Bus Travel Time Prediction Model for Dynamic Operations Control and Passenger Information Systems

Amer Shalaby, Ph.D., P.Eng.
Assistant Professor, Department of Civil Engineering
University of Toronto
35 St. George Street
Toronto, Ontario, Canada M5S 1A4
Tel: (416) 978-5907
Fax: (416) 978-5054
e-mail: amer@ecf.utoronto.ca

Ali Farhan, M.A.Sc.
Graduate Student, Department of Civil Engineering
University of Toronto
35 St. George Street
Toronto, Ontario, Canada M5S 1A4
Tel: (416) 978-5049
Fax: (416) 978-5054
e-mail: ali.farhan@utoronto.ca

Paper submitted for presentation at the

Word Count = 6523
ABSTRACT
Automatic Vehicle Location (AVL) and Automatic Passengers Counters (APC) systems have been increasingly implemented by transit agencies for the real time monitoring of transit vehicles and automatic counting of passengers boarding and alighting at bus stops. As a result, a vast amount of potentially online data related to transit operation could be obtained from these systems.

The emphasis of this research effort was on using AVL and APC dynamic data to develop a bus travel time model capable of providing real time information on bus arrival times to passengers, via traveler information services and to transit controllers for the application of proactive control strategies. The developed model is based on two Kalman filter algorithms for the prediction of running times and dwell times alternately in an integrated framework. The AVL and APC data used were obtained for a specific bus route in Downtown Toronto. The performance of the developed prediction model was tested using “hold out” data and other data from microsimulation representing different scenarios of bus operation along the investigated route using the “VISSIM” microsimulation software package. The Kalman Filter algorithm outperformed all other developed models in terms of accuracy, demonstrating the dynamic ability to update itself based on new data that reflected the changing characteristics of the transit-operating environment.

A user-interactive system was developed to provide continuous information on the expected arrival time of buses at downstream stops, hence the expected deviations from schedule. The system enables the user to assess in real time transit stop-based control actions to avoid such deviations before their occurrence, hence allowing for proactive control, as opposed to the traditional reactive control, which attempts to recover the schedule after deviations occur.

Key Words: Advanced Public Transit Systems, Automatic Vehicle Location, Automatic Passenger Counters, bus arrival times, bus travel times, operations control
INTRODUCTION
Recently, a growing interest has been developing in various Advanced Public Transportation Systems (APTS) solutions that mainly aim at maximizing transit system efficiency and productivity using emerging technologies. Examples of such advanced technologies include Automatic Vehicle Location (AVL) and Automatic Passenger Counter (APC) systems.

Several researchers (1, 2, and 3) have used AVL (and less often APC) data to develop models specifically for bus travel time prediction. The motivation for developing these models was mostly for providing information to transit riders on expected bus arrival times with virtually no sensitivity of such models to operations control strategies. Thus, these models included very simple independent variables such as historical link travel times, upstream schedule deviations, and headway distributions, in addition to the current location of the next bus.

This study aims at developing a dynamic bus arrival time prediction model, using AVL and APC information, for dynamic operations control and dissemination of real-time transit information. The study is part of a larger project that aims at developing an integrated system for dynamic operations control and real-time transit information. Currently, almost all transit operators implement control strategies, such as bus holding and expressing, after detecting schedule/headway deviations in the system, hence reactive in nature. The proposed system (shown in Figure 1) takes a proactive approach to operations control that would enable the controller to implement preventive strategies before the actual occurrence of deviations. This innovative approach requires the use of arrival time models sensitive to the considered control strategies (mainly stop-based strategies). This research study focuses on developing a model of such characteristics.

DATA
The data used for this study were collected from bus route number 5 in the downtown Toronto area in May 2001. The route length is approximately 6.5 km, spanning 27 bus stops in each direction, 6 of which are time-point stops located at points of high passenger demand (e.g. major intersections). The route starts at the Eglinton subway station stop in the north and ends at the Treasury stop in the south during the morning peak period. At the other times of the day, the route ends further south at the Elm stop. There are 21 signalized intersections along the route, 10 of which are actuated SCOOT system signals. The schedule headway during the peak periods is 12 min, increasing to 30 min during off peak. For the duration of the study (5 weekdays in May 2001), the TTC (Toronto Transit Commission) assigned to the route 4 buses, each fitted with a GPS (Global Positioning System) receiver and an APC (Automatic Passenger Counter). Each time the bus stopped, the bus location was recorded using the GPS receiver. Also, the numbers of passengers boarding and alighting at bus stops were recorded using the APC. The route was segmented into 5 links in each direction, with each link starting and ending at two consecutive time-point stops. The links range from 0.40 to 1.7 km in length depending on the spacing between the time-point stops and may include 2 to 8 bus stops. This study focused on modeling travel time along those links for the morning peak hour bus operation.

APPROACH
As implied earlier, most of the models found in the literature (e.g. 1, 2, 3 and 4) have included bus dwell times along any link in the travel time of that link (i.e. link travel time includes running time plus dwell times). As such, these models cannot consider explicitly the effect of late or early bus arrivals at bus stops on the dwell times at those stops and vice versa. Ignoring such relationship yields these models insensitive to the effects of variations at upstream bus stops, such as demand surge, bus holding strategy, and bus expressing strategy, etc., on downstream bus arrivals and subsequent dwell times. The approach taken in this study addresses this issue.

Conceptual Framework
The link running time and bus dwell time are modeled separately in this study but in a consistent single modeling framework. It is assumed that real-time information on vehicle location, passenger boarding and alighting at bus stops and bus arrival and departure times is known from AVL and APC systems. The prediction modeling system consists of two separate algorithms, each based on Kalman Filter techniques. The first algorithm is “Link Running Time Prediction Algorithm” which makes use of the last three-day historical data of the bus link running time for the instant of prediction (k+1) as well as the bus link running time for the previous bus on the current day at the instant (k) to predict the bus running time at k+1. The study used data for the previous 3 days only as this was deemed practical given the limited historical data available for the study. Obviously in real-world application the algorithm can make use of longer range of historical data. The second
algorithm, “Passenger Arrival Rate Prediction Algorithm”, employs similar historical data on passenger arrival rate. In order to predict the dwell time, the predicted arrival rate is simply multiplied by the predicted headway (i.e. last departure time minus predicted arrival time) and by the passenger boarding time (assumed 2.5 sec/passenger).

Separating the bus dwell time prediction from bus running time prediction in this modeling framework enhances the model’s ability to capture the effects of lateness or earliness of bus arrivals at stops on the bus dwell time at those stops, and hence the bus departure from such stops. In other words, bus dwell time at a stop is affected by the arrival time of the bus at that stop (i.e. the later the bus is the longer the dwell time will be and vice versa). In addition, since the model treats dwell time separately, it is sensitive to stop-based control strategies such as bus holding and expressing.

In order to better understand the prediction-modeling framework, Figure 2 shows a schematic of a hypothetical transit route. The route is divided into a number of links between bus stops. When the transit bus (n) leaves the stop (i), the departure time is known from the AVL system. At this instant the Kalman Filter prediction algorithm for link running times will predict the next link running time $RT_n(i,i+1)$. Subsequently, the predicted arrival time of the bus at the downstream bus stop ($i+1$) can be determined.

Assuming the bus n is currently at stop (i)

$$AT_n(i+1) = DT_n(i) + RT_n(i,i+1)$$

Where:

- $AT_n(i+1)$: predicted arrival time of bus n at stop ($i+1$)
- $RT_n(i,i+1)$: predicted running time between (i) and ($i+1$) from Kalman Filter prediction algorithm
- $DT_n(i)$: actual departure time of bus n from stop (i)

This predicted arrival time $AT_n(i+1)$ is used to predict the dwell time for bus n at stop ($i+1$) based on the passenger arrival rate and the average passenger boarding time at stop ($i+1$).

$$DWT_n(i+1) = \lambda_{(i+1)} * [AT_n(i+1) - AT_{n-1}(i+1)] * \rho_{avg}(i+1)$$

Where:

- $DWT_n(i+1)$ is the predicted dwell time for bus n at stop ($i+1$)
- $\lambda_{(i+1)}$: predicted passenger arrival rate at stop ($i+1$) from Kalman Filter prediction algorithm
- $AT_{n-1}(i+1)$: actual arrival time of previous bus n-1 at stop ($i+1$)
- $[AT_n(i+1) - AT_{n-1}(i+1)]$: is the predicted headway for bus n at stop ($i+1$)
- $\rho_{avg}(i+1)$: average passenger boarding time at stop ($i+1$), assumed to be 2.5 sec/passenger

In equation (2) above, the assumption is that the boarding passengers at each bus stop have the significant effect on bus dwell time at that stop compared with alighting passengers. It is noteworthy to mention that the time points used in this study, for which equation 2 applies, are located at high demand spots (i.e. subway station and major intersections) where stop skipping because of no passenger demand is extremely rare. If stop skipping at a particular time point were frequent, equation (2) would need to be modified to address this problem.

Having the arrival time and dwell time for bus n at stop ($i+1$) predicted, it is now easy to calculate the predicted departure time for bus n from stop ($i+1$) by adding the predicted arrival time to the predicted dwell time at stop ($i+1$).

$$DT_n(i+1) = AT_n(i+1) + DWT_n(i+1)$$

Where:

- $DT_n(i+1)$ is the predicted departure time for bus n from stop ($i+1$). This departure time prediction $DT_n(i+1)$ is a function of both arrival time prediction and dwell time prediction at stop ($i+1$). Hence, the effect of any variations in bus operation (i.e. bus early or late) could be captured in this step and will be reflected on all downstream bus stops.

Similarly, predictions of arrival times and departure times at all downstream stops can be computed while the bus is still at stop (i). This process is updated every time the bus leaves or arrives at a new stop.

**Kalman-Filter Prediction Algorithms**

As mentioned above, two Kalman Filter algorithms are used for the prediction modeling system. The main assumption is that the patterns of the link running time and passenger arrival rate at stops are cyclic for a specific period of day. In other words, knowledge of link travel time and number of passengers waiting for a specific bus in a certain period of time will allow one to predict these variables for the next bus during the same
period so long as conditions remain steady. When conditions change (e.g. demand surge at a stop and/or an incident occurred at a link) the model can update the effect of the new conditions on the predictions so long as the new conditions persist for a sufficient length of time.

The Kalman Filter algorithm works conceptually as follows. The “last three days” historical data of actual running times between links at the instant \( k+1 \) plus the last running time observation at the instant \( k \) on the current day are used to predict the bus running time at the instant \( k+1 \). Similarly, the passenger arrival rates of the previous three days at the instant \( k+1 \) plus the passenger arrival rate at the instant \( k \) on the current day are used to predict the passenger arrival rate at the instant \( k+1 \). The historical passenger arrival rate is obtained from the (APC) data as follows: at a bus stop it is the number of the on-passengers at that stop divided by the most recent headway (i.e. the arrival time of the previous bus minus the arrival time of the current bus).

Below are the equations used for both Kalman Filter prediction algorithms.

**Link Running Time Prediction Algorithm**

Generally, a Kalman-Filter algorithm for bus link running time has the following structure (5):

\[
g(k+1) = \frac{e(k) + \text{VAR}_{\text{data_{out}}}}{\text{VAR}_{\text{data_{in}}} + \text{VAR}_{\text{data_{out}}} + e(k)}
\]

\[
a(k + 1) = 1 - g(k + 1)
\]

\[
e(k + 1) = \text{VAR}\left[\text{data_{in}}\right] \cdot g(k + 1)
\]

\[
P(k + 1) = a(k + 1) \cdot \text{art}(k) + g(k + 1) \cdot \text{art}_1(k + 1)
\]

where:
- \( g \) = filter gain,
- \( a \) = loop gain,
- \( e \) = filter error,
- \( p \) = prediction,
- \( \text{art}(k) \) = actual running time of the previous bus at instant \( k \),
- \( \text{art}_1(k+1) \) = actual running time of the previous day at instant \( k+1 \),
- \( \text{VAR}_{\text{data_{out}}} \) = prediction variance, and
- \( \text{VAR}_{\text{data_{in}}} \) = last three days “\( \text{art}_3(k+1) \), \( \text{art}_2(k+1) \) and \( \text{art}_1(k+1) \)” variance.

The input variance \( \text{VAR}_{\text{data_{in}}} \) is calculated at each instant \( k + 1 \) using the actual running time values for the last three days: \( \text{art}_1(k + 1) \), \( \text{art}_2(k + 1) \) and \( \text{art}_3(k + 1) \):

\[
\text{VAR}_{\text{data_{in}}} = \text{VAR}[\text{art}_1(k + 1), \text{art}_2(k + 1), \text{art}_3(k + 1)]
\]

where:
- \( \text{art}_1(k+1) \): actual running time of the bus at instant \( k+1 \) on the previous day
- \( \text{art}_2(k+1) \): actual running time of the bus at instant \( k+1 \) two days ago
- \( \text{art}_3(k+1) \): actual running time of the bus at instant \( k+1 \) three days ago

The definition of the variance for a random variable is:

\[
\text{VAR}[X] = E[(X - E[X])^2]
\]

\[
E(X) = \text{Avg}(\text{art}) = \frac{\text{art}_1(k + 1) + \text{art}_2(k + 1) + \text{art}_3(k + 1)}{3}
\]

Now the variance can be calculated as shown in equation (14):
\[ \Delta_1 = [art_1(k+1) - \text{avg}(art)]^2 \]  
\[ \Delta_2 = [art_2(k+1) - \text{avg}(art)]^2 \]  
\[ \Delta_3 = [art_3(k+1) - \text{avg}(art)]^2 \]  
\[ \text{VAR} [\text{data}_{\text{in}}] = \frac{\Delta_1 + \Delta_2 + \Delta_3}{3} \]  

\[ \text{VAR} [\text{data}_{\text{out}}] \text{ is based on the model prediction output and the corresponding future observation. Both pieces of data are not available since the prediction is not made yet and the future trip has not been made either. Ideally, } \text{VAR} [\text{data}_{\text{out}}] \text{ should be equal to } \text{VAR} [\text{data}_{\text{in}}] \text{ for good prediction performance (6). Now, a new variance is introduced } \text{VAR} [\text{local data}]. \text{ It is equal to the variance of the input and output data.} \]  

\[ \text{VAR} [\text{local data}] = \text{VAR} [\text{data}_{\text{in}}] = \text{VAR} [\text{data}_{\text{out}}] \]  

and equations (4) and (6) become:

\[ g(k+1) = \frac{e(k) + \text{VAR}[\text{local data}]}{e(k) + 2 \cdot \text{VAR}[\text{local data}]} \]  

\[ e(k+1) = \text{VAR}[\text{local data}] \cdot g(k+1) \]  

Now it becomes easy is to implement the actual Kalman Filter algorithm to predict the bus running times along the links. The order of appearance of the equations in the program should be (16), (5), (17) and (7). Table 1 provides example calculations for the running time prediction along a particular link. Suppose we want to predict the bus running time for the next trip at 7:15, instant (k+1), and that the observed running time art(k) of the previous bus (i.e. at time 7:00) is 125 seconds, which could be known directly from the AVL system, and the historical actual running times of the last three days at instant k+1 (i.e. at time 7:15) are art1 =135, art2 =145 and art3 =116 seconds. The mean and variance are calculated as shown in the table. Now, to predict the running time for the current day bus P(k+1) (i.e. at time 7:15), the Kalman Filter equations (16), (5), (17) and (7) are used. First, to calculate the value of g(k+1) from equation (16), we need the value of the VAR [local data] for the instant of prediction k+1 (i.e. 7:15), which is equal to 144.67, and the value of the e(k), which is equal to 25.72 (from the previous bus prediction at time 7:00) as shown in Table 1. The value of a(k+1) is then calculated using equation (5). Now we calculate the value of e(k+1) so that it can be used for the next step prediction (k+2). The value of e(k+1) is equal to g(k+1) multiplied by Var [local data]. With the above calculations made, the predicted running time at instant k+1 (i.e. 7:15) can be calculated using equation (7). The value of P(k+1) is mainly based on the running time value of the previous bus art(k), (i.e. the 7:00 bus) on the current day and the value of the running time art1(k+1), (i.e. the 7:15 bus) on the previous day, with each value multiplied respectively by the dynamic weights g(k+1) and a(k+1), which are updated in each new prediction step.

Passenger Arrival Rate Prediction Algorithm

An algorithm with a similar structure was developed to predict the passenger arrival rate at time instant (k+1) using data from the APC system.

MODEL PERFORMANCE EVALUATION

In order to assess the predictive performance of the Kalman-Filter model, it is compared with three previously developed models for the same route. They include historical average model, regression analysis model, and an artificial neural network (specifically, Time Lag Recurrent Network). A more detailed exposition of the models can be found elsewhere (4). Similar to most models found in the literature, the regression and TLRN models predict individual link travel times, which include running plus dwell times. As mentioned earlier, the AVL and APC data for the study route were available for 5 consecutive days. The regression and TLRN models were developed using data of 4 days only, with the fifth day’s data held out for testing.
Four testing data sets were used for the comparison; the first set includes the “hold out” data of the fifth day of observations, while the remaining three sets include artificial data collected from three different bus operation scenarios representing: normal-condition bus operation, special event scenario when there is demand surge on transit service, and lane closure scenario when one lane on a specific link is blocked due to accident or maintenance work. In contrast to the “hold out” data and “normal condition” scenario, the “lane closure” and “demand surge” scenarios represent atypical conditions. Because real-world data of such conditions are hard to obtain, the “VISSIM” traffic microsimulation software was used to simulate these scenarios. The simulation of the entire corridor, the calibration results and the simulation of the scenarios are described elsewhere (7). After each simulation run, all the necessary data required for model testing was extracted and analyzed. Three prediction error measurements were computed for all developed models to test the model performance (8). These error indices include:

**Mean Relative Error** ($\epsilon_{\text{mean}}$), which indicates the expected error as a fraction of the measurement

$$\epsilon_{\text{mean}} = \frac{1}{N} \sum \left( \frac{X_{\text{true}}(t) - X_{\text{pred}}(t)}{X_{\text{true}}(t)} \right)$$

**Root Squared relative Error** ($\epsilon_{\text{rs}}$), which captures large prediction errors

$$\epsilon_{\text{rs}} = \sqrt{\frac{1}{\sum X_{\text{true}}(t) \sum \left( \frac{X_{\text{true}}(t) - X_{\text{pred}}(t)}{X_{\text{true}}(t)} \right)^2 X_{\text{true}}(t)}}$$

**Maximum relative error** ($\epsilon_{\text{max}}$), which capture the maximum prediction error

$$\epsilon_{\text{max}} = \max \left( \frac{X_{\text{true}}(t) - X_{\text{pred}}(t)}{X_{\text{true}}(t)} \right)$$

Where $N$ = the number of samples (here $N=50$) $X_{\text{true}}(t)$ = measured value at time $t$; $X_{\text{pred}}(t)$ = predicted value at time $t$.

Table 2 shows the three error measures $\epsilon_{\text{mean}}$, $\epsilon_{\text{rs}}$, $\epsilon_{\text{max}}$ for the “hold out” data set, while Figure 3(a,b,c) summarize the performance of the three prediction models for each simulated scenario. Obviously, the lower the error, the better the model performance.

**Discussion of Results**

The results summarized in Table 2 show that for all links, the Kalman Filter model provides the minimum value for the error measures $\epsilon_{\text{mean}}$, $\epsilon_{\text{rs}}$, $\epsilon_{\text{max}}$ pointing to the fact that its performance was the best compared with the other prediction models, except for link # 4 where the regression and TLRN models perform slightly better than the Kalman Filter model.

Table 2 and Figure 3(a,b,c) show there is no significant difference in the performance of the regression and Artificial Neural Network models for the three different scenarios. Both models give similar performance results for each scenario; their accuracy performance decreased for the special event and lane closure scenarios. But in general the artificial neural network model always gives lower values for the relative error indices.

The Kalman Filter model shows the best prediction performance in all three scenarios. Its performance was similar to the regression and ANN models in the normal condition scenario, but it showed superior performance to the other models in the special event and lane closure scenarios. These results showed the superior performance of the Kalman Filter model compared with other prediction models in terms of the relative error, and it also demonstrates how this model can capture dynamic changes due to different bus operation characteristics.

In addition to its highly accurate performance in dynamic environments, the model has the advantage of capturing the effects of control strategies such as holding and expressing at upstream bus stops. For example, if the
bus is currently at a time-point where it will be held for additional 1 min, the model appropriately captures the effect of this extra time on the dwell time at that stop (which is function of number of passengers waiting at that bus stop when the bus is predicted to arrive) and the arrival time at the next bus stop and so forth for the prediction of arrival time and dwell times at the subsequent bus stops.

USER-INTERACTIVE DYNAMIC CONTROL SYSTEM
The developed arrival time prediction model is used to develop a dynamic control system. This system simply uses the scheduled arrival time at time-points to develop a time profile for each scheduled trip along the route, i.e. ‘schedule travel time profile’ (Figure 4), which is done for each bus at the start of its journey. Another ‘prediction travel time profile’ is constructed using the Kalman filter prediction model. The prediction travel time profile is updated dynamically every time the bus arrives and departs from a time point.

By using these two travel time profiles, we are able to predict if the bus is running early or late at each time point. This is shown in Figure 4 as \( \Delta \). A positive value of \( \Delta \) means that the prediction profile is currently lying above the scheduled one (bus will reach the downstream time points late), while a negative value of \( \Delta \) occurs when the prediction profile lies below the schedule (bus will be ahead of schedule). In these cases, implementation of a corrective proactive control strategy is required to prevent expected schedule deviation downstream. A value of 0 refers to the compliance of the bus to the schedule. The value of \( \Delta \) is the key factor for deciding what type of control strategy to be implemented. If \( \Delta \) is positive, an expressing type of control is required to be applied while a negative value of \( \Delta \) indicates implementation of some type of holding strategy.

System Design and Architecture
The proposed system, shown in Figure 5, is an interactive program developed using the Visual Basic programming language. This program effectively utilizes AVL and APC data for dynamic bus arrival information and performance analysis at downstream bus stops for the purpose of applying real-time, proactive control strategies.

The program records the dynamic actual location and time of the transit vehicle when it arrives and leaves all time points along the transit route based on the AVL data. Also, it records the number of passengers alighting and departing the bus at each bus stop. Real time AVL and APC information is transferred instantly from the server to the program. For a specific bus trip, at the departure instant of bus from the terminal station, the prediction algorithms will automatically be activated to predict the arrival and departure times of the bus at all downstream time points (prediction profile). At the same time, the associated bus departure schedule for all time points along the route is also displayed (schedule profile). The difference between the predicted and the schedule departure (\( \Delta \) information) is automatically computed for all time points. If the value of the predicted bus time deviation \( \Delta \) is within an accepted range (e.g. 0 to 2 minutes), the predicted departure times are within the schedule, and no control strategies are required to be implemented. In such a case, the program will display black font color with white background for \( \Delta \) value labels. On the other hand, when \( \Delta \) values are more than 2 minutes (i.e. bus expected to depart late at downstream time points), the font color will display red. The transit controller can interact with the program to assess the effect of bus expressing at one or more downstream time points (by setting dwell times at 0 for such time points) so as to as reduce predicted deviations. If the \( \Delta \) value is less than 0, (i.e. bus expected to be ahead of schedule) then the \( \Delta \) value label background is displayed red and the font color is black. The controller can assess the effect of bus holding at one or more downstream time points (by increasing the dwell times at those time points).

The prediction algorithms of the system will be always dynamically updated based on the AVL data. As soon as the bus arrives or departs a new time point, new arrival and departure time predictions and new \( \Delta \) values for the remaining time points downstream will be processed.

At the end of each trip, the system records the observed AVL arrival and departure times as well as the real APC data regarding the number of passenger boardings and alightings at each bus stop. These data are used to update the system database (link running time and passenger arrival rate at time points) to be used for the kalman filter prediction algorithms.

In addition, the system computes the on-time route performance for each trip by comparing the actual bus arrival, departure with the schedule arrival, departure times for all time points. The average on-time performance is automatically calculated and displayed on the screen for all trips. This feature provides the transit management with an easy tool to evaluate the route level of service in terms of on-time performance.
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS
An innovative model was developed for dynamic bus arrival time prediction. The model is based on two Kalman filter algorithms for the prediction of running times and dwell times alternately in an integrated framework. As such, the model can capture the interaction between the 2 variables (i.e. the effect of one on the other). The model was shown to outperform other traditional models (regression and Neural Network models) in terms of predictive ability when tested on a “hold out” real-world data. More importantly, the superiority of the model was even more prominent when tested on two simulated scenarios representing passenger demand surge (for example because of a special event) and lane closure (for example because of an incident). This is primarily due to the continuous updating of the model parameters based on dynamic real-time data, as opposed to traditional models which are typically calibrated using historical data, with infrequent recalibration of the model, if any.

Because dwell time is predicted separately and its effect on bus arrival times at downstream stops is accounted for, the model can be used for assessing transit stop-based dynamic control actions (e.g. bus holding, bus expressing). A user-interactive system was developed to provide continuous information on the expected arrival time of buses at downstream stops, hence the expected deviations from schedule. The system enables the user to assess in real time transit stop-based control actions to avoid such deviations before their occurrence, hence allowing for proactive control, as opposed to the traditional reactive control which attempts to recover the schedule after deviations occur.

The model developed here was based on data from one bus route in downtown Toronto. However, the same modeling approach is applicable to other medium- to low-frequency routes where schedule control and dissemination of expected arrival times are relevant.

Further work can improve the model developed here in several ways. For example, better representative distributions of passenger arrivals at bus stops could be attempted instead of the implied uniform distribution assumed here. Also, further investigation is required to develop predictive models for overlapping routes that serve the same bus stops. In such cases, special consideration should be given to dwell time prediction. Finally, the assessment of the model developed here would be greatly enhanced if tested in the field under both normal and atypical conditions.

REFERENCES


2. Lin, Wei-Hua and Zeng, Jian “ Experimental Study of Real-Time Bus Arrival Time Prediction with GPS Data”. Transportation Research Record 1666, PP 101-109, 1999


LIST OF TABLES

TABLE 1 Kalman Filter Algorithm Calculations for Link Running Time
TABLE 2 Relative Error Results of the Prediction Models Using “Hold Out” Data

LIST OF FIGURES

FIGURE 1 Integrated Operations Control and Information-Dissemination System
FIGURE 2 Schematic of a Bus Route with Several Stops
FIGURE 3 Relative Error Results of the Prediction Models Using Artificial Data
FIGURE 4 Predicted and Schedule Travel Time Profiles
FIGURE 5 Illustration of the Interactive Information and Control System
### TABLE 1: Kalman Filter Algorithm Calculations for Link Running Time

<table>
<thead>
<tr>
<th>time</th>
<th>art3</th>
<th>art2</th>
<th>art1</th>
<th>average</th>
<th>Δ 3</th>
<th>Δ 2</th>
<th>Δ 1</th>
<th>VAR [local]</th>
<th>g(k+1)</th>
<th>a(k+1)</th>
<th>e(k+1)</th>
<th>P(K+1)</th>
<th>observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:12:00</td>
<td>72</td>
<td>66</td>
<td>85</td>
<td>74.33</td>
<td>5.44</td>
<td>69.44</td>
<td>113.78</td>
<td>62.89</td>
<td>0.58</td>
<td>0.42</td>
<td>36.66</td>
<td>82.00</td>
<td>86</td>
</tr>
<tr>
<td>6:27:00</td>
<td>92</td>
<td>86</td>
<td>90</td>
<td>89.33</td>
<td>7.11</td>
<td>11.11</td>
<td>0.44</td>
<td>6.22</td>
<td>0.87</td>
<td>0.13</td>
<td>5.43</td>
<td>89.49</td>
<td>95</td>
</tr>
<tr>
<td>6:39:00</td>
<td>105</td>
<td>95</td>
<td>112</td>
<td>104.00</td>
<td>1.00</td>
<td>81.00</td>
<td>64.00</td>
<td>48.67</td>
<td>0.53</td>
<td>0.47</td>
<td>25.62</td>
<td>103.95</td>
<td>107</td>
</tr>
<tr>
<td>6:51:00</td>
<td>265</td>
<td>110</td>
<td>132</td>
<td>169.00</td>
<td>9216.00</td>
<td>3481.00</td>
<td>1369.00</td>
<td>4688.67</td>
<td>0.50</td>
<td>0.50</td>
<td>2350.72</td>
<td>119.53</td>
<td>95</td>
</tr>
<tr>
<td>7:03:00</td>
<td>118</td>
<td>127</td>
<td>115</td>
<td>120.00</td>
<td>4.00</td>
<td>49.00</td>
<td>25.00</td>
<td>26.00</td>
<td>0.99</td>
<td>0.01</td>
<td>25.72</td>
<td>114.78</td>
<td>125</td>
</tr>
<tr>
<td>7:15:00</td>
<td>116</td>
<td>145</td>
<td>135</td>
<td>132.00</td>
<td>256.00</td>
<td>169.00</td>
<td>9.00</td>
<td>144.67</td>
<td>0.54</td>
<td>0.46</td>
<td>78.24</td>
<td>130.41</td>
<td>129</td>
</tr>
<tr>
<td>7:27:00</td>
<td>155</td>
<td>195</td>
<td>165</td>
<td>171.67</td>
<td>277.78</td>
<td>544.44</td>
<td>44.44</td>
<td>288.89</td>
<td>0.56</td>
<td>0.44</td>
<td>161.67</td>
<td>149.15</td>
<td>185</td>
</tr>
<tr>
<td>7:39:00</td>
<td>241</td>
<td>168</td>
<td>179</td>
<td>196.00</td>
<td>2025.00</td>
<td>784.00</td>
<td>289.00</td>
<td>1032.67</td>
<td>0.54</td>
<td>0.46</td>
<td>553.82</td>
<td>181.78</td>
<td>165</td>
</tr>
<tr>
<td>7:51:00</td>
<td>185</td>
<td>172</td>
<td>210</td>
<td>199.00</td>
<td>16.00</td>
<td>289.00</td>
<td>441.00</td>
<td>246.67</td>
<td>0.76</td>
<td>0.24</td>
<td>189.64</td>
<td>199.35</td>
<td>190</td>
</tr>
<tr>
<td>8:03:00</td>
<td>236</td>
<td>248</td>
<td>236</td>
<td>240.00</td>
<td>16.00</td>
<td>64.00</td>
<td>16.00</td>
<td>32.00</td>
<td>0.87</td>
<td>0.13</td>
<td>27.97</td>
<td>239.20</td>
<td>240</td>
</tr>
<tr>
<td>8:15:00</td>
<td>212</td>
<td>269</td>
<td>210</td>
<td>230.33</td>
<td>33.11</td>
<td>1495.11</td>
<td>413.44</td>
<td>748.22</td>
<td>0.51</td>
<td>0.49</td>
<td>380.97</td>
<td>224.72</td>
<td>235</td>
</tr>
<tr>
<td>8:27:00</td>
<td>305</td>
<td>225</td>
<td>265</td>
<td>265.00</td>
<td>1600.00</td>
<td>1600.00</td>
<td>0.00</td>
<td>1066.67</td>
<td>0.58</td>
<td>0.42</td>
<td>614.15</td>
<td>252.27</td>
<td>233</td>
</tr>
<tr>
<td>8:39:00</td>
<td>214</td>
<td>235</td>
<td>210</td>
<td>219.67</td>
<td>32.11</td>
<td>235.11</td>
<td>93.44</td>
<td>120.22</td>
<td>0.86</td>
<td>0.14</td>
<td>103.31</td>
<td>213.24</td>
<td>232</td>
</tr>
<tr>
<td>8:51:00</td>
<td>188</td>
<td>269</td>
<td>215</td>
<td>224.00</td>
<td>1266.00</td>
<td>2025.00</td>
<td>81.00</td>
<td>1134.00</td>
<td>0.52</td>
<td>0.48</td>
<td>591.70</td>
<td>223.13</td>
<td>190</td>
</tr>
<tr>
<td>9:08:00</td>
<td>215</td>
<td>236</td>
<td>165</td>
<td>205.33</td>
<td>93.44</td>
<td>940.44</td>
<td>1626.78</td>
<td>886.89</td>
<td>0.63</td>
<td>0.37</td>
<td>554.37</td>
<td>174.37</td>
<td>180</td>
</tr>
</tbody>
</table>
TABLE 2: Relative Error Results of the Prediction Models Using “Hold Out” Data

<table>
<thead>
<tr>
<th>link</th>
<th>error indices</th>
<th>Historical Average</th>
<th>Regression</th>
<th>TLRN</th>
<th>Kalman Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ε&lt;sub&gt;mean&lt;/sub&gt;</td>
<td>0.268</td>
<td>0.099</td>
<td>0.076</td>
<td>0.069</td>
</tr>
<tr>
<td>link1</td>
<td>ε&lt;sub&gt;rs&lt;/sub&gt;</td>
<td>0.322</td>
<td>0.115</td>
<td>0.094</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;max&lt;/sub&gt;</td>
<td>0.900</td>
<td>0.192</td>
<td>0.168</td>
<td>0.117</td>
</tr>
<tr>
<td>link2</td>
<td>ε&lt;sub&gt;mean&lt;/sub&gt;</td>
<td>0.537</td>
<td>0.134</td>
<td>0.063</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;rs&lt;/sub&gt;</td>
<td>0.603</td>
<td>0.149</td>
<td>0.075</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;max&lt;/sub&gt;</td>
<td>3.172</td>
<td>0.224</td>
<td>0.131</td>
<td>0.077</td>
</tr>
<tr>
<td>link3</td>
<td>ε&lt;sub&gt;mean&lt;/sub&gt;</td>
<td>0.369</td>
<td>0.200</td>
<td>0.165</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;rs&lt;/sub&gt;</td>
<td>0.543</td>
<td>0.220</td>
<td>0.166</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;max&lt;/sub&gt;</td>
<td>0.801</td>
<td>0.466</td>
<td>0.342</td>
<td>0.227</td>
</tr>
<tr>
<td>link4</td>
<td>ε&lt;sub&gt;mean&lt;/sub&gt;</td>
<td>0.173</td>
<td>0.079</td>
<td>0.071</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;rs&lt;/sub&gt;</td>
<td>0.208</td>
<td>0.092</td>
<td>0.088</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;max&lt;/sub&gt;</td>
<td>0.502</td>
<td>0.228</td>
<td>0.153</td>
<td>0.232</td>
</tr>
<tr>
<td>link5</td>
<td>ε&lt;sub&gt;mean&lt;/sub&gt;</td>
<td>0.186</td>
<td>0.084</td>
<td>0.104</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;rs&lt;/sub&gt;</td>
<td>0.181</td>
<td>0.076</td>
<td>0.120</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>ε&lt;sub&gt;max&lt;/sub&gt;</td>
<td>0.502</td>
<td>0.228</td>
<td>0.240</td>
<td>0.122</td>
</tr>
</tbody>
</table>
FIGURE 1: Integrated Operations Control and Information-Dissemination System

FIGURE 2: Schematic of a Bus Route with Several Stops
FIGURE 3: Relative Error Results of the Prediction Models Using Artificial Data
FIGURE 4: Predicted and Schedule Travel Time Profiles
FIGURE 5: Illustration of the Interactive Information and Control System