CTA Purple Line Delay Propagation Analysis

By Charlie Marshall
Northwestern University
TRANS 310
21 May 2021

Abstract

The Purple Line is the Chicago Transit Authority's main route serving the northern suburbs of Evanston and Wilmette. The Purple Line Express provides a quick and reliable mode of transit too and from the city's central business district for many working professionals in these areas. Fifteen weekdays of afternoon Purple Line Express runs were collected and analyzed to understand the effect of delay propagation on the express service. This dataset includes the arrival times at stations of nearly 330 runs. Overall, delayed, early, and on-time departures from Linden all arrived at Linden at the end of their run earlier than scheduled on average. Headways stay relatively consistent throughout the entirety of a run, no matter if the run starts early, delayed, or on-time. Run time analysis showed that the trip's status as delayed, on-time, or early when departing Linden had no effect on its ability to arrive back to Linden quicker than scheduled run time. These results are most likely an impact of low ridership caused by the COVID-19 pandemic. Operators likely deal with less delays caused by passengers attempting to board the train and can spend significantly less dwell time in the station. Overall, it is unclear whether consistent early trains, which is a frequent occurrence during the captured period, is good for overall quality of service. Passengers who come to expect a train to arrive at a specific scheduled time might be dissatisfied when that train has departed early.

1 Introduction

Since 1908, the Purple Line has provided service to riders in Evanston and Wilmette. The modern Purple Line provides two distinct service configurations. During non-peak hours and weekends, the Purple Line is a shuttle between Linden Station in Wilmette and the Howard Terminal in Evanston, allowing riders to connect to Yellow and Red Line trains at Howard Terminal. During morning and evening weekday rush hours, the Purple Line provides an express service to Chicago's central business district, "The Loop." The Purple Line runs express on the North Side Main Line between Howard Terminal and Belmont Station, only stopping at Wilson Station, and provides service around The Loop. During this trip, a Purple Line train has many opportunities to interact with other CTA "L" routes including the Yellow and Red Lines at Howard Terminal, The Brown Line at Belmont Station, and the Pink, Green, Orange, and Brown Lines at Tower 12 and Tower 18 in The Loop. These interactions with other trains at critical junctions in a Purple Line Express run have the potential to cause deviations from scheduled arrival times.

This research paper analyzes whether delays caused by interactions with other routes continue to propagate throughout the rest of the trip on the CTA Purple Line. Propagating delays have effects on the quality and consistency of service that can be provided to Purple Line passengers.

This paper specifically analyzes delay propagation in the PM Peak period for the Purple Line Express. The PM peak is preferred to the AM peak when considering delay propagation analysis because the majority of PM ridership is boarding in The Loop. Therefore, PM Purple Line Express trains do half of a trip before picking up much of their ridership, making the perceived quality of service much more dependent on propagating delays. In the AM peak period, passengers are mainly boarding the train early in its trip, so their service quality is not as sensitive to delay propagation.

Along with delays, another metric for service quality is the difference between the scheduled headway and actual headway. A headway is the time difference between trains on a route. Ideal service involves headways which are close to scheduled, allowing sufficient gaps between trains to accommodate for demand. However, when a train is delayed, it can cause the headway to the train in front to balloon and the headway to the train behind to shrink. There will be excess demand for the delayed train and low demand for the train following behind. This paper analyzes whether expanding headways also propagate throughout the system.

Lastly, the run time of purple line trains was analyzed to add context of the results of the previous delay and headway analysis. In general, run times are an important metric for operations and scheduling, especially for a long run like the Purple Line Express because it encapsulates the entire system performance.

2 Literature Review

Literature on delay propagation, including the metrics used in different industries informed my research and analysis. Ghofrani, F. *et al.* (2018) was a great starting point to understand recent applications of big data analytics in railway transportation. This survey paper set out different applications of big data analytics, especially safety, operations, and maintenance and set forth various analytics techniques used. My focus was on descriptive analytics in operations, which was most useful to my own research and intentions.

Yuan, J. (2006) is a PhD Thesis paper written by a PhD candidate Jianxin Yuan at Delft University. The first part of Yuan's thesis included background research on analysis and modeling of train delays and delay propagation. Yuan sets forth the negative exponential distribution as the main distribution used to model non-negative headways and non-negative delays in past research, first defined in Schwanhauber (1974). Additionally, this thesis set forth much of the established jargon around delays, such as primary and secondary delays. Yuan's own research involved analyzing delay propagation at the Hauge HS Station in the Netherlands, including performing analysis on 9 different routes in both directions at the station. In general, this paper set a good framework for what I should focus on in my own research, including the actual distribution of delays instead of just averages. In Yuan's research and background info, non-negative delays and the exponential distribution was used because it is assumed that trains would operationally not leave earlier than scheduled. However, with real CTA Purple Line data the non-negative assumption proved to be false.

From the aviation perspective, Mitsokapas, E. *et al.* (2021) presented a statistical analysis of airline arrival delays from UK airports between 2018 and 2020. Unlike Yuan (2006), Mitsokapas also analyzed and fit distributions to negative delays, meaning planes which arrived earlier than expected and considered the consequences of these delays along with positive delays. Additionally, this study considered the effects of the pandemic on delays and operations, which would be important to my own research which is affected by the pandemic.

Carey, M. and Kwieciński, A. (1994) considered the effect of knock-on delays and varying headway times on the overall trip time. The shorter the scheduled headway between trains, the greater is the expected knock-on delay and hence the greater trip times of the following trains. This literature was important in justifying the analysis of headways and run time in my own analysis along with delays.

3 Data Sources

Data was collected from the CTA Train Tracker API, a tool which publishes real-time estimated arrival times for all trains running in the 'L' system. This tool is most commonly used by developers wishing to make customer-facing applications to display real-time train waiting times for trains. The Train Tracker API, specifically the Locations API, publishes the real time location of all trains on a chosen 'L' route. For each train on the route, the API gives an estimated arrival time at the next station on the route. The CTA claims that train arrival predictions average less than one minute difference from the actual arrival times (CTA Developer Center). Additional data including the direction of the train is headed and whether it is "due" in the station is also included. Using a python script and a cronjob, I logged minute-by-minute location data for three weeks (15 weekdays) from April 26th until May 14th for the Purple Line during the PM rush period from 2pm until 9pm. This data was subsequently cleaned by removing duplicate predicted arrival times and organized into trips, which include the predicted arrival times at each station during a Purple Line run from leaving Linden southbound to arriving at Linden northbound.

The data collection and Train Tracker API does have its shortcomings. Of the twenty-eight scheduled daily PM rush trips, on average one to three trips were fully missed or only partially logged each day. Additionally, some stations such as Foster and Noyes southbound and Howard and Davis northbound consistently did not have predicted arrival times. Both these issues can be contributed to failures in the Train Tracker API or with the ability of track sensors to detect the presence of rolling stock. Additionally, inconsistencies in a trip, for instance one or two station arrival times missing in an entire trip, can potentially be contributed to the data collection methodology. Specifically, some predictions might have been missed because data were only logged once a minute. Despite the shortcomings, 321 complete Purple Line loop trips were logged over the 15-day period, a diverse and complete representation of Purple Line Express service.

Additionally, scheduled delays, headways, and run times were calculated using the Weekday Rail Supervisors Guide for the Purple Line. Scheduled times are only published for select locations along a route. For the Purple Line, these locations include: Linden station, Davis station, Howard Terminal, Thorndale Interlocking, Montrose Interlocking, Clark Junction, Fullerton Station, Tower 18, and Tower 12. Many of these locations are not stations, but nearby stations were used to approximate, including: Wilson station used to approximate Montrose Interlocking, Belmont station for Clark Junction, Merchandise Mart for Tower 18, and Adams/Wabash for Tower 12.

All data analysis for the project was done in Microsoft Excel. Distribution fitting was done using @Risk Software, a risk analysis and simulation Microsoft Excel Add-in.

4 Analysis

4.1 Initial Delay Analysis

Delays were calculated by subtracting the real time predicted arrival time at a location from the scheduled arrival times presented in the Supervisors Guide. Delays were calculated in minutes. A positive delay represents a trip which is behind schedule, whereas a negative delay represents a trip which is ahead of schedule.

The delays at each station for all 15 days were analyzed and descriptive statistics are displayed in Table 1.

Table 1: Descriptive Statistics for Purple Line PM Arrival Delays at Select Stations

	Leaving Linden SB	Davis SB	Howard SB	Wilson SB	Belmont SB	Fullerton SB	Merchandise Mart SB	Adams/Wabash SB	Merchandise Mart NB	Fullerton NB	Belmont NB	Wilson NB	Howard NB	Davis NB	Linden NB
Average Delay (min.)	0.52	-0.83	-1.19	-4.04	-2.27	-2.68	-3.64	-3.34	-2.07	-4.60	-5.20	-3.85	-4.88	-5.79	-5.31
Std. Dev. (min.)	3.23	3.25	3.25	3.32	3.54	3.44	3.67	4.13	4.43	4.69	4.93	4.89	5.79	6.30	5.95
Min. Delay (min.)	-5.17	-6.08	-6.45	-9.50	-9.42	-10.00	-12.47	-14.50	-15.38	-19.35	-20.82	-17.48	-27.72	-20.35	-20.37
Max Delay (min.)	16.45	14.60	13.05	10.58	10.55	10.00	10.58	14.65	16.28	14.05	13.30	15.57	15.98	15.07	21.18
# of Runs	321	322	321	327	284	289	328	321	319	291	285	172	241	142	324

On average, trains leaving Linden were delayed 0.52 minutes compared to their scheduled departure time. However, since this is within the one-minute buffer, we can assume that on average trains leave Linden station on time. At each of the fourteen additional locations, trips arrived earlier than their scheduled time on average. As the trip continues, the trains tend to arrive earlier, continuing to gain time compared to the schedule. By the time the trip enters The Loop at Merchandise Mart, an average trip is 3.64 minutes earlier than scheduled. A train arrives at back at Linden station on average 5.31 minutes earlier than scheduled.

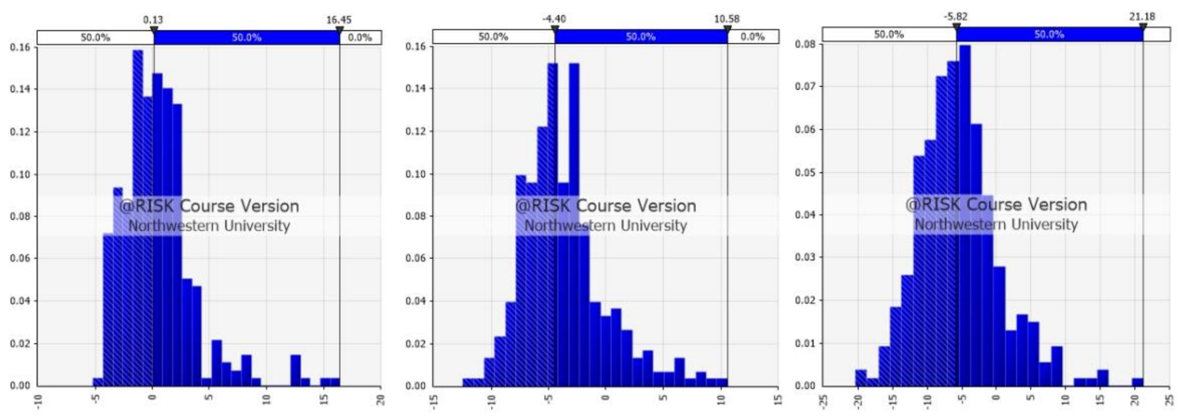

Figure 1: Probability Distribution of Delays on PM Peak Runs at Selected Stations (Left: Leaving Linden SB, Middle: Merchandise Mart SB, Right: Arriving at Linden NB)

Furthermore, Table 1 displays that as the average run continues, the standard deviation of the delays continues to expand. This is especially true on the second half of the express trip when the train is returning from the loop. On the first half of the trip, from Linden to Adams/Wabash, the standard deviation grows by 27.9%. On the second half of the trip, from Adams/Wabash to Linden, the standard deviation grows by 44.1%.

Additionally, Table 1 shows that the range of delays seen at each station also increases throughout the trip. When leaving Linden, the range of delays was 21.62 minutes. Entering The Loop at Merchandise Mart station, the range of delays was 23.05 minutes. Getting back to Linden station, the range of delays was 41.55 minutes.

The probability distributions in Figure 1 confirm the range and standard deviation data from Table 1. Additionally, the mean is greater than the median, so the probability distributions of delays at the selected stations are skewed right by large, positive delays which are outliers in the data. This skew rules out a normal distribution as a good fit to describe the distribution of delays. Using @Risk software, a Log-logistic distribution is found to best fit the distribution of arrival delays.

4.2 Delay Propagation Based on Initial Delay

Further delay analysis was done to understand how staring delayed, early, or on-time might affect delay propagation through a run. The initial hypothesis was that if a trip started delayed it would be delayed further through its run because of unscheduled interactions with other trains. Figure 2 displays the delay propagation of trains which were delayed when leaving Linden station, trains which were on-time when departing Linden station, and trips which were early when departing Linden station. The aggregate trips are also included for comparison. Delayed trips are defined as leaving at least 3 minutes after the scheduled departure time from Linden. On-time trips depart Linden station within 3 minutes of their scheduled departure time. Lastly, early trips depart Linden at least 3 minutes before their scheduled departure time. The 3-minute threshold was carried over from the analysis done in Yuan, J. (2006). Trendlines are included in Figure 2 to display the general delay propagation through a trip.

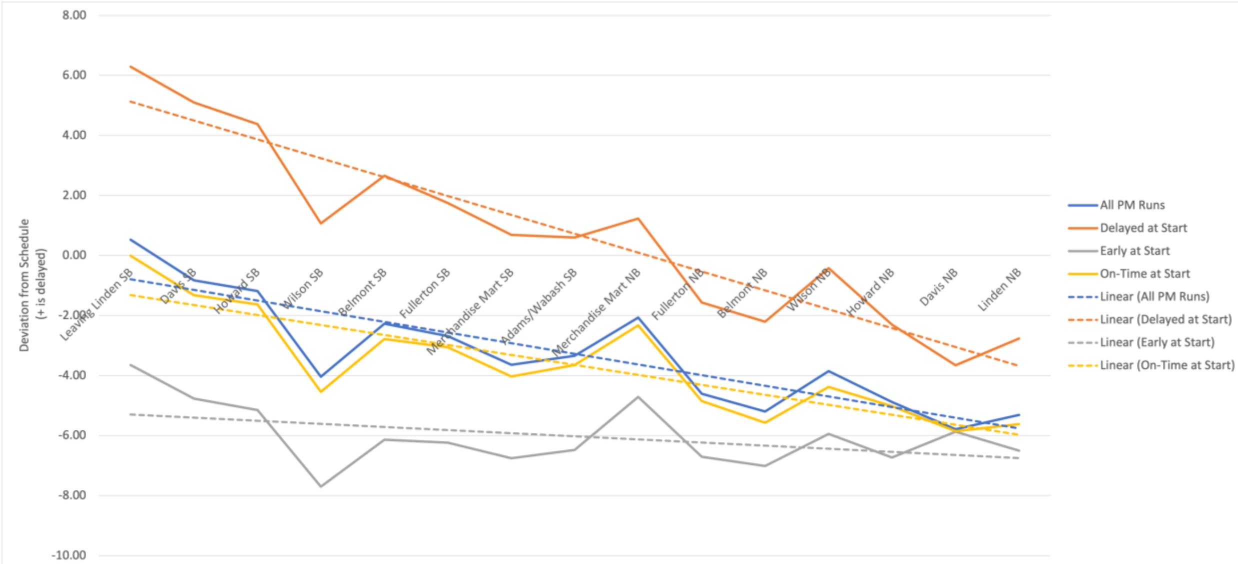

Figure 2: Graphical Representation of Delay Propagation through PM Purple Line Express Run at Selected Stations

From Figure 2, trips with delayed departures are on average 6.29 minutes late leaving Linden. However, these trips make up significant time throughout the run and arrive at Linden station northbound 2.77 minutes earlier than scheduled. Therefore, over this 3-week period, on average, trips which started delayed did not continue to get more delayed throughout the trip. Additionally, delayed trips had the largest difference between starting and ending delay of all three categories of trips, making up 9.06 minutes on average over the entirety of the run.

Trips with early departures leave Linden station 3.65 minutes early and arrive at Linden 6.50 minutes earlier than scheduled. Early departing trips had the smallest different between starting and ending delay, 2.85 minutes.

Since most trips are considered "on-time" leaving Linden, these trips generally follow the same trend as the aggregate delay data. On average on-time trips depart Linden exactly on time and arrive to Linden 5.62 minutes before their scheduled arrival.

4.3 Subset of Delays

Further analysis was performed by looking at only a subset of runs which arrived more than 5 minutes late to Linden station at the end of their trip. Figure 3 displays the delay propagation for these runs. The solid lines represent trips which initially departed Linden station late, more than 3 minutes after the scheduled time. The dotted lines represent trips which initially departed Linden station "on-time", within 3 minutes of their scheduled departure time.

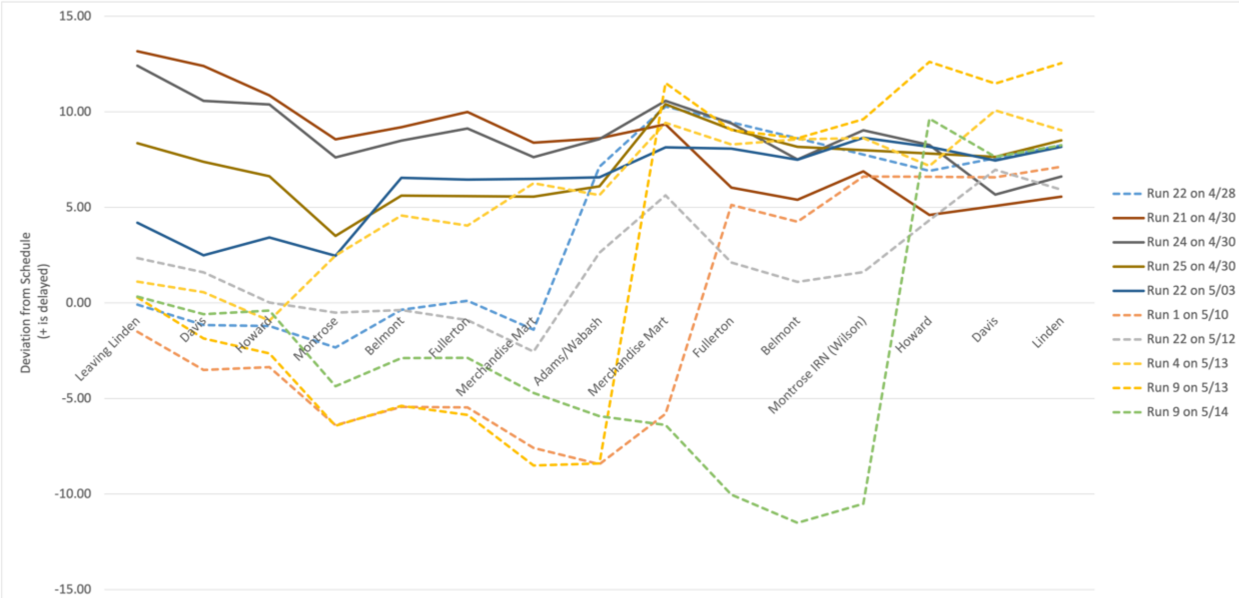

Figure 3: Subset of Runs with Large Delays Arriving at Linden NB

In Figure 3, nearly all the trips which initially departed on-time (dotted lines) experience a major delay at a point in the trip which deviates them significantly from the schedule. For instance, Run 9 on 5/13 (orange dotted line) experienced a major delay between Adams/Wabash and Merchandise Mart which changed its trajectory from 8.40 minutes ahead of schedule at Adams/Wabash to 11.52 minutes behind schedule at Merchandise Mart.

In Figure 3, trips which departed Linden station more than 3 minutes after their scheduled time (solid lines) did not experience major delays. These trips started late and were not able to make up time on their run. From Section 4.2 of the Analysis, this is not usually the case for trips which depart Linden delayed.

4.4 Headway Analysis

Scheduled headways were calculated by subtracting the scheduled arrival time for the leading train from the scheduled arrival time of the following train. The real-time headways were calculated by subtracting the real time predicted arrival time of the leading train from the real time predicted arrival time of the following train. The scheduled headway is not constant throughout the peak period. At its max, the scheduled headway is 15 minutes at the end of the peak period. At the minimum, the scheduled headway is 4 minutes in the middle of the peak period. To compare headways of different length, the actual headway was presented as a proportion of the scheduled headway for the trip.

The headways at select station for all 15 days were analyzed and descriptive statistics were displayed in Table 2.

Table 2: Descriptive Statistics for Purple Line PM Headways at Select Stations

	Leaving Linden SB	Davis SB	Howard SB	Wilson SB	Belmont SB	Fullerton SB	Merchandise Mart SB	Adams/Wabash SB	Merchandise Mart NB	Fullerton NB	Belmont NB	Wilson NB	Howard NB	Davis NB	Linden NB
Headway	1.04	1.04	1.04	1.02	1.03	1.02	1.01	1.03	1.03	1.03	1.03	0.93	1.14	1.07	1.04
Std. Dev.	0.27	0.29	0.30	0.33	0.35	0.36	0.37	0.42	0.43	0.43	0.47	0.47	0.54	0.55	0.57
Min	0.38	0.46	0.17	0.10	0.25	0.13	0.14	0.19	0.11	0.10	0.09	0.18	0.19	0.16	0.15
Max	2.14	2.23	2.25	2.14	2.41	2.52	2.56	2.60	2.81	2.42	2.50	2.08	2.93	2.41	3.08
Runs	240	235	230	234	181	189	230	222	214	176	173	63	121	60	216
# Below Headway	119	112	109	113	85	92	115	104	107	89	81	35	54	31	107
# Above Headway	120	121	121	120	95	95	115	118	107	87	91	28	67	29	109

On average, trains leaving Linden were had a headway 4% greater than their scheduled headway. At each of the fourteen additional locations, trips had similarly longer than scheduled headways on average, staying relatively consistent throughout the trip. By the time the trip enters The Loop at Merchandise Mart, the average headway was 1% greater than scheduled. On average, each trip arrived at Linden station with a headway 4% greater than scheduled. For the longest scheduled headway, 15 minutes, a 4% greater than scheduled is a 36 second difference between actual and scheduled headway

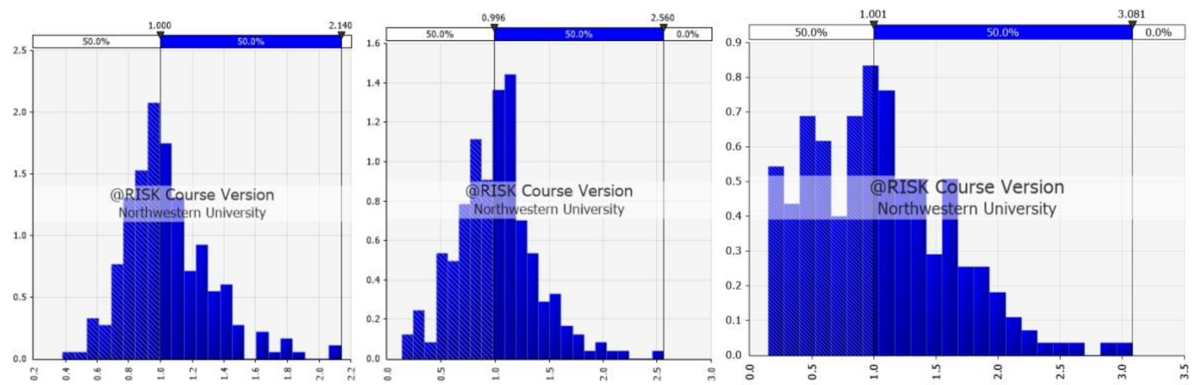

Figure 4: Probability Distribution of Headways on PM Peak Runs at Selected Stations (Left: Leaving Linden SB, Middle: Merchandise Mart SB, Right: Arriving at Linden NB)

Furthermore, Table 2 displays that as the average run continues, the standard deviation of the headways continues to expand. Unlike delays, the headway standard deviation grows relatively larger in the first half of the express. On the first half of the trip, from Linden to Adams/Wabash, the standard deviation grows by 59.3%. On the second half of the trip, from Adams/Wabash to Linden, the standard deviation grows by 35.7%.

Additionally, Table 2 shows that the range of headways seen at each station also increases throughout the trip. When leaving Linden, the range of headways was 1.76. Entering The Loop at Merchandise Mart station, the range of headways was 2.42. Getting back to Linden station, the range of headways was 2.93.

The probability distributions in Figure 4 confirm the range and standard deviation data from Table 2. Additionally, the mean is greater than the median, so the probability distributions of headways at the selected stations are skewed right by large, positive headways which are outliers in the data. This skew rules out a normal distribution as a good fit to describe the distribution of delays. Using @Risk software, a Log-logistic distribution is found to best fit the distribution of arrival delays.

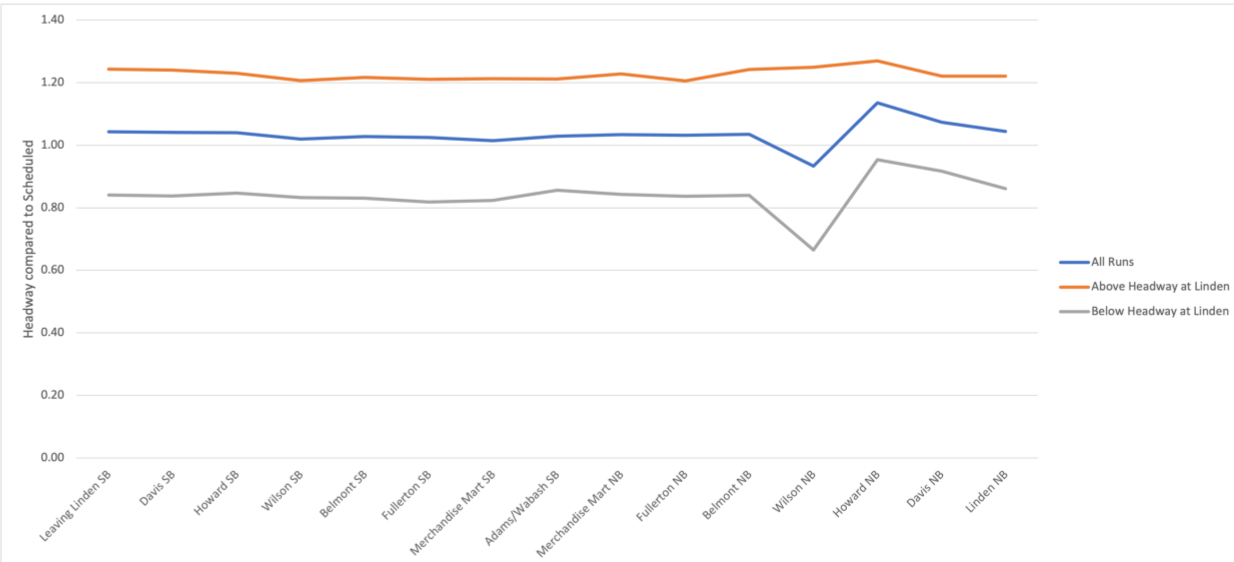

Figure 5: Headway Propagation through PM Purple Line Express Run at Selected Stations

Figure 5 displays a graphical representation of how headways propagate through the route, including headways that were longer than scheduled, headways shorter than scheduled, and an aggregate of all headways. In all cases, the average headway stayed relatively unchanged throughout the run.

4.5 Headway Propagation Based on Initial Delay

Additional headway propagation analysis was performed based on whether the train was initially delayed, early, or on-time when leaving Linden. On-time trains are defined as within 3 minutes of scheduled departure time. Figure 6 displays a graphical representation of the headway propagation based on initial departure delay.

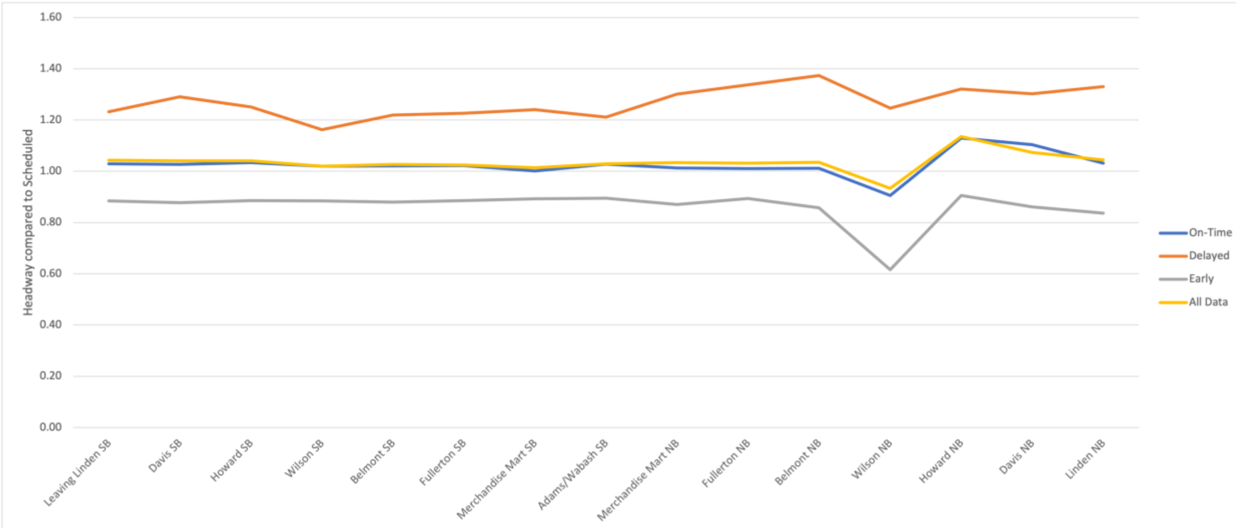

Figure 6: Headway Propagation through PM Purple Line Express Run at Selected Stations Based on Initial Delay at Linden

From Figure 6, runs which were delayed (more than 3 minutes behind schedule) when departing Linden on average had a larger than scheduled headway. Additionally, this headway increased by 7.9% on average by the end of the trip. Runs which were early (more than 3 minutes ahead of schedule) when departing Linden had a shorter than scheduled headway on average. Additionally, this headway decreased by 5.4% on average by the end of the trip. Because most trips are on-time departing Linden, the on-time trips closely follow the aggregate trips and have a headway very near scheduled on average. This headway does not change by the end of the trip, however, Figure 6 displays fluctuations in the headway around Wilson and Howard northbound.

4.6 Run Time Analysis

The scheduled run time of a trip is calculated by subtracting the scheduled arrival time returning to Linden from the scheduled time departing Linden station. The actual run time of a trip is calculated by subtracting the actual arrival time returning to Linden from the actual time departing Linden station. The run time is important because there is limited rolling stock and later express runs depend on the earlier runs to consistently arrive at Linden within a certain window to depart on-time. Because of the length of the run and number of interactions with other routes, the Purple Line Express has the highest run time in the entire CTA ‘L’ System. During the PM Peak period, the shortest scheduled run time is 105 minutes and the longest is 112 minutes. Since scheduled run times are not constant for each trip, actual run times were expressed as a proportion of their scheduled times. Figure 7 displays the probability distribution of all run times during the 3-week period.

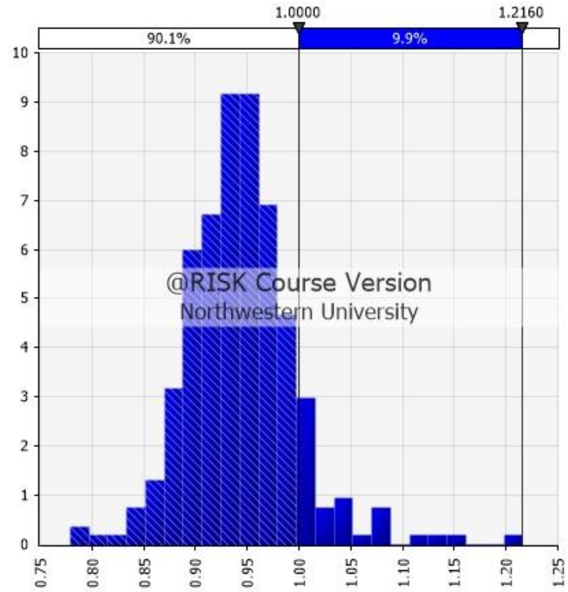

Figure 7: Probability Distribution of PM Run Times During 3-week Period

During the three-week period, the average run time took 94% of the scheduled time. Additionally, 90.1% of all runs were shorter than scheduled. Additionally, the mean and median are close together, suggesting the data is not skewed in either direction. However, @Risk distribution fitting software favors a Log-logistic distribution over a Normal distribution as the best-fit distribution.

4.7 Run Time Based on Initial Delay

Additional run time analysis was done based on initial delays departing Linden. Run times were compared for runs which started on-time, delayed, and early when departing Linden. Like other sections, runs which departed within three minutes of their scheduled departure time were considered on-time. Figure 8 displays the findings from this analysis.

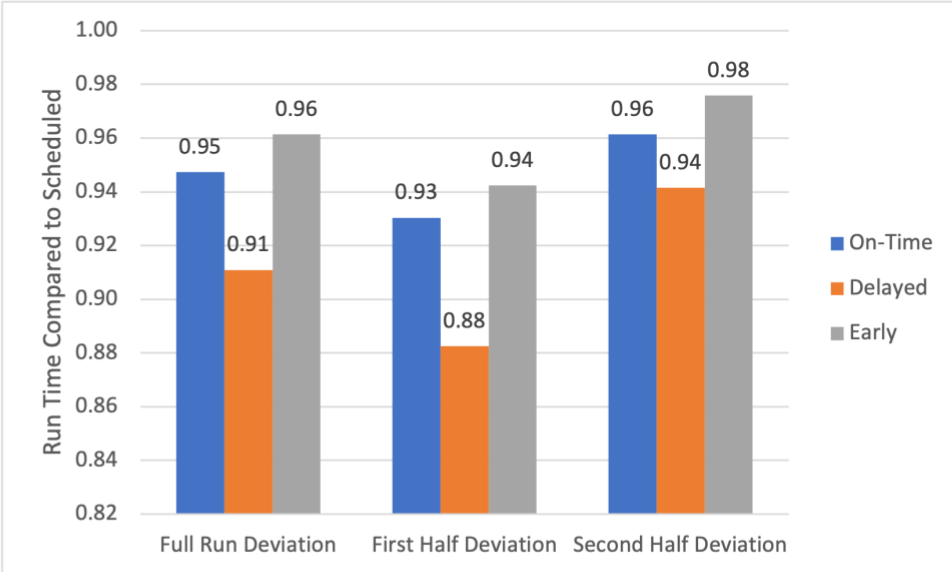

Figure 8: Run Time Compared to Scheduled Based on Initial Linden Departure Delay

From Figure 8, runs which departed delayed from Linden had shorter run times than runs which departed on-time or early. Runs which departed delayed from Linden had on average 9% shorter overall run times and 12% shorter run times than scheduled in the first half of the run (from Linden to Adams/Wabash). Overall, early, on-time, and delayed trains all had shorter run times than scheduled on average.

5 Discussion of Findings

In summary, fifteen weekdays of PM Purple Line Express runs were collected and analyzed. This dataset includes the arrival times at stations of nearly 330 runs, which can be formed into delay, headway, and run time data.

Overall, starting a trip delayed did not lead to the delay propagating and expanding throughout the trip. On the contrary, trips with delayed departures ended their run earlier than scheduled on average. In fact, during the three-week period, delayed, early, and on-time departures from Linden all arrived earlier than scheduled on average. This phenomenon is most likely an impact of low ridership caused by the COVID-19 pandemic. Because of low ridership at stations, operators likely deal with less delays caused by passengers attempting to board the train and can spend significantly less dwell time in the station. Empirically this is supported by Figure 8, showing that delayed runs have lower overall, first half, and second half run times relative to the schedule than on-time and early runs. This combination has led to significantly lower mean delays.

However, there are some aspects of rail operations which have not been impacted by the pandemic, mainly the heavy influence of large delays. This influence can be seen in the right skew of the probability distributions in Figure 1. As seen in Figure 3, large delays can have a significant impact on service quality. Even while ridership returns to normal, these large delays will continue to threaten the quality of service which can consistently be provided by the CTA. Operationally, the focus needs to be on mitigating these delays, so they are as short and infrequent as possible.

Next, headway analysis proves that although run times are below scheduled, headways are consistently near their scheduled time on average. Additionally, headways stay relatively consistent throughout the entirety of a run, no matter if the run starts early, delayed, or on-time. Consistent headways close to their scheduled time is a mark of good service. Passengers can rely on a train within a reasonable amount of time. However, headways for delayed trains tend to stay consistently above scheduled and not improve as the trip continues. Therefore, a delayed train often causes passengers to wait longer than scheduled for their train.

Run time analysis showed that the trip’s status as delayed, on-time, or early when departing Linden had no effect on its ability to arrive back to Linden quicker than scheduled run time. In fact, runs which were delayed leaving Linden had were 12% quicker than scheduled on average and quicker than early and on-time trips. This finding explains how trips delayed leaving Linden were able to return to Linden ahead of schedule. The significant improvement on the scheduled run time is most likely an effect of low ridership due to the COVID-19 pandemic. Operators likely have significantly quicker dwell times in station and can save time compared to the schedule by leaving the station quickly.

Lastly, although the headway is consistent and close to scheduled, it is unclear whether consistent negative delays are good for overall quality of service. Passengers who might memorize when a particular train comes could be disappointed when it has already departed.

6 Conclusions

Overall, the COVID-19 pandemic has had a large effect on the service operations of the Chicago Transit Authority. Overall, Purple Line run times are much lower than scheduled, leading to trips consistently returning to Linden before their scheduled time. Most importantly, the presence of a delayed departure from Linden does not have knock-on effects in terms of further delays, longer run times, and significantly larger headways as would be expected. Low ridership allows for delays to be mitigated and for operators to make up time compared to the schedule along the route with shorter in-station dwell times. However, despite the diminishing presence of knock-on delays due to the pandemic, large primary delays still have a large effect on operations and are the currently the main cause of delayed trains returning to Linden.

With vaccination rates currently increasing and low COVID-19 infection rates, a rebound in ridership is inevitable. It is important to understand how an increase in ridership will likely increase the effect of knock-on delays throughout the service and not make it possible to recover lost time with a quicker stop in the station. Therefore, mitigating and quickly correcting large delays will be even more of a priority because they will have the ability to have a much further reaching effect on the overall service. In conclusion, although pandemic data and low ridership could be considered a limitation of this research, this paper helps to understand the operational impact of an important facet of public transportation: the passenger.

Works Cited

Carey, M. and Kwieciński, A. (1994) 'Stochastic approximation to the effects of headways on knock-on delays of trains', *Transportation Research Part B: Methodological*, 28(4), pp. 251–267. doi: 10.1016/0191-2615(94)90001-9.

CTA Developer Center - Train Tracker Real-Time Arrival and Locations API (no date) *CTA*. Available at: https://www.transitchicago.com/developers/traintracker/ (Accessed: 23 May 2021).

Ghofrani, F. *et al.* (2018) 'Recent applications of big data analytics in railway transportation systems: A survey', *Transportation Research Part C: Emerging Technologies*, 90, pp. 226–246. doi: 10.1016/j.trc.2018.03.010.

Mitsokapas, E. *et al.* (2021) 'Statistical characterization of airplane delays', *Scientific Reports*, 11(1), p. 7855. doi: 10.1038/s41598-021-87279-8.

Yuan, J. (2006) 'Stochastic modelling of train delays and delay propagation in stations'. Available at: https://repository.tudelft.nl/islandora/object/uuid%3Acaa72522-26b1-4088-afc0-59c6e5c346f6 (Accessed: 23 May 2021).