Predicting Bus Arrival Time on the Basis of Global Positioning System Data

Dihua Sun, Hong Luo, Liping Fu, Weining Liu, Xiaoyong Liao, and Min Zhao

The ability to obtain accurate predictions of bus arrival time on a real-time basis is a vital element to both bus operations control and passenger information systems. Several studies had been devoted to this arrival time prediction problem; however, few resulted in completely satisfactory algorithms. This paper presents a new system that can be used to predict the expected bus arrival times at individual bus stops along a service route. The proposed prediction algorithm combines real-time location data from Global Positioning System receivers with average travel speeds of individual route segments, taking into account historical travel speed as well as temporal and spatial variations of traffic conditions. A geographic information system–based map-matching algorithm is used to project each received location onto the underlying transit network. The system is implemented as a finite state machine to ensure its regularity, stability, and robustness under a wide range of operating conditions. A case study on a real bus route is conducted to evaluate the performance of the proposed system in terms of prediction accuracy. The results indicate that the proposed system is capable of achieving satisfactory accuracy in predicting bus arrival times and perfect performance in predicting travel direction.

Accurate prediction of the expected arrival time of the next bus at individual bus stops is of significant value to both transit operators and users. With this arrival time information, transit operators can promptly respond to unexpected service interruptions and delays by introducing various bus control strategies, such as holding and stop-skipping (*1*). In the long term, transit operators could spot problematic routes and shifts that chronically run late and promptly take managerial and technical actions—such as training operators, adjusting schedules, and implementing transit signal priority—for improved performance. The overall quality of transit service could therefore be improved, which would ultimately make the public transportation mode more user-friendly and attractive as compared with other transportation modes such as driving.

Bus arrival time information can also be disseminated to travelers through various mediums such as electronic boards installed at bus stops, Internet, cable TV, and cellular phones (*2, 3*). Such information could reduce passengers' anxieties while waiting for buses and save them travel time because they can time their arrival at the bus stops more close to the schedule. Better information therefore helps retain existing passengers and attracts other trip makers who would otherwise not use transit.

D. Sun, H. Luo, X. Liao, and M. Zhao, College of Automation, and W. Liu, College of Computer Science, Chongqing University, Chongqing, China 400044. L. Fu, Department of Civil Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Corresponding author: D. Sun, d3sun@163.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 2034, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 62–72.
DOI: 10.3141/2034-08

LITERATURE REVIEW

A number of studies have been initiated in the past to address the bus arrival time prediction problem. These efforts have resulted in three types of prediction models: (*a*) models based on historical data, (*b*) multilinear regression models, and (*c*) artificial neural network models.

The first type of prediction models infers the current and future travel time of a bus based on the historical travel time of the same bus or other buses. Lin and Zeng (*4*) proposed a set of bus arrival time prediction algorithms for a transit traveler information system implemented in Blacksburg, Virginia. Four algorithms were introduced with different assumptions on input data and were shown to outperform several algorithms from the literature. Their algorithms, however, did not consider the effect of traffic congestion and dwell time at bus stations. Kidwell (*5*) presented an algorithm for predicting bus arrival times based on real-time vehicle location. The algorithm worked by dividing each route into zones and recording the time that each bus passed through each zone. Predictions were based on the most recent observation of a bus passing through each zone. However, this algorithm was not suitable for large cities where both travel time and dwell time could be subject to large variations. Generally speaking, these models are reliable only when the traffic pattern in the area of interest is relatively stable. One of their main limitations is that it requires an extensive set of historical data, which may not be available in practice, especially when the traffic pattern varies significantly over time.

The second approach is applying mathematical models to predict the expected travel times between stops and then the expected bus arrival times at individual stops. These models are usually established by regressing travel times against a set of independent variables, such as traffic conditions, ridership, number of intermediate stops, and weather condition. Patnaik et al. (*6*) developed a set of regression models to estimate bus arrival times with data collected by automatic passenger counters installed on buses. The results obtained were promising and indicated that the developed models could be used to estimate bus arrival times under various conditions. However, this approach is reliable only when such equations can be established, which may not be possible for many application environments where many of the system variables are typically correlated.

The third approach is applying artificial neural networks (ANN) that are capable of capturing complex nonlinear relationships. Jeong and Rilett (*7*) proposed an ANN model for predicting bus arrival times and demonstrated its superior performance as compared with other methods. However, ANN models require extensive training and testing in order to find the right network structure and determine the best parameter values.

OBJECTIVES AND SCOPE

In this research the authors have developed a new bus arrival time prediction algorithm that combines Global Positioning System (GPS) data with real-time estimates of interstation travel speeds. The proposed algorithm is implemented in an intelligent system that can automatically detect the running route and direction of a bus and predict its arrival times at the downstream stops under any operating conditions. This paper details the structure, function, and performance of the proposed prediction system along with the development and validation of the proposed model.

ALGORITHM OF BUS ARRIVAL TIME PREDICTION

As discussed previously, a number of studies have been carried out in the past on the problem of bus arrival time prediction in the area of intelligent transportation systems. Past studies have indicated that bus travel times and thus arrival times at the downstream stops could be affected by a wide range of factors, such as route structure and schedule, traffic conditions, passenger boarding and alighting activities, and weather (*6*). Accordingly, an accurate and reliable algorithm for predicting bus arrival time must take into account the effects of these factors. This section describes an algorithm aimed at achieving this goal.

Route Linearization

To determine the trajectory of a moving bus and its distances to the downstream bus stations, the GPS readings of each equipped bus need to be projected onto the underlying transit network. In a digital transit network model, bus routes are represented by a sequence of line features as an approximation to their true geographical composition. Such straight line approximations are usually not accurate enough for tracking purposes. To ensure representation accuracy, a transit route is segmented into a set of sequential straight lines (called links) less than 100 m in length (*4*). The end points of each link, also called nodes, are specified by their longitudes and latitudes. All links and nodes are numbered according to the sequence in which the bus passed, and then they are recorded into a file for later use.

Route linearization provides a bus network model that is required in the later steps of map matching and distance calculation. In a situation when a same location is covered by a bus multiple times, link number needs to be used in order to determine the exact status of the bus (*8*).

Model of Bus Arrival Time Prediction

The algorithm on bus arrival time prediction in this paper includes two components. The first component consists of a real-time bus-tracing model with the purpose of processing the GPS data, projecting them onto the electronic map and then obtaining the distance to each bus station. The second component is a bus arrival time prediction model used to estimate the time to downstream bus station in real time on the basis of the output of the first component and various other factors.

Real-Time Bus Tracing

The first step in a bus arrival time prediction procedure is to determine the exact location of a bus on the linearized route based on its GPS data. Consider that a particular position with its coordinates denoted by (x_c, y_c) and its closeness to a given route link i, denoted by its end nodes (x_{ui}, y_{ui}) and (x_{di}, y_{di}), can be defined as follows:

$$D_i = L_{cu} + L_{cd} - L_{ud} \qquad (1)$$

where

D_i = measure representing the closeness of the current bus location (x_c, y_c) to the link i,
L_{cu} = curve distance between current bus location (x_c, y_c) and upstream point (x_{ui}, y_{ui}) of link i,
L_{cd} = curve distance between current bus location (x_c, y_c) and downstream point (x_{di}, y_{di}) of link i, and
L_{ud} = curve distance between upstream point (x_{ui}, y_{ui}) and downstream point (x_{di}, y_{di}) of link i.

The link with the smallest D_i value is considered as the matched link i. Once a link is matched, the GPS coordinates are projected onto the link and the distance to each downstream station could therefore be determined accordingly on the basis of the length of the individual links along the route.

Real-Time Bus Arrival Time Estimating

Because of the uncertain nature of traffic conditions, the travel speed of a bus on route segment normally fluctuates around the average speed of the route segment. To take into account this inherent variation, the average speeds of individual segments are obtained from a dedicated traffic information system which monitors traffic using GPS-equipped vehicles as probes. Travel speeds are grouped by time of day and are updated continuously to reflect changes in traffic congestion. Therefore, it is taken into account in the bus arrival time prediction. Figure 1 shows the time–space diagram of buses along a service route by four different time periods. It is found that the average speeds of route segments are quite stable in most time periods except for the peak hour (8:30 to 9:30 a.m.) when traffic is most congested.

Before being able to predict the arrival time of a bus at a given downstream station, its average travel speed to that station is needed. Li et al. (*9*) proposed an algorithm that estimates this speed based on historical bus travel speed along the route segment and the current travel speed of the bus derived from GPS data, that is

$$v = \frac{\sum_{i=1}^{n-1} v_{ai} + v_r}{n} \qquad (2)$$

where

v_{ai} = historical average speed of route segment i ($i = 1,2, \ldots, n-1$),
v_r = current speed of the bus obtained from GPS data, and
$n - 1$ = number of route segments to the bus station of interest.

The implication of this algorithm is that when the bus is far away from the station, its predicted speed v would depend primarily on its historical average speed along the route v_{ai} ($i = 1, \ldots, n-1$) rather than its current speed v_r. In real operating conditions, however, the current speed of the bus is usually a more important factor influencing

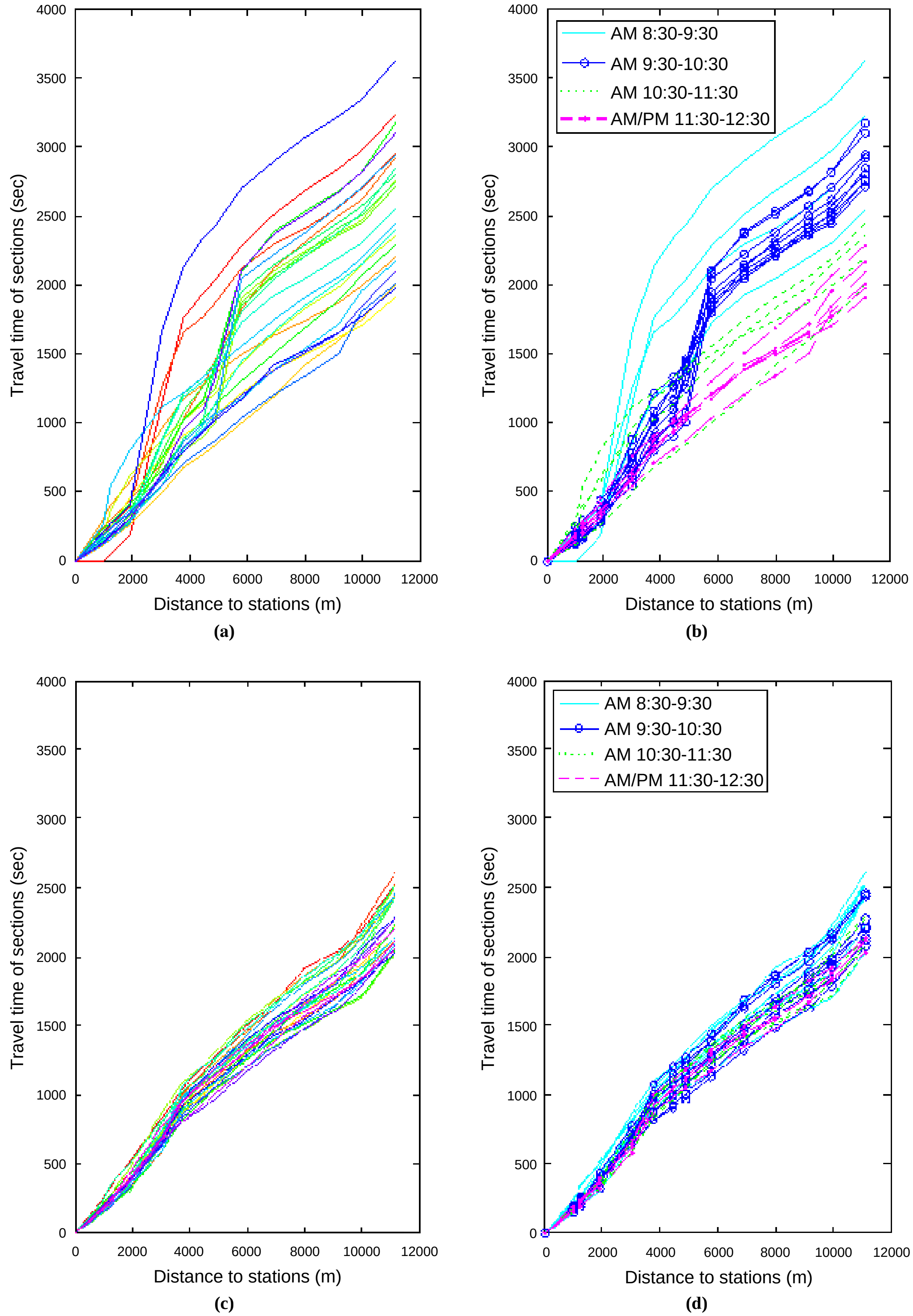

FIGURE 1 Bus time–space diagrams: (*a*) travel time, December 8, 2005; (*b*) travel time clustered by time period, November 8, 2005; (*c*) travel time, March 16, 2006; (*d*) travel time clustered by time period, March 16, 2006.

(continued)

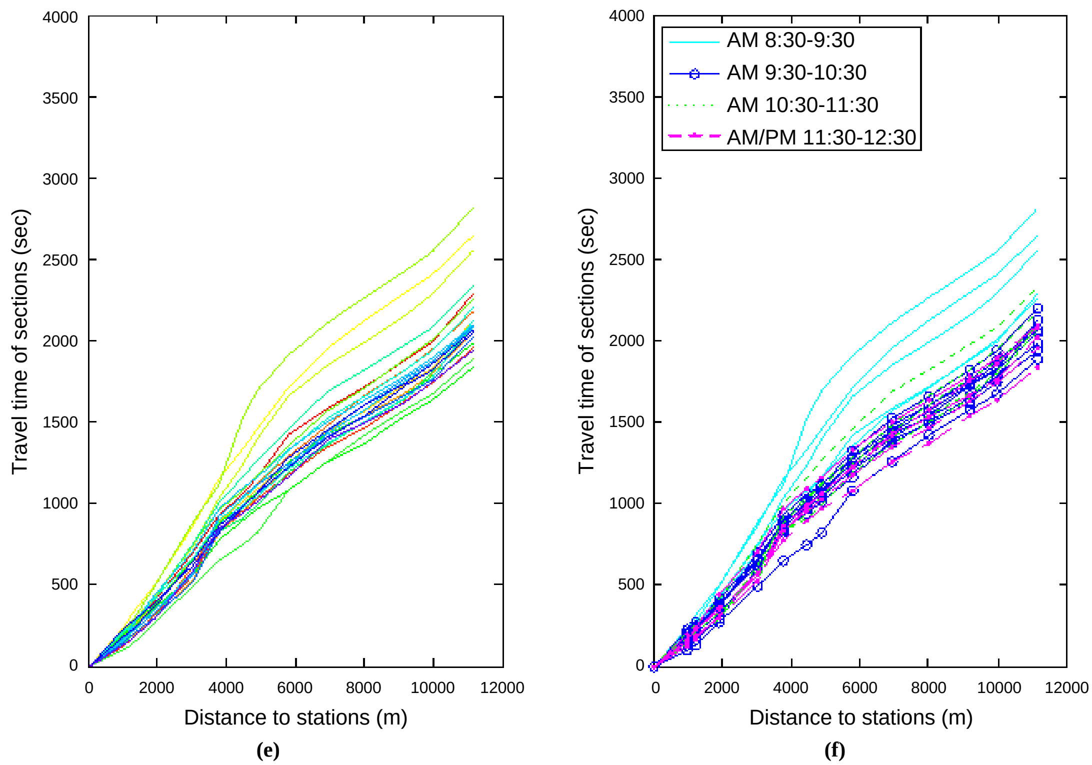

FIGURE 1 *(continued)* **Bus time–space diagrams: (*e*) travel time, April 4, 2006; and (*f*) travel time clustered by time period, April 4, 2006.**

how fast the bus will travel over the distance to the station of interest. Furthermore, this method would predict a nonzero speed even when $v_r = 0$.

This model could therefore perform poorly in applications where the bus speed changes frequently, the GPS speed is not sufficiently accurate, or the historical average speed of the route does not adequately reflect the current speed of the traffic. As a result, this paper proposed a new scheme for predicting arrival time to a downstream stop. The proposed method first estimates the average travel speed to the immediate downstream route segments, denoted by i, based an average speed estimate updated dynamically as follows:

$$v_i = \frac{av_r + bv_{ai}}{a+b} \tag{3}$$

where

v_i = predicted speed of route segment i (to the immediate downstream route segments),
v_r = current speed of bus derived from GPS data,
v_{ai} = historical average speed of route segment i in current time period, and
a,b = weighting factors defined in Equation 4.

$$a = S_{if} \qquad b = S_{ib} \tag{4}$$

where S_{if} is the distance from the current bus location to the end of the route segment i, and S_{ib} is the distance from the beginning of the route segment i to the current bus location. Based on Equations 3 and 4, the bus arrival time to downstream stations can be calculated as shown in Equation 5.

$$T = \frac{S_i}{v_i} + t_{i+1} + t_{i+2} + \cdots + t_{i+n} + t_d \tag{5}$$

where

S_i = distance from the current bus location to the end of route segment i;
t_{i+n} = travel time of each segment, estimated based on the average travel speeds of individual route segments in current period;
n = number of route segments before reaching the station of interest; and
t_d = sum of average dwell time at each bus station that the bus will pass by, which could be estimated on the basis of historical dwell time data for different time of day.

It is more reasonable for the proposed algorithm to develop a prediction model based on the average speeds of route segments rather than the historical average velocity. It also leads to a superior result, because this model only estimates the speed to the end of the route segment instead of the speed to downstream bus stations, when the bus is far from the station.

INTELLIGENT SYSTEM FOR PREDICTING BUS ARRIVAL TIME

In order to predict the arrival time, the bus's running direction must first be determined. In many existing studies, this issue is circumvented by assuming that bus running direction can be derived from bus schedule and consecutive GPS data. However, in a real application environment, a bus could turn around in the middle of a trip and GPS data could give erroneous indication of bus travel direction due to the well-known "backward data" problem (*10*). Wrong estimation on running direction could lead to totally incorrect prediction of arrival times. As a result, a bus arrival time prediction system is required to identify the running route and direction automatically, respond to unexpected incidents in time, and predict the expected arrival times for all the buses in service. To meet these requirements, this research proposed an intelligent prediction system based on the concept of finite state machine (also known as finite state automata). A system of finite state machine, denoted by M, is defined as follows:

$$M = \left(Q, \Sigma, \delta, q_0, F\right) \tag{6}$$

where

Q = finite state set, $\forall q \in Q$;
Σ = input data;
δ = transition function, $\delta : Q \times \Sigma \rightarrow Q$;
q_0 = initial state, $q_0 \in Q$; and
F = final state set, $F \subseteq Q$.

At any point of time, the system could be in one of the four states, namely, Q equals {beginning state, prediction state (F), terminus state, unknown state (q_0)}. The relationship among the four states is shown in Figure 2. In the beginning state, the system would try to identify the route that a bus is currently servicing and the direction that it is traveling. The underlying procedures are shown in Figures 3 and 4, respectively. If there is no incident, δ (beginning state, beginning state is normal) equals prediction state; otherwise, δ (beginning state, unexpected incident occurs) equals an unknown state. When the system is in the prediction state, it would continually update its prediction on arrival times at the downstream stations unless an unexpected incident has occurred, which would lead the system into the unknown state. When a bus reaches the terminus of its servicing route, it enters the terminus state, which could be returning to garage or getting ready for the next trip. The terminus state is included to ensure that the system would go into the unknown state directly after processing, namely, δ (terminus state, Σ) equals an unknown state. The unknown state represents the situation in which the system is disrupted by unexpected events such as missing GPS data, map-matching failure, and vehicle breakdown. Under the unknown state, the system would try to return to the beginning state through initialization.

CASE STUDY

The proposed arrival time prediction system was implemented on a trunk bus route (Route 871), passing through one of the busiest districts in the city of Chongqing, China, as shown in Figure 5. The route includes a total of 14 bus stations, inbound from BaShan Station to Guan Yinyan Station and outbound from Guan Yinyan Station to BaShan Station, with a total distance of 22.8 km. The interstation distances are given in Table 1.

A total of 50 buses operated on this route, of which 36 are equipped with GPS receivers. The transit operated at a head of 10 min with a cycle time of 80 min. Two weeks of GPS data were collected with each day operating from 8:00 a.m. to 7:30 p.m. The GPS data were logged every 26 s, including device ID, time, longitude, latitude, and speed. The proposed system was set up in such a way so that it would predict bus arrival time whenever the system receives the GPS data. The performance of the prediction algorithm is evaluated using mean absolute percentage error (MAPE), defined as

$$\text{MAPE} = \frac{1}{n}\sum_{i}^{n}\frac{\left|t_i - t_o\right|}{t_o} \times 100\%$$

where

t_i = predicted arrival time,
t_o = observed true arrival time, and
n = number of predictions.

To predict bus arrival time, the GPS positions of the buses were first matched to the route and then used to calculate the distances from the current location of these buses to their downstream bus stations,

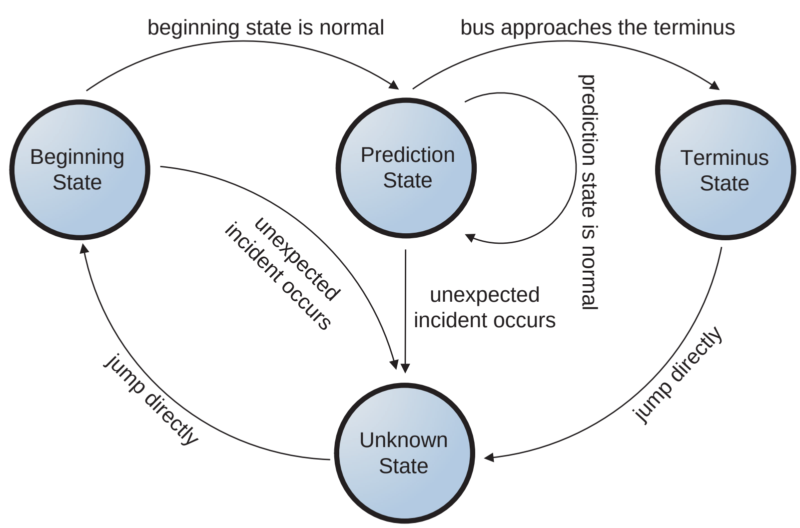

FIGURE 2 The four states of the proposed intelligent system for predicting bus arrival time.

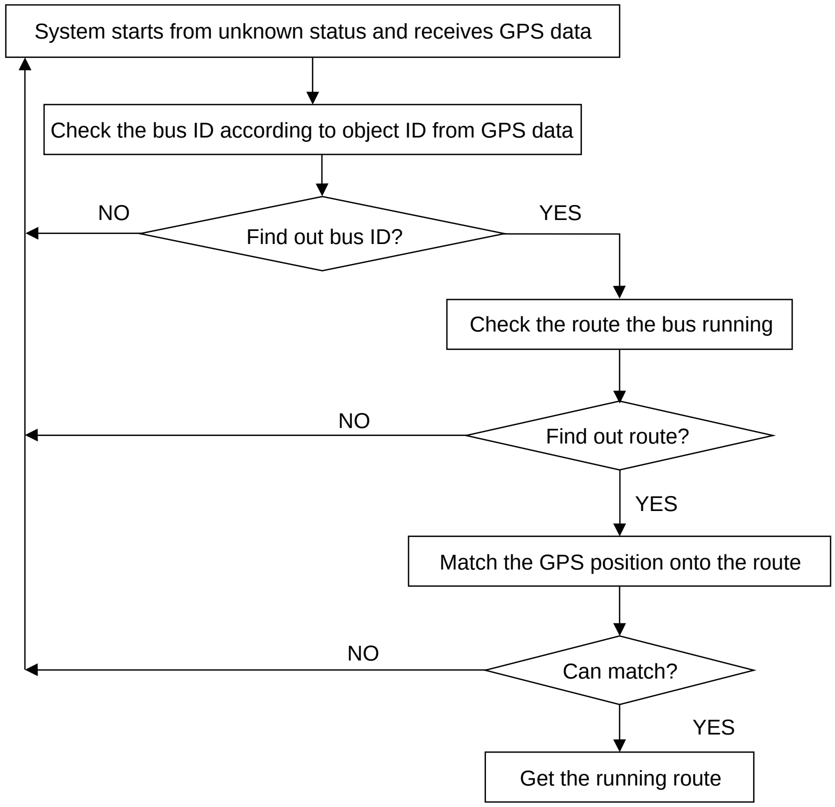

FIGURE 3 Algorithm for determining bus route.

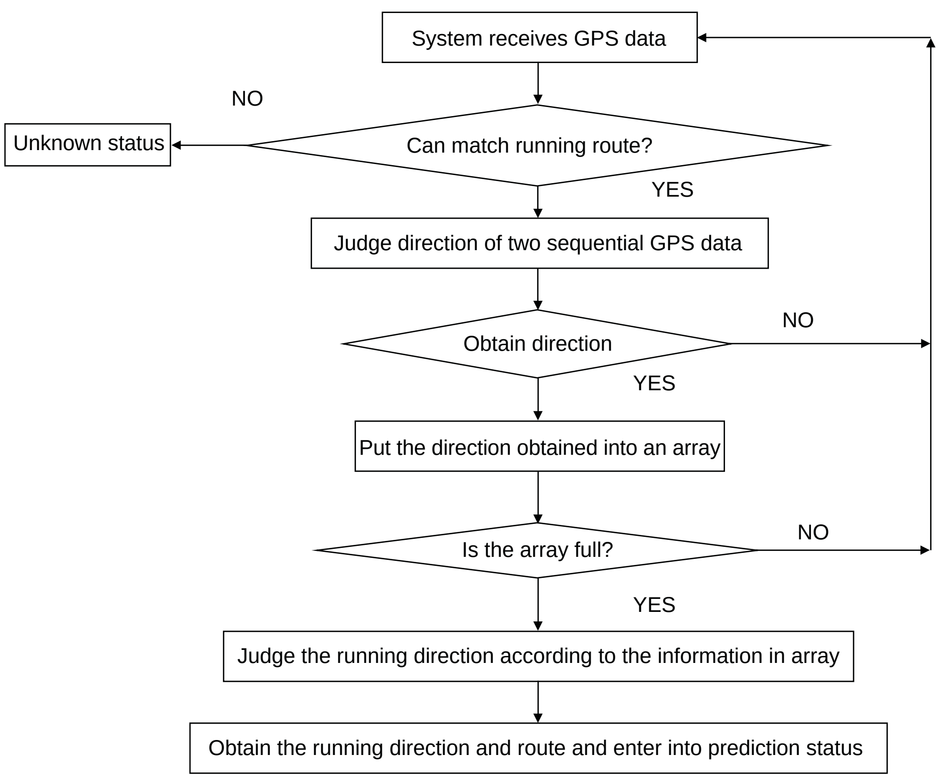

FIGURE 4 Algorithm for determining bus running direction.

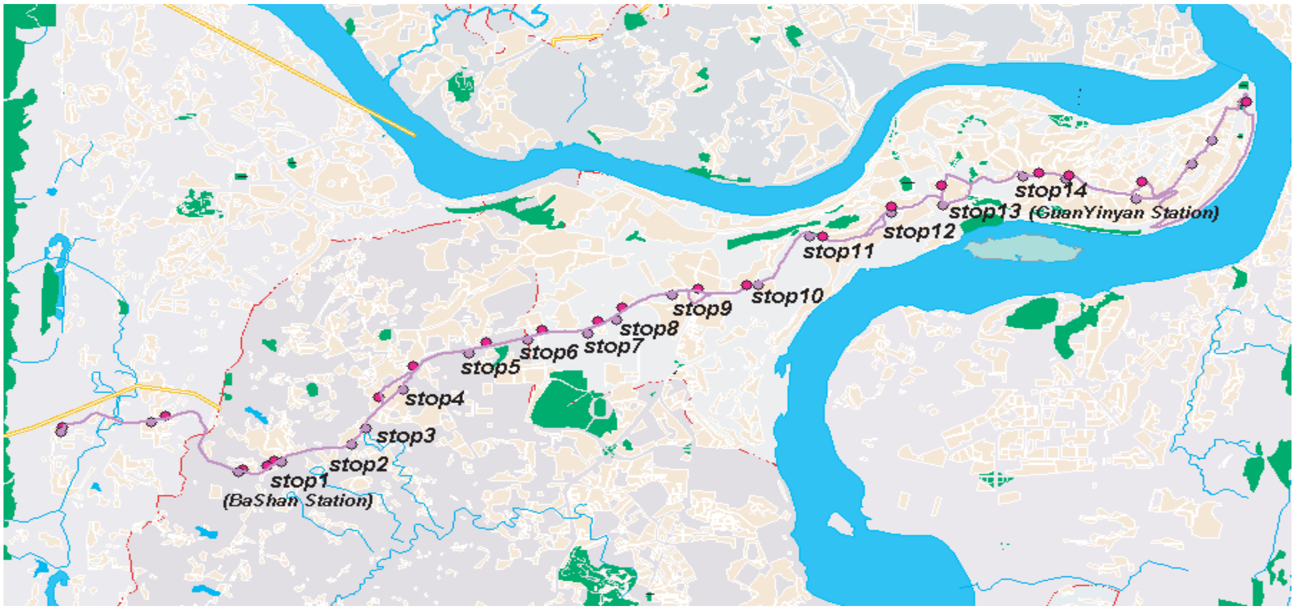

FIGURE 5 Layout of Bus Route 871 in Chongqing, China.

as described previously. The test results showed that the maximal absolute estimation error was less than 5%.

This research also compared the performance of the proposed algorithm with the algorithm proposed by Li et al. (*9*) in terms of prediction accuracy. Figure 6 shows the estimated arrival times by the two algorithms along with the actual arrival times as well as the corresponding estimation errors. The results indicate that the proposed algorithm had much less average and maximum estimation errors compared with Li et al.'s algorithm. This performance improvement may be attributed to timely detection of traffic congestion with continuous location information.

TABLE 1 Interstation Distance of Route 871

Inbound		Outbound	
Bus Station	Distance (meters)	Bus Station	Distance (meters)
Stop 1 (BaShan Station)	0	Stop 1 (Guan Yinyan Station)	0
Stop 2	993.2	Stop 2	1,418.7
Stop 3	230.5	Stop 3	851.3
Stop 4	708.3	Stop 4	1,107.2
Stop 5	1,101.2	Stop 5	1,310.4
Stop 6	729.5	Stop 6	646.6
Stop 7	711.3	Stop 7	978.1
Stop 8	423.0	Stop 8	352.8
Stop 9	871.2	Stop 9	741.3
Stop 10	1,153.0	Stop 10	695.3
Stop 11	1,075.7	Stop 11	985.6
Stop 12	1,208.1	Stop 12	890.1
Stop 13	774.8	Stop 13	386.2
Stop 14 (Guan Yinyan Station)	1,182.8	Stop 14 (BaShan Station)	1,109.6

Table 2 gives the performance statistics of the proposed algorithm by time period. It can be observed that the average estimation error is the highest at the early morning period (8:30 to 9:30 a.m.) when the bus route was most congested. This is because variation in traffic conditions and thus speed increases as the level of congestion increases, which means travel times to downstream stations become less predictable.

Figure 7 shows the MAPE of the predicted arrival times at individual bus stations for different time periods. As shown in the figure, there is a large variation in prediction accuracy with respect to time period and station. For example, the MAPE of arrival time at Station 13 was very high for all periods due to the combined effects of traffic lights and dwell time along the route segment leading to this station. For the same reason, the MAPE for the period from 10:30 to 11:30 a.m. was the smallest.

The performance of the bus arrival time prediction algorithm is also expected to change as the prediction horizon or the distance to the downstream station increases. To quantify this relationship, a total of 500 GPS readings, located between Station 1 and Station 2, were used. Figure 8 shows the MAPE as a function of the distance. Interestingly, while the absolute prediction error increases as the prediction horizon increases, the relative prediction error actually decreases as the leading time increases. The relative error levels off when the prediction horizon exceeds 10 min.

Lastly, this research evaluated the performance of the proposed intelligent prediction system in terms of route determining and direction detection. Two GPS-equipped buses were tested with runs involving many direction changes, and the results are summarized in Table 3. The prediction system was perfect in detecting travel direction and had a success rate of over 98% in matching the service route. The route matching errors are expected due to possible errors in GPS location data and inaccuracy of the transit network model. These matching errors can be reduced by using higher-quality GPS receivers and more accurate digital street maps.

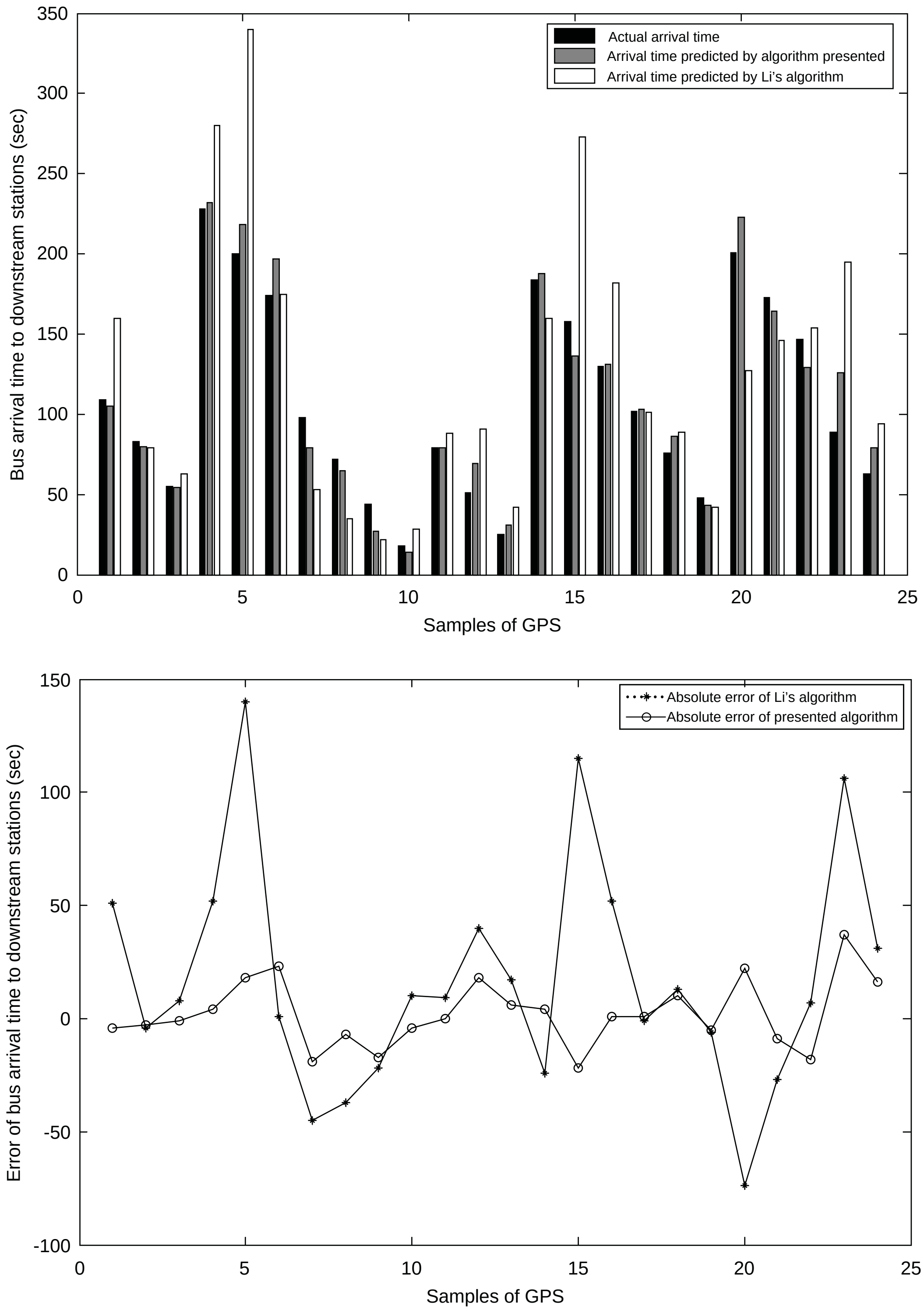

FIGURE 6 Performance of the proposed algorithm for bus arrival time prediction.

TABLE 2 Performance of Bus Arrival Time Prediction by Time of Day

	8:30–9:30 a.m.	9:30–10:30 a.m.	10:30–11:30 a.m.	11:30 a.m.–12:30 p.m.
Number of samples	1,181	918	1,113	516
Maximal error (s)	186	110	152	170
Minimal error (s)	0	0	0	0
Average error (s)	13.2	10.0	9.1	10.2
MAPE (%)	19.07	15.06	13.76	15.69

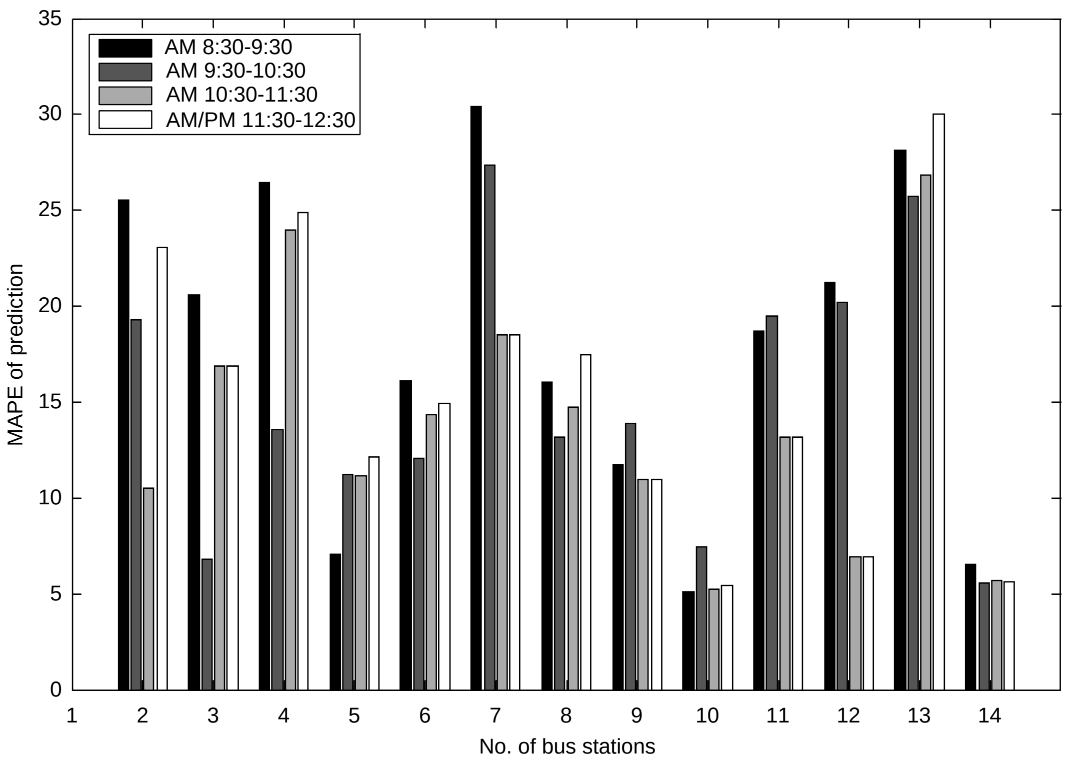

FIGURE 7 Prediction error by bus station.

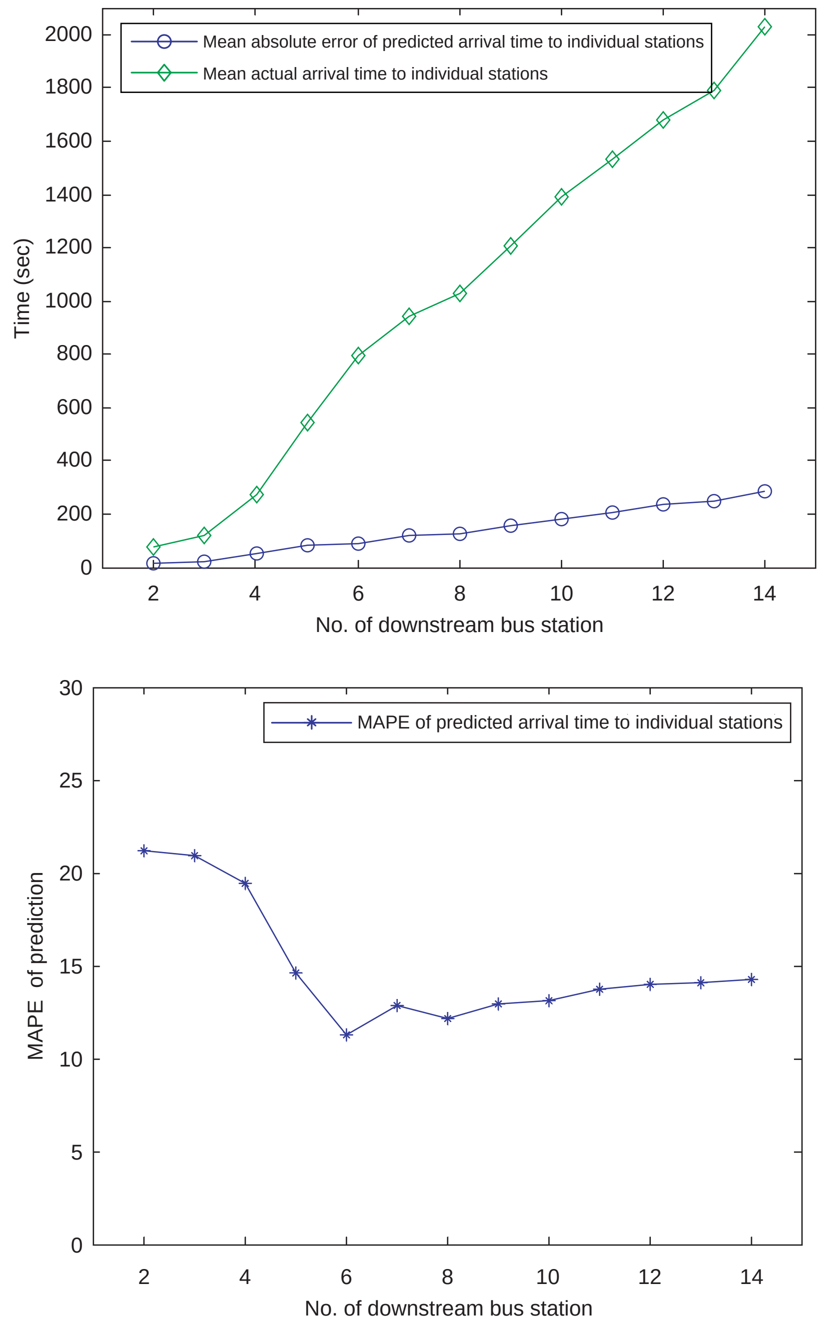

FIGURE 8 Prediction error by length of prediction.

TABLE 3 Performance of the Proposed System for Detecting Service Route and Running Direction

Bus ID	Number of GPS Records	Number of Successful Map Matchings	Rate of Successful Route Matching (%)	Rate of Successful Detection of Running Direction (%)
Bus 1	1,113	1,106	99.4	100
Bus 2	440	435	98.9	100

CONCLUSIONS

Accurate prediction of bus arrival time can not only help passengers time their departure times from work places and homes and make successful transfers by reducing waiting times at stops, but also help transit agencies manage and operate their systems in a more responsive manner such as real-time dispatching and scheduling. This paper has presented an algorithm for real-time prediction of bus arrival time, which has been implemented in an intelligent prediction system. The system is capable of tracking a large number of buses simultaneously, detecting their service routes and directions automatically, and predicting their arrival time to downstream stations with an acceptable accuracy.

REFERENCES

1. Fu, L., and X. Yang. Design and Implementation of Bus-Holding Control Strategies with Real-Time Information. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1791,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 6–12.
2. Schweiger, C. L. *TCRP Synthesis 48: Real-Time Bus Arrival Information Systems: A Synthesis of Transit Practice.* Transportation Research Board of the National Academies, Washington, D.C., 2003.
3. Sun, D., and L. Fu. Cellular Phone Based Real-Time Bus Arrival Information System. *Proc., Eighth International Conference on Applications of Advanced Technologies in Transportation Engineering,* Beijing, May 2004.
4. Lin, W.-H. and J. Zeng. Experimental Study on Real-Time Bus Arrival Time Prediction with GPS Data. In *Transportation Research Record: Journal of the Tranportation Research Board, No. 1666*, TRB, National Research Council, Washington, D.C., 1999, pp. 101–109.
5. Kidwell, B. *Predicting Transit Vehicle Arrival Times.* GeoGraphics Laboratory, Bridgewater State College, Bridgewater, Mass., 2001.
6. Patnaik, J., S. Chien, and A. Bladikas. Estimation of Bus Arrival Times Using APC Data. *Journal of Public Transportation*, Vol. 7, No. 1, 2004, pp. 1–20.
7. Jeong, R., and L. R. Rilett. Bus Arrival Time Prediction Using Artificial Neural Network Model. *Proc., IEEE Intelligent Transportation Systems Conference,* Washington, D.C., 2004, pp. 988–993.
8. Lin, W., and R. L. Bertini. Modeling Schedule Recovery Processes in Transit Operations for Bus Arrival Time Prediction. *Proc., IEEE 5th International Conference on Intelligent Transportation Systems,* Singapore, 2002, pp. 857–862.
9. Li, W., M. W. Koendjbiharie, R. C. Juca, Y. Yamashita, and A. Maciver. Algorithms for Estimating Bus Arrival Times Using GPS Data. *Proc., IEEE 5th International Conference on Intelligent Transportation Systems,* Singapore, 2002, pp. 868–873.
10. Jeong, R. The Prediction of Bus Arrival Time Using Automatic Vehicle Location Systems Data. txspace.tamu.edu/bitstream/1969.1/1458/1/etd-tamu-2004C-CVEN-Jeong.pdf. Texas A&M University, College Station, 2004.

The Bus Transit Systems Committee sponsored publication of this paper.