een_vraagafhankelijke_dienstregeling_voor_ns_1_2021.pdf. (In Dutch).

Deschaintres, E., Morency, C., Trépanier, M., 2019. Analyzing transit user behavior with 51 weeks of smart card data. Transp. Res. Rec.: J. Transp. Res. Board 2673 (6), 33–45.

Durán-Micco, J., Vansteenwegen, P., 2022. A survey on the transit network design and frequency setting problem. Public Transp. 14 (1), 155–190.

El Mahrsi, M.K., Come, E., Oukhellou, L., Verleysen, M., 2017. Clustering smart card data for urban mobility analysis. IEEE Trans. Intell. Transp. Syst. 18 (3), 712–728.

Eltved, M., Breyer, N., Ingvardson, J.B., Nielsen, O.A., 2021. Impacts of long-term service disruptions on passenger travel behaviour: A smart card analysis from the Greater Copenhagen area. Transp. Res. C 131, 103198.

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques. Elsevier.

Henrion, V., Ingvardson, J.B., Nielsen, O.A., 2023. The influence of the COVID-19 pandemic on travel behaviour in the Greater Copenhagen area. J. Adv. Transp. 2023, 1–12.

Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM Comput. Surv. (CSUR) 31 (3), 264–323.

Ji, Y., Mishalani, R.G., McCord, M.R., Goel, P.K., 2011. Identifying homogeneous periods in bus route origin-destination passenger flow patterns from automatic passenger counter data. Transp. Res. Rec. 2216, 42–50.

Johnson, D., Shires, J., Nash, C., Tyler, J., 2006. Forecasting and appraising the impact of a regular interval timetable. Transp. Policy 13, 349–366.

Kennisinstituut voor Mobiliteitsbeleid, De Haas, M., Hamersma, M., Faber, R., 2021. Thuiswerken tijdens en na de coronacrisis. (In Dutch).

Kieu, L.M., Bhaskar, A., Chung, E., 2015. Passenger segmentation using smart card data. IEEE Trans. Intell. Transp. Syst. 16 (3), 1537–1548.

Le Cam, L., Yang, G.L., 2000. Asymptotics in Statistics: Some Basic Concepts. Springer Science & Business Media.

Lu, A., Reddy, A., 2012. Strategic look at Friday exceptions in weekday schedules for urban transit. Transp. Res. Rec. 2274, 30–51.

Ma, X., Wu, Y.-J., Wang, Y., Chen, F., Liu, J., 2013. Mining smart card data for transit riders' travel patterns. Transp. Res. C 36, 1–12.

Mahmoudzadeh, A., Wang, X.B., 2020. Cluster based methodology for scheduling a university shuttle system. Transp. Res. Rec. 2674, 236–248.

McCord, M., Mishalani, R., Hu, X., 2012. Grouping of bus stops for aggregation of route-level passenger origin-destination flow matrices. Transp. Res. Rec. 2277, 38–48.

Muller, T., Furth, P., 2001. Trip time analyzers: Key to transit service quality. Transp. Res. Rec. 1760, 10–19.

Mützel, C.M., Scheiner, J., 2022. Investigating spatio-temporal mobility patterns and changes in metro usage under the impact of COVID-19 using Taipei Metro smart card data. Public Transp. 14 (2), 343–366.

Nederlandse Spoorwegen, 2020. Nederlandse Spoorwegen, 2019 Annual Report. URL: https://2019.nsannualreport.nl/annual-report-2019/introduction/2019-in-a-nutshell/.

Pelletier, M.-P., Trépanier, M., Morency, C., 2010. Smart card data use in public transit: A literature review. Transp. Res. C 19, 557–568.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.

Schöbel, A., 2012. Line planning in public transportation: Models and methods. OR Spectrum 34 (3), 491–510.

Serafini, P., Ukovich, W., 1989. A mathematical model for periodic scheduling problems. SIAM J. Discrete Math. 2 (4), 550–581.

Ton, D., Arendsen, K., De Bruyn, M., Severens, V., Van Hagen, M., Van Oort, N., Duives, D., 2022. Teleworking during COVID-19 in the Netherlands: Understanding behaviour, attitudes, and future intentions of train travellers. Transp. Res. A 159, 55–73.

Van Hagen, M., De Bruyn, M., Ton, D., Severens, V., Duives, D., Van Oort, N., 2021. Covid-19 and train travel behavior. In: Proc. of the 49th European Transport Conference.

Van Oort, N., Brands, T., De Romph, E., 2015. Short-term prediction of ridership on public transport with smart card data. Transp. Res. Rec. 2535, 105–111.

Wardman, M., Shires, J., Lythgoe, W., Tyler, J., 2004. Consumer benefits and demand impacts of regular train timetables. Int. J. Transp. Manag. 2, 39–49.