

Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data

Amer Shalaby, University of Toronto

Ali Farhan, City of Calgary

Abstract

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 and departure times to passengers (via traveler information services) and to transit controllers for the application of proactive control strategies. The developed model is comprised of 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 a microsimulation model representing different scenarios of bus operation along the investigated route using the VISSIM microsimulation software package. The Kalman filter-based model outperformed other conventional 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 and departure times 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.

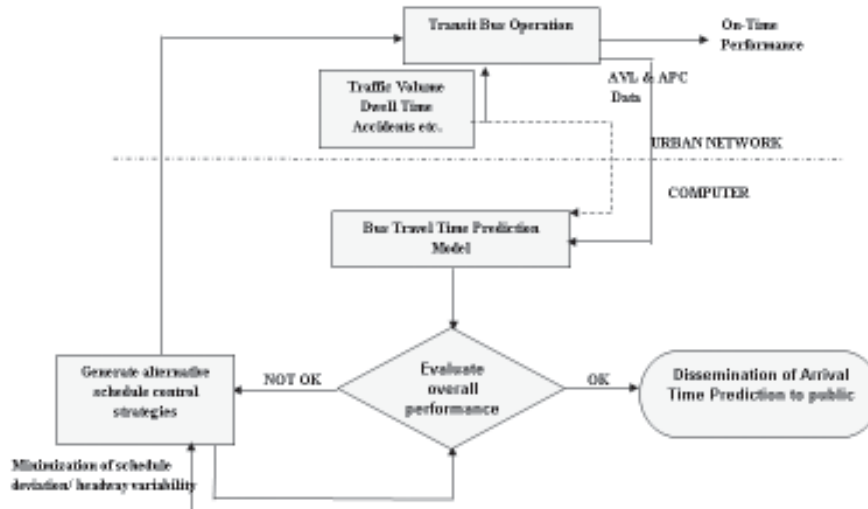
Introduction

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 Counting (APC) systems.

Several researchers (Kalaputapu and Demetsky 1995; Lin and Zeng 1999; Wall and Dailey 1999) 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 develops a dynamic bus arrival/departure 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/departure time models sensitive to the considered control strategies (mainly stop-based strategies). This research study focuses on developing a model of such characteristics.

Figure 1. Integrated Operations Control and Information-Dissemination System



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 AM and PM peak periods is 12 minutes, increasing to 30 minutes during off peak. For the duration of the study (five weekdays in May 2001), the Toronto Transit Commission (TTC) assigned 4 buses to the route, 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 minor bus stops. This study focused on modeling travel times along those links for the morning peak-hour bus operation.

Approach

As implied earlier, most of the models found in the literature (e.g., Kalaputapu and Demetsky 1995; Lin and Zeng 1999; Wall and Dailey 1999; Farhan et al. 2002) have included bus dwell times along any link in the travel time of that link (i.e., link travel time includes running 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, numbers of boarding and alighting passengers at bus stops, and bus arrival and departure times is known from the AVL and APC systems. The prediction modeling system consists of two separate algorithms, each based on the Kalman filter method. To predict the bus running time along a particular link at instant $k+1$, the first algorithm, *Link Running Time Prediction Algorithm*, 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 . The study used data for the previous three days only as this was deemed practical, given the limited historical data available for the study. Obviously, in real-world applications, the algorithm can make use of longer ranges of historical data. The second algorithm, *Passenger Arrival Rate Prediction Algorithm*, employs similar historical data of passenger arrival rate. To predict the dwell time at a particular stop, the predicted arrival rate is simply multiplied by the predicted headway (i.e., the actual arrival time of the last bus minus the predicted arrival time of the next bus) and by the passenger boarding time (assumed 2.5 seconds per 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. This is simply because bus dwell time at a stop is affected by the actual arrival time of the bus at that stop, particularly for high frequency transit routes where passengers are expected to arrive at a nearly constant rate (i.e., the later the bus, the longer the dwell time 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 stop i , the actual 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.

Figure 2. Schematic of a Bus Route with Several Stops



Assuming that bus n is currently at stop i

$$AT_{n(i+1)} = DT_{n(i)} + RT_{n(i, i+1)} \quad (1)$$

Where:

$AT_{n(i+1)}$ is the predicted arrival time of bus n at stop $i+1$

$RT_{n(i, i+1)}$ is the predicted running time between i and $i+1$ from Kalman Filter prediction algorithm

$DT_{n(i)}$ is the 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.

$$DWT_{n(i+1)} = \delta_{(i+1)} * [AT_{n(i+1)} - AT_{n-1(i+1)}] * D_{avg(i+1)} \quad (2)$$

Where:

$DWT_{n(i+1)}$ is the predicted dwell time for bus n at stop i+1

$\delta_{(i+1)}$ represents the predicted passenger arrival rate at stop i+1 from Kalman filter prediction algorithm

$AT_{n-1(i+1)}$ is the actual arrival time of the 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

$D_{avg(i+1)}$ represents average passenger boarding time at stop i+1, assumed to be 2.5 seconds/passenger.

In equation (2), the assumption is that the boarding passengers at each bus stop have a significant effect on bus dwell time at that stop, compared with alighting passengers. 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.

$$DT_{n(i+1)} = AT_{n(i+1)} + DWT_{n(i+1)} \quad (3)$$

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 stop 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 time-point stop.

Kalman-Filter Prediction Algorithms

As mentioned above, the prediction modeling system consists of two Kalman filter algorithms. In general, the Kalman filter is a linear recursive predictive update algorithm used to estimate the parameters of a process model. Starting with initial estimates, the Kalman filter allows the parameters of the model to be predicted and adjusted with each new measurement. Its ability to combine the effects of noise of both the process and measurements, in addition to its easy computational algorithms, has made it very popular in many research fields and applications, particularly in the area of autonomous and assisted navigation (for further information on Kalman filters, see Maybeck 1979).

The main assumption used in developing the Kalman filters 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 the 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 (at least one headway length).

The Kalman filter algorithm works conceptually as follows. The last three-day historical data of actual running times along a particular link 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, 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 in this fashion: The number of on-passengers at a bus stop is 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 the Kalman filter prediction algorithms.

Link Running Time Prediction Algorithm

Generally, a Kalman filter algorithm for bus link running time has the following structure (Reinhoudt and Velastin 1997):

$$g(k + 1) = \frac{e(k) + \text{VAR}[\text{data}_{\text{out}}]}{\text{VAR}(\text{data}_{\text{in}}) + \text{VAR}[\text{data}_{\text{out}}] + e(k)} \quad (4)$$

$$a(k + 1) = 1 - g(k + 1) \quad (5)$$

$$e(k + 1) = \text{VAR}[\text{data}_{\text{in}}] \mathbf{A}g(k + 1) \quad (6)$$

$$P(k + 1) = a(k+1) \mathbf{A}art(k) + g(k+1) \mathbf{A}art_1(k + 1) \quad (7)$$

where:

- g equals the filter gain
- a is the loop gain
- e represents filter error
- p equals prediction
- art(k) is actual running time of the previous bus at instant (k)
- art₁(k+1) is actual running time of the previous day at instant (k+1)
- VAR[data_{out}] equals the prediction variance
- VAR[data_{in}] is the last three days “art3(k+1), art2(k+1) and art1(k+1)” variance

The input variance $VAR[data_{in}]$ is calculated at each instant $k + 1$ using the actual running time values for the last three days: $art_1(k + 1)$, $art_2(k + 1)$ and $art_3(k + 1)$:

$$VAR[data_{in}] = VAR[art_1(k + 1), art_2(k + 1), art_3(k + 1)] \quad (8)$$

where:

- $art_1(k+1)$ is the actual running time of the bus at instant $(k+1)$ on the previous day
- $art_2(k+1)$ is the actual running time of the bus at instant $(k+1)$ two days ago
- $art_3(k+1)$ is the actual running time of the bus at instant $(k+1)$ three days ago

The definition of the variance for a random variable is:

$$VAR[X] = E[(X - E[X])^2] \quad (9)$$

$$E(X) = Avg(art) = \frac{art_1(k + 1) + art_2(k + 1) + art_3(k + 1)}{3} \quad (10)$$

Now the variance can be calculated as shown in equation (14):

$$\Delta_1 = [art_1(k+1) - avg(art)]^2 \quad (11)$$

$$\Delta_2 = [art_2(k+1) - avg(art)]^2 \quad (12)$$

$$\Delta_3 = [art_3(k+1) - avg(art)]^2 \quad (13)$$

$$VAR[data_{in}] = \frac{\Delta_1 + \Delta_2 + \Delta_3}{3} \quad (14)$$

$VAR[data_{out}]$ 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, $VAR[data_{out}]$ should be equal to $VAR[data_{in}]$ for good prediction performance (Maybeck 1979). Now, a new variance is introduced $VAR[local_{data}]$. It is equal to the variance of the input and output data.

$$VAR[local_{data}] = VAR[data_{in}] = VAR[data_{out}] \quad (15)$$

and equations (4) and (6) become:

$$g(k+1) = \frac{e(k) + VAR[local_{data}]}{e(k) + 2 \cdot VAR[local_{data}]} \quad (16)$$

$$e(k+1) = VAR[local_{data}] Ag(k+1) \quad (17)$$

Now it becomes easy to implement the actual Kalman filter algorithm to predict the bus running times along the links. The order of applying the equations should be (16), (5), (17), and (7).

Passenger Arrival Rate Prediction Algorithm

A Kalman filter algorithm was also developed to predict the passenger arrival rate using data from the APC and AVL systems. At the prediction instant $k+1$, the historical passenger arrival rates (at instant $k+1$ of the previous three days and instant k of the current day) at a particular stop are computed based on the number of corresponding boardings divided by the actual previous headway. Similar equations to those of the running time Kalman filter were developed and used to predict the passenger arrival rate at instant $k+1$.

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 model, and an artificial neural network (specifically, Time Lag Recurrent Network [TLRN]). A more detailed exposition of the models can be found elsewhere (Farhan et al. 2002). Similar to most models found in the literature, the regression and TLRN models predict individual link travel times, which include running plus dwell times. Another distinct feature of those models is their static nature, in that the model parameters are not updated with new available data. As mentioned earlier, the AVL and APC data for the study route were available for five consecutive days. The regression and TLRN models were developed using data of four days only, with the fifth day's data held out for testing.

Four testing datasets 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 where there is a demand surge on transit service, and lane-closure scenario where one lane on a specific link is blocked (e.g., due to an accident or construction work). In contrast to the hold out data and normal condition scenario, the lane closure and special event 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. Simulation of the entire corridor, calibration results and simulation of the scenarios are described elsewhere (Farhan 2002). 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 (Okutani and Stephanedes 1984). These error indices include:

Mean relative error (g_{mean}), which indicates the expected error as a fraction of the measurement

$$E_{mean} = \frac{1}{N} \sum_t \left| \frac{X_{true}(t) - X_{pred}(t)}{X_{true}(t)} \right| \quad (18)$$

Root squared relative error (\mathbf{g}_{rs}), which captures large prediction errors

$$\mathcal{E}_{rs} = \sqrt{\frac{1}{\sum_t X_{true}(t)} \sum_t \left\{ \frac{X_{true}(t) - X_{pred}(t)}{X_{true}(t)} \right\}^2 X_{true}(t)} \quad (19)$$

Maximum relative error (\mathbf{g}_{max}), which captures the maximum prediction error

$$\mathcal{E}_{max} = \max \left\{ \left| \frac{X_{true}(t) - X_{pred}(t)}{X_{true}(t)} \right| \right\} \quad (20)$$

where:

N is the number of samples (here N=50) $X_{true}(t)$ = measured value at time t

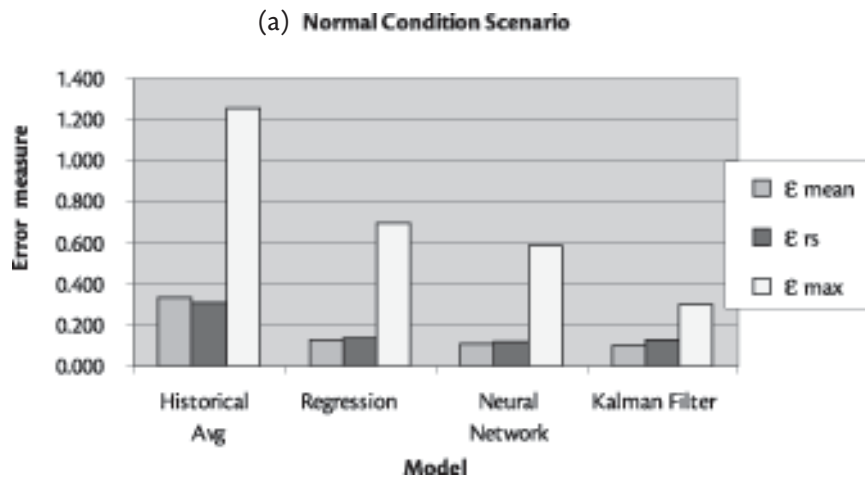
$X_{pred}(t)$ is the predicted value at time t.

Table 1. Relative Error Results of the Prediction Models Using "Hold Out" Data

Link	Error Indices	Historical Avg	Regression	NN	Kalman Filter
Link 1	E mean	0.268	0.099	0.076	0.069
	E rs	0.332	0.115	0.094	0.074
	E max	0.900	0.192	0.168	0.117
Link 2	E mean	0.537	0.134	0.063	0.028
	E rs	0.603	0.149	0.075	0.036
	E max	3.172	0.224	0.131	0.077
Link 3	E mean	0.369	0.200	0.165	0.076
	E rs	0.543	0.220	0.166	0.093
	E max	0.801	0.466	0.342	0.227
Link 4	E mean	0.130	0.079	0.071	0.087
	E rs	0.208	0.092	0.088	0.109
	E max	0.502	0.228	0.153	0.232
Link 5	E mean	0.168	0.084	0.104	0.044
	E rs	0.181	0.076	0.120	0.055
	E max	0.502	0.228	0.240	0.122

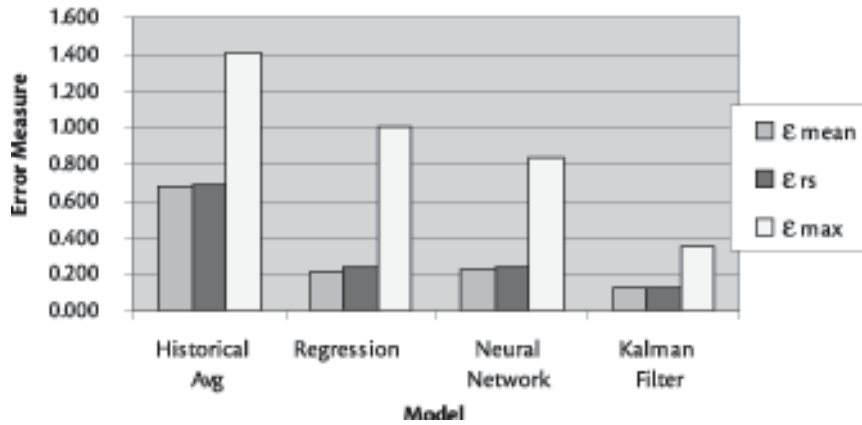
Table 1 shows the three error measures g_{mean} , g_{rs} , g_{max} for the hold out dataset, while Figure 3 (a, b, c) summarize the performance of the three prediction models for each simulated scenario. Obviously, the lower the error, the better the model performance.

Figure 3. Relative Error Results of the Prediction Models Using Artificial Data (a, b, c)



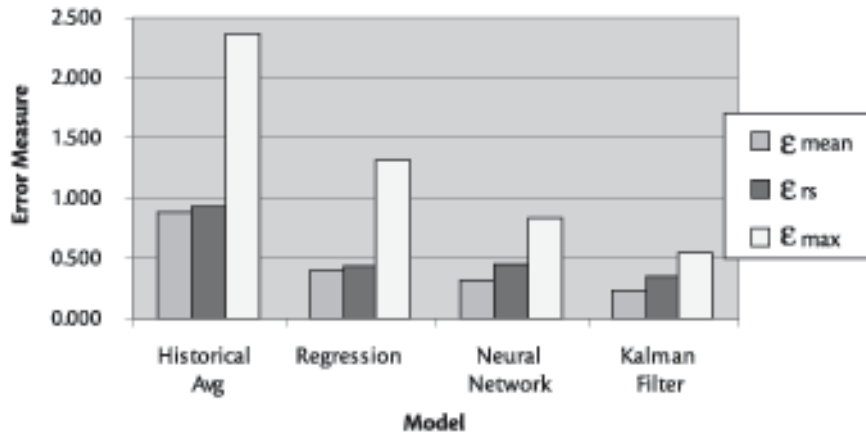
Normal Condition Scenario				
Total	Historical Avg	Regression	Neural Network	Kalman Filter
ϵ_{mean}	0.330	0.124	0.111	0.097
ϵ_{rs}	0.305	0.132	0.120	0.124
ϵ_{max}	1.253	0.695	0.584	0.299

(b) Special Event Scenario



Special Event Scenario				
Total	Historical Avg	Regression	Neural Network	Kalman Filter
ϵ_{mean}	0.679	0.218	0.220	0.123
ϵ_{rs}	0.685	0.240	0.239	0.127
ϵ_{max}	1.411	0.998	0.830	0.349

(c) Lane Closure Scenario



Lane Closure Scenario				
Total	Historical Avg	Regression	Neural Network	Kalman Filter
ϵ_{mean}	0.881	0.392	0.316	0.232
ϵ_{rs}	0.933	0.428	0.442	0.349
ϵ_{max}	2.362	1.324	0.830	0.547

Discussion

The results summarized in Table 1 show that for all links, the Kalman filter model provides the minimum value for the error measures g_{mean} , g_s , g_{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 1 and Figure 3 (a, b, c) shows 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.

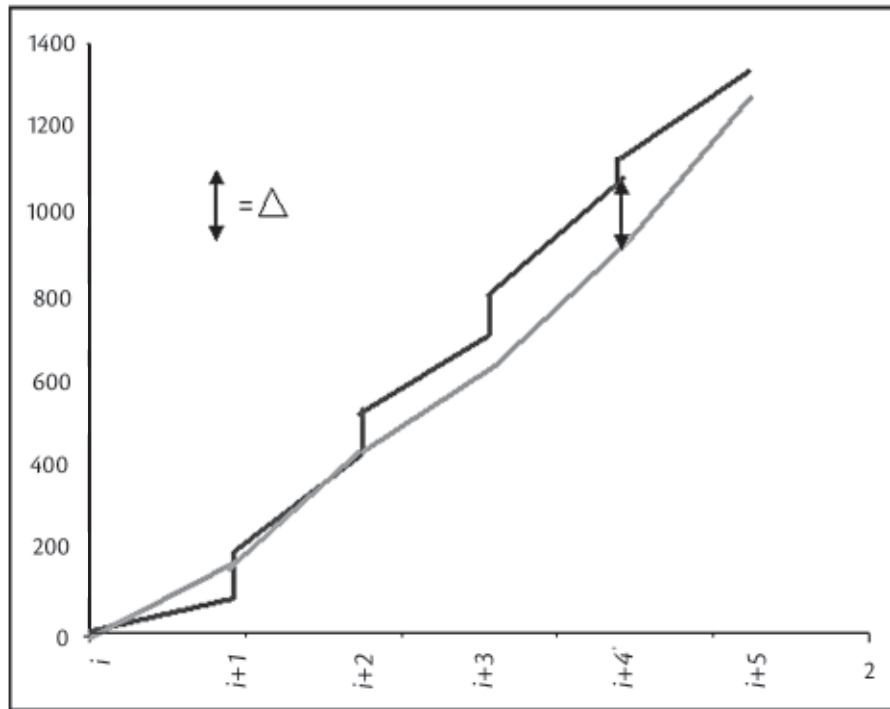
The Kalman filter model shows the best prediction performance in the simulated scenarios. In particular, it showed superior performance to the other models in the special event and lane closure scenarios. These results show the superior performance of the Kalman filter model compared with other prediction models in terms of relative error. The results also demonstrate 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 an additional one minute, the model appropriately captures the effect of this extra time on the arrival time at the next bus stop and the dwell time at the next stop (which is function of number of passengers waiting at that bus stop when the bus is predicted to arrive) and so forth for the prediction of arrival and dwell times at subsequent stops.

User-Interactive Decision Support System

The developed arrival/departure time prediction model was used to build a DSS (Decision Support System) for dynamic control. This system simply uses the timetable of the time points to develop a time profile for each scheduled trip along the route (i.e., schedule travel time profile; see 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.

Figure 4. Schedule Travel Time Profile

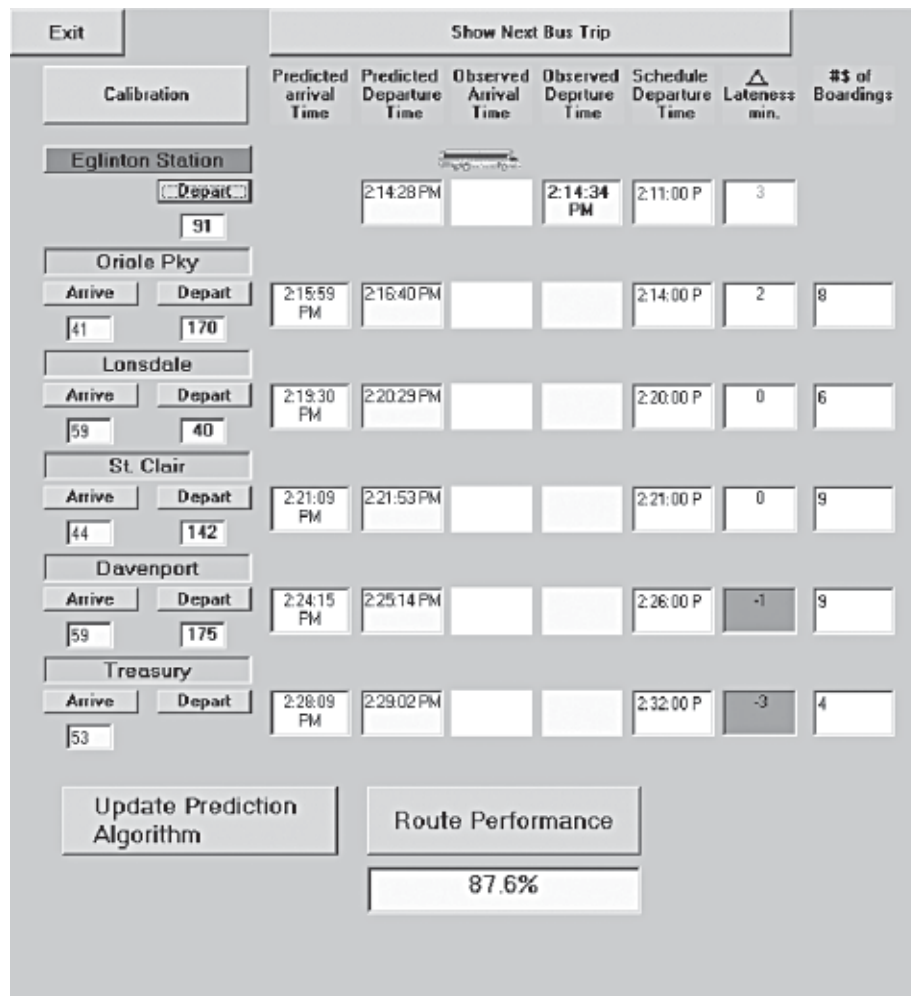


By using these two travel-time profiles, we are able to predict if the bus is running early or late at each time point. This is shown in Figure 4 as Δ . A positive value of Δ means that the prediction profile is currently lying above the scheduled one (bus will reach the downstream time points late), while a negative value of Δ 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 Δ is the key factor for deciding what type of control strategy to implement. If Δ is positive, an expediting type of control is required to be applied, while a negative value of Δ 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/departure information and performance analysis at downstream bus stops for the purpose of applying real-time, proactive control strategies.

Figure 5. Illustration of the Interactive Decision Support System



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 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 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 (Δ information) is automatically computed for all time points. If the value of the predicted bus time deviation Δ 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 Δ value labels. On the other hand, when Δ values are more than 2 minutes (i.e., bus expected to arrive/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 reduce predicted deviations. If the Δ value is less than 0 (i.e., bus expected to be ahead of schedule), then the Δ 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 interaction with the program could be done by setting values under the "arrive" or "depart" button of the selected current/downstream time point (as shown on the left hand side of Figure 5), then clicking on the corresponding "arrive" or "depart" button to update model predictions.

The prediction algorithms of the system will be dynamically updated based on the AVL and APC data. As soon as the bus arrives or departs a new time point, new arrival and departure time predictions and new Δ 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 historical database (link running time and passenger arrival rate at time points) to be used for the Kalman filter predictions of future trips.

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. The information and analysis provided by this system could possibly be used for updating and adjusting schedules.

Conclusion

An innovative model was developed for dynamic bus arrival and departure 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 two 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 “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 (e.g., because of a special event) and lane closure (e.g., because of an incident). This is primarily due to 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 DSS was developed to provide continuous information on the expected arrival and departure times 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, 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, a 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.

Acknowledgments

The research reported in this article was funded by the Geomatics for Informed Decisions (GEOIDE) network, a Network of Centres of Excellence funded by the Government of Canada.

References

- Farhan, A., A. Shalaby, and T. Sayed. 2002. Bus travel time prediction using GPS and APC. ASCE 7th International Conference on Applications of Advanced Technology in Transportation, Cambridge, Massachusetts (August).
- Farhan, Ali. Bus arrival time prediction for dynamic operations control and passenger information systems. 2002. Unpublished Thesis of Masters of Applied Science, Department of Civil Engineering, University of Toronto.
- Kalaputapu, Ravi, and Michael J. Demetsky. 1995. Modeling schedule deviations of buses using Automatic Vehicle Location data and artificial neural networks. *Transportation Research Record 1497*, 44-52.
- Lin, Wei-Hua, and Jian Zeng. 1999. Experimental study of real-time bus arrival time prediction with GPS data. *Transportation Research Record 1666*, 101-109.
- Maybeck, Peter S. 1979. *Stochastic models, estimation and control*. Vol. 1, Academic Press.
- Okutani, Iwao, and Yorgos J. Stephanedes. 1984. Dynamic prediction of traffic volume through Kalman filter. *Transportation Research Board 18B(1)*, 1-11.

Reinhoudt, Edwin M., and S. A. Velastin. 1997. A dynamic predicting algorithm for estimating bus arrival time. 8th IFAC/IFIP/IFORS Symposium, Vol. 3.,1295-8. Crete, Greece (June 16-18).

Wall, Z., and D. J. Dailey. 1999. An algorithm for predicting the arrival time of mass transit vehicle using Automatic Vehicle Location data. Transportation Research Board Paper No. 990870. 78th Annual Meeting, Washington, D.C.

About the Authors

AMER SHALABY (*amer@ecf.utoronto.ca*) received his master and Ph.D. degrees from the Department of Civil Engineering, University of Toronto (U of T) in 1991 and 1997, respectively. He was a postdoctoral fellow at U of T's Joint Program of Transportation in 1996-1997 and an NSERC Industrial Research Fellow at IBI Group in 1997-1998. Between 1998 and 2000, he was an assistant professor at Ryerson University, where he also directed the Vehicle Safety Research Centre. Since January 2001, Dr. Shalaby has been an assistant professor at the University of Toronto. He is specialized in the area of Intelligent Transportation Systems, with emphasis on transit applications. His research interests also include transportation system planning and road safety.

ALI FARHAN (*alimfarhan@hotmail.com*) received his master's degree from the Department of Civil Engineering, University of Toronto (U of T) in 2002. Subsequently, he worked as a research associate at the Joint Program in Transportation at the University of Toronto, then as a transportation engineer with Maricopa Association of Governments, Phoenix, Arizona. Currently, Mr. Farhan works for the City of Calgary.