

# Data Analysis to Study Sub-threshold Delays Incurred by Tyne and Wear Metro Trains

Daniel Screen<sup>1</sup> · James Parkinson<sup>2</sup> · Christopher Shilton<sup>2</sup> · Aleksandrs Rjabovs<sup>3</sup> · Marin Marinov<sup>4</sup> 

Received: 20 June 2019 / Revised: 24 November 2019 / Accepted: 6 January 2020  
© The Author(s) 2020

**Abstract** The performance of Tyne and Wear Metro system in the UK is measured on a headway basis, and gaps in service that are 4 min or more in excess of scheduled gaps are investigated and the cause documented. The metro system has a number of infrastructure constraints including single-line sections, junctions and level crossings, all of which have to be taken account of when constructing the timetable, in order to avoid trains being held by the signalling system, causing delays. The objective of this study is to analyse delays less than 4 min, which are not investigated or attributed to a cause, known as sub-threshold delays. The purpose of the analysis is to identify regularly occurring issues which are due to the timetable, in order to recommend changes. Two different data sets were used. The first data set explored specific trains, areas and times of day where delays were highest. The second data set allowed us to drill down on each of those in greater detail by studying station departure times for each train. A number of options to resolve the issues identified during the analysis are proposed. Whilst the results are specific to the Tyne and Wear Metro system, the methodology is suitable for use by other urban rail transit systems. The study identified several areas of future work including resolving data recording issues, carrying out further investigation of trains at peak times in particular scenarios,

and automating the analysis through the use of other software.

**Keywords** Urban rail transit · Punctuality · Performance analysis · Metro · Metro trains · Delays · Timetables

## 1 Introduction

The Tyne and Wear Metro is a light rail network in North East England. The network consists of 60 stations which form two lines, the yellow line and the green line. During the daytime, a 12-min frequency is provided on each line, with additional trains added to run between selected stations during weekday morning and afternoon peaks. A 15-min frequency runs on each line during evenings and Sundays.

The objective of this study is to analyse the operation data and investigate sub-threshold delays incurred by Tyne and Wear Metro trains. The performance of Tyne and Wear Metro is measured on a headway basis, with delays being attributed to causes when the actual headway is 4 min or more greater than scheduled. Sub-threshold delays are defined as delays of less than 4 min, and are not investigated or attributed in the current performance regime of the metro system in question.

Performance of metro systems and train delays have been studied by Marinov and Viegas [1], Rjabovs et al. [2], Rjabovs and Palacin [3], Wales and Marinov [4], Dampier and Marinov [5], Darlton and Marinov [6], Rjabovs and Palacin [7], Rjabovs and Palacin [8], and Powell et al. [9, 10].

Marinov and Viegas noted that because of the “upstream downstream” flow nature of the rail system, a malfunctioning rail subsystem such as a junction may

---

✉ Marin Marinov  
m.marinov@aston.ac.uk

<sup>1</sup> Newcastle University, Newcastle upon Tyne, UK

<sup>2</sup> Tyne and Wear Metro, Newcastle upon Tyne, UK

<sup>3</sup> Network Rail, London, UK

<sup>4</sup> Aston University, Birmingham, UK

---

Communicated by Waessara Weerawat.

compromise the entire service, causing upstream delays and downstream idle periods. In such awkward situations, the final product of the train service rapidly deteriorates, leading to low customer satisfaction and a significant increase in operating costs on average over the long term [1].

Wales and Marinov developed a variety of tactics to reduce primary and secondary delays by introducing both a delay information system and measures to reduce travel times. Strategies to address ahead-of-schedule operations were also developed and evaluated using an event-based simulation modelling methodology [4].

Rjabovs et al. studied the performance of metro drivers in terms of safety-related incidents and what performance-shaping factors could affect them. They concluded that there is a clear relationship between locations and incident propagation. The operational environment and physical system constraints were major contributors to this relationship. For example, the risk of signal passed at danger occurrences is higher at locations where conflicting train movements might occur, such as near sidings and double junctions [2].

Rjabovs and Palacin concluded that drivers have good route knowledge that should in theory translate into near ideal timetable keeping if there are no other influences. One can claim that irregularities in a timetable causing conflicting movements and delays is one of the examples of such influence [3].

Marinov et al. provided an overview of the general principles of constructing a timetable. This paper, however, is interested in issues with metro performance due to the assumptions used in constructing the timetable that do not reflect the actual situation when trains run. Therefore, this analysis is not interested in large, one-off delays such as a technical fault. Specifically, this paper is developed with the aim of amalgamating data for a time period of 16 months, ranking metro trains by lateness, investigating reasons for regular metro train delays, and recommending changes to reduce this issue [11].

## 2 Methodology

A methodology was developed that can be used in the future to repeat the analysis. Specifically, macros were created in Excel so that this analysis would be easy to repeat. A stepwise approach was employed and applied to the Tyne and Wear Metro:

- Identify any operational and infrastructure-related constraints to be taken into consideration when analysing train delays.

- Ensure robust real-world data sets, preferably raw data of the behaviour of trains.
- Check the data collected, and if inaccurate records are detected, cleanse the data.
- Check whether the cleansed data can be described by normal distribution.
- Complete the data cleansing.
- Create a new data set, set and run macros to format and analyse the data.
- Draw interim conclusions based on what the data analysed so far have identified.
- Outline the areas of the system where large delays are occurring.
- Call these areas “areas of large delays” (ALD).
- Take a sample and continue the data analysis for the outlined ALDs.
- Identify the sections where more time can be given to particular trains within each ALD.
- Identify the metro trains with the highest average train delays and rank them.
- Select the trains with the greatest delay and perform an in-depth analysis for each of them.
- Identify possible causes for each of the most delayed trains.
- Draw conclusions and suggest improvements.
- Advise on future work.

## 3 Tyne and Wear Metro Network

### 3.1 Layout and Stations

The layout of the Tyne and Wear Metro operating network is shown in Fig. 1, with the stations at which performance is routinely measured, known as monitoring points, shown in red. Performance is measured in both directions at each station, with the exception of Longbenton, where performance is measured in the direction towards South Gosforth only. For the purpose of both simplicity and presentation, a list of metro station abbreviations is added, as shown in Table 1.

The Metro owns and controls most of the infrastructure used, with the exception being Pelaw to South Hylton, which is controlled by Network Rail. The yellow and green lines combine between Pelaw and South Gosforth, which is the busiest part of the network. Trains are timetabled to provide regular service frequency between Pelaw and South Gosforth; as a result, on a 12-min service for the yellow and green line, this gives a 6-min frequency between Pelaw and South Gosforth. Peak short trains run between Pelaw and Monkseaton and between Pelaw and Regent Centre during weekday morning and evening

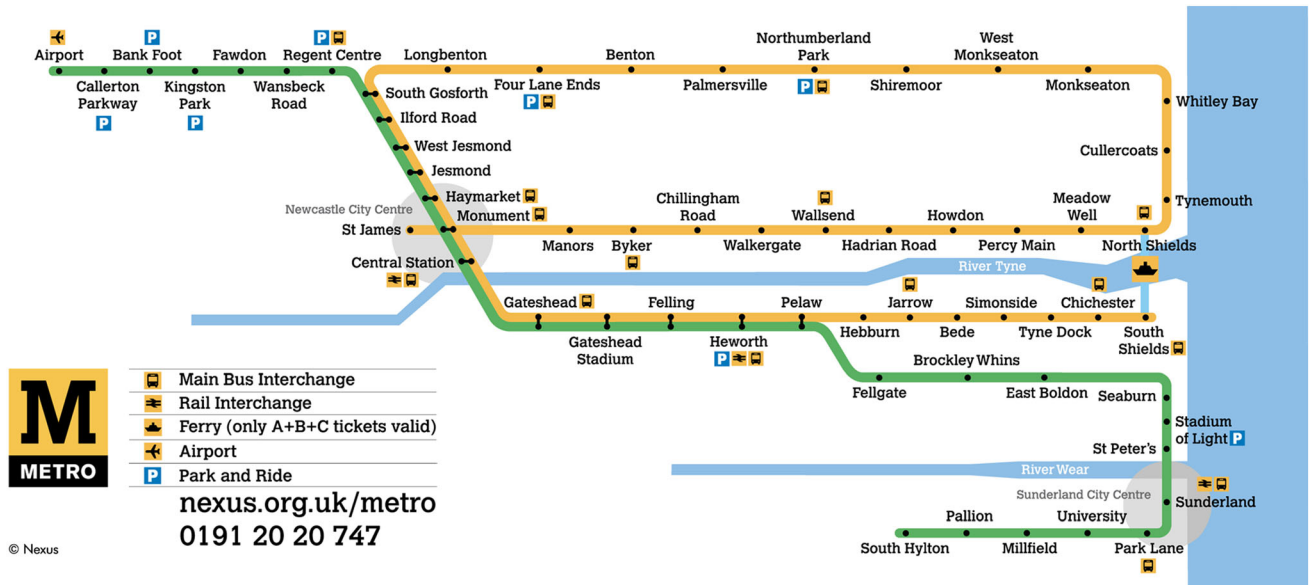


Fig. 1 Illustration of the layout of the Tyne and Wear Metro operating network Source: Nexus

peaks. These peak short trains result in a 3-min frequency between Pelaw and South Gosforth, and as they run between the services that run to the end of the yellow and green line, this results in them running within 3 min of these trains between South Gosforth and Regent Centre, and South Gosforth and Monkseaton, in both directions.

### 3.2 Network Constraints

There are several network constraints that must be taken into consideration when analysing the Tyne and Wear Metro train delays. Firstly, there is a single line section connecting the four stations between Pelaw and Bede, with passing loops at stations. This means that only one train can travel in either direction at any given time, and trains must be timetabled so that there are no clashes caused by this single line section.

Secondly, there are junctions where the green and yellow lines combine at Pelaw and South Gosforth. At these junctions, if two trains are ready to enter the same platform from the separate lines, one must wait for the other. In a similar manner, the timetable must be constructed to avoid trains crossing each other's paths at the junction, for example at South Gosforth a train departing to Longbenton at the same time as a train arriving at South Gosforth from Regent Centre.

There are also five level crossings on the system. They are located at Howdon, Fawdon, Kingston Park, Bank Foot and Callerton Parkway. Only one train can use a level crossing at one time, meaning the timetable must avoid trains arriving at level crossings at the same time. When

peak short services terminate at Monkseaton or Regent Centre, they must cross the other line to access the sidings, creating an opportunity for clashes and delays. At Pelaw this does not present a problem, as the siding is located between the two tracks.

The Network Rail-controlled infrastructure also has constraints. There are junctions at Pelaw and Sunderland, and this section of track interfaces with other train services such as LNER, Grand Central, Northern Rail and freight trains.

These constraints are mitigated in the current timetable; however, the margins between trains at these constraints can be very small. Whilst options to improve the margins have been examined in the past, improving a margin at one constraint can decrease the margin or create a conflict at another constraint. Therefore, as the margins can be very small, ensuring that the actual train running times reflect the timetable assumptions is important, in order to prevent delays.

All these network constraints mean that the effects of a small delay to a train can multiply. For example, if a train is slightly delayed, it can miss its allotted slot to enter a junction and become delayed additional minutes, or it can cause delays to other trains if it is late into a junction, making the other train wait. As the signalling system used in Tyne and Wear Metro is based on a fixed block, a late-running train can create a "domino effect" of small delays to trains behind it, especially when running through the central section between South Gosforth and Pelaw at a peak time.

**Table 1** Tyne and Wear Metro station names and abbreviations

Tyne and Wear Metro stations			
Station abbreviation	Station name	Station abbreviation	Station name
APT	Airport	TYN	Tynemouth
CAL	Callerton Parkway	CUL	Cullercoats
BFT	Bank Foot	WTL	Whitley Bay
KSP	Kingston Park	MSN	Monkseaton
FAW	Fawdon	WMN	West Monkseaton
WBR	Wansbeck Road	SMR	Shiremoor
RGC	Regent Centre	NPK	Northumberland Park
SGF	South Gosforth	PMV	Palmersville
ILF	Ilford Road	BTN	Benton
WJS	West Jesmond	FLE	Four Lane Ends
JES	Jesmond	LBN	Longbenton
HAY	Haymarket	FGT	Fellgate
MTS	Monument (North and South lines)	BYW	Brockley Whins
CEN	Central Station	EBO	East Boldon
GHD	Gateshead	SBN	Seaburn
GST	Gateshead Stadium	SFC	Stadium of Light
FEL	Felling	MSP	St. Peters
HTH	Heworth	SUN	Sunderland
PLW	Pelaw	PLI	Park Lane Interchange
MTW	Monument (East and West lines)	UNI	University
MAN	Manors	MIL	Millfield
BYK	Byker	PAL	Pallion
CRD	Chillingham Road	SHY	South Hylton
WKG	Walkergate	HEB	Hebburn
WSD	Wallsend	JAR	Jarrow
HDR	Hadrian Road	BDE	Bede
HOW	Howdon	SMD	Simonside
PCM	Percy Main	TDK	Tyne Dock
MWL	Meadow Well	CHI	Chichester
NSH	North Shields	SSS	South Shields

## 4 Data Analysis

### 4.1 Data Sets

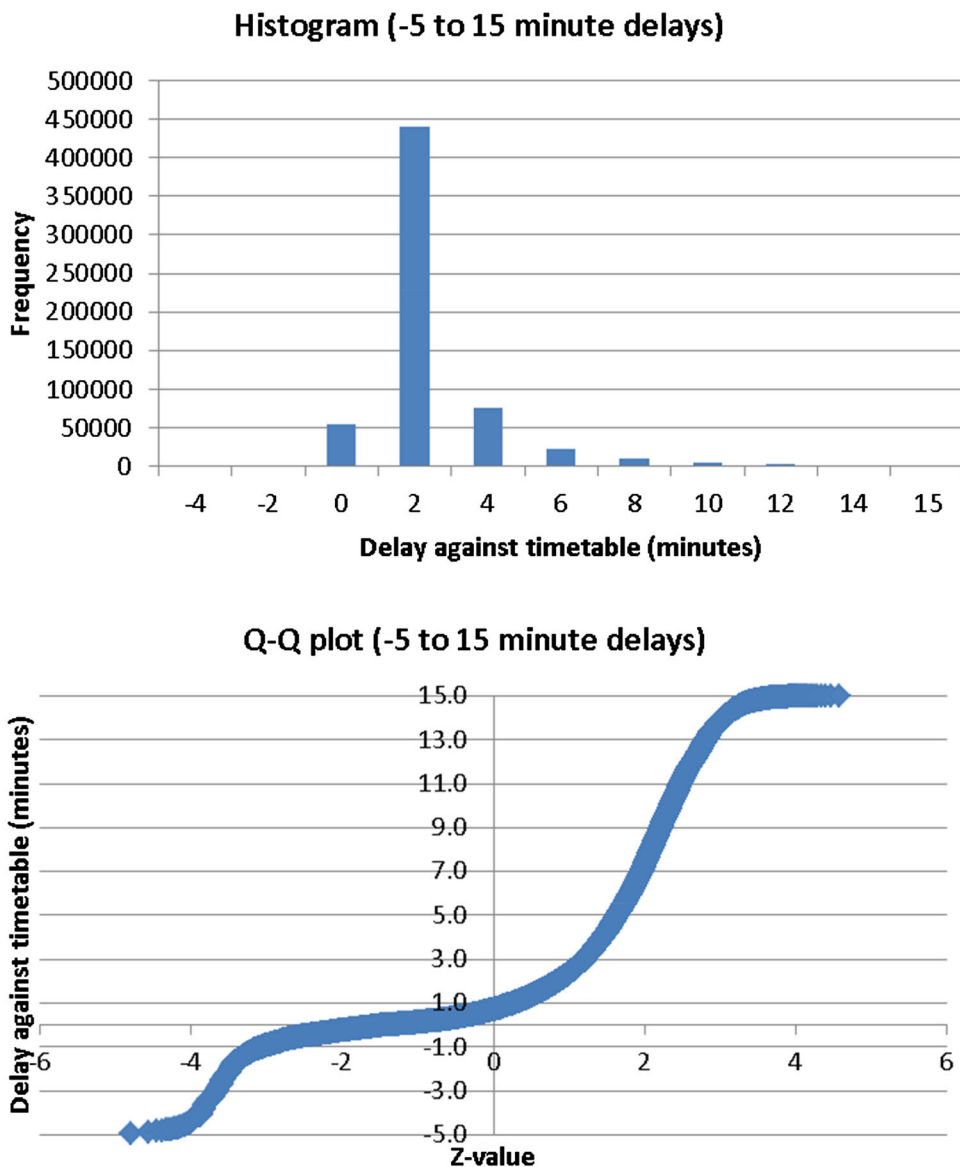
Two data sets, provided by Tyne and Wear Metro, were studied for the purpose of the analysis. The first data set, *dataset1*, consisted of records from the 15 monitoring points shown in Fig. 1, including a large amount of information about the performance of each train. Specifically, it contained 17 columns including the time that trains departed from the 15 monitoring points around the Tyne and Wear Metro system, the delay at each location, train number, timeslot (time of day), direction of travel and other information regarding headway and attributed events where excess headway was 4 min or more. *Dataset1* contained 894,527 rows of data.

The second data set, *dataset2*, was similar to *dataset1*, though it included more data. Specifically, *dataset2* contained the time that each train departed each station as opposed to each monitoring point. It also contained the time that the train approached, changed ends, pressed “ready to start” and stood in the station. This created significantly more data, including 14 columns and 20,299,755 rows of data in total.

### 4.2 Data Cleansing

Data records were checked and cleansed carefully. Certain dates were removed due to events which would skew the results of the analysis. The date column was filtered to remove data coinciding with such events. Specifically, data cleansing resulted in removing days when a different timetable was in place, which included all weekends and

**Fig. 2** Q–Q plot and histogram of delay data



bank holidays and periods of engineering works. Days where performance was affected by other factors were also removed, including periods of low rail adhesion in autumn and records of disruptions caused by extreme weather conditions, for example large storms and flooding.

The data was checked for any patterns of delay that would cause a day or records to be an outlier, skewing the results. This led to the identification and exclusion of large jumps in delays between stations, or large value delays, as these were considered to be one-off events.

There were several delay values of around  $-33,000$  s due to a technical fault with the train tracking software. Such values were removed as well. Large delays caused by one-off incidents were removed. Several normality checks were performed to determine which range of delay data to

analyse. Histograms and Q–Q plots for several different ranges of data were plotted. The delay range closest to a normal distribution was the range  $-5$  min to  $10$  min, as shown in Fig. 2. These plots show that this data range fits the description of a normal distribution but do not confirm that the data is normally distributed.

It was also observed that when a train arrives early, its delay time is recorded as a negative value. Hence, early train arrivals were excluded from the analysis of *dataset1*, as negative values would cancel out some positive delay values and skew the results. Early train arrivals were included in the analysis of *dataset2*.

The cleansed data was then used as a new data set. For formatting and analysis of the new data set, macros were set and run in Excel.

### 4.3 Data Reformatting

The cleansed raw data was re-formatted using custom made macros in Excel. Apart from removing unnecessary columns, improvements to layout of the data, and converting data into a required format, the macros were used for:

- Displaying the data by a timeslot (time of day)
- Displaying the data by a monitoring point
- Displaying the data by train
- Displaying the data by train at each monitoring point
- Colour-coding delays

### 4.4 Sample

To maximise accuracy and produce representative results, a sample of 49 days from a 16-month period were selected. Before collecting the sample, all days deemed to be outliers were excluded. One day from each week was then chosen—for example, Monday 6 February, Tuesday 14 February, Wednesday 22 February 2017—and cycled to account for daily changes, and this was continued up to the end of June 2018. This provided both a more manageable and a more representative data set.

### 4.5 Macros

Macros were used to amalgamate data from different days and calculate the average delay change between stations compared with the time allowed in the timetable. Information on the time allowed in the timetable was also added. Macros were written to present the information in two different ways. The first was by amalgamating data for all trains between stations, as in Table 2. Data across the network was analysed in the same manner and other similar changes identified.

Individual trains, e.g. T101–T111, T125, T155, were studied next in order to determine an average delay per location and a delay change from the last monitoring point. In order to speed up the investigation process, colour coding was employed for this analysis, as can be seen in Figs. 3, 4 and 5. This colour coding was added both for the average delay against timetable, and delay change between stations.

## 5 Results and Discussion

### 5.1 General Findings

The table of delay by train, at each monitoring point, is shown for a sample of trains in Fig. 3. Studying such diagrams for all trains produced a number of findings.

Firstly, trains were most delayed during Peak timeslots, e.g. 7 am to 9 am. The peaks are the most congested times in the system when headways are considerably reduced, especially through the central section, and volumes of passengers are highest. Any late running will have a significant knock-on effect. One can note that for the evening peaks (3:30 pm to 5:30 pm) the knock-on effect continued well past those times, with many monitoring points highlighted as red from 6 pm onwards. This can be attributed to a timetable change after 6 pm when running times between stations are reduced, corresponding to the change in service frequency for the evening when there are fewer passengers. This results in less time available for drivers to recover any delays. Hence, it would be of interest to consider changing the time when the change to running times occurs in order to mitigate the peak knock-on effects.

Secondly, trains were most delayed at certain monitoring points. These were typically those towards the end of the line, such as Callerton Parkway OUT and Monument OUT. This suggests that trains pick up delay on each journey, which is typically recovered at a terminus station. It is possible that trains are affected by the constraints of the system, e.g. not arriving on time for a timetabled slot for a level crossing, towards the end of their running diagrams. Future work should look into rearranging any slack time available in the timetable towards the end of each journey to mitigate such trends.

Thirdly, trains pick up the largest delays travelling in certain sections of the line, notably around Seaburn (refer to Fig. 3).

Finally, Table 2 shows that on average, each metro train that travelled between Pelaw and Fellgate became delayed by 11.1 s more than they had been. In addition, between Brockley Whins and East Boldon, each train was delayed by an additional 66.7 s. However, trains decreased their delays between Fellgate and Brockley Whins, and East Boldon and Seaburn.

Assuming that any decrease in delay will have a positive effect on performance, it can be proposed that the

**Table 2** PLW–SBN, table showing average delay change from PLW to SBN and time allowed in the timetable to travel between each station

Station-to-station departure	PLW–FGT	FGT–BYW	BYW–EBO	EBO–SBN
Average delay change compared with timetable (seconds)	11.126	−3.564	66.712	−16.004
Time allowed in timetable (minutes: seconds)	04:45	02:00	03:00	03:00





SchedTime	Location	Average Delay	Average Delay Change
14:34:45	WMMN_OUT	62.45	
14:36:45	MSN_OUT	42.09	-20.36
14:38:30	WTL_OUT	53.64	11.54
14:40:00	CUL_OUT	53.64	0.00
14:42:45	TYN_OUT	43.78	-9.86
14:45:45	NSH_OUT	39.31	-4.47
14:47:45	MWL_OUT	45.69	6.38
14:49:00	PCM_OUT	68.00	22.31
14:51:00	HOW_OUT	64.37	-3.63
14:53:15	HDR_OUT	76.22	11.85
14:55:00	WSD_OUT	77.00	0.78
14:57:15	WKG_OUT	73.78	-3.22
14:58:45	CRD_OUT	77.16	3.38
15:00:45	BYK_OUT	73.91	-3.25
15:03:15	MAN_OUT	70.56	-3.34
15:05:00	MTW_OUT	88.51	17.95
15:13:30	SJM_IN	23.03	-65.48
15:15:15	MTW_IN	18.72	-4.31
15:16:45	MAN_IN	15.29	-3.43
15:19:15	BYK_IN	21.73	6.44
15:21:15	CRD_IN	30.70	8.98
15:22:45	WKG_IN	36.12	5.42
15:25:15	WSD_IN	35.97	-0.15
15:27:00	HDR_IN	35.29	-0.68
15:29:30	HOW_IN	48.47	13.19
15:31:30	PCM_IN	64.69	16.22
15:33:00	MWL_IN	64.69	0.00
15:35:15	NSH_IN	45.76	-18.94
15:38:15	TYN_IN	32.97	-12.79
15:41:00	CUL_IN	46.09	13.12
15:42:30	WTL_IN	64.94	18.84
15:44:15	MSN_IN	54.22	-10.72
15:46:15	WMMN_IN	56.85	2.63
15:48:45	SMR_IN	64.00	7.15
15:50:45	NPK_IN	65.70	1.70
15:53:00	PMV_IN	58.46	-7.24
15:55:30	BTN_IN	60.54	2.09
15:57:00	FLE_IN	62.64	2.10
15:58:30	LBN_IN	76.91	14.28
16:01:30	SGF_IN	46.44	-30.47
16:03:00	ILF_IN	57.92	11.47
16:04:30	WJS_IN	48.17	-9.75
16:06:30	JES_IN	51.69	3.52
16:08:15	HAY_IN	48.67	-3.02
16:10:00	MTS_IN	51.39	2.72
16:11:45	CEN_IN	40.24	-11.15
16:14:00	GHD_IN	50.74	10.49
16:15:45	GST_IN	52.15	1.42
16:17:30	FEL_IN	81.88	29.72
16:19:15	HTH_IN	60.00	-21.88
16:21:15	PLW_IN	60.06	0.06
16:25:30	HEB_IN	68.50	8.44
16:28:30	JAR_IN	66.28	-2.22
16:31:00	BDE_IN	64.56	-1.72
16:32:30	SMD_IN	67.22	2.66
16:34:30	TDK_IN	92.25	25.03
16:36:30	CHI_IN	78.22	-14.03
16:42:30	SSS_OUT	49.50	-28.72
16:44:30	CHI_OUT	62.88	13.38
16:46:30	TDK_OUT	56.06	-6.83
16:48:30	SMD_OUT	49.20	-6.86
16:50:00	BDE_OUT	41.58	-7.62
16:52:45	JAR_OUT	70.76	29.17
16:55:30	HEB_OUT	70.47	-0.29
16:59:45	PLW_OUT	68.85	-1.62
17:01:30	HTH_OUT	53.72	-15.13
17:03:15	FEL_OUT	48.76	-4.96
17:05:00	GST_OUT	48.76	0.00
17:07:00	GHD_OUT	49.60	0.84
17:09:15	CEN_OUT	59.88	10.28
17:11:00	MTS_OUT	61.97	2.09
17:12:30	HAY_OUT	71.56	9.58
17:14:15	JES_OUT	65.79	-5.77
17:16:15	WJS_OUT	62.60	-3.19
17:17:45	ILF_OUT	75.27	12.67
17:20:00	SGF_OUT	96.81	21.53
17:22:30	LBN_OUT	118.83	22.03
17:24:00	FLE_OUT	119.25	0.42
17:25:30	BTN_OUT	129.97	10.72
17:28:00	PMV_OUT	138.86	8.88
17:30:15	NPK_OUT	151.56	12.71
17:32:15	SMR_OUT	148.50	-3.06
17:34:45	WMMN_OUT	202.78	54.28
17:36:45	MSN_OUT	187.88	-14.91
17:38:30	WTL_OUT	187.88	0.00
17:40:00	CUL_OUT	187.88	0.00
17:42:45	TYN_OUT	199.77	11.89
17:45:45	NSH_OUT	189.39	-10.38
17:47:45	MWL_OUT	193.14	3.75

**Fig. 4** Formatted table of average delays for the 125 metro train, with average delay of 60 s or more in yellow and 150 s or more in red, and average delay change of 0–15 s in yellow, 16–29 s in amber, and 30 s or more in red. Negative delay changes are shown in green

timetabled allowance in that section of the system be redistributed, in particular by reducing the timetabled allowance from East Boldon to Seaburn by 15 s and increasing it by 15 s between Brockley Whins and East Boldon. Even though this would not account for all the delay incurred between these two stations, it should reduce the risk of a train being affected by system constraints.

### 5.2 Performance of Particular Trains

Train 125 increased its delay from 40 to 215 s between 16:50 at BDE and 17:34:45 at West Monkseaton (WMMN) on average (see Fig. 4). It was also observed that Train 125 maintained an average delay of around 100 s or more from



SchedTime	Location	Average Delay	Delay Change
16:25:30	RGC_IN	21.4	
16:28:30	SGF_IN	71.40740741	50.01
16:30:00	ILF_IN	87.07407407	15.67
16:31:30	WJS_IN	77.22222222	-9.85
16:33:30	JES_IN	117	39.78
16:35:15	HAY_IN	90.77272727	-26.23
16:37:00	MTS_IN	80.65384615	-10.12
16:38:45	CEN_IN	63.19047619	-17.46
16:41:00	GHD_IN	59.6	-3.59
16:42:45	GST_IN	59.6	0.00
16:44:30	FEL_IN	84.20833333	24.61
16:46:15	HTH_IN	87.80952381	3.60
16:48:15	PLW_IN	144.6190476	56.81
16:53:12	DEP_PLW	144.6190476	0.00
16:56:45	PLW_OUT	55.18181818	-89.44
16:58:30	HTH_OUT	32.59259259	-22.59
17:00:15	FEL_OUT	38.5	5.91
17:02:00	GST_OUT	38.5	0.00
17:04:00	GHD_OUT	45.65384615	7.15
17:06:15	CEN_OUT	55.90909091	10.26
17:08:00	MTS_OUT	44	-11.91
17:09:30	HAY_OUT	51.55555556	7.56
17:11:15	JES_OUT	57.09090909	5.54
17:13:15	WJS_OUT	62.81818182	5.73
17:14:45	ILF_OUT	61.86363636	-0.95
17:17:00	SGF_OUT	64.18518519	2.32
17:19:30	LBN_OUT	85.03703704	20.85
17:21:00	FLE_OUT	80.77777778	-4.26
17:22:30	BTN_OUT	94.03703704	13.26
17:25:00	PMV_OUT	96.8	2.76
17:27:15	NPK_OUT	110.61111111	13.81
17:29:15	SMR_OUT	105.3636364	-5.25
17:31:45	WMN_OUT	120.6363636	15.27
17:33:45	MSN_OUT	159.1363636	38.50

Fig. 5 Formatted table of average delays for the 155 metro train

18:26:30 to 21:02. Several potential causes of this train’s high average delay are listed in Table 3.

Data analysed for other trains following the same route showed the same patterns as Train 125. What was observed was that large delays were frequently picked up travelling from Manors (MAN) to Monument (MTW), South Gosforth (SGF) to Longbenton (LBN), and Howdon (HOW) to

Table 3 Potential causes of the high average delays of Train 125

★	Possibly held at West Monkseaton (WMN) waiting for peak train to clear following signal
★	Frequent delays picked up travelling from MAN to Monument (MTW)
★	Frequent delays picked up travelling from South Gosforth (SGF) to Longbenton (LBN). Possibly caused by waiting at junction
★	Frequent delays picked up travelling from Howdon (HOW) to Hadrian Road (HDR)
★	Possibly a recording error or consistent delay increase from Simonside (SMD) to Tyne Dock (TDK) then decrease from Tyne Dock (TDK) to Chillingham Road (CHI)

Table 4 Potential causes of the high average delays for Train 155

★	Delays at junctions
★	Delays at Monkseaton (MSN) and Pelaw (PLW). This is caused by the 155 train needing to enter sidings at these stations

Hadrian Road (HDR). To resolve this, it is recommended that the timetabled allowance between these stations be increased.

It was also observed that large delays were picked up at West Monkseaton (WMN) during the evening peak. Large delays did not occur at West Monkseaton during any other period of the day, implying that this delay was caused by peak shorts entering sidings late at Monkseaton.

Analysis was carried out on other trains which were most delayed. The analysis of Train 155 (Fig. 5) shows that on average, the train arrived at its first station 21 s late and then gained another 50 s of delay before reaching South Gosforth. It is important to note that Train 155 is an evening peak service. Two potential causes were proposed to explain this pattern (see Table 4). One of those is drivers leaving the depot late, thus arriving at the South Gosforth junction late and missing their slot at the junction, hence picking up 70 s of delay before leaving the first timetabled station.

The analysis further shows that Train 155 is delayed by 1 min or more for its entire journey. This provides strong support to an assumption that peak services outside the central corridor can significantly affect regular trains, when delayed. When entering sidings at Pelaw and Monkseaton, the train is delayed on average 170 s and 146 s, respectively. As peak shorts are only 3 min ahead of normal service trains, Train 155 would cause delays to trains behind it. This is likely to occur with other peak shorts as well. One way to mitigate this is to increase the gap between peak and regular trains to 4 min. However, further investigation is needed to understand how such a change would affect timetables and interfaces between trains in other parts of the system, particularly between Pelaw and South Gosforth.

## 6 Conclusions

In this work, a data analysis was conducted to investigate the sub-threshold delays incurred by Tyne and Wear Metro trains. The analysis first made use of cleansed data from 15 monitoring points to identify trains, areas and times of interest where delays most frequently occur. This allowed use of data from each station to home in on these items of interest to more precisely identify the area in which the delay occurred and, in conjunction with network and service pattern knowledge, the potential reason for this delay.

The analysis revealed that there are several changes that could be made to the timetable which would reduce the size and number of sub-threshold delays.

Firstly, the analysis of the data provided showed that there were consistent large delays from 6 pm to 8 pm throughout the network. The timetables for most trains change at roughly 6 pm, giving each metro train less time to travel between stations. This change does not allow the metro trains to decrease their delays, meaning that this causes the trains to remain delayed until 8 pm or later. To prevent this from occurring in the future, it is suggested that this timetable change is implemented at 7 pm instead of 6 pm.

The data analysis suggested several changes to the timetabled allowance for trains between stations, to better match the actual running times achieved.

In addition, this analysis suggested that the timetable for peak shorts should change. The large delays occurring at West Monkseaton during the peak periods and the fact that Train 155 and other peak trains are on average 3 min late at some points in their journey prove that peak trains cause delays for regular service trains. To reduce these delays, this analysis has suggested running peak shorts 4 min in front of regular service trains.

Another option to reduce these delays at West Monkseaton would be adding another signal halfway between West Monkseaton and MSN. The journey from West Monkseaton to Monkseaton has a timetabled allowance of 2 min. This means that while the peak train is entering sidings at Monkseaton, the regular service train could depart from West Monkseaton and travel for roughly 1 min before needing to wait at the signal. The added delay to regular service trains is typically 40 s, which implies that by the time the train reached the new signal, the peak will have cleared the station at Monkseaton, and the regular service train will not need to stop. This solution would also cause less congestion than changing peaks to 4 min in front, as they would be further away from the green line trains while travelling between Pelaw and South Gosforth.

To reduce delays during the peak period, it may be of interest to run coast peak trains 3 min behind coast trains

and add 6 min to the timetabled allowance for the trains to travel to Monkseaton and back. This would mean that the peaks are always on time and would not hold up the regular service train when entering sidings at Monkseaton. When they arrive at South Gosforth they would be 3 min in front of a coast train and would run as normal to Pelaw. When at Pelaw, the train would turn around and have 6 min less time in sidings, and it would depart Pelaw 3 min behind a coast train. For peak shorts that travel to Regent Centre, the procedure would be the same but with a 6-min extra wait at Regent Centre instead of travelling to Monkseaton.

As peak trains are leaving the depot late, this analysis suggests that train drivers should make sure that they leave on time for service and monitor their performance.

Whilst the infrastructure constraints and service pattern will vary in other metro systems, the methodology can be applied to other metro systems, as it will capture similar data on the running of trains.

## 7 Future Work

To increase the accuracy of data analysis, future studies can be done to more precisely investigate data recording errors. Resolving these errors would make it possible to gain greater knowledge of the delays currently experienced.

It is worth analysing the pattern of peak trains in greater detail. Sometimes during peak periods there is a 6-min gap between two trains through the central corridor followed by two gaps of 3 min. This could be changed in the future to be more flexible, and instead of these gaps we could create three gaps of 4 min. This would ease the pressure on junctions and cause fewer knock-on delays. It would be worthwhile to investigate the junction margins to see whether there are any changes that could be made to reduce the likelihood of trains waiting to enter the junction.

Further work should also be carried out to analyse the performance of the Tyne and Wear Metro system when peak trains do not run. There are some days when peak shorts do not run due to a shortage in drivers or operational trains. These days could be analysed and compared with days where all peak shorts do run, to show exactly how much the delay of regular trains is reduced by peak shorts.

In addition, there have been temporary speed limit reductions in place over several sections of track, and further analyses could be carried out here to determine the repercussions of these speed limits.

Whilst the macros automate the data analysis, they are constructed to work with a specific timetable. As the timetable can change permanently on a 6-monthly or annual basis, the macros have to be updated to work with

the new timetable. Automation of the process through a data analytics and business intelligence platform such as Qlik would make it easier to carry out analysis in the future.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Marinov M, Viegas J (2011) A mesoscopic simulation modelling methodology for analyzing and evaluating freight train operations in a rail network. *Simul Model Pract Theory* 19:516–539
2. Rjabovs A, Palacin R, Robinson M (2014) Cab and system design influence on metro drivers' performance: preliminary study. In: *Transport research arena*, April 2014, Paris
3. Rjabovs A, Palacin R (2015) Attitudes of metro drivers towards design of immediate physical environment and system layout. *Urban Rail Transit* 1(2):104–111
4. Wales J, Marinov M (2015) Analysis of delays and delay mitigation on a metropolitan railway network using event based simulation. *Simul Model Pract Theory* 52:55–77
5. Dampier A, Marinov M (2015) A study of the feasibility and potential implementation of metro-based freight transportation in Newcastle upon Tyne. *Urban Rail Transit* 1(3):164–182
6. Darlton A, Marinov M (2015) Suitability of tilting technology to the Tyne and Wear Metro system. *Urban Rail Transit* 1(1):47–68
7. Rjabovs A, Palacin R (2016) The influence of system design-related factors on the safety performance of metro drivers. *Proc Inst Mech Eng Part F J Rail Rapid Transit* 231:317–328
8. Rjabovs A, Palacin R (2016b) Design and layout of the physical environment in a metro system: appraisal of Tyne and Wear Metro drivers' perceptions. In: *WCTRS*, July 2016, Shanghai
9. Powell JP, Fraszczyk A, Cheong CN, Yeung HK (2016) Potential benefits and obstacles of implementing driverless train operation on the Tyne and Wear Metro: a simulation exercise. *Urban Rail Transit* 2(3):114–127
10. Powell JP, Batty P, González-Gil A, Palacin R (2017) Determining system-wide energy use in an established metro network. *Proc Inst Mech Eng Part F Rail Rapid Transit* 231(5):570–577
11. Marinov Marin, Şahin İsmail, Ricci Stefano, Vasic-Franklin Gordana (2013) Railway operations, time-tabling and control. *Res Transp Econ* 41(1):59–75