Unsupervised approach to bunching swings phenomenon analysis

Degeler, Viktoriya; Heydenrijk-Ottens, Leonie; Luo, Ding; van Oort, Niels

Publication date
2018

Document Version
Accepted author manuscript

Published in
Proceedings of CASPT 2018

CASPT 2018 Full Paper

Unsupervised approach to bunching swings phenomenon analysis

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

Abstract We perform analysis of public transport data from March 2015 from The Hague, the Netherlands, combined from three sources: static network information, automatic vehicles location (AVL) and automated fare collection (AFC) data. We highlight the effect of bunching swings, and show that this phenomenon can be extracted using unsupervised machine learning techniques, namely clustering. We show different cases of bunching swings, some of which can persist for a considerable time. We also show the correlation of bunching rate with passenger load, and bunching probability patterns for working days and weekends. We show, how formations of bunching swings can be extracted, and clustered into four different types, which we name "high passenger load", "whole route", "evening late route", "long duration". We analyse each bunching swings formation type in detail.

Keywords: Public transport · Machine learning · Clustering · Bunching · Passenger load · Bunching probability

1 Introduction

Increasing amount and complexity of data describing public transport (PT) services allows us to better explore the detection methods and analysis of different phenomena of PT operations. One such phenomenon is bunching. 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

Viktoriya Degeler · Léonie Heydenrijk-Ottens · Ding Luo · Niels van Oort · Hans van Lint
Department of Transport and Planning, Delft University of Technology
Delft, the Netherlands
Email: {V.Degeler, L.J.C.Heydenrijk-Ottens-1, D.Luo, N.vanOort, J.W.C.vanLint}@tudelft.nl

more. The next vehicle, even though starting according to schedule, has fewer passengers to collect, therefore is able to travel faster, further decreasing the headway with the delayed vehicle. Bunching has been shown to severely negatively affect the operations of PT (Osuna & Newell, 1972; Chapman & Michel, 1978) and different techniques were designed to deal with it (Daganzo, 2009; Feng & Figliozzi, 2011; Moreira-Matias et al., 2016).

By computing passenger occupancy rate of public transport vehicles from Automated Fare Collection data, (Yu et al. 2016) showed that supervised learning techniques such as Support Vector Machines can be used to predict headways.

In this paper, we show that it is possible to extract and detect a bunching phenomenon by using fully unsupervised techniques, namely clustering. That confirms that bunching is one of the fundamental behaviours of the PT vehicles. Moreover, this technique allows us to highlight, how bunching propagates over time, and, specifically, highlight the interesting phenomenon of "bunching swings": When one vehicle is delayed, the next one runs ahead of schedule and has low number of passengers, while the next vehicle is delayed again, and so the pattern repeats. When investigating data, we regularly observed 5 or more pairs of vehicles forming these "bunching swings", without returning to normal scheduled times for nearly two hours or even longer. We looked further into formation of bunching swings, and in this paper we present a way to detect and extract a bunching swings formations from the PT data, and show that these formations can be split into four different types, roughly defined as (1) "very high passenger load", (2) "affecting the whole route for a few consecutive trips", (3) "happening late in route, evening and weekends" and (4) "long duration, affecting many trips".

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

2 Case study and data description

For this study, we used 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.

Static data includes information about the transportation network, its geographical structure, stops, routes, and schedules. It is provided in GTFS format. Dynamic data comes from two different sources. One is Automatic Vehicle Location (AVL) data (Hickman, 2004): 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), also known as Smart Card data (van Oort et al. 2016), which contains the tap-in / tap-out times

of personalized smart cards (which are extremely prevalent in the Netherlands over other types of payment), and the exact vehicles in which these transactions happened. Using the tap-in and tap-out times of the smart cards, the passenger load (or occupancy) of a vehicle can be estimated. (Luo et al. 2018) describe, how the load profiles were computed for this dataset.

3 Situation profiles via clustering

Occupancy data combined with automatic vehicles location data provide us with an opportunity to construct profiles of different typical situations in which the PT vehicles can be found. We use unsupervised clustering to find these profiles.

We look at every line and its direction separately. We prepare the dataset by 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 because of different routes or times when they occurred. 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 times on stops;
- delay of arrival;
- passenger load;
- previous AVL headway;
- next AVL 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 AVL headways are unknown. This happens in around 1.0% of the whole dataset. In order to keep these points in our dataset, we use an imputer to fill the missing values with their probable values, in this case we use the scheduled headway.
All features are vectorized and normalized, and we perform K-means clustering in order to find the situational profiles. The results of the clustering with different number of clusters are shown in Table 1. All values are reported in seconds.

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

1. **"Normal"** situations: Characterized by average dwell times; low delay (half a minute on average); average passenger load; and equal headways with previous and next vehicles.

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

(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.

Feature	Cluster 1 "Delayed" (17.5%)	Cluster 2 "Normal" (67%)	Cluster 3 "Bunched" (15.5%)
dwell	30.3 ± 16	27.9 ± 15	25.9 ± 18
delayArr	229.0 ± 138	28.1 ± 73	-64.5 ± 117
load	34.8 ± 24	22.9 ± 18	22.2 ± 16
preAvlHw	835.4 ± 203	697.5 ± 166	369.9 ± 262
nextAvlHw	433.1 ± 168	713.6 ± 169	841.4 ± 302

(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

Feature	Cluster 1 (13.6%)	Cluster 2 (49.8%)	Cluster 3 (30.3%)	Cluster 4 (6.6%)
dwell	30.5 ± 16	28.1 ± 15	27.4 ± 16	25.0 ± 19
delayArr	258.5 ± 140	57.2 ± 71	-29.4 ± 81	-72.1 ± 136
load	35.8 ± 24	23.6 ± 19	23.5 ± 17	19.1 ± 15
preAvlHw	847.4 ± 207	739.9 ± 166	582.0 ± 148	207.2 ± 309
nextAvlHw	391.2 ± 158	679.5 ± 165	780.0 ± 180	866.5 ± 403

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

Feature	Cluster 1 (7.3%)	Cluster 2 (20.2%)	Cluster 3 (48%)	Cluster 4 (20%)	Cluster 5 (4.7%)
dwell	30.7 ± 16	29.7 ± 16	27.6 ± 15	27.0 ± 18	24.5 ± 17
delayArr	320.5 ± 141	130.7 ± 88	24.2 ± 61	-48.3 ± 94	-66.9 ± 141
load	37.3 ± 25	31.0 ± 21	21.4 ± 17	24.3 ± 17	17.8 ± 14
preAvlHw	875.6 ± 218	780.7 ± 179	707.4 ± 162	518.9 ± 137	137.4 ± 339
nextAvlHw	313.4 ± 145	562.6 ± 139	724.6 ± 166	803.9 ± 191	874.8 ± 453

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 considerable smaller than the headway with the next one.

The interesting effect can be observed when changing the number of clusters. It can be seen that the profiles described above are always created, no matter the number of clusters. 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.

4 “Bunching swings” phenomenon

Further in this section, we use three clusters, due to the following: (1) most situations are regarded as “normal” (67%), with slights deviations being ignored, which makes it easier to concentrate on well-defined cases of bunching swings; (2) there is one cluster for each fundamental type of situation (normal / delayed / bunched). However, further in Section 5 we are interested in more pronounced cases of bunching swings, so we use the clustering on four clusters, but combine two middle ones into one to obtain a bigger, more relaxed, “normal” situation.

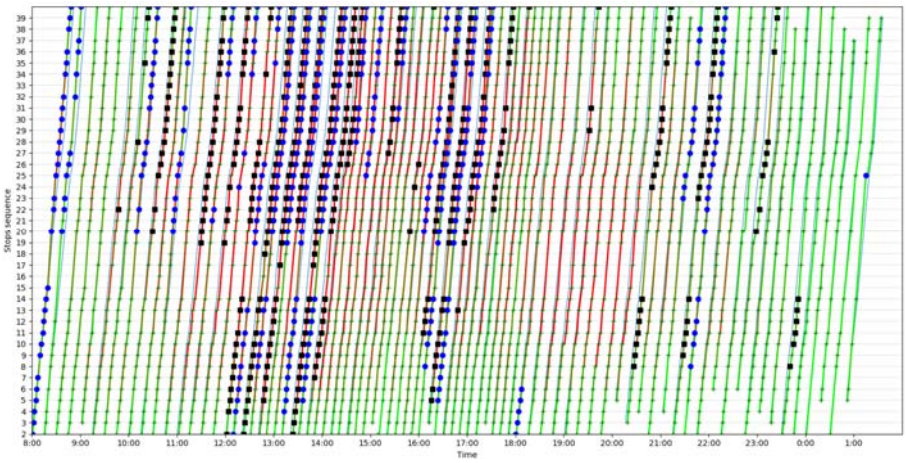
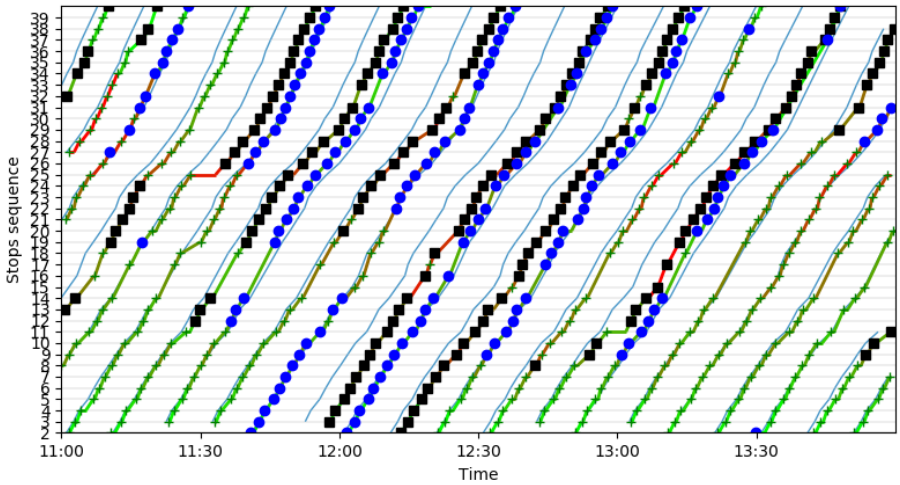


Fig. 1 Clustering results with observed “bunching swings” (Line 1, March 1, Sunday). Red-green line represents AVL routes, with the colour matching actual passenger load (green is low, red is high). Clusters are marked with green cross (normal), black square (delayed), blue circle (bunched).

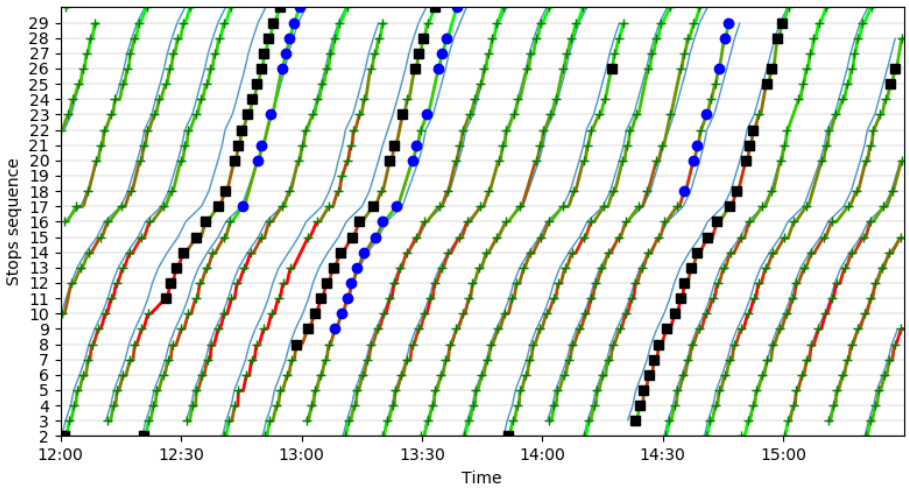
An example situation of tram operations can be seen in Figure 1, which represents the whole day of operations of the Hague’s tram line 1 on the 1st of March, 2015, a rather busy Sunday. Every line represents a trip of a single tram, in time (x-axis) and space (y-axis, representing stops). The line varies its colour depending on the relative occupancy rate of the tram. The markers on stops represent belonging of this particular event (a tram arriving, serving and leaving a stop) to a particular cluster, with green crosses representing the normal situation, black cubes – a “delayed” cluster, blue circles – a “bunched” cluster.

Here, a clear phenomenon of “bunching swings” can be observed. By “bunching swings” we mean cases where several consecutive PT vehicles in a row alternate between “delayed” and “bunched” clusters, not returning to a “normal” state. A very clearly marked case of such a formation can be observed in Figure 2-a, from line 1 on March 20. One tram got delayed at a stop 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 an hour and a half. Figure 2-b 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.



(a) Case of several consecutive bunching swings. (Line 1, March 20)



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

Fig. 2 Different cases of bunching swings formation. Pale blue line represents expected schedule.

The clustering allows us to construct patterns of bunching probability, as shown in Figure 3. We calculate the bunching probability as a percentage of trams clustered into “delayed” or “bunched” clusters, compared to all trams of the period. Bunching patterns differ noticeably between working days and weekends.

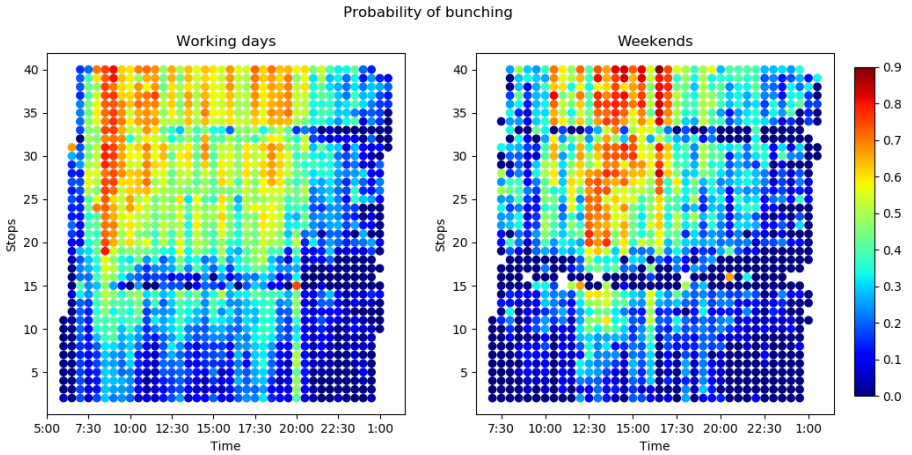


Fig. 3 Probability of bunching, Line 1. Bunching patterns are different on working days and weekends.

5 Bunching swings formation

We want to look in detail at the different types of consecutive bunching swings formations, such as those that are shown in Figure 2. 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 a situation evolution.

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

5.1 Formations extraction

Each day there are usually several bunching swings formations occurring, therefore we need to be precise when extracting an single interlinked formation, to avoid combining into one formation two or more separate bunching swings occurrences.

We first need to define precisely, what we are interested in.

1. We are not interested in cases of a single tram being delayed/early, when it is not 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.

2. 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 happened, continues until the end of a particular trip, it also happens that the delay or early arrival are rectified enroute. 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.

3. During our data analysis we could see some situations, where one of the trams in the middle of a bunching swings pattern runs on schedule, however, 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 bunching formations before and after the tram in question, or (2) as a single bunching formation 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 bunching formation. 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 4 (4th sub-figure) and Figure 7. 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 bunching swings formation detection is the following:

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 Section 3. Each data point is assigned a particular cluster type (“delayed” / “normal” / “bunched”).
3. During bunching swings formation (BSF) extraction, regard each day separately. Extract all data points, related to this line, direction, date into a current dataset.
4. While 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 is 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 BFS formation. 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.

5.2 Formations clustering and profiling

Once we have separate bunching swings formations extracted, we want to look carefully at their parameters. In this analysis, we used the following parameters when looking at bunching swings formations:

Bunching Swings Formations Parameters:

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, obviously, 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 that change trips density over time and on different days), so any one of them can be used in further analysis, depending on the preference. We used both separately and didn't 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.

Table 2 Clustering results for bunching swings formation types extraction

Feature	**Cluster 1 (36,5%)**	**Cluster 2 (25,1%)**	**Cluster 3 (19.6%)**	**Cluster 4 (18.7%)**
Average passenger load	**35.2 ± 9**	22.2 ± 7	**15.5 ± 5**	29.1 ± 5
Trips involved	4.5 ± 3	5.0 ± 3	**3.5 ± 2**	**15 ± 6**
Duration	1h ± 30m	1h20m ± 30m	50m ± 30m	**3h ± 1h**
Average starting stop	18 ± 5	**8 ± 4**	**22 ± 5**	14 ± 4
Time when starts	12h ± 3h	14h ± 4h	**19h30m ± 4h**	12h ± 3h
Average length in stops	20 ± 5	**31 ± 5**	18 ± 4	26 ± 4
Day type	work 86.3% weekend 13.7%	work 81.8% weekend 18.2%	**work 69.8% weekend 30.2%**	work 80.5% weekend 19.5%
Until route end	**Yes 86.2% No 13.8%**	Yes 98.2% No 1.8%	Yes 95.4% No 4.6%	Yes 100% No 0%
Explanation	Very high average passenger load Medium length during route More often returns to normal schedule before the trip ends	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 Often occurs during weekend	Very long duration with many consecutive trips affected Rather high average passenger load
Nickname	**"High passenger load"**	**"Whole route"**	**"Evening late route"**	**"Long duration"**

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 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 bunching formations 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.

We looked at using different number of clusters K for K-means bunching swings types extraction, and settled at using K=4. We found, that three clusters do not lead to enough distinction between different situations, and result in common occurrences of different formations being assigned to the same cluster. With K=5 and K=6, we found that some clusters had very slight distinction with each other, so extra clusters were not producing extra meaningful insights. However, we would like to note that since we used data for one month for one line only in this analysis, we expect that by increasing the order of data volume used, we may use the same approach with a bigger cluster number to get more nuanced cluster results.

6 Results

6.1 Clustering results

As described in Section 5, we detected, extracted, and clustered bunching swings formations in four different types. You can see the types combined in Table 2. We highlighted the most important differences for each cluster.

1. **"High passenger load"** – The most common type of bunching swings 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. It has to be noted, that this is the only cluster, that contains considerable number of bunching swings formations that resolve before the end of the route, although this number is still rather small, at 13.8%. However, for other clusters it stands at 1.8%, 4.6%, 0%, so by far the vast majority of bunching swings formations cannot be resolved mid-route. Examples can be seen in Figure 4.

2. **"Whole route"** – Bunching swings of this type usually start very early in the route and last for the whole duration of the trip. They have average number of trips involved and average passenger load. Examples can be seen in Figure 5.

3. **"Evening late 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 on weekends much more often than other types. The bunching swings usually start very late in the route, noticeably later than for other clusters. However, by far the most interesting factor of this cluster is the average passenger load, as it is very small, considerably smaller than for other bunching swings formation types. And, very importantly, 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 shown in Section 5.2 and by previous research works, bunching effect in itself correlates considerably with high passenger load. The fact that there is a type of bunching swings formations that consistently happen with low passenger number is, therefore, very interesting, and deserves further investigation into external factors of why this type of bunching swings occurs. Examples can be seen in Figure 6.

4. **"Long duration" –** This formations type contains mainly very long and heavy bunching swings occurrences, lasting for a long time with many trips involved. 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. Examples can be seen in Figure 7.

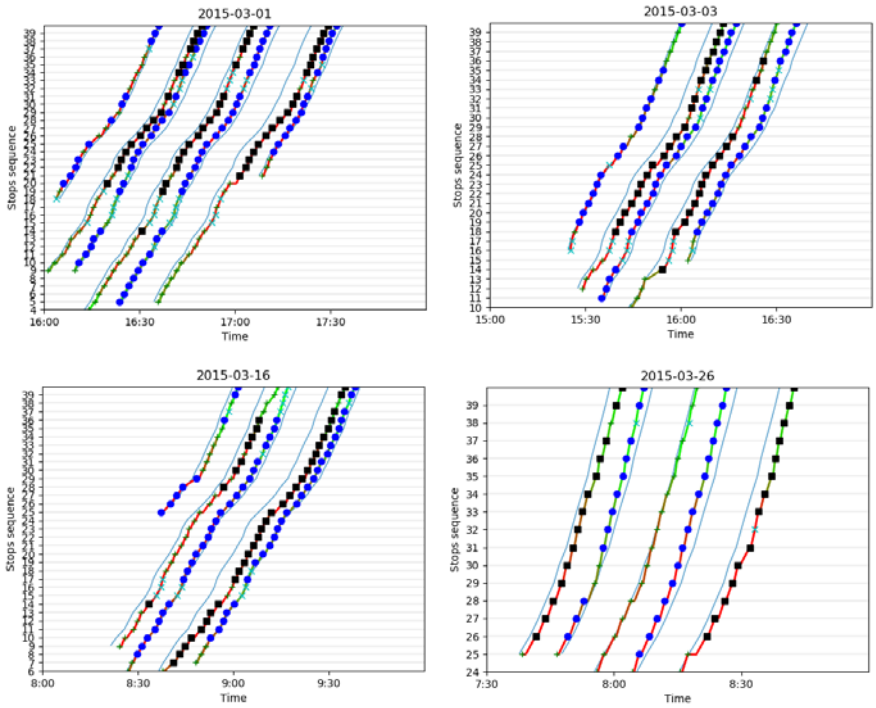


Fig. 4 Examples of bunching swings formations from cluster 1 “High passenger load”

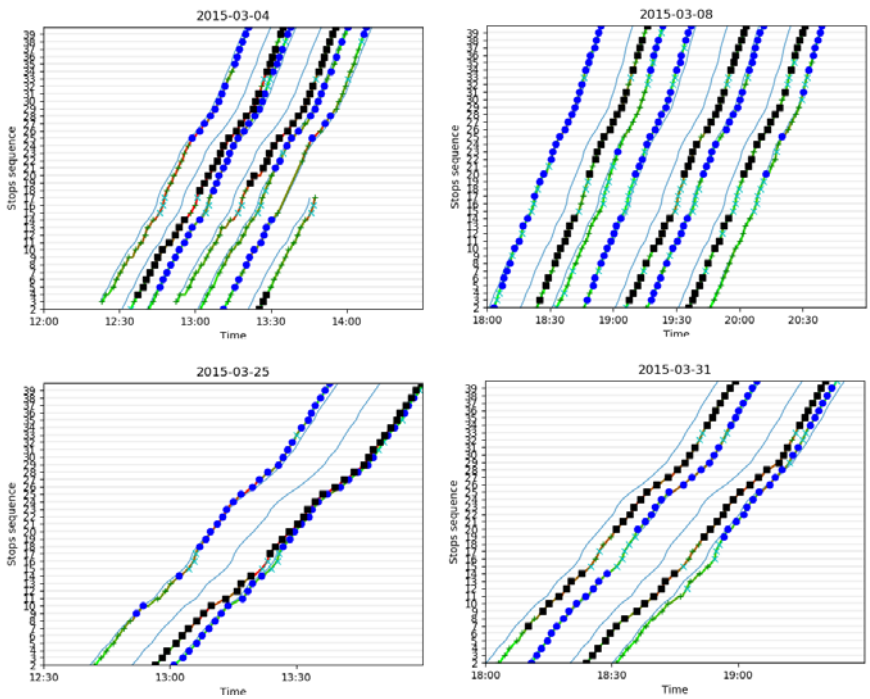


Fig. 5 Examples of bunching swings formations from cluster 2 “Whole route”

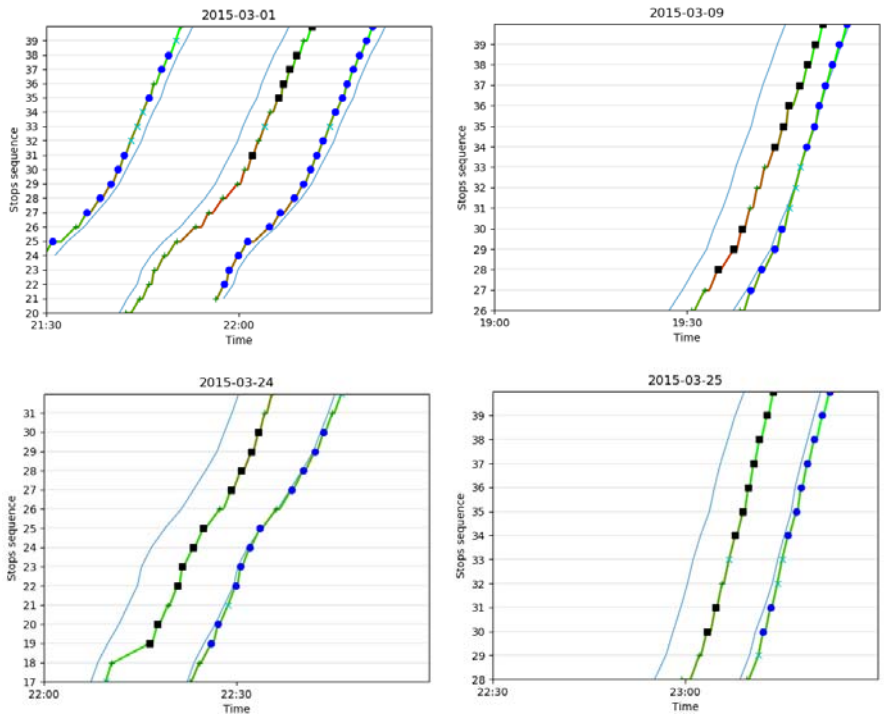


Fig. 6 Examples of bunching swings formations from cluster 3 “Evening late route”

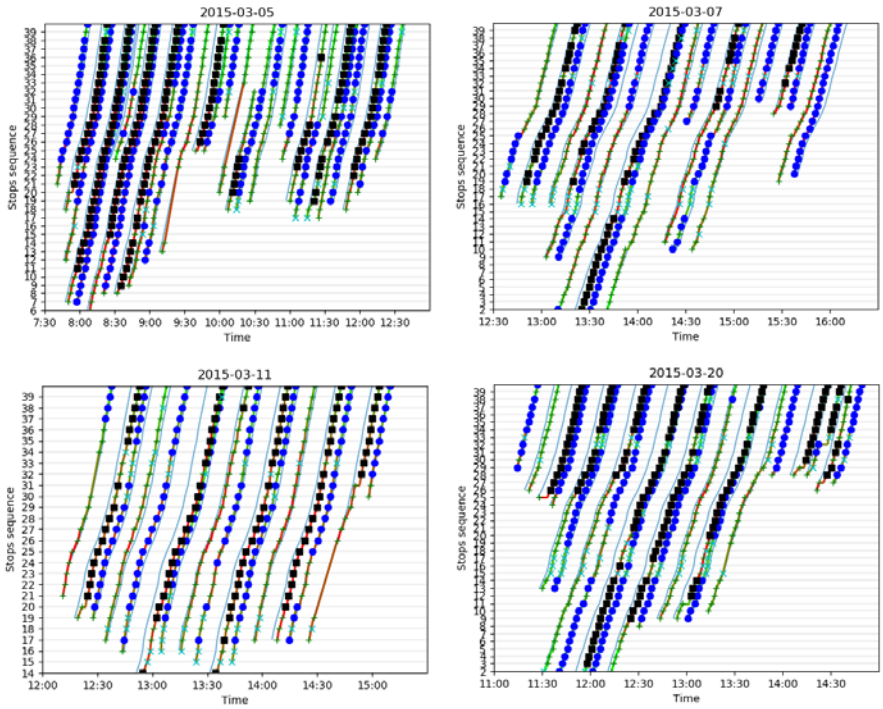


Fig. 7 Examples of bunching swings formations from cluster 4 “Long duration”

6.2 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 Figure 1.

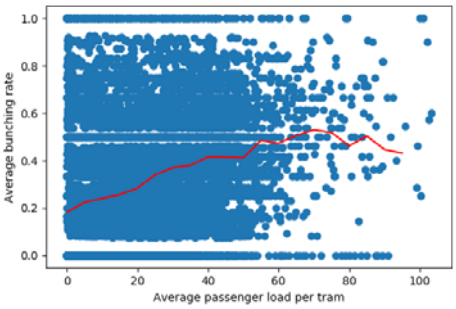


Fig. 9 Bunching rate occurrences depending on passenger load

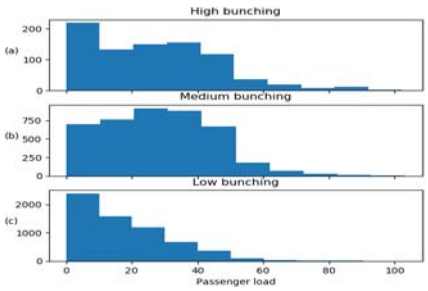


Fig. 8 Histograms of average passenger load for (a) high (x>0.7); (b) medium (0.3<x≤0.7); or (c) low (x≤0.3) bunching rates

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 hours. 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.

Figure 4 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, with the Spearman coefficient between the two values being $\rho=0.86$, and Pearson coefficient being $r=0.88$, which clearly shows a high correlation of passenger load and bunching. In Figure 5 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 for every rate. It can be seen that low bunching rate corresponds to lower passenger load.

7 Conclusions

In this paper, we showed that clustering techniques can be used to extract three fundamental types of a PT vehicle's situation (normal, delayed or bunched). We showed that clustering clearly highlights the 'bunching swings' phenomenon, which sometimes lasts for a considerable time. By varying the number of clusters, we can tune the severity bunching patterns that we extract. We also showed a clear

correlation between passenger load and bunching rate. Clustering results allow us to perform further analysis 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 can be clustered into four types of situations: "high passenger load", "whole route", "evening late route", "long duration".

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. We also aim to use this information to look at how the evolution of bunching swings formation can be predicted in real time.

Acknowledgements: This research was supported by H2020 project My-TRAC (Grant No. 777640). We also would like to thank HTM and Stichting OpenGeo for providing the AFC and AVL datasets, respectively.

References

Yu, H., Chen, D., Wu, Z., Ma, X., & Wang, Y. (2016). Headway-based bus bunching prediction using transit smart card data. Transportation Research Part C: Emerging Technologies, 72, 45-59.

Luo D., Bonnetain L., Cats O. and van Lint H. (2018). Constructing Spatiotemporal Load Profiles of Transit Vehicles with Multiple Data Sources. Proceedings of the 97th Transportation Research Board Annual Meeting, Washington DC.

Daganzo, C. F. (2009). A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. Transportation Research Part B: Methodological, 43(10), 913-921.

Feng, W., & Figliozzi, M. (2011). Empirical findings of bus bunching distributions and attributes using archived AVL/APC bus data. In ICCTP 2011: Towards Sustainable Transportation Systems (pp. 4330-4341).

Osuna, E. E., & Newell, G. F. (1972). Control strategies for an idealized public transportation system. Transportation Science, 6(1), 52-72.

Chapman, R. A., & Michel, J. F. (1978). Modelling the tendency of buses to form pairs. Transportation Science, 12(2), 165-175.

Moreira-Matias, L., Cats, O., Gama, J., Mendes-Moreira, J., & de Sousa, J. F. (2016). An online learning approach to eliminate Bus Bunching in real-time. Applied Soft Computing, 47, 460-482.

Van Oort, N., T. Brands, E. de Romph, M. Yap (2016), Ridership Evaluation and Prediction in Public Transport by Processing Smart Card Data:A Dutch Approach and Example, Chapter 11, Public Transport Planning with Smart Card Data, eds. Kurauchi F., Schmöcker, J.D., CRC Press.

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