Unsupervised approach towards analysing the public transport bunching swings formation phenomenon

Degeler, Viktoriya; Heydenrijk-Ottens, Léonie; Luo, Ding; van Oort, Niels; van Lint, Hans

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Unsupervised approach towards analysing the public transport bunching swings formation phenomenon

Viktoriya Degeler[1] · Léonie Heydenrijk-Ottens[2] · Ding Luo[2] · Niels van Oort[2] · Hans van Lint[2]

Abstract

We perform an analysis of public transport data from The Hague, the Netherlands, combined from three sources: static network information, automatic vehicles location and automated fare collection data. We highlight the effect of bunching swings, and show that this phenomenon can be extracted using unsupervised machine learning techniques, namely clustering. We also show the correlation between bunching rate and passenger load, and bunching probability patterns for working days and weekends. We present the approach for extracting isolated bunching swings formations (BSF) and show different cases of BSFs, some of which can persist for a considerable time. We applied our approach to the tram line 1 of The Hague, and computed and presented four different patterns of BSFs, which we name "high passenger load", "whole route", "evening, end of route", "long duration". We analyse each bunching swings formation type in detail.

✉ Viktoriya Degeler
v.degeler@rug.nl

Léonie Heydenrijk-Ottens
L.J.C.Heydenrijk-Ottens-1@tudelft.nl

Ding Luo
D.Luo@tudelft.nl

Niels van Oort
N.vanOort@tudelft.nl

Hans van Lint
J.W.C.vanLint@tudelft.nl

[1] Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands

[2] Department of Transport and Planning, Delft University of Technology, Delft, The Netherlands

1 Introduction

One of the most important quality aspects of public transport (PT) is service reliability, being the match of operations and planning. In PT systems worldwide, passengers consider this aspect both important and yet not sufficient (van Oort 2014; Diab et al. 2015). An increasing amount and complexity of data describing PT services allows us to better explore the detection methods and analysis of different phenomena of PT operations, related to service reliability, e.g. AVL data (Hickman 2004) and smartcard data (Pelletier et al. 2011). One such phenomenon is the bunching of PT vehicles, which is characterised by uneven deviations of headways between vehicles from the designed value. These disturbances are magnified over time, until PT vehicles travel in pairs, instead of evenly spaced (Pilachowski 2009). The cause of this phenomenon is described as follows: the delay of a vehicle compared to its expected schedule (and resulting increase of headway with the previous vehicle) causes more passengers to gather at PT stops, which increases the vehicle's dwell times, which in turn increases the delay of that vehicle even more. The next vehicle, even though starting according to schedule, has fewer passengers to collect, and, therefore, is able to travel faster, further decreasing the headway with the delayed vehicle, and so forth.

The bunching phenomenon was first identified by Newell and Potts (1964) and has been extensively studied by many scholars in the past decades. For instance, Mandelzys and Hellinga (2010) propose a method for identifying the causes of performance issues in bus schedule adherence using both Automatic Vehicle Location (AVL) and Automatic Passenger Count (APC) data. Fonzone et al. (2015) conclude that unexpected passenger demands are the root cause of bunching. Sun and Schmöcker (2018) model the choice behavior of passengers when there is more than one bus serving a stop. Yu et al. (2016) show that supervised learning techniques such as Support Vector Machines can be employed to predict bus headways and bunching by using the information available from transit smart cards. When it comes to the solutions to this problem, many have attempted to develop various control strategies for the vehicle headway, such as Daganzo (2009), Daganzo and Pilachowski (2011), Bartholdi and Eisenstein (2012), Moreira-Matias et al. (2016), Varga et al. (2018), etc. Bunching has been shown to severely negatively affect the operations of PT (Osuna and Newell 1972; Chapman and Michel 1978) and different techniques have been proposed designed to deal with the adverse effects of bunching (Daganzo 2009; Feng and Figliozzi 2011; Moreira-Matias et al. 2016). Most strategies revolve around holding and headway control, e.g. Zhang and Lo (2018) proposed two-way-looking self-equalizing headway control. Despite the vast research effort that has been spent on this topic, not many studies are available that investigate bunching (patterns) using real data sets. This may be attributed to the scarcity of observations on PT system operations on a sufficiently large spatiotemporal scale. This, however, is changing rapidly, with the increased data availabilty of (real-time) scheduling, vehicle location and fare collection systems. Andres and Nair (2017) propose to use linear regression for data-driven prediction of headways and use it in combination with the existing dynamic holding strategy for corrective control.

To the best of our knowledge, our study is the first attempt to explore and examine—by combining these new data sources—bunching patterns on different levels of scale (i.e. not only at the level of a vehicle or pairs of vehicles, but also over larger spatiotemporal periods for a particular service) using machine learning techniques. Specifically, we show that it is possible to extract and detect single instances of bunching by using fully unsupervised techniques (clustering). Furthermore, the same technique allows us to identify and track how bunching propagates over time, and, specifically, to uncover the *bunching swings* phenomenon, defined as repeating patterns of pairs of delayed and bunched vehicles, without 'normal' vehicles, which are running according to schedule with evenly spaced headways, in between. When investigating data, we regularly observed 5 or more pairs of vehicles forming bunching swings, without returning to normal scheduled times for nearly two hours or even longer.

Our main contribution in this paper is looking at the whole formations of these bunching swings, isolated in time and space by periods of normal (i.e. those that conform to the schedule) operations. We call these patterns *bunching swings formations (BSF)*, and define them as sequences of PT vehicles on the one axis with sequences of PT stops on the other axis, where these vehicles are affected by either delays or bunching, surrounded by the normal PT operations. The precise formal definition of a BSF is given in Sect. 5. We looked further into the formations of bunching swings, and additionally present a way to detect and extract these BSFs directly from the PT data. To demonstrate the methods we apply them to a densely used tram service in the The Hague area, the Netherlands, and we show that the BSFs can be categorised into four distinct types, that differ in passenger load, and in spatial and temporal extent. We argue that the presented methods and findings are relevant for both operational and tactical planning of PT services.

The paper is outlined as follows. In Sect. 2, we define our case study and describe the data that we used in our analysis. In Sect. 3, we describe how the clustering can be used to extract delayed or bunched situations. In Sect. 4, we discuss the bunching swings phenomenon. Section 5 formally defines bunching swings formations, and shows, how BSFs can be extracted, and which parameters can be used for finding different types of these formations. Section 6 discusses the results of clustering for formation types extraction, and discusses each of four types in detail. Finally, Sect. 7 concludes the paper.

2 Case study and data description

For this study, we use a dataset containing static and dynamic information for each stop of the public transport network in The Hague, the Netherlands, which consists of 12 tram lines and 8 bus lines. The dataset covers the period of one month, March 2015.

The static data includes information about the transportation network, its geographical structure, stops, routes, and schedules. It is derived from General Transit Feed Specification (GTFS) data. The dynamic data comes from two different sources. One is Automatic Vehicle Location (AVL) data (Hickman 2004;

van Oort et al. 2015a): which contain actual times of arrival/departure of vehicles, headways, delays, etc. Arrival ahead of schedule is represented as a negative value of delay. The second type of dynamic data is the Automated Fare Collection (AFC) data, also known as Smart Card data (Pelletier et al. 2011; van Oort et al. 2015b), which contain the tap-in/tap-out times of personalised smart cards (which are extremely prevalent in the Netherlands over other types of payment), and the exact vehicles in which these transactions took place. Using the tap-in and tap-out times of the smart cards, the passenger load (or occupancy) of a vehicle can be estimated with an intricate set of tools presented in Luo et al. (2018).

Main analysis in this paper is performed specifically on tram line number 1. This is the longest line in The Hague, having 41 stops; going from the west side of the city (Scheveningen, the beach district, nearby the sea) through the city center; continuing to the south-east; and then south into the center of the nearby city of Delft. Figure 1 shows the plan of this line.

Important feature of this route is that there is no visible control from the side of the operator with regards to the holding patterns (e.g. holding the next vehicle if the first one becomes delayed). Therefore, the analysis in the following sections relates to the behavior of PT vehicles in an uncontrolled situation.
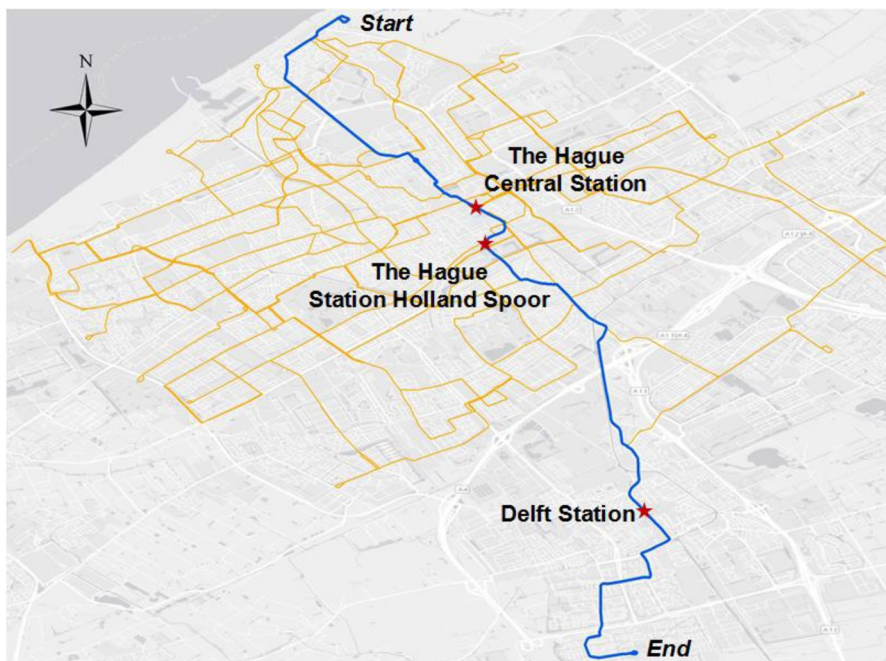
Fig. 1 Tram line 1 from Scheveningen Noorderstrand to Delft Tanthof

3 Situation profiles via clustering

In this section, we apply unsupervised techniques to look for recurring patterns we call "situation profiles" in single trips (of PT vehicles). We do this using occupancy data combined with automatic vehicle location data.

We prepare the dataset by first removing all time/place/route signifying information. This includes time of the day, date, line number, stop ID, trip ID, and so on. The reason for removal of this information is that when constructing situation profiles, we want to look at traffic conditions, and we want to avoid clustering two situations with similar conditions differently, just because they occurred on different routes or times. The features that we use are, therefore, all related directly to the traffic conditions, and are obtained per every stop on every tram route:

- *dwell*—dwell times on stops;
- *delayArr*—delay of arrival;
- *load*—passenger load;
- *preHw*—previous headway;
- *nextHw*—next headway.

It has to be noted, that the original dataset contains some missing periods of data, which sometimes produce data points, where either previous or next headways are unknown. This happens in around 1% of the whole dataset. In order to keep these points in our dataset, we use an imputation heuristic to fill the missing values with their probable values. In this case, we use the scheduled headway.

Line 1, which we investigated, utilises two different headways in scheduling. In the majority of cases (from 7:00 to 20:00 on weekdays and from 12:00 to 18:00 on weekends), the scheduled headway between the vehicles is 10 min; however, it becomes 15 min in the very early and late hours. In order to control for this difference in the initial unsupervised investigation, we performed this analysis only on vehicles with 10 min (600 s) scheduled headways. However, further analysis in Sect. 5 will use all vehicles of the line due to its generalised features.

All features are vectorised and normalised, and we perform K-means clustering in order to find the situational profiles. The results with different numbers of clusters are shown in Table 1. All values, except passenger load, are reported in seconds.

It can be seen that there are three fundamental types of situations:

1. "**Normal**" situations: Characterised by average dwell times; low delay (half a minute on average); average passenger load; and equal headways with previous and next vehicles.
2. "**Delayed**" situations: Increased dwell times; considerable delay; considerably increased passenger load; the headway with previous vehicle is considerably larger than the headway with the next one.
3. "**Bunched/early**" situations: Decreased dwell times; no delay or early arrival; low to medium passenger load; the headway with previous vehicle is considerably smaller than the headway with the next one.

Table 1 Clustering results with (a) three; (b) four; (c) five clusters

(a) Feature	Cluster 1 "delayed" (18.1%)	Cluster 2 "normal" (59.9%)	Cluster 3 "bunched" (22.0%)
dwell	30.6 ± 17	29.1 ± 16	26.6 ± 19
delayArr	239.6 ± 136	33.5 ± 70	-66.4 ± 104
load	37.1 ± 24	27.6 ± 19	24.5 ± 17
preHw	804.1 ± 167	608.2 ± 83	407.7 ± 138
nextHw	388.5 ± 135	605.6 ± 88	768.8 ± 161

(b) Feature	Cluster 1 (10.6%)	Cluster 2 (30.2%)	Cluster 3 (45.0%)	Cluster 4 (14.3%)
dwell	31.0 ± 17	30.1 ± 17	28.4 ± 16	25.9 ± 19
delayArr	296.2 ± 137	99.6 ± 79	-2.8 ± 64	-80.2 ± 111
load	38.5 ± 25	32.0 ± 21	25.8 ± 18	22.7 ± 17
preHw	850.9 ± 180	674.9 ± 101	567.0 ± 76	355.2 ± 135
nextHw	323.5 ± 128	537.8 ± 82	650.5 ± 95	794.3 ± 174

(c) Feature	Cluster 1 (7.7%)	Cluster 2 (18.6%)	Cluster 3 (41.1%)	Cluster 4 (24.3%)	Cluster 5 (8.3%)
dwell	30.5 ± 16	30.6 ± 17	28.9 ± 16	28.0 ± 17	24.9 ± 20
delayArr	327.6 ± 138	146.1 ± 90	31.5 ± 56	-35.7 ± 84	-93.6 ± 116
load	39.1 ± 26	34.9 ± 22	26.4 ± 19	27.3 ± 18	19.8 ± 15
preHw	874.6 ± 181	721.8 ± 124	609.6 ± 69	508.4 ± 88	288.3 ± 120
nextHw	283.7 ± 119	495.7 ± 90	602.2 ± 73	711.6 ± 122	815.6 ± 188

(a) Three clusters produce a good distinction between three fundamental types of vehicle conditions: normal operation; being late with increased passenger load, being early and bunched with a previous vehicle.

(b) Four clusters provide a further distinction in "normal" situations (clusters 2 and 3), dividing them on "slightly late" and "early". Delayed (clusters 1) and bunched (cluster 2) situations are more pronounced

(c) Five clusters further split the situation. Note the last cluster 5, which now shows extremely bunched trams, with just over 2 min headway time on average and very low passenger load.

An interesting effect occurs when changing the number of clusters. When comparing Table 1 (a–c) (with 3, 4, and 5 clusters, respectively) it can be seen that the three profiles described above are always created. However, the bigger the number of clusters, the more fine-grained these clusters are, further discriminating between low delays/high delays, or low passenger load to medium passenger load. These results suggest that bunching has the most pronounced effect on the difference in the situation profiles, and led us to further investigate this phenomenon.

4 The bunching swings phenomenon

In this section, we zoom out and investigate the larger spatiotemporal context in which, particularly, bunching occurrences take place. Further in this paper, we use clustering of points with four clusters, which are shown in Table 1b. We combine

two middle clusters (clusters 2 and 3) into one "normal" cluster, which now contains about 80% of all situations, while marking cluster 1 as "delayed", and cluster 4 as "bunched".

An example of such larger scale tram operations can be seen in Fig. 2, which represents the whole day of operations of The Hague's tram line 1 on Sunday, the 1st of March, 2015. In this and the following figures, red-green line represents AVL routes, with the colour matching the actual passenger load (green is low, red is high). Clusters are marked with green crosses (normal), black squares (delayed), and blue circles (bunched). Every line represents a trip of a single tram, in time (*x*-axis) and space (*y*-axis, representing stops). The line varies its colour depending on the relative occupancy rate of the tram. The markers on stops indicate to which situation cluster this particular event (a tram arriving, serving and leaving a stop) belongs, with green crosses representing the normal situation, black cubes—a "delayed" situation, blue circles—a "bunched" situation.

In Fig. 2 a clear phenomenon of *bunching swings* can be observed. We define it as follows:

Bunching swings is the phenomenon of several consecutive PT vehicles in a row alternating between "delayed" and "bunched" states, not returning to a "normal" state.

A more close-up and very clearly marked case of such bunching swings can be observed in Fig. 3a, from line 1 on March 20. One tram (that left < 11:00) got delayed at a stop (nr. 25) for a considerable time, with 5 pairs of trams afterwards alternating between being delayed with a high number of passengers and being early with a low number of passengers, a situation that lasted for almost two hours. It can also be seen how bunching got progressively worse over time (e.g.
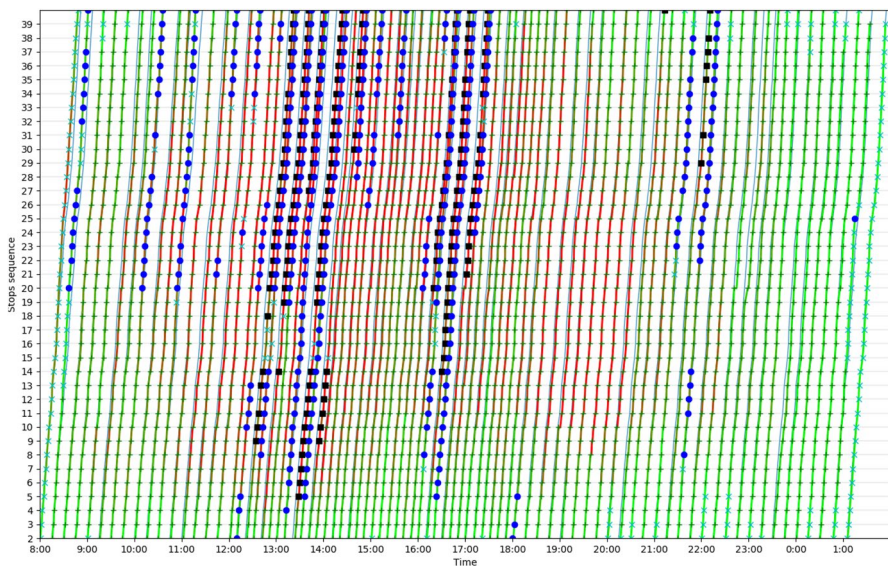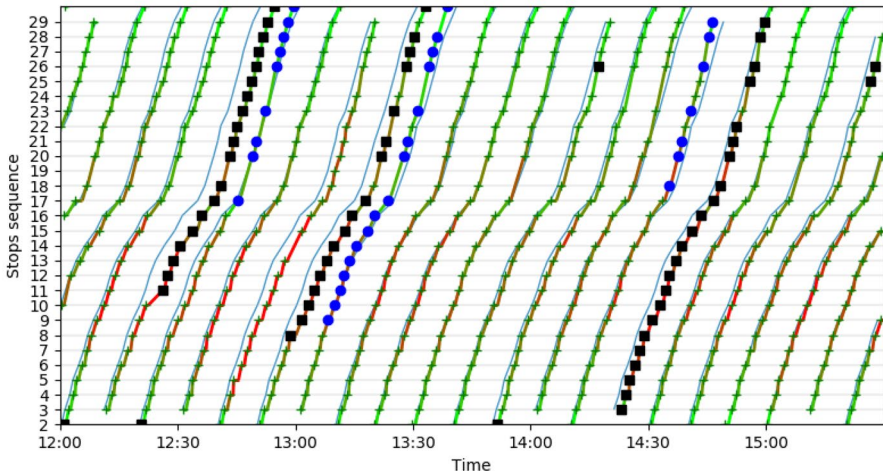

Fig. 2 Clustering results with observed "bunching swings" (Line 1, March 1, Sunday)
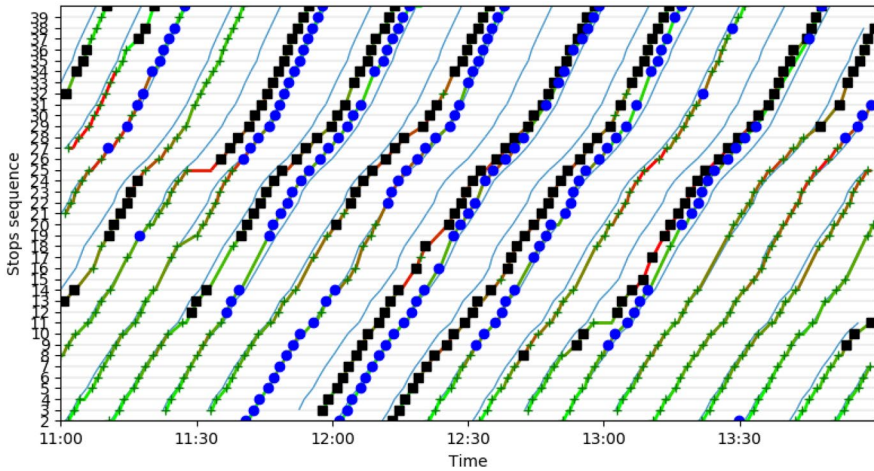
(a) Case of several consecutive bunching swings. (Line 1, March 20)

(b) Case of single bunching swings. (Line 9, March 4)

Fig. 3 Different cases of bunching swings. Pale blue line represents expected schedule. Clusters are marked with green crosses (normal), black squares (delayed), and blue circles (bunched) (colour figure online)

each next pair of vehicles were closer to each other than the previous pair), before being diminished around 13:30 and returning to a more or less normal schedule after this. Figure 3b shows a different kind of situation, from line 9 on March 4, with three separate cases of a single swing, where two times swings are started by a delayed tram, and one time by an early tram. In these cases there is one vehicle

that clearly got out of its schedule, which caused some issues to the neighbouring vehicles, but did not affect the route efficiency in the long term.

The clustering allows us to construct patterns of bunching probability, as shown in Fig. 4. We calculate the bunching probability as a percentage of trams clustered into "delayed" or "bunched" clusters, compared to all trams during the same period. Bunching patterns differ noticeably between working days and weekends. As can be seen, weekdays have two clear peaks of bunching rate increase: a huge one in the morning, and a more moderate one in the evening. Weekends have a considerable increase in bunching rate in the middle of the day. In all cases, bunching is very low to almost non-existent at the beginning of the route, but steadily increases during the route, and is at its heaviest by the end of the route.

5 Bunching swings formations

We now look in detail at the different types of consecutive bunching swings formations, such as those that are shown in Fig. 3. The formation as a whole represents a tightly interlinked situation, where early schedule irregularities may be still having an effect on bunching/delays and uneven passenger distribution two or more hours later. Therefore, understanding the types of formations and conditions, under which they occur, leads to a better anticipation of how a situation will evolve.

Before we dig deeper into the analysis of bunching formations, we want to define precisely, what a bunching swings formation is. We define it as follows:

A bunching swings formation (BSF) is a consecutive sequence of public transport vehicles each serving a consecutive sequence of stops (a part of the PT route), in which all or a majority of vehicles are either being delayed or being bunched compared to the previous vehicle on the corresponding affected parts of the route, such that this formation is isolated and surrounded in space (PT stops) and time
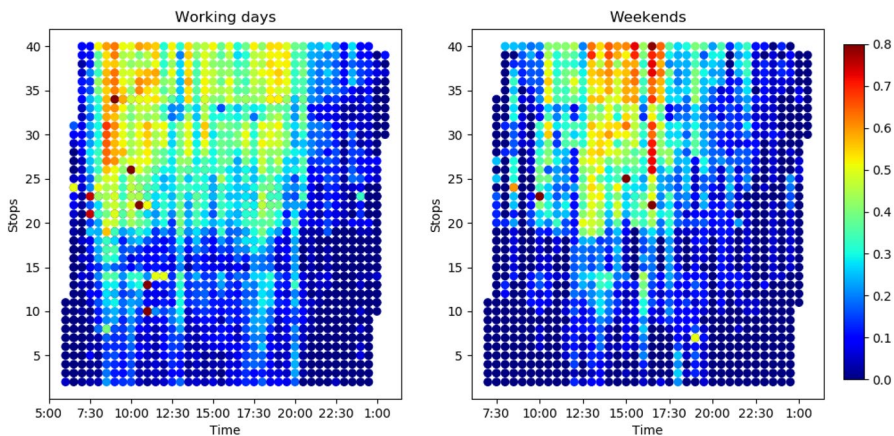
Fig. 4 Probability of bunching, Line 1. Bunching patterns are different on working days and weekends

(consecutive vehicles on route) by the "normal" operations of PT vehicles, e.g. operations according to schedule.

We perform the following steps to analyse BSFs. First, we extract the linked formations and look at each formation separately. Then, we extract important features of each formation, in order to be able to cluster them by the formation type. In the ensuing two subsections we describe each step in detail.

5.1 Formations extraction

Each day there are usually several BSFs occurring; therefore, we need to be precise when extracting a single interlinked formation, to avoid combining into one formation two or more separate bunching swings occurrences. To this end, we are not interested in cases of a single tram being delayed/early, unless it is followed by a discrepancy with the schedule in the following trips. Therefore, we only look at formations that have at least two bunched/delayed trips (a single bunching swing) or more.

More precisely, we are interested only in the part of the route where bunching occurs. Earlier stops in the route should be excluded from the formation. Although it is a common situation that bunching, once it occurs, continues until the end of a particular trip, it also happens that the delay or early arrival are resolved en-route. We will later see that some bunching cases are interesting due to the fact that they happen in the middle of the route with a potential to be resolved during further stops.

During our data analysis we observed some situations, where one of the trams in the middle of a bunching swings pattern runs on schedule, whereas the trams before it and after it are both involved in a bunching pattern. This situation can be treated in two different ways: (1) as a two different BSFs before and after the tram in question, or (2) as a single BSF with the tram involved in-between bunched/delayed trams being regarded as participating in the formation as well. There are arguments for both types of treatment, and in our analysis we looked at formation clustering with both of these types, and we found that it does alter further clustering results. Further in this paper, we report the results based on (2), treating such situation as a single BSF. The reason is that, based on the situations that we looked at, such bunching swings usually represent a single unfolding situation, see, for example, Figure 7 (4th sub-figure) and Figure 10. However, if at least two consecutive trams run on schedule in between observed bunching swings, this does cause the creation of two different bunching swings formations.

The algorithm for detecting bunching swings formation reads as follows:

1. Regard each line and direction separately. Extract a collection of data points for the line and direction in question. Data points are represented by a list of AVL locations at each stop (missing information can be handled). Each data point should contain the following information: date, line number, line direction, stop ID, stop order in the route sequence, trip ID, timestamp, dwell time on a stop, delay of arrival in time units, passenger load, headway to the previous vehicle, headway to the next vehicle, previous trip ID, next trip ID

2. Perform clustering of data points as defined in Sect. 3. Each data point is assigned a particular cluster type ("delayed"/"normal"/"bunched").

3. During BSF extraction, regard each day separately. Extract all data points related to this line, direction, date into a current dataset.

4. While there exists an un-investigated "delayed" or "bunched" point in the current dataset:

    4.1. Create a new unique *potential BSF ID*, and put the point in question into the queue of points for this ID.

    4.2. Take the next point from the queue for the current potential BSF, and mark it as investigated. Extract neighbours of this data point: neighbours are data points that correspond both to the neighbouring trips (the trip in question, the previous trip or the next trip) and neighbouring stops (the stop in question and the certain number of stops before and after this stop, we used 3 stops before and after in our analysis). If at least 20% of the neighbouring data points belong to non-normal clusters, mark the current data point with the unique current "potential BSF" marker and add all its still un-investigated neighbours to the queue. Remove investigated point from the queue, and repeat this full step, while the queue is not empty.

    4.3. Extract all points marked with the current potential BSF marker, and perform the checks on the current potential BSF. Remove leading and trailing normal trips. Split the BSF into two or more, if it contains at least 2 "normal" trips in between ("normal" trips are those that have less than the predefined threshold of non-normal AVL points, in our case: 3). Check that it contains at least the minimum number of trips (in our case: 2). If all checks pass, a new BSF is detected and added to the list of BSFs.

We can now analyse the data of different days in terms of bunching swings formations, rather than separate bunching occurrences. For example, after the application of this algorithm to Line 1 on March 1, the occurrences of bunching swings that we visualised earlier in Fig. 2 can now be represented in BSFs, as shown in Fig. 5. Three different BSFs were extracted, each independent from others in space and time, with varying duration, severity, affected stops and other parameters. In this figure, unlike in previous and following figures, we use blue, red and purple colours to visualise distinct extracted BSFs on one figure; green still marks the normal situations. Note that there are also some short-lived occurrences of bunching (marked with green), that do not form BSFs due to normally having only one affected vehicle. Therefore, the BSF detection algorithm will correctly ignore them.

5.2 Formations clustering and profiling

Once we have separate bunching swings formations extracted, we carefully examine their parameters. In this analysis, we use the following parameters:

Bunching Swings Formations Parameters:
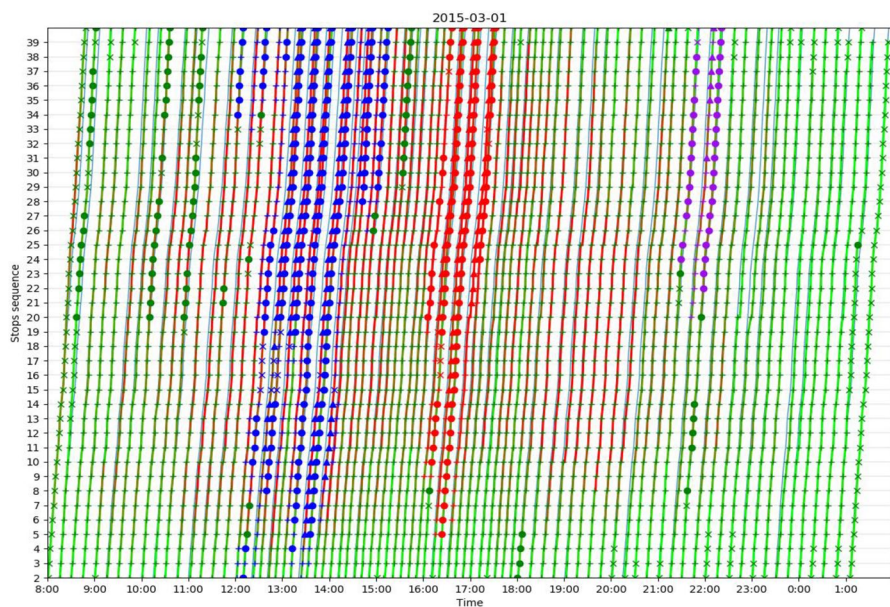
Fig. 5 Three separate bunching swings formations, extracted for Line 1 on March 1, marked with blue, red and purple colours (colour figure online)

1. **Average passenger load**—we average passenger load for the whole formation, mainly due to the fact that in two consecutive trams in a formation, one being bunched and one being early, the load can differ significantly.
2. **Number of trips involved**—the total number of trams that were affected
3. **Total duration**—Total duration, in hours, of the bunching swings occurrence. It has to be noted that this variable is considerably correlated with the number of trips involved (Pearson's $r = 0.96$ for the line 1 that we used in our analysis; however, it will be different for other lines depending on variations in planned headways density over time and on different days), so any one of them can be used in further analysis, depending on the preference. In our case, we used them together and each separately and did not find any meaningful difference in reported results.
4. **Average starting stop**—when in the sequence of stops the bunching effect starts to occur.
5. **Average length in stops**—how long during the route the bunching effect lasts.
6. **Time of day when the bunching swings formation starts**
7. **Day type**—work day or weekend
8. **Lasts until route end?**—yes or no, depending on whether bunching is resolved mid-route, or lasts until the end of the route.

Once we extract these factors from each detected bunching swings formation, we can use them to perform a second layer of clustering, in order to combine formations by type.

One of the main concerns when doing this type of analysis, is the inability to combine bunching swings from different lines into one common type extraction. The geographical differences of lines, different stops being a part of central/busy areas, different schedule and frequency, different coverage by neighbouring lines providing feasible alternatives for passengers to avoid taking delayed trams, and many other external factors can all influence the bunching swings formations and evolution differently. In our future research analysis, it is our goal to add such external factors to our dataset and specifically look at differences in BSFs on different lines and in different cities. In this paper, however, we control all those factors by looking at bunching swings formation types within one line, namely line 1 in The Hague.

6 Results: exploring bunching swings formations

6.1 Clusters number

The first question to be asked in the analysis of BSF types, is how to decide on the number of clusters. In order to do this, we performed K-means clustering, varying the number K of clusters from 2 to 7, and performed the silhouette analysis for each K, to visually assess the quality of clusters. The silhouette analysis allows to see, how similar the points within the cluster are with the centroid point, and how different they are from the points of different clusters (Rousseeuw 1987). Each point in a cluster obtains a silhouette score on a scale $[-1, 1]$, which indicates, how much more similar this point is to the points in its own cluster than to the points in different clusters. Here, 1.0 indicates complete equality of the point with all points in its own cluster, and difference with points in others, while numbers below 0 indicate that the assignment of a cluster for this point may have been wrong, as it is more close to the points outside of the cluster rather than in its own cluster. With a good cluster number, there are no clusters, whose silhouette score is considerably lower than for others.

Figure 6 shows the silhouettes for cluster numbers K from 2 to 7, while Fig. 7 shows the average silhouette score obtained for each K.

As can be seen, with $K=2$ and $K=3$, we can get a good separation between clusters, but the variability of points inside the clusters does not allow us to have a good description of which situations each cluster represents due to common occurrences of different situational types of formations being assigned to the same cluster.

With $K=5$, we have one of the clusters with considerably lower quality than others (cluster number 1 in the figure), and the situation stays the same for $K=6$ and $K=7$. We did not investigate the number of clusters being larger than 7, since with the one month of data a larger number would likely result in overfitted clusters, i.e. those representing particular situations of this exact time frame rather than general formation patterns.

$K=4$ has the best average silhouette score, with no cluster being clearly worse than others. Therefore, out of all investigated possibilities, we chose $K=4$ as providing the best number of clusters.
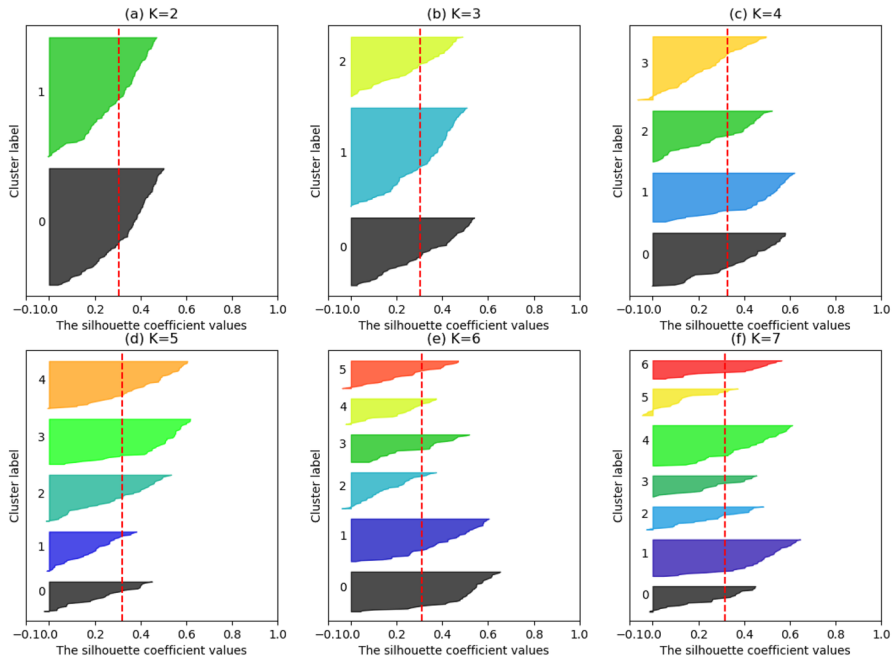
Fig. 6 Silhouettes of bunching formations clusters for K from 2 to 7
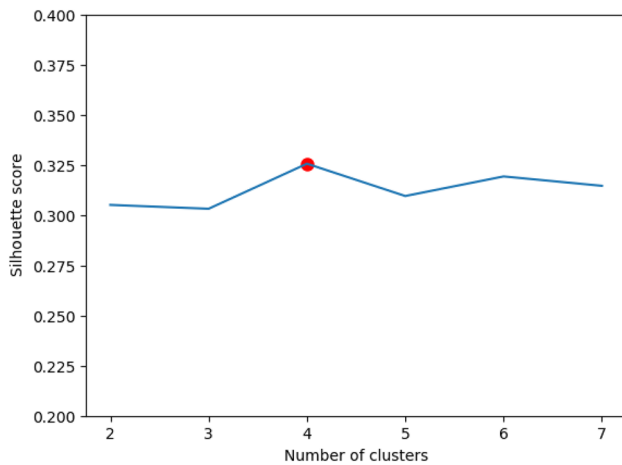
Fig. 7 Average silhouette scores for number of clusters K from 2 to 7. The red point marks K=4, the chosen number of clusters

Table 2 Clustering results for bunching swings formation types extraction

| Feature	Cluster 1 (29.2%)	Cluster 2 (23.6%)	Cluster 3 (22.7%)	Cluster 4 (24.5%)
Average passenger load	**32.1 ± 10**	25.9 ± 9	**15.9 ± 6**	30.9 ± 8
Trips involved	3.8 ± 2	4.7 ± 2	3.3 ± 2	**15 ± 5**
Duration	50 m ± 20 m	1h20m ± 30 m	50 m ± 30 m	**3 h ± 1 h**
Average starting stop	21 ± 5	**8 ± 4**	**20 ± 6**	15 ± 4
Time when starts	12 h ± 3 h	14 h ± 4 h	**20 h 30 m ± 2.5 h**	12 h ± 3 h
Average length in stops	18 ± 5	**31 ± 5**	18 ± 5	25 ± 4
Day type	**Work 88.9% Weekend 11.1%**	Work 80.4% Weekend 19.6%	**Work 75.5% Weekend 24.5%**	Work 77.3% Weekend 22.6%
Until route end	Yes 88.9% No 11.1%	Yes 96.1% No 3.9%	Yes 85.7% No 14.3%	**Yes 100% No 0%**
Explanation	Very high average passenger load; Medium length during route	Starts very early on the route; Lasts for the whole route duration	Low trips number; Evening bunching; Starts late in the route; Very low passenger load	Very long duration with many consecutive trips; Rather high average passenger load
Nickname	**"High passenger load"**	**"Whole route"**	**"Evening, end of route"**	**"Long duration"**

6.2 Clustering results

As described in earlier sections, we detected, extracted, and clustered bunching swings formations. In total, we extracted 216 BSF occurrences within one month, and clustered them in four different types. You can see the types combined in Table 2. We highlight in bold the most significant features that distinguish each cluster from other clusters and are used to explain the type of BSFs that belong to it.

1. "**High passenger load**"—The most common type of BSFs and it is specified by very high average passenger load for the whole duration of the swings formation. It often starts in the middle of the route and more often than other types can be experienced on work days. Examples can be seen in Fig. 8. It is worth mentioning that although the calculated average starting time of this bunching cluster is in the middle of the day, most of the occurrences during work days start during the morning peak hours (8:00–9:00), or during the evening peak hours (15:00–18:00). Therefore, we conclude that this type of bunching swings formations is clearly associated with heavy demand on the public transport route, specifically the demand that happens during peak hours. This conclusion is also fully consistent with the increased occurrences of this cluster during week days (89%).

2. "**Whole route**"—Bunching swings formations of this type usually start very early in the route and last for the whole duration of the trip. They have an average number of trips involved and an average passenger load. Examples can be
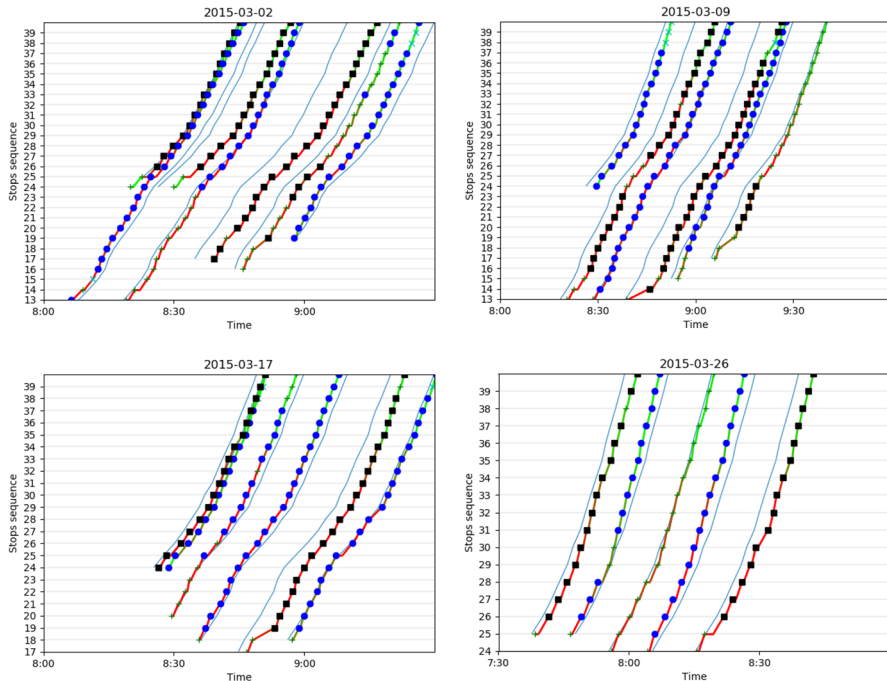
Fig. 8 Examples of bunching swings formations from cluster 1 "High passenger load"
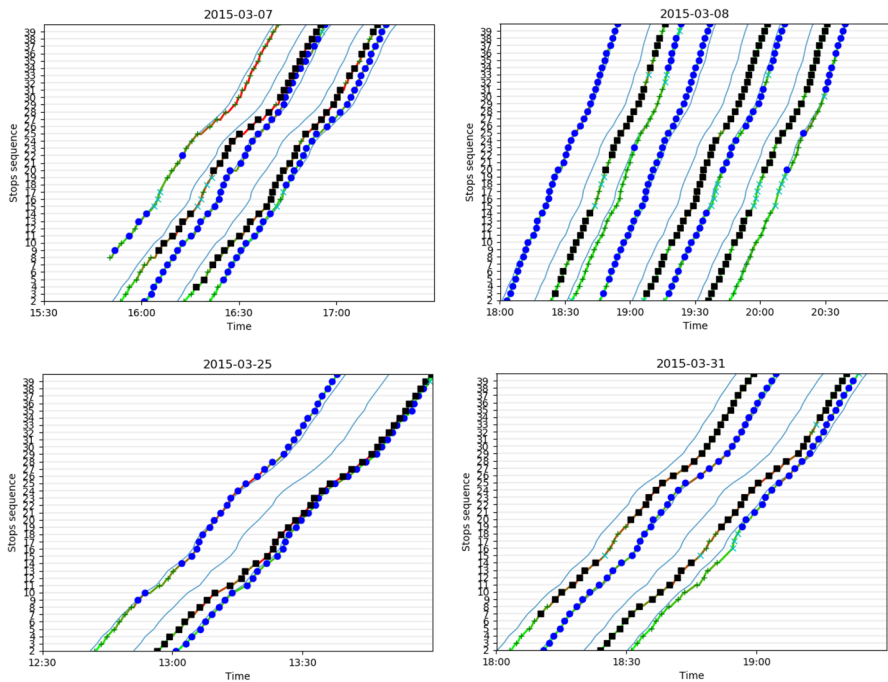
Fig. 9 Examples of bunching swings formations from cluster 2 "Whole route"

seen in Fig. 9. When looking closely at this cluster, we found that in a bit more than half (53%) of all BSFs in this cluster the delay started at the very beginning of the route, e.g. the first observable stop (marked as 2 in the stops sequence in the figures) already had either delays (most of the times), or the tram left early (the minority of the times). This irregularity in the dispatch of trams seems to originate from a circular nature of the route, so, given the absence of extra buffer times, the tram that arrives late to the last station has to start its next trip in the opposite direction late. In the remaining half of the cases, the bunching phenomenon developed very early, normally within the first ten stops.

3. "**Evening, end of route**"—This is a somewhat unique formation type in terms of many factors involved. First of all, the time of day and the day type when it happens: it usually starts late in the evening and can be observed a bit more often on weekends. The bunching swings usually start very late in the route, noticeably later than for other clusters, but can still be resolved in 14.3% of the cases. However, the most interesting factor of this cluster by far is the average passenger load, as it is very small, considerably smaller than for other bunching swings formation types. And, very important, this number is low even if we consider all trips, not only trips that are involved in bunching. On the one hand, this correlates very well with the fact, that this type of bunching swings usually happens on evenings and weekends, as at these times and days passenger numbers generally are much lower than average.

However, as was shown by previous research works and will be shown in Sect. 6.3, the bunching effect in itself correlates considerably with a high passenger load. The fact that there is a type of bunching swings formations that consistently happen with low passenger numbers is, therefore, very interesting and deserves further investigation into external factors of why this type of bunching swings occurs. One of the plausible explanations lies in the fact that the scheduling during evening and weekends already takes into account decreased passenger load, thus assumes faster dwell and en-route times, and leaves less margin to mitigate minor delays. Examples can be seen in Fig. 10.

4. "**Long duration**"—This formation type contains mainly very long and heavy bunching swings occurrences, lasting for a long time with many trips involved. Examples can be seen in Fig. 11. Passenger load stays rather high for the duration of such a formation. Bunching swings of this magnitude have no chance to be resolved mid-route, as clearly shown by the fact that exactly 100% of such bunching swings formations in this cluster lasted until the end of the route. This is also the cluster that is the most likely to have the most extreme occurrences of bunching, i.e. two vehicles coming into very close proximity to each other (rather than still being apart, but with decreased headway between them). The severity and irregular nature of occurrences of this cluster points out at potential external (to the data available in our analysis) rather than internal (self-inflicting) cause factors for these severe bunching swings formations. That assumption is also consistent with the fact, that some bunching swings occurrences are seemingly
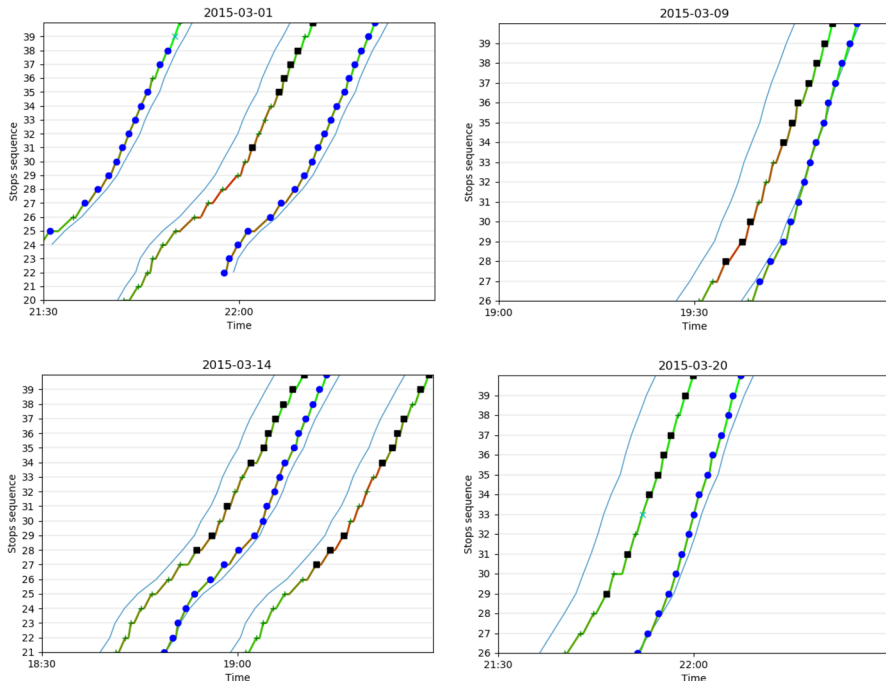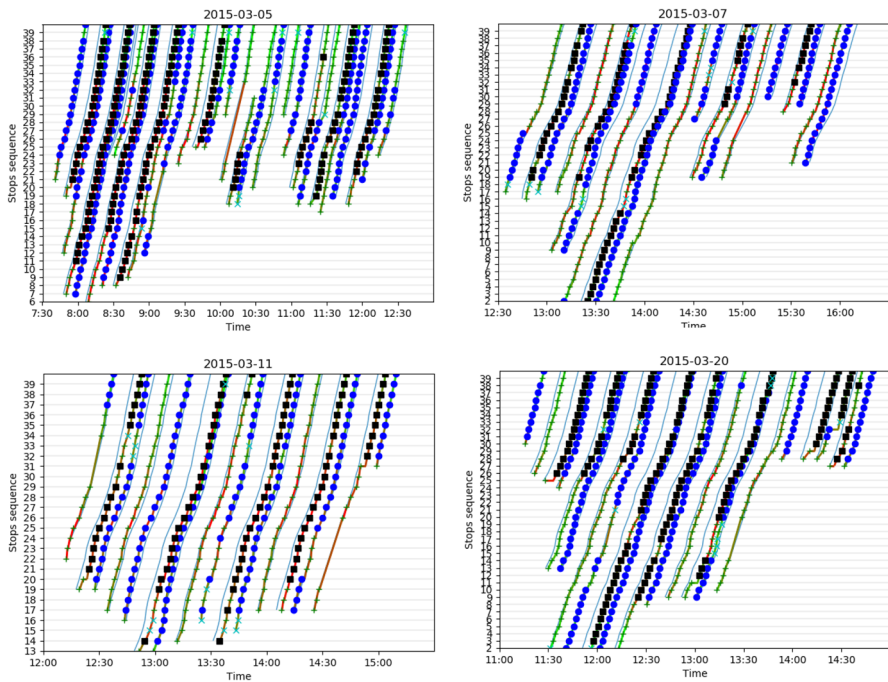

Fig. 10 Examples of bunching swings formations from cluster 3 "Evening late route"

**Fig. 11** Examples of bunching swings formations from cluster 4 "Long duration"

being fully or partially resolved en-route for some vehicles, only to be seen reoccurring again for the next vehicle (such examples can be seen in Fig. 11). This can be caused by any of the numerous external factors, for example, a car traffic jam on a road, which a tram may need to follow or cross. In any case, we believe that further analysis and potential comparison with more external data sources can be useful for this cluster.

6.3 Passenger load effect on bunching

It has been shown in previous research (Yu et al. 2016) that the number of passengers and changing load are one of the culprits of public transport bunching. In our analysis we can clearly see some cases of increased passenger load that nevertheless do not result in emergence of a bunching pattern, e.g. in Fig. 2.

In order to investigate the effect of increased passenger load on a bigger scale, we need to analyse the average rate of bunching pattern emergence over time. We look at all stops of our dataset, and split the full operations at each stop on periods of 2 h. We want to obtain the average passenger load per tram (i.e. all transported passengers divided by a number of trams), and the bunching rate (percentage of bunched/delayed trams) during these particular periods.
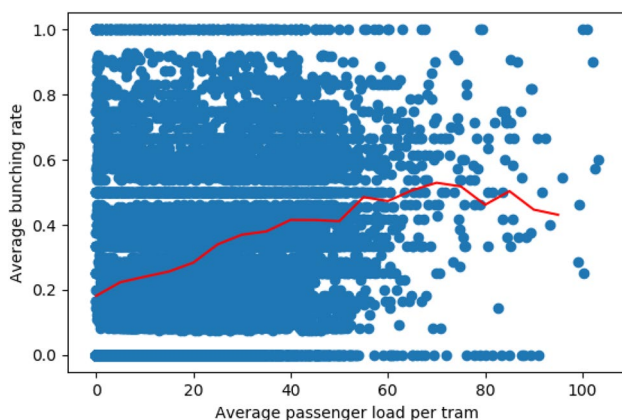
Fig. 12 Bunching rate occurrences depending on passenger load

Figure 12 shows the combined data of all occurrences of average load (*x*-axis) vs. bunching rate (*y*-axis) for the whole month for one direction of tram 1. The red line shows the average bunching rate depending on average passenger load values. The average bunching rate clearly goes up until an average load of about 70 people per tram. The Pearson coefficient, which measures the linear dependency between the two variables (bunching rate and passenger load) is $r = 0.88$. The Spearman coefficient, which measures a monotonic—but not necessarily linear—correlation between these two variables is $\rho = 0.86$. This clearly shows a high correlation of passenger load and bunching. In Fig. 13 we split bunching rates on three categories: high (rates over 0.7), low (rates lower than 0.3), medium (between 0.3 and 0.7), and draw histograms of average passenger load
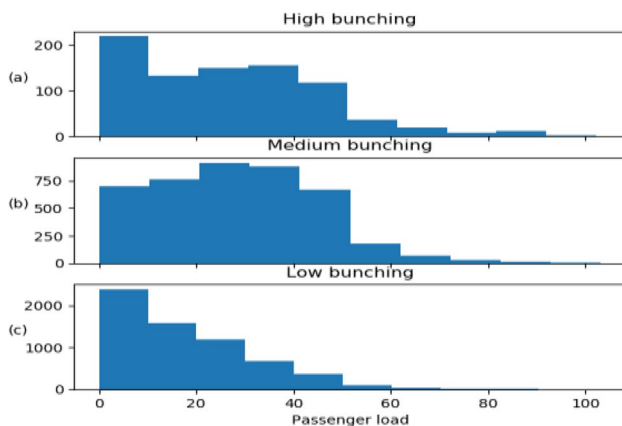
Fig. 13 Histograms of average passenger load for **a** high ($x > 0.7$); **b** medium ($0.3 < x \leq 0.7$); or **c** low ($x \leq 0.3$) bunching rates

for every rate. It can be seen that a low bunching rate corresponds to a lower passenger load.

7 Conclusions

In this paper, we showed that clustering techniques can be used to extract three fundamental types of situations in which PT vehicles may be observed (normal, delayed or bunched). Using these labels on each line and stop we showed that clustering also unravels so-called 'bunching swings' phenomena, which sometimes last for a considerable time. By varying the number of clusters, we can tune the characteristics and severity of the bunching patterns we extract. We also showed a clear correlation between passenger load and bunching rate. Clustering results allow us to perform further analyses on bunching swings in an uncontrolled environment, e.g. their characteristics and conditions under which the swings return to normal or intensify. We showed, how the formations of bunching swings can be extracted, and that they in turn can be clustered into four types: "high passenger load", "whole route", "evening late route", "long duration".

These explorative data-driven analyses may hold important benefits for PT planning and operations management. First, the techniques demonstrated can support detecting smaller scale bunching patterns automatically in large service networks. Second, as these smaller scale bunching patterns evolve over time and the network, detecting common bunching swings formations may support the prediction of these patterns, which allows operators to take appropriate measures faster, and users to be informed better.

In our further research, we plan to investigate to abstract the parameters of bunching swings formation from the specific characteristics of a particular line, by parameterizing running frequency and other differences in schedule, and by including the information about the geographical location and other external parameters into the model. The reason why we are looking at the parameterization is to be able to combine different lines and routes and analyse them in a uniform manner. This includes using relative values (e.g. percentage of delay with respect to the planned schedule as opposed to its value in seconds, percentage of a route affected, etc.) However, we have to be careful with parameterization, because some of the features still affect the outcome in absolute values, e.g. we are not only interested in the percentage of actual passenger load with respect to the average load on a line, but we are also interested in the absolute number of people to embark/disembark, since that affects the dwell time. Geographical parameters of the stops are also important, e.g. if a stop is in the suburb, business neighbourhood or the city center, and what are the distances between consecutive stops.

This analysis can be useful in control strategies. In particular, we aim to use this information to look at how the evolution of bunching swings formation can be predicted in real time. When processing real-time information, the onset of a bunching swings formation can be detected as early as with the second or third vehicle. The initial features (severity of delay, passenger load, etc.) can then

be extracted and updated in real time. These features can be used in predictive machine learning algorithms, to predict the duration and the severity of the ongoing bunching swings formation. We expect the accuracy of prediction to grow with the total time passing since the onset of the bunching swings formation. This is especially important for clusters such as cluster 2 ("whole route") and cluster 4 ("long duration"), due to their longer duration and relatively high severity. The biggest hurdle is the real-time estimation of the passenger load, as this feature is quite important for the analysis of BSFs. The smart card data that we used for analysis in this paper cannot be obtained in real time at the moment, as it is normally gathered in batches and becomes available for previous time periods (e.g. the previous day). Other passenger load estimation techniques should be employed. Real-time detection of ongoing BSFs is important not only for the calculation of its total expected duration and severity, but also for the analysis of which particular vehicles are likely to become delayed, and which of them are likely to become bunched. Since in most cases bunching patterns unfold in pairs, this becomes a somewhat straightforward task in the majority of cases. However, there are cases where the pattern of going in pairs breaks by having an odd vehicle not following the pattern (as was discussed in step 4.3 of the algorithm in Sect. 5.1), which should be further investigated to increase the prediction accuracy for any control strategy that uses the predicted bunched/delayed pattern of a particular vehicle.

References

Andres M, Nair R (2017) A predictive-control framework to address bus bunching. Transp Res Part B Methodol 104:123–148

Bartholdi III, John J, Eisenstein DD (2012) A self-coördinating bus route to resist bus bunching. Transp Res Part B Methodol 46(4):481–491

Chapman RA, Michel JF (1978) Modelling the tendency of buses to form pairs. Transp Sci 12(2):165–175

Daganzo CF (2009) A headway-based approach to eliminate bus bunching: systematic analysis and comparisons. Transp Res Part B Methodol 43(10):913–921

Daganzo CF, Pilachowski J (2011) Reducing bunching with bus-to-bus cooperation. Transp Res Part B Methodol 45(1):267–277

Diab EI, Badami MG, El-Geneidy AM (2015) Bus transit service reliability and improvement strategies: integrating the perspectives of passengers and transit agencies in North America. Transp Rev 35(3):292–328. https://doi.org/10.1080/01441647.2015.1005034

Feng W, Figliozzi M (2011) Empirical findings of bus bunching distributions and attributes using archived AVL/APC bus data. In ICCTP 2011: Towards Sustain Transp Syst. (pp. 4330–4341)

Fonzone A, Schmöcker JD, Liu R (2015) A model of bus bunching under reliability-based passenger arrival patterns. Transp Res Proced 7:276–299

Hickman M (2004) Evaluating the benefits of bus automatic vehicle location (AVL) systems. In: Levinson D, Gillen D (eds) Assessing the benefits and costs of intelligent transportation systems, vol 5. Kluwer, Boston

Luo D, Bonnetain L, Cats O, van Lint H (2018) Constructing spatiotemporal load profiles of transit vehicles with multiple data sources. Transp Res Rec. https://doi.org/10.1177/0361198118781166

Mandelzys M, Hellinga B (2010) Identifying causes of performance issues in bus schedule adherence with automatic vehicle location and passenger count data. Transp Res Rec J Transp Res Board 2143:9–15

Moreira-Matias L, Cats O, Gama J, Mendes-Moreira J, de Sousa JF (2016) An online learning approach to eliminate Bus Bunching in real-time. Appl Soft Comput 47:460–482

Newell GF, Potts RB (1964) Maintaining a bus schedule. Australian Road Research Board (ARRB) Conference, 2nd, Melbourne

Osuna EE, Newell GF (1972) Control strategies for an idealized public transportation system. Transp Sci 6(1):52–72

Pelletier M-P, Trépanier M, Morency C (2011) Smart card data use in public transit: a literature review. Transp Res Part C Emerg Technol 19:557–568. https://doi.org/10.1016/j.trc.2010.12.003

Pilachowski JM (2009) An approach to reducing bus bunching, Doctoral dissertation, University of California, Berkeley

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

Sun W, Schmöcker JD (2018) Considering passenger choices and overtaking in the bus bunching problem. Transp B Transp Dyn 6(2):151–168

van Oort N (2014) Incorporating service reliability in public transport design and performance requirements: international survey results and recommendations. Res Transp Econ 48:92–100

van Oort N, Sparing D, Brands T, Goverde RMP (2015a) Data driven improvements in public transport: the Dutch example. Public Transp 7:369–389. https://doi.org/10.1007/s12469-015-0114-7

van Oort N, Brands T, de Romph E (2015b) Short-term prediction of ridership on public transport with smart card data. Transp Res Rec J Transp Res Board 2535:105–111. https://doi.org/10.3141/2535-12

Varga B, Tettamanti T, Kulcsár B (2018) Optimally combined headway and timetable reliable public transport system. Transp Res Part C Emerg Technol 92:1–26

Yu H, Chen D, Wu Z, Ma X, Wang Y (2016) Headway-based bus bunching prediction using transit smart card data. Transp Res Part C Emerg Technol 72:45–59

Zhang S, Lo HK (2018) Two-way-looking self-equalizing headway control for bus operations. Transp Res Part B Methodol 110:280–301