

Florida International University FIU Digital Commons

Department of Mathematics and Statistics

College of Arts, Sciences & Education

2011

Using AVL Data to Improve Transit On-Time Performance

Fabian Cevallos

Lehman Center for Transportation, Florida International University, fabian.cevallos@fiu.edu

Xiaobo Wang

Lehman Center for Transportation, Florida International University, xiaobo.wang@fiu.edu

Zhenmin Chen

Department of Mathematics and Statistics, Florida International University, chenzh@fiu.edu

Albert Gan

Department of Civil and Environmental Engineering, Florida International University, gana@fiu.edu

Follow this and additional works at: http://digitalcommons.fiu.edu/math_fac

 Part of the [Engineering Commons](#), and the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Cevallos, Fabian, et al. 2011. Using AVL Data to Improve Transit On-Time Performance. *Journal of Public Transportation*, 14 (3): 21-40. DOI: <http://dx.doi.org/10.5038/2375-0901.14.3.2> Available at: <http://scholarcommons.usf.edu/jpt/vol14/iss3/2>

This work is brought to you for free and open access by the College of Arts, Sciences & Education at FIU Digital Commons. It has been accepted for inclusion in Department of Mathematics and Statistics by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

Using AVL Data to Improve Transit On-Time Performance

Fabian Cevallos, Xiaobo Wang, Zhenmin Chen, and Albert Gan
Florida International University

Abstract

This paper describes an approach for improving on-time performance at transit agencies. It takes advantage of the schedule adherence information from an AVL system. A methodology that can be used to update the bus timetables by using AVL schedule adherence data is described. Using statistical analysis, the main goal is to maximize the density area of the on-time performance range. From this distribution, the optimal value is obtained and used to update the times in the timetables. Then, a comparison process is used to assess the on-time performance improvements. In addition, a simulation process is presented to provide a different perspective than the statistical methodology. This approach also presents possibilities for further on-time performance improvements. To demonstrate the applicability of this research, a case study using data from Miami-Dade Transit is included. The on-time performance calculations for Routes 99 and 57 also are presented.

Introduction

To passengers, schedule adherence is a matter of service quality. From the service provider perspective, schedule adherence reflects the quality of the service plan (the schedule) and the operations control (Furth et al. 2003). Researchers have long noticed the importance of schedule adherence information contained in Automatic Vehicle Location (AVL) systems. Lee et al. (2001) studied the effect of an AVL system on schedule adherence and operator behavior and willingness to

keep on schedule. In addition, Hammerle et al. (2005) pointed out that some transit agencies would like to use Automatic Passenger Counter (APC) and AVL data to inform service planning and management and ultimately provide more reliable service. Methods for extracting information from these data were developed to compute service reliability indicators. Also, some schedule adherence properties were observed and reported in their research. These studies show a general interest in improving schedule adherence.

It is important to clarify the difference between schedule adherence and on-time performance. Schedule adherence refers to the difference between real time and scheduled times of arrival or departures times, usually presented in minutes. On-time performance, on the other hand, is a percentage value used to indicate buses arriving or departing late, on time, or early. Depending on the AVL system and the transit agency, on-time performance can be calculated using arrivals, departures, or possibly a combination of both.

AVL systems are computer-based vehicle tracking systems that function by measuring the real-time position of each vehicle and relaying this information back to a central location. Many researchers also see the potential uses of analyzing AVL or APC data to improve service quality. A study that uses data from Tri-Met in Portland, Oregon, shows that scheduling can be improved through performance monitoring using AVL data and that very useful information has been retrieved (Kimpel et al. 2004). Shalaby and Farhan (2004) made efforts to use AVL and APC data to develop a bus travel time model capable of providing real-time information on bus arrival and departure times to passengers (via traveler information services) and to transit controllers for the application of proactive control strategies.

One continual question asked by researchers is how to use AVL data to improve on-time performance. The importance of on-time performance to both the transit customer and the transit providers has been discussed in many research projects. For instance, New York City transit established a customer-oriented bus performance indicators program to measure on-time performance. The program contains two schedule adherence indicators that measure different aspects of service performance: route on-time performance and service regularity. The purpose of this program is to measure the quality of service experienced by the customer (Nakanishi 1997).

This research attempts to fill the gap in the understanding of AVL data and presents another perspective on how the data can be used to improve on-time performance. With the availability of AVL data, it is possible to improve on-time

performance by modifying the scheduled times in the timetables. Changes in the timetables will take effect in the following line-up period (also known as mark-up, shake-up, sign-up, bid, or pick). This paper makes use of the AVL schedule adherence data (measured in minutes) to calculate on-time performance expressed as a percentage of buses arriving or departing within an acceptable window. In other words, it uses disaggregate information to derive the aggregate information.

Research Methodology

This research focuses on the idea that transit agencies can maximize on-time performance by adjusting their published schedule timetables. First, the data from an AVL system need to be studied, cleaned, and prepared for use on this on-time performance methodology (OTPM). Second, the distribution of the data needs to be determined by using statistical analysis. Once the statistical distribution is known, the optimal value is calculated and used to update the times in the timetables. Finally, a computerized iterative validation and comparison process is provided to ensure that the updating takes effect and on-time performance is improved. In addition, an on-time performance simulation (OTPS) process is included to better understand the relationship between the time adjustments and their effect on on-time performance.

The distribution of the schedule adherence times (the difference between actual and scheduled arrival or departure times at time points) is used to make schedule adjustments that can maximize the on-time performance. The methodology assumes that the distribution is normal (which can be justified by the fact that normality tests fail to reject the normality assumption). The objective is to adjust the timetables such that the probability of on-time performance is maximized, where the probability of on-time performance is the area under the density function between the acceptable schedule adherence values.

The data process for the improvement of on-time performance is described in Figure 1. It starts with the AVL data to determine the statistical distribution of the schedule adherence data. Then the first adjustment, using the mean of normal distribution, is made, and the schedule adherence values are brought to the center (mean) of the on-time performance parameters. The details on how to determine the adjustment are described in the following sections. Since the values in the timetables are rounded to the nearest minute, a round function is used to change the adjustment values that can be tested; this is done for practical purposes. Tran-

sit agencies present the times in the timetables in minutes. In theory, using continuous values (decimals) or smaller intervals (e.g., ½ minutes), rather than discrete values (integers), could improve on-time performance even further. Then, on-time performance is calculated, using the rounded adjustments, and it is determined if the modifications increased on-time performance. If yes, then the timetables will be updated.

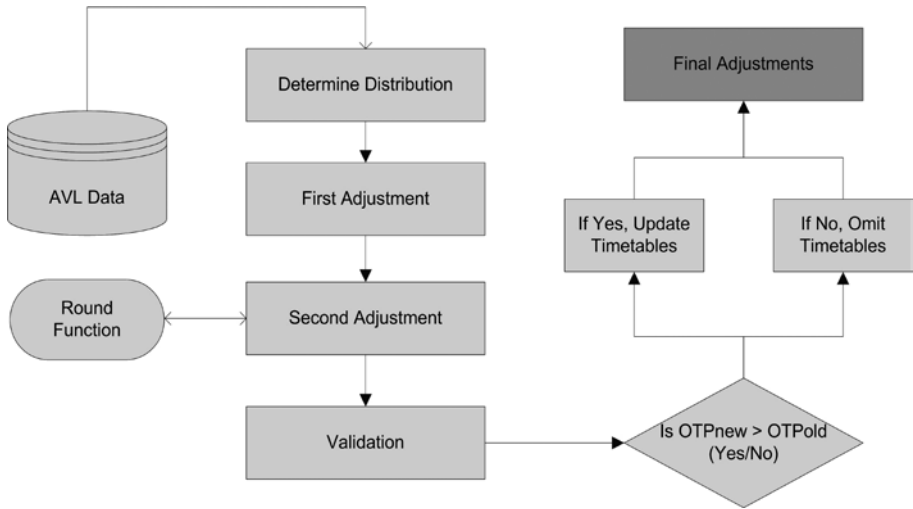


Figure 1. Data flow for the on-time performance improvement process

On-Time Density Area

As previously mentioned, the strategy is to maximize the on-time density area in order to improve on-time performance. Figure 2 displays two distribution curves. One is before the adjustments and the other is after the adjustments. The two shaded areas, which correspond to the on-time performance density areas, are different under the two curves.

As can be seen, the shaded area on the top is smaller than the shaded area on the bottom. For this example, the on-time performance parameters are assumed to go from -5 to 2, which are the limits of the density areas. Buses arriving within 2 minutes early to 5 minutes late are considered to be on time. To maximize on-time performance, the sample mean (0) of the curve on the left is shifted to $[(2-5)/2 = -1.5]$, which is the center of the on-time range and depicted on the right curve.

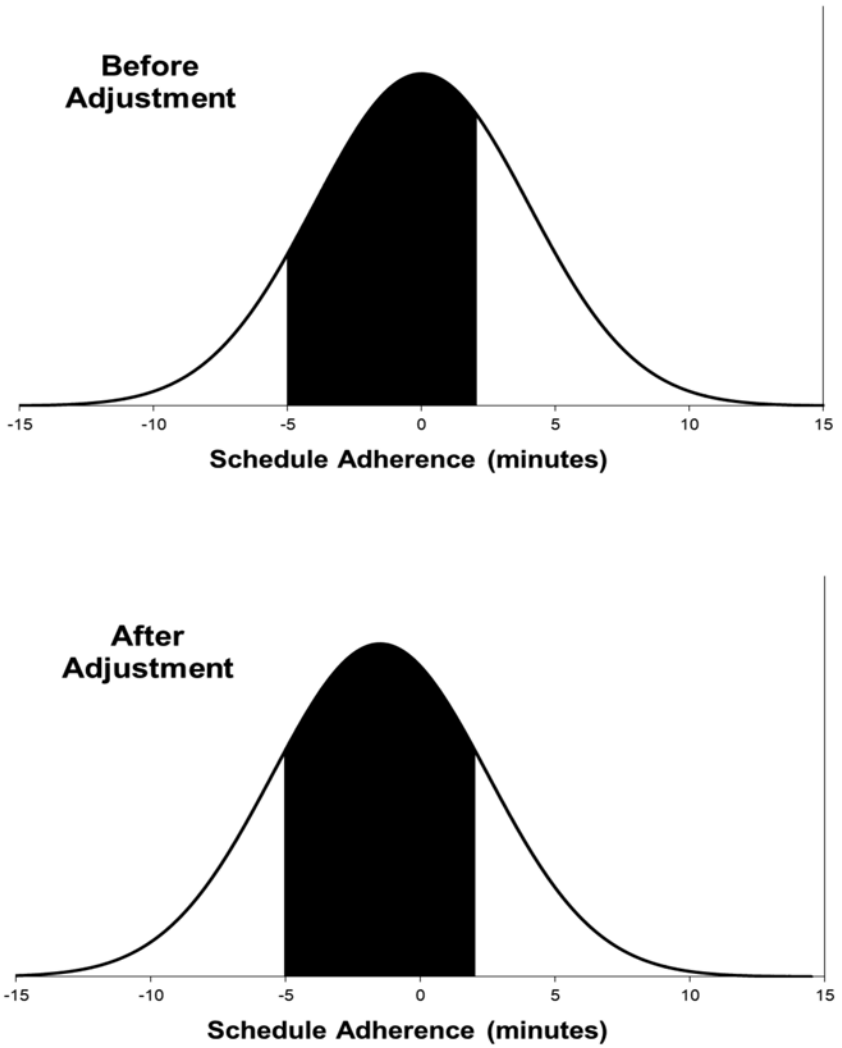


Figure 2. Density areas before and after adjustments

The explanation of why the density area will be maximized when the sample mean is in the center of on-time range is described below for the case that the distribution is normal.

Define:

x_1 = Lower value of the on-time parameters.

x_2 = Upper value of the on-time parameters.

c : = $x_2 - x_1$.

μ = Mean of the schedule adherence (ADH) data.

σ = Standard Deviation of the ADH data.

$D(x)$ = Cumulative distribution function for the normal distribution.

The density area in the on-time range is

$$G(x_2) = D(x_2) - D(x_2 - c) = \int_{x_2 - c}^{x_2} \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ -\frac{(t - \mu)^2}{2\sigma^2} \right\} dt.$$

The first and second derivatives of $G(x_2)$ are

$$G'(x_2) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ -\frac{(x_2 - \mu)^2}{2\sigma^2} \right\} - \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ -\frac{(x_2 - c - \mu)^2}{2\sigma^2} \right\}$$

and

$$G''(x_2) = -\frac{x_2 - \mu}{\sqrt{2\pi} \sigma^3} \exp \left\{ -\frac{(x_2 - \mu)^2}{2\sigma^2} \right\} + \frac{x_2 - c - \mu}{\sqrt{2\pi} \sigma^3} \exp \left\{ -\frac{(x_2 - c - \mu)^2}{2\sigma^2} \right\},$$

respectively. Letting $G'(x_2) = 0$, then $x_2 - \mu = \frac{c}{2}$. Note that

$$G''(x_2) \Big|_{x_2 = \mu + \frac{c}{2}} = -\frac{c}{\sqrt{2\pi} \sigma^3} \exp \left\{ -\frac{c^2}{8\sigma^2} \right\} < 0.$$

This implies that $G(x)$ reaches its maximum at $x_2 = \mu + \frac{c}{2}$. It means that the density area in the on-time range is maximized when x_1 and x_2 are symmetrically allocated about μ , and, finally, we obtain that when $\mu = \frac{x_1 + x_2}{2}$, the density area of the on-time range is maximized.

Another finding about the on-time density area is described in Figure 3. It is clear that the on-time density area is related to the data variance. We use kurtosis to measure the peakedness or flatness of the dataset distribution: a positive kurtosis indicates a relatively peaked distribution, while a negative kurtosis indicates a relatively flat distribution. Higher kurtosis is an indication that more of the variance is due to infrequent extreme deviations.

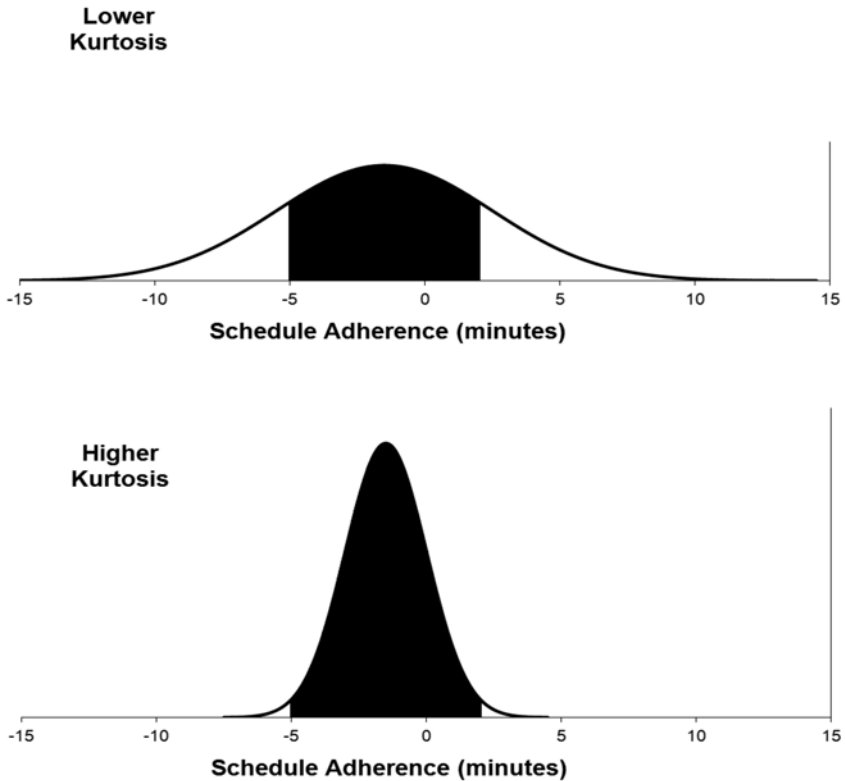


Figure 3. Data with different kurtoses (on-time parameters between -5 to 2)

In Figure 3, the bottom figure shows a higher kurtosis than the top one. From this figure, it is clear that the dataset with higher kurtosis will have more area within the on-time parameters area than dataset with lower kurtosis. Thus, one approach to increase on-time performance is to increase the dataset kurtosis value—in other words, reduce its variance.

As high variances have a negative impact on on-time performance, transit agencies can use different operational strategies to reduce this variability. For instance, the early and late arrival tails can be addressed with a combination of field and dispatch supervision as well as optimal timetable adjustments. Thus, improving on-time performance may require both better operations control and scheduling strategies.

Data Source and Distribution

The schedule adherence dataset used in this research comes from the Miami-Dade Transit (MDT) CAD/AVL system. MDT buses transmit data at a 2-minute poll interval. However, at the timepoints, a record is generated at the exact time of the event. These records, in combination with data from the Transit Operations System (TOS), are used to calculate schedule adherence. The schedule adherence dataset includes GPS location, date, time, and operational data such as Route, Direction, Run Number, Employee Number, Vehicle Number, and Time Point information (Cevallos et al. 2008).

Even though the schedule adherence data from an AVL system can be assumed to be normally distributed, the data can be skewed to the right or to the left (it most likely follows the log normal distribution), depending on how schedule adherence is calculated by the AVL system. This is dependent of the sign (i.e., + or -) used to represent lates or earlyies.

The normality assumption, as justified by the normality tests, is very useful for practical purposes. Using the normal distribution, the timetables adjustments to maximize on-time performance can be determined easily. The goal of this strategy is to maximize the density area of the schedule adherence distribution, within the on-time performance parameters. Therefore, adjustments to the AVL data in the database are made to maximize the density area. By doing this, more buses will arrive or depart within the density area, which improves on-time performance. This process also reduces the number of buses arriving or departing early, which is an undesirable condition for both the transit agency and the passengers.

QQ plot diagrams were used as a graphical tool to diagnose differences between the probability distribution of a statistical population from which a random sample has been taken and a comparison distribution. To demonstrate the applicability of this research, Route 99 was selected. This is a simple route with only one pattern using the same running times throughout the day. Figure 4 shows the normal QQ plot diagrams for the time period of 16:00 - 19:00 at the 199S47AV (NW 199 St. & 47 Ave) time point in Route 99. According to the shapes, with the exception of lower and upper end points, it can be assumed that the statistical distribution is approximately normal.

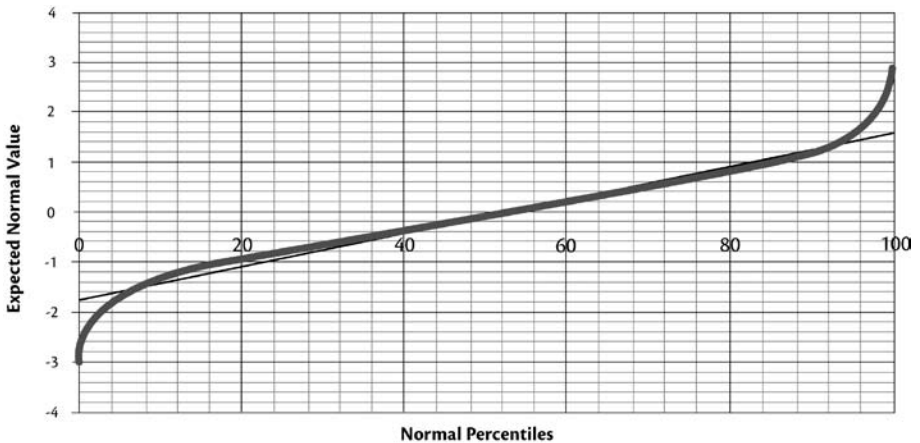


Figure 4. Normal QQ plot for schedule adherence for Route 99 at NW199 St. and 47 Ave.

In this research, the schedule adherence dataset used is considered to follow the normal distribution. This is supported by the shapes of the QQ plots. However, this may not be absolutely guaranteed for other datasets in the AVL database. Even though the normality assumption of the dataset plays an important role in this research, the key point in this paper is how to make the on-time performance area maximum by shifting the ADH values. In fact, even if the dataset does not follow normal distribution, it still is possible to find the approximate center of the schedule adherence values and shift the ADH dataset to improve on-time performance. If the dataset is close only to the normal distribution, the on-time performance area also is close to the maximum area possible. In this situation, simulation can be used to make sure that the adjustments made maximize the on-time performance area. This is further discussed in the Simulation Process section.

Depending on the AVL system, the data can be skewed to the right or to the left. Data generated from AVL systems that use negative numbers to represent late buses will tend to show data skewed to the right because it is expected that the majority of values be towards the left, whereas if late buses correspond to positive numbers and early buses to negative numbers, then the distribution is likely to be skewed to the left. Further, this distribution is likely to follow the log normal distribution (Cevallos and Zhao 2006). This is due to the fact that buses are likely to arrive or depart late rather than early most of the time. Early arrivals and departures is an undesirable condition, and it is controlled by most transit agencies. The reason for this is obvious. Transit agencies usually work hard to avoid early departures, as this can be very detrimental to passengers.

Hypothetical Transit Route

To better understand the framework of our research, a hypothetical transit route to better demonstrate the process is presented. Figure 5 is used for this purpose. The diagram depicts a schematic of a hypothetical transit route. The route is divided into a number of timepoints with a bus traversing along the route. When the transit bus arrives at stop $i+1$, the actual arrival time is known from the AVL system, and the schedule adherence is calculated at that particular location. If the bus arrival time falls outside the range of the on-time parameters, which are defined by the transit agency, the bus will be regarded as either late or early. This is depicted in Figure 5 under Actual Arrival Time, which shows a bus arriving early. However, if the schedule time is shifted a little earlier, chances are that the number of early arrivals is reduced. The question is, how much should the scheduled time be moved to improve on-time performance?

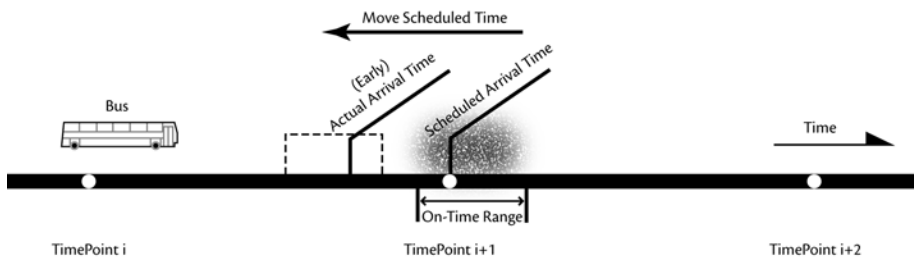


Figure 5. Moving scheduled times in a hypothetical transit route

Adjustment of Timetable

Before adjusting the timetables, some terminology needs to be defined first:

- *ADH*: Schedule adherence.
- *Before.Time*: The scheduled time before adjustments.
- *After.Time*: The new scheduled time after adjustments.
- *Adj.First*: First raw adjustment of the schedule adherence.
- *Adj.Second*: The second adjustment of the schedule adherence by applying round function.
- *On-Time.Period*: $-5 \leq ADH \leq 2$. A bus is considered to be on time if it is within 2 minutes early and 5 minutes late. These are the on-time performance parameters used by MDT. Different agencies may have a different definition for the on-time period. Other agencies may use 0 and -5 or 1 and -5. There appears to be no industry standards for on-time performance.
- *Adj.ADH*: This is a constant, based on the statistical distribution and time period. In the normal distribution, this corresponds to the middle value of the on-time range. In other distributions, it should be the number that maximizes the density area of the on-time range.
- *Layover/Recovery Time*: 10% of the trip time. The time used at the end of a trip (Pine et al. 1998).

To minimize the complexity of this methodology, some assumptions need to be made:

- *Assumption 1*: The schedule adherence follows the normal distribution. Since the schedule adherence distribution is close to the normal distribution, this assumption seems reasonable and it is supported by QQ plot diagrams.
- *Assumption 2*: Shifting the schedule does not change behavior. Operator behavior is not expected to change with the new schedule—that is, vehicle operators will drive as usual.

The method used in this research is to adjust the times in the timetables by shifting the *ADH* to *Adj.ADH* values, so that density area for the on-time range is maximized. For this, the following formulas were developed:

$$Adh.First = Mean.ADH - Adj.ADH$$

$$Adh.Second = Function Round (Changes.First)$$

$$\begin{aligned} \text{After.TIME} &= \text{Changes.Second} + \text{Before.TIME} \\ \text{After.ADH} &= \text{Changes.Second} + \text{Before.ADH} \end{aligned}$$

The Round function is applied to the first change from a decimal to an integer value, because the minimum unit is one minute. For example, in the case of time-point NW 199 St & 47 Ave in Route 99 in Table 1, $\text{Adj.First} = 4.23$. After applying the simple round function, Adj.Second will be 4, which means that the schedule time should be updated from 05:23 PM to 05:19 PM.

Adj.ADH is a very important constant in this model. Though the data are considered to follow the normal distribution, other distributions can be applied (like the lognormal distribution). In any case, the goal should be to maximize the area of the statistical distribution considering the on-time performance parameters. According to the on-time performance parameters used in this research, the vehicle is on-time if and only if ADH is less than or equal to 2 or ADH is more than or equal to -5. Since it is assumed that the data follows the normal distribution, the center of the on-time performance area is -1.5, which is the adjustment value that can maximize the on-time performance area. Therefore, in this particular case the Adj.ADH will be -1.5.

Figure 6 shows the fit curve of the schedule adherence on the right, compared to the ideal curve that maximizes on-time performance on the left. The solid line display the fitted normal distribution based on a sample data in this research. The discrete data are approximated and treated as normal distribution. The center of the solid curve is the mean of the sample data. According to the methodology, the mean of the schedule adherence values is shifted to -1.5. This value (-1.5) is the center of the on-time range [-5, 2]. This is depicted in Figure 6, where the right solid normal distribution curve is shifted to the left dotted line. After the adjustments, the on-time performance area is maximized.

After the second adjustments are calculated, the whole timetable is adjusted. The sample data selected consisted of seven timepoints in Route 99 in the Eastbound direction. Table 1 displays information of these time points. The first row shows the stop name of each time point, the second row shows the name of time points. The third row shows the schedule time before adjustments. The data collected is from the time period of 16:00 - 19:00. The fourth row shows the mean of the schedule adherence at the timepoints. The fifth row is the standard deviation of the schedule adherence data. The sixth row is the first adjustment and the seventh row is the second adjustment. The eighth row shows the schedule time after adjustments.

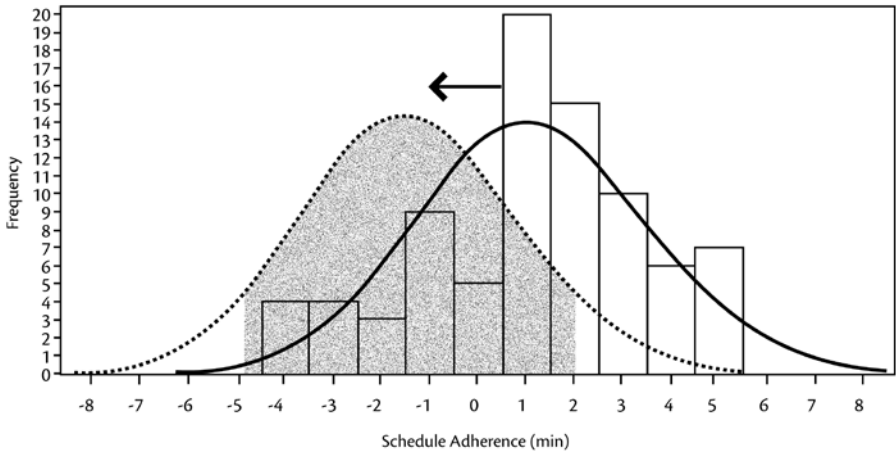


Figure 6. Adjustment diagram

Table 1. Adjustment of Timetable in Route 99

| Stop Name | NW 199 St & 47 Ave | NW 199 St & 27 Ave | NW 215 St & 2 Ave | NE 199 St & 2 Ave | NE 199 St & 10 Ave | NE 203 St & 20 Ave | Aventura Mall |
|---------------|--------------------|--------------------|-------------------|-------------------|--------------------|--------------------|---------------|
| 1 Time Point | 199S47AV | 199S27AV | U441215S | NE2A199S | 199S10AV | 203S20AV | AVEN-MALL |
| 2 Before-Time | 05:23 PM | 05:29 PM | 05:45 PM | 05:53 PM | 05:58 PM | 06:06 PM | 06:15 PM |
| 3 Mean (adh) | 2.73 | -1.72 | -0.05 | -2.06 | -0.94 | -0.73 | -0.67 |
| 4 SD (adh) | 5.41 | 3.05 | 3.55 | 3.85 | 3.76 | 3.28 | 3.93 |
| 5 Adj. First | 4.23 | -0.22 | 1.45 | -0.56 | 0.55 | 0.77 | 0.82 |
| 6 Adj. Second | 4 | 0 | 1 | 0 | 1 | 1 | 1 |
| 7 After Time | 05:19 | 05:29 | 05:44 | 05:53 | 05:57 | 06:05 | 06:14 |

For instance, the mean of the schedule adherence at NE 203 St & 20 Ave is -0.73. This uses the -1.5 constant, which is the center of the on-time period of -5 and 2. The first adjustment (*Adj.First*) should be 0.77, which is the result of $-0.73 - (-1.5)$.

After applying the round function, the second adjustment (*Adj.Second*) equals 1. Therefore, the new time would be 06:05 PM (06:06 PM – 1).

The final results, based on the adjustments implementation, are presented in Figure 7. By applying this methodology, the on-time performance for Route 99 in the Eastbound direction increased from 72.3 percent to 80.4 percent (i.e., an 8.1% increase) during the time period from 16:00 to 19:00.

Simulation Process

Though the timetable approach is described and the effect is proved by the on-time performance improvement results above, these results are based on a normal assumption and a rounding procedure. Ideally, a better solution could be obtained by using the best possible statistical distribution for the schedule adherence data as well as using continuous or smaller time intervals, instead of discrete integer values for the timetable adjustments. Nevertheless, a good solution can be obtained through simulation to get all possible outcomes for on-time performance using different schedule adherence adjustments. This is accomplished by introducing different adjustments until the best on-time performance value is obtained. The simulation uses schedule adherence adjustments at 1-minute intervals, since this is the basic unit used in the timetables and schedule adherence data.

A Structured Query Language (SQL) statement is built to perform the simulation. The output data contains columns that include old on-time performance, new on-time performance, average of *ADH*, variance of *ADH* and the changes of *ADH*. It calculates all possible on-time performance by changing the *ADH* values. Table 2 presents a partial outcome of the simulation test. According to this test, the highest on-time performance value at timepoint SOUTMIAM from Route 57 is reached when the *ADH_CHANGE* value equals to -2 and the *AVG_ADH* is -1.41. As it can be seen, this average value is close to the constant of -1.5.

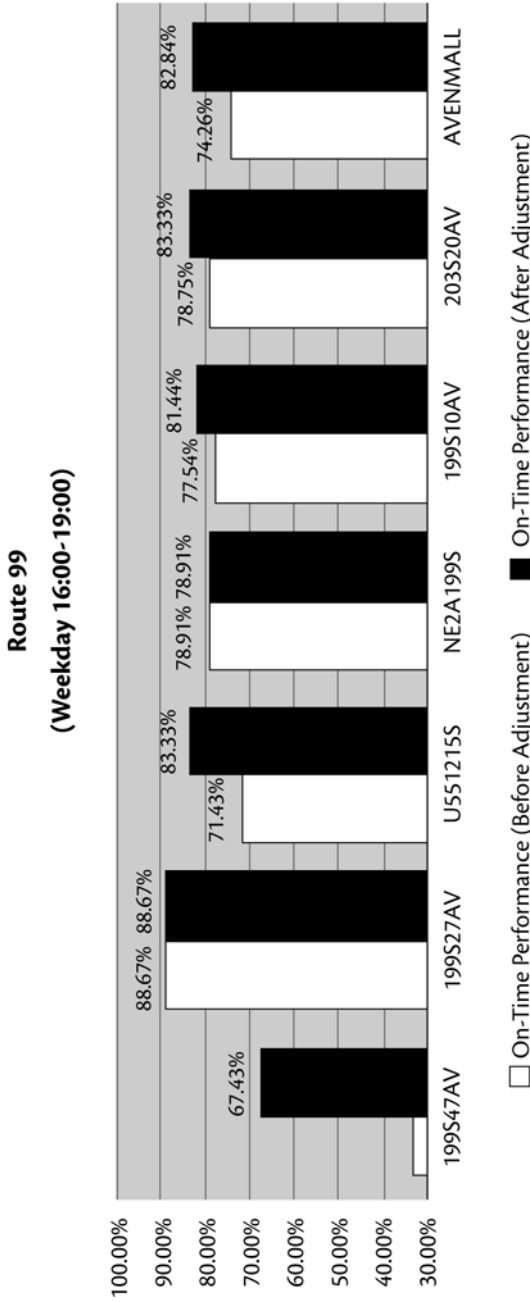


Figure 7. On-time performance comparisons of intermediate time points

Table 2. Partial Dataset from Simulation Process*

| TP | ROUTE | AVG_ADH | PERF1 | PERF2 | ADH_CHANGE |
|----------|-------|---------|-------|-------|------------|
| SOUTMIAM | 57 | -4.41 | 47.85 | 61.21 | -5 |
| SOUTMIAM | 57 | -3.41 | 47.85 | 64.35 | -4 |
| SOUTMIAM | 57 | -2.41 | 47.85 | 64.55 | -3 |
| SOUTMIAM | 57 | -1.41 | 47.85 | 65.19 | -2 |
| SOUTMIAM | 57 | -0.41 | 47.85 | 57.48 | -1 |
| SOUTMIAM | 57 | 0.59 | 47.85 | 47.85 | 0 |
| SOUTMIAM | 57 | 1.59 | 47.85 | 40.98 | 1 |
| SOUTMIAM | 57 | 2.59 | 47.85 | 34.55 | 2 |
| SOUTMIAM | 57 | 3.59 | 47.85 | 28.58 | 3 |
| SOUTMIAM | 57 | 4.59 | 47.85 | 23.70 | 4 |

*TP: time point; Route: route number; AVG_ADH: average of ADH values after changes; PERF1: on-time performance value before any changes; PERF2: on-time performance after ADH changes; ADH_CHANGE: changes of ADH value

Based on the methodology presented in this paper, the optimal on-time performance is obtained when the mean of the schedule adherence values is shifted to -1.5. Note that using simulation, the optimal value is the same as the one obtained from the previous methodology. However, it is not guaranteed that the optimal values that are obtained from both methodologies will be always the same.

Figure 8 presents the on-time performance comparison between the solutions of On-Time Performance Metrology (OTPM) and On-Time Performance Simulation (OTPS) for a 24-hour period. The chart in Figure 8 shows that there are no differences in four of the five timepoints. However, in time point U441215S, the solution by using OTPS increased the on-time performance. For practical purposes, there is no major difference from either method. Yet, the OTPS process produced a slightly higher on-time performance improvement.

As depicted in Figure 8, the OTPS approach improves the on-time performance of timepoint U441215S approximately 2 percent; the other timepoints remain the same. What this means is that using the simpler OTPM approach, good results can be obtained without having to go through the more elaborated OTPS process. Usually, the OTPS process generates better results, because it tries all possible adjustments. However, it is a more complicated than OTPM process. The changes obtained by either process can be used to adjust the time in the timetables, which

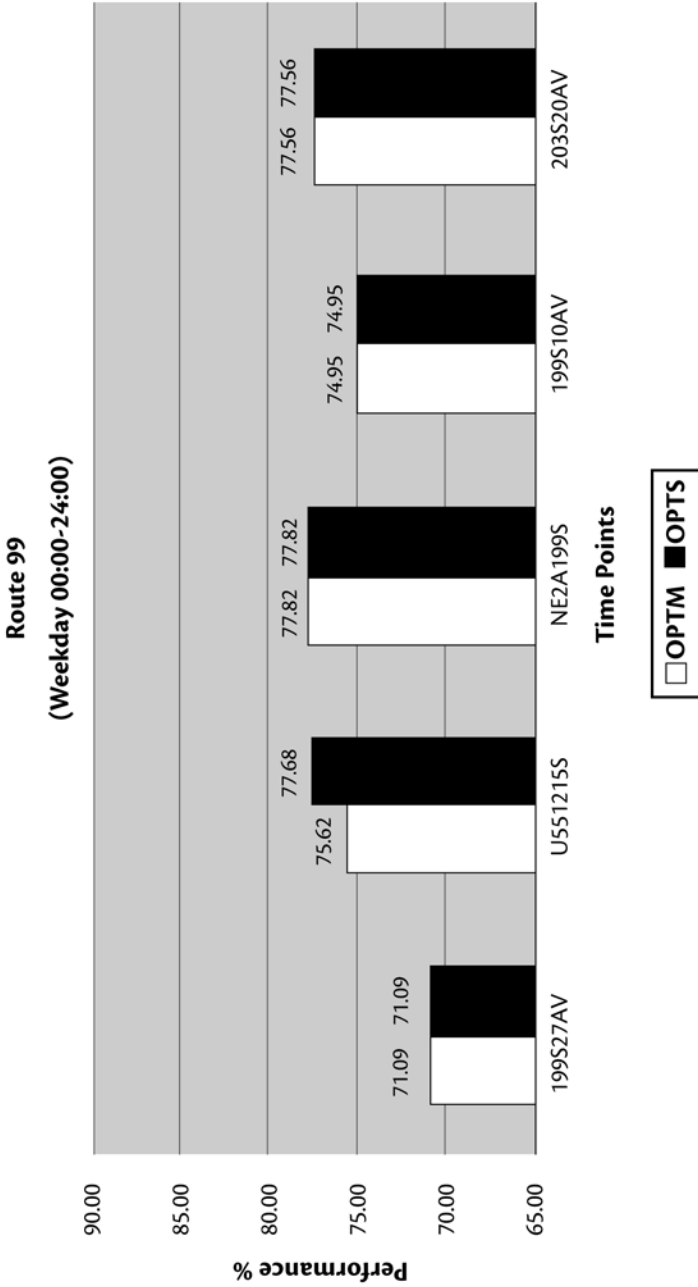


Figure 8. On-time performance results from OPTM and OPTS approaches

can take effect in the next line-up period. Based on this research, it is expected that adjusting the timetables on-time performance can be improved.

Summary and Conclusion

This paper presents a perspective on how to improve transit on-time performance by using schedule adherence data. It takes advantage of data from an AVL system, in particular, schedule adherence information. To demonstrate the applicability of this research, a case study using data from MDT is included, and on-time performance calculations for Routes 99 and 57 also are presented.

A methodology that can be used to update the times in the bus timetables by using schedule adherence data is described. The goal of this methodology is to maximize the on-time density area. The density area is maximized when the mean of the schedule adherence data is in the center of the on-time range. This is supported by a mathematical proof. A validation process is provided to ensure that the updating takes effect. Further, a comparison process is used to assess the on-time performance improvements, before and after the adjustments.

Before this methodology can be applied, the AVL data need to be cleaned, manipulated, and stored in a database to allow the processing of the data. Using statistical analysis, the distribution of the data is determined, and the mean and standard deviation are calculated. Once the statistical distribution is known, the main goal is to maximize the density area of the on-time performance range, which is based on the on-time performance parameters. From this distribution, the optimal value that can maximize on-time performance is obtained and used to update the times in the timetables. The updated timetable will take effect in the next line-up period, and the on-time performance is improved. This process also reduces the number of early arrivals and departures.

There were two assumptions in this research: normal distribution and operator behavior. The assumption of normal distribution is supported based on QQ plots and normality tests. Operator behavior can be attributed to many different factors, and it is assumed to be unchanged after the timetable modifications.

A simulation process is presented to demonstrate additional possibilities. Simulation is a good solution to obtain all possible outcomes for on-time performance using different schedule adherence adjustments and to finally obtain the optimal values. This is accomplished by introducing different adjustments until the best on-time performance value is achieved.

There are still many opportunities for improvements and optimization on this subject. For instance, the peakedness or flatness of the dataset distribution can be measured using the kurtosis statistic. This can be useful for developing strategies that reduce the variance of the data. Studying the relationship between the standard deviation (variability of arrivals/departures) and the schedule adherence adjustments would be beneficial. Including the standard deviation in the methodology may influence the outcome of the on-time performance results. In addition, developing a computerized application could assist with the automation of the described methodology.

In particular, there is potential for generating a more accurate distribution model. The statistical distribution of the schedule adherence data is important to this research. Therefore, better statistical techniques can be applied if a more accurate distribution model is obtained. Different distribution models will produce different adjustment factors, which makes the density area of the sample data maximized.

References

- Cevallos, F., and F. Zhao. 2006. Minimizing transfer times in a public transit network with a genetic algorithm. *Transportation Research Record* 2034: 74-79.
- Cevallos, F., K. Kirwin, and R. Pearsall. 2008. Using CAD/AVL data for performance management. *Proceedings of the 10th International Conference on Applications of Advanced Technologies in Transportation*, ASCE, Athens, Greece, May 27- 31.
- Furth, P., B. Hemily, T. Muller, and J. Strathman. 2003. Using archived AVL-APC data to improve transit performance and management. Transit Cooperative Research Program.
- Hammerle, M., M. Haynes, and S. McNeil. 2005. Use of automatic vehicle location and passenger count data to evaluate bus operations. *Transportation Research Record* 1903: 27-34.
- Kimpel, T., and J. Strathman. 2004. Improving scheduling through performance monitoring using AVL and APC data. Submitted to University of Wisconsin-Milwaukee as a Local Innovations in Transit project report under the Great Cities University Consortium.

Lee Y., K. Chon, D. Hill, and N. Desai. 2001. Effect of automatic vehicle location on schedule adherence for mass transit administration bus system. *Transportation Research Record* 1760: 81-90.

Nakanishi, Y. 1997. Bus performance indicators-on-time performance and service regularity. 1997. *Transportation Research Record* 1571: 1–13.

Pine, R., J. Niemeyer, and R. Chisholm. 1998. Report 30 transit scheduling: basic and advanced manuals. Transportation Research Board of the National Academies, Washington, D.C.

Shalaby, A., and A. Farhan. 2004. Prediction model of bus arrival and departure times using AVL and APC data. *Journal of Public Transportation* 7(1).

About the Authors

FABIAN CEVALLOS (fabian.cevallos@fiu.edu) is the Transit Program Director at the Lehman Center for Transportation Research. He holds a Ph.D. in Civil (Transportation) Engineering from Florida International University. His research interests include Advanced Public Transportation Systems (APTS), transit planning and operations, traffic safety, driving simulation, business intelligence, and information technology.

XIAOBO WANG (xiaobo.wang@fiu.edu) is a Software Engineer at the Lehman Center for Transportation Research. He holds an M.S. from Beijing University of Aeronautics and Astronautics and an M.S. in Transportation Engineering from Florida International University. His areas of interests include transportation software research and development.

ZHENMIN CHEN (chenzh@fiu.edu) is a Professor at Florida International University. He holds a Ph. D. in Statistics from the University of Texas at Dallas. His research interests include applied statistics, reliability, and clustering analysis.

ALBERT GAN (gana@fiu.edu) is an Associate Professor at Florida International University. He holds a Ph.D. in Civil (Transportation) Engineering from the University of Florida. His research interests include public transit, traffic simulation and control, ITS, highway safety, access management, and information technology.