2004 IEEE Intelligent Transportation Systems Conference
Washington, D.C., USA, October 3-6, 2004

Bus Arrival Time Prediction Using Artificial Neural Network Model

Ranhee Jeong, and Laurence R. Rilett

Abstract— **A major component of ATIS is travel time information. The provision of timely and accurate transit travel time information is important because it attracts additional ridership and increases the satisfaction of transit users. The objectives of this research are to develop and apply a model to predict bus arrival time using Automatic Vehicle Location (AVL) data. In this research, the travel time prediction model considered schedule adherence and dwell times. Actual AVL data from a bus route located in Houston, Texas was used as a test bed. A Historical data based model, Regression Models, and Artificial Neural Network (ANN) Models were used to predict bus arrival time. It was found that ANN Models outperformed the historical data based model and the regression models in terms of prediction accuracy.**

I. Introduction

ONE component of Intelligent Transportation Systems (ITS) is Advanced Traveler Information Systems (ATIS) and a major component of ATIS is travel time information. The provision of timely and accurate transit travel time information is important because it attracts additional ridership and increases the satisfaction of transit users. In addition, transit operators can identify vehicles that fall behind schedule and react in a proactive way.

Automatic Vehicle Location (AVL) Systems, which is a part of ITS, have been adopted by many transit agencies and allows them to track their transit vehicles in real-time. While the provision of real-time information, such as bus location, is relatively straightforward, forecasting transit information, such as when a bus will arrive at a particular location, is significantly more complex. Because of the complexity of the inputs there is a definite requirement for a robust statistical approach. While there has been some preliminary work in this area it is still not clear which statistical method would work best.

In order to predict travel time, in an accurate and timely manner, the consideration of traffic condition is essential, including traffic congestion, dwell time at stops, etc. However, previous research did not explicitly consider traffic congestion and/or dwell time at stops. The objectives of this research are to develop and apply a model to predict bus arrival time using AVL data. In this research, therefore, the arrival time prediction model will consider traffic congestion and dwell times. It is anticipated that this forecast information can be provided to help travelers with their trip decision making.

Manuscript received July 30, 2004.

R. Jeong is with the Texas Transportation Institute, College Station, TX 77840 USA (corresponding author to provide phone: 979-845-1599; fax: 979-845-6481; e-mail: r-jeong@ ttimail.tamu.edu).

L. R. Rilett, was with the Department of Civil Engineering, Texas A&M University, College Station, TX 77840 USA. He is now with the Department of Civil Engineering, University of Nebraska-Lincoln, P. O. Box 880531, Lincoln, NE 68588-0531 USA (e-mail: lrilett2@unl.edu).

II. Background

The accurate prediction of link travel time is critical to ITS transit applications. With the development of Advanced Travelers Information Systems (ATIS) the importance of the short-term travel time prediction has increased markedly. A number of prediction models, including historical data based models, regression models, time series models and neural network models, have been developed over the years by various transit agencies.

Recently, there have been some studies conducted on predicting transit travel time using AVL data. Wall and Dailey used a combination of both AVL data and historical data to predict bus arrival time in Seattle, Washington [1]. They used a Kalman filter model to track a vehicle location and to predict bus travel time. They did not explicitly deal with dwell time as an independent variable.

Chien et al developed an artificial neural network model to predict dynamic bus arrival time [2]. They stated that the back-propagation algorithm, which is the most used algorithm for transportation problems, is hard to apply on-line due to the lengthy learning process. Consequently, they developed an adjustment factor to modify travel time prediction with new input of real-time data. They used generated data to predict bus arrival time, and they did not consider dwell time and scheduled data. They used simulated data from CORSIM including volume and passenger demand.

Shalaby and Farhan developed bus travel time prediction model using Kalman filtering technique [3]. They used downtown Toronto data collected with 4 buses equipped AVL and APC. They insisted that Kalman filtering techniques outperformed historical model, regression mode, and time lag recurrent neural network model. They used 5 weekday data in May 2001. A week data would be cyclical and with these data time series model and historical model would predict well. However, larger data including more variability in data set should need a model to explain the variability and uncertainty of the data. They developed two Kalman filtering algorithm to predict running times and dwell times separately. However, when they developed historical average model, regression model, and time leg recurrent neural network model, they include dwell time in link travel time. They defined that a link is between two time check point stops and the link include 2 to 8 bus stops. Consequently, they predicted dwell time only at time check point.

In summary, prediction model considering traffic congestion and dwell time at bus stops are required in urban congested areas.

III. OBJECTIVES AND SCOPE

The objectives of this research are to develop and apply a model to predict bus arrival time using AVL data. The predicted arrival time can be provided to travelers to help in their decision making and can be used by transit operators to help their operations. This research focused on forecasting bus arrival time based on traffic congestion and dwell times at stops. Different model formulations including a historical data based model, regression models, and an artificial neural network model were used. The measure of effectiveness used to qualify accuracy is the difference (mean absolute percentage error) between the observed and predicted arrival times.

IV. CASE STUDY

AVL data collected in Houston, Texas was used for the test bed. The Houston data was collected by Houston Metro buses equipped with DGPS receiver at 5 second intervals. The data was collected over 6 months in 2000 (from June to November). The test bed was Route 60 which is highly congested in the morning and afternoon peaks and only the southbound direction was studied. This DGPS provides time, speed, heading, etc as well as bus location.

There are two test bed sites; a downtown area corridor and a north area corridor. The first corridor has 9 bus stops and is 1.6 kilometers long. Stop 1 and 9 are used as time check point for schedule adherence. The second corridor has 25 bus stops and is 4.26 kilometers long. Stop 6 and 20 are used as time check point. The schedule headway during the weekday peak period is 30 minutes and during the weekday non peak period and weekends is one hour. 340 buses data were used. A total of 240 buses were used for calibrating models and 100 buses data were used for evaluating models.

V. MODEL DEVELOPMENT

A number of modeling techniques were used in this study including a simple statistical model (historical data), a regression model, and an artificial neural network model. The input variables are arrival time, dwell time, and schedule adherence at each stop. In order to consider traffic congestion, the schedule adherence was calculated by subtracting the scheduled arrival time from the actual arrival time. A positive value of schedule adherence means that bus was delayed at the stop while a negative value means that the bus arrived early. To consider traffic congestion, the link travel time was clustered by the time period. The output variable is arrival time at each stop.

The transit schedule and congestion for the weekday peak hour, the non peak hour, the evening, and the weekend are different. Intuitively it would be expected that dwell time and link travel time would also be different. To account for these differences, data was clustered by the time of the week and the time of the day. In this research, weekday means Monday through Friday, and weekend means Saturday and Sunday. Weekday peak period data included the bus data arrived at the first bus stop during 6:15 A.M. ~ 8:15 A.M. and 4:15 P.M. ~ 6:10 P.M. Weekday non peak period data included the bus data arrived before 6:15 A.M., 8:15 A.M. ~ 4:15 P.M., and 6:10 P.M. ~ 7:15 P.M. Weekday evening period data include the bus data arrived after 7:15 P.M. Fig. 1 shows the pattern of arrival time, dwell time and schedule adherence by the different time periods. Not surprisingly these variables are a function of time of day and there is a wide variation in values. In general, the variability of downtown data is larger that that of north area data.

It is unexpected that the variability of dwell time is larger than that of arrival time. That means the influence of dwell time on predicted bus arrival time would be greater than that of arrival time. In downtown area, stop 1 and stop 9 are time check points. However, drivers stayed longer at stop 1 and stop 5. This could result from various reasons, including more demand at stop 5, intersection delay, and the driver staying longer at these locations to stay on schedule. Our expectation is that bus drivers will stay longer at time check point if they arrive early. However, the pattern of dwell time in Fig. 1 shows that drivers tend to stay when they have passenger demand or intersection delays. In other words, they stay when and where they stopped rather than they stay at time check point to keep the schedule. Therefore, we can not assume that bus will wait at time check point when the bus arrives early.

In the downtown area, the test bed for this research has 9 bus stops. Consequently, eight separate models were developed for each of the three techniques analyzed. For example, model 1 uses the input data from stop 1 (i.e. arrival time, dwell time, and schedule adherence) and

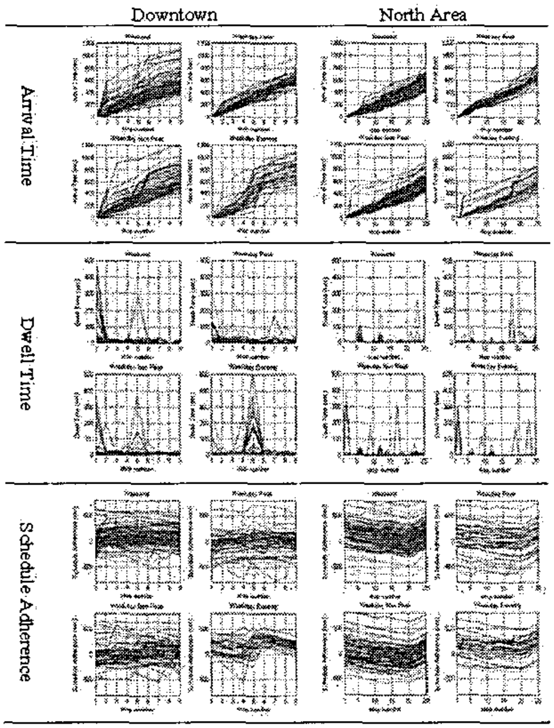

Fig 1. Arrival Time, Dwell time, and Schedule Adherence by Time Period.

predicts the arrival times at stops 2 through 9. In contrast, model 5 uses the input data from the first five stops and predicts arrival times at stops 6 through 9. In the north area, there are 25 bus stops and consequently 24 models developed using the same approach.

A. Historical Data Based Model

The historical data based model is shown in (1) and (2). First the link travel time between transit stops is calculated. It can be seen in (1) that this is a function of the difference in the average time of arrival of the downstream stops and the average departure time (i.e. arrival time + dwell time) at the upstream stop. Subsequently a recursive formula is used to predict the arrival time at the remaining stops as shown in (2). Link travel time does include stopped delay at intersections but does not include dwell times. The arrival time calculations are done only at transit stops and only when the bus first arrive a given stop. These constraints could be generalized but they were useful for limiting the number of models that were calibrated in this study.

$$T_{it} = A_{i+1,t} - (A_{it} + D_{it}),\ \forall_i = M,N,\ \forall_y = 1,T \quad (1)$$

where,

T_{it} = Estimated Link travel time from stop i to stop i+1 for bus departing time period t;

A_{it} = Average arrival time at stop i for bus departing during time period t;

D_{it} =Average dwell time at stop i for bus departing during time period t;

$A_{i+1,t}$ =Average arrival time at stop i+1 for bus departing during time period t;

T =Number of time period. For the test bed, this is equal to 4, weekend, weekday peak, weekday off-peak, and weekday evening;

M =Current bus stop. i.e. from 1 to N-1

N =Last bus stop. For the downtown test bed, this is equal to 9, and for the north area, this is equal to 25.

$$A_{jtk} = A_{Mtk} + \sum_{i=M}^{N-1} T_{it} + \sum_{i=M}^{N-1} D_{it} \qquad (2)$$
$$\forall_j = M,N, \quad \forall_t = 1,T$$

where,

A_{jtk} =Forecast arrival time at stop j for bus departing during time period t for bus k

A_{Mtk} =Observed arrival time at current stop M for bus departing during time period t for bus k

B. Regression Models

Five multiple linear regression specifications shown in (3) through (7) were tested in this research after analyzing stepwise regression and correlation coefficient. The distance from stop M to each of the subsequent stops i, bus schedule adherence, and dwell time were used as independent variables. The correlation coefficient of the three independent variables was all less than 0.15 and these independent variables can be used for regression models. However, dwell time waw not significant statistically and it was not used to develop regression models. Consequently, distance and schedule adherence were used as independent variables and arrival time at each stops were used as dependent variable. The five model specifications are as follows:

Reg. 1: $T_{Mik} = b_0 + b_1 D_{Mk}$ (3)

Reg. 2: $T_{Mik} = b_0 + b_1 D_{Mk}^{\ 2}$ (4)

Reg. 3: $T_{Mik} = b_0 + b_1 D_{Mk}^{\ 2} + b_2 S_{Mk}$ (5)

Reg. 4: $T_{Mik} = b_0 + b_1 D_{Mk}^{\ 2} + b_2 S_{Mk}^{\ 2}$ (6)

Reg. 5: $T_{Mik} = b_0 + b_1 D_{Mk} + b_2 D_{Mk}^{\ 2} + b_3 S_{Mk}$ (7)

where,

T_{Mik} =Travel time from current stop M to stop i for bus k, i = M, N

D_{Mk} =Distance from stop M to stop i for bus k

S_{Mk} =Bus schedule adherence at stop M for bus k. Equal to observed arrival time at current stop M (A_{Mk}) minus the schedule arrival time.

$$A_{ik} = A_{Mk} + T_{Mik} \qquad (8)$$

A_{ik} =Arrival time of bus k at node i

C. Artificial Neural Network Models

ANNs emulate the learning process of human brain. They are good at pattern recognition, prediction, classification, etc. ANNs are calibrated using two steps, training and testing. During the training stage, the ANN uses inductive learning principles to learn from a training data set. There are two types of learning techniques used in ANN development: unsupervised and supervised. In unsupervised learning, the network attempts to classify the training set data into different groups based on input patterns. In supervised learning, the desired output from output layer neurons is known, and the network adjust weight of connections between neurons to produce the desired output. During this process, the error in the output is propagated back from one layer to the previous layer by adjusting weights of the connections [8]. This is called the back-propagation method, which is the most frequently used techniques in transportation application. The learning process of ANNs can be continuous so that the models can adapt to the change of environmental characteristics. Therefore ANN models can be considered dynamic prediction models because they can be updated and/or modified using new online data [2].

The ANN architecture used in this paper had three layers: an input layer, a hidden layer, and an output layer. The weights and parameters associated with the hidden layer were identified during the calibration process. The optimal values were based on minimizing the prediction error. Fifteen different number of hidden neuron, from one to fifteen inclusive, were tested. Fig. 2 shows the MAPE by the different number of hidden neuron. After ANN models were tested with the fifteen different neurons, the best number of neuron was selected for each ANN models. It means that the downtown area could have different number of neuron than that of the north area and that the weekend time period could have difference number of neuron than that of weekday peak time period. However, the prediction results from the fifteen different neurons were not significantly different from each other.

The backpropagation algorithm and the Hyperbolic Tangent Sigmoid transfer function were used in the model

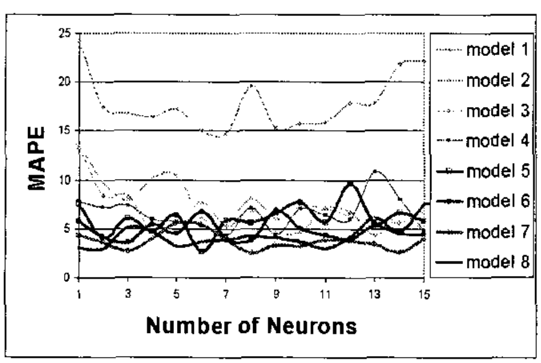

Fig. 2. MAPE by the different hidden neuron

TABLE I
AVERAGE MAPE BY TRAINING FUNCTIONS

Functions	Average MAPE by Time period				
	NC	END	PK	NPK	EVN
Batch training with weight and bias learning rule	109.46	134.35	146.59	136.10	118.91
BFGS quasi-Newton backpropagation	12.28	11.11	9.80	12.96	11.19
Bayesian regularization	5.09	4.12	4.82	7.88	5.08
Powell-Beale conjugate gradient backpropagation	11.64	11.55	9.99	13.84	10.38
Fletcher-Powell conjugate gradient backpropagation	12.94	11.49	9.29	13.92	10.60
Gradient descent backpropagation	146.68	156.29	163.00	145.96	144.74
Gradient descent with adaptive learning rate backpropagation	27.95	29.02	25.79	26.53	26.65
Levenberg-Marquardt backpropagation	5.18	4.28	5.33	8.65	7.18
One step secant backpropagations	14.83	13.17	12.16	15.73	13.34
Random order incremental update	19.86	18.85	17.06	22.07	25.65
Resilient backpropagation	12.66	11.92	10.26	15.24	12.74
Sequential order incremental update	294.74	309.56	301.65	265.92	185.52
Scaled conjugate gradient backpropagation	13.17	11.96	10.67	13.96	11.77

NC: Non Clustering, END: Weekend, PK: Weekday peak, NPK: Weekday non peak, EVN: Weekday Evening

development. After testing thirteen different training functions, the Levenberg-Marquardt optimization algorithm was choose for training function. Table I shows the average MAPE by these thirteen different training functions. The average MAPE of the Bayesian Regularization training function is slightly less than that of the Levenberg-Marquardt Backpropagation training function. However, the running time of the Bayesian Regularization training function was far more than that of the Levenberg-Marquardt Backpropagation training function. Therefore, the Levenberg-Marquardt Backpropagation training function is chosen as the best training function for this research in terms of efficiency and accuracy.

After testing fourteen different learning functions, a Perceptron Weight and Bias learning function was used for learning function. Table II shows the average MAPE by these fourteen different learning functions. The results from these fourteen different learning functions were not significantly different. A complete description of the ANN training and testing process may be found elsewhere [4]. Similar to the previous techniques arrival time, dwell time and schedule adherence were used as input as shown in (9).

$$T_{Mik} = f(A_{Mk}, W_{Mk}, S_{Mk}) \tag{9}$$

where,

T_{Mik} =Travel time from current stop M to stop i for bus k, i = M, N;

TABLE II
AVERAGE MAPE BY LEARNING FUNCTIONS

Functions	Average MAPE by Time period				
	NC	END	PK	NPK	EVN
Conscience Bias Learning Function	5.28	4.14	5.91	8.05	7.57
Gradient Descent Weight/bias Learning Function	5.19	4.03	5.26	8.61	7.33
Gradient Descent with Momentum Weight/bias Learning Function	5.16	5.20	5.20	8.37	8.24
Hebb Weight Learning Function	5.17	4.55	5.48	8.51	7.50
Hebb with Decay Weight Learning Function	4.72	4.25	5.15	8.66	8.00
Instar Weight Learning Function	5.33	4.32	5.63	8.97	7.66
Kohonen Weight Learning Function	5.27	3.85	5.16	8.58	7.82
LVQ1 Weight Learning Function	5.18	4.28	5.33	8.65	7.18
LVQ2 Weight Learning Function	5.31	4.33	5.30	8.79	7.77
Outstar Weight Learning Function	5.19	4.21	5.18	8.62	7.60
Perceptron Weight and Bias Learning Function	5.13	4.31	5.04	7.87	7.17
Normalized Perceptron Weight and Bias Learning Function	5.13	4.70	5.22	8.16	8.61
Self-organizing Map Weight Learning Function	4.98	4.25	5.30	8.26	7.42
Widrow-Hoff Weight and Bias Learning Rule	5.04	4.55	5.48	8.57	8.65

NC: Non Clustering, END: Weekend, PK: Weekday peak, NPK: Weekday non peak, EVN: Weekday Evening

A_{Mk} =Arrival time of bus k at node M;
W_{Mk} =Dwell Time of bus k at stop M ; and
S_{Mk} =Bus schedule adherence at stop M for bus k. Equal to observed arrival time at current stop M (A_{Mk}) minus the schedule arrival time.

VI. MODEL EVALUATION

All three model architectures were calibrated using the training and testing data sets. The Mean Absolute Percentage Error (MAPE) was used as the measure of effectiveness (MOE) in this paper. The MAPE is shown in (1). It basically represents the average percentage difference between the observed value (in this case arrival time at a transit stop) and the predicted value (in this case predicted arrival time at a transit stop).

$$MAPE = \frac{1}{n}\sum_{i}^{n}\frac{|y_i - y_o|}{y_o} \times 100\% \qquad (10)$$

where,
y_i =Predicted value (i.e. arrival time at given transit stop)
y_o =Observed value (i.e. arrival time at given transit stop)

Table III, fig. 3, and fig. 4 show the average MAPE for five time periods of three prediction models. It can also be seen that the clustering the data leads to a smaller MAPE in the historical data based model and regression models. This would not be unexpected because the clustering explicitly accounts for different congestion and demand levels associated with different parts of the day. In contrast to the previous two techniques the clustering resulted in poorer results than the non-clustering option in the artificial neural network models. It is hypothesized that the ANN, as a universal function approximator, was able to identify the non-linear relationships associated with the different clusters. While in general the clustering should not do worse than the non-clustering option it is hypothesized that there may not have been enough observations to adequately fit the functions. If more observations were available the results between the two approaches might have been more similar.

Interestingly, the lowest MAPE of historical model of downtown area was for the weekday peak. It is hypothesized that the congestion reduces the variability in travel times and this makes the historic model more accurate for this time period. Interestingly, the use of real time schedule adherence data did not improve much the results. It is hypothesized that there is a non-linear relationship between arrival time and schedule adherence and this caused the relatively poor results. In addition, the various non-linear model specifications were unable to capture this phenomenon. For this test bed the historic model gave superior results, in terms of MAPE, in comparison to the MLR results. The most important point to note is that the ANN had the lowest MAPE as compared to the historic model and the MLR model. On average the ANN models had a 54.24 percent improvement in downtown area and 48.61 percent in north area improvement with respect to the best historic model and a 71.01 percent improvement in downtown area and 76.53 percent in north area compared to the best MLR models. It is hypothesized that the use of historic data (representing congestion) coupled with the real time schedule adherence data (representing real-time congestion and demand inputs) resulted in the better performance of the ANN model.

TABLE III
MAPE OF PREDICTION MODELS

Models	site	NC	END	PK	NPK	EVN
Historical data based models	D	14.63	13.03	10.77	14.34	12.82
	N	7.58	7.27	7.02	8.74	6.43
Regression Models	D	Reg. 1	Reg. 1	Reg. 5	Reg. 1	Reg. 1
		24.85	22.81	14.88	23.48	19.97
	N	Reg. 5	Reg. 5	Reg. 1	Reg. 5	Reg. 1
		15.25	14.54	11.57	17.56	13.92
ANN models	D	5.07	4.13	5.20	8.36	7.17
	N	2.88	1.96	4.58	4.46	4.87

NC: Non Clustering, END: Weekend, PK: Weekday peak, NPK: Weekday non peak, EVN: Weekday Evening
D: Downtown, N : North Area

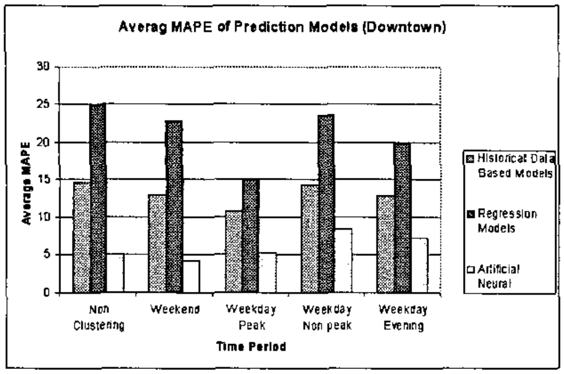

Fig 3. Average MAPE of prediction models for the downtown area

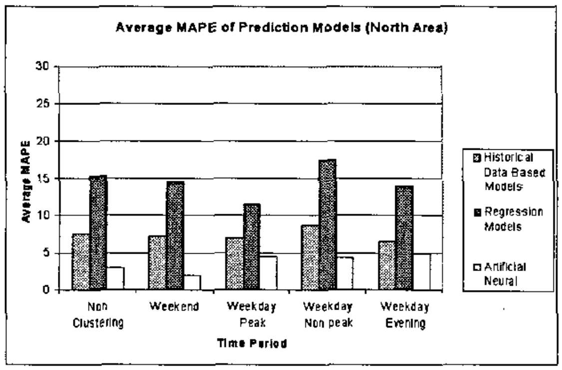

Fig 4. Average MAPE of prediction models for the north area

CONCLUSION REMARKS

This paper described the results of three bus travel time prediction algorithms which were calibrated and tested on a transit route in Houston, Texas. The input to the models consisted of historic data (i.e. link travel time and dwell time) and real-time schedule adherence data. It was found that the Artificial Neural Network models (used without clustering of the data) performed considerably better than either a historic data based model or MLR models. It was hypothesized that the ANN was able to identify the complex non-linear relationship between travel time and the independent variables and this led to the superior results.

While the results are encouraging there are still a number of extensions to the model that should be studied. It is hypothesized that if other real time data were available, such as the demand at any given bus stop, traffic congestion measures, incident information etc. the arrival time prediction could be improved. Note that this type of information is typically unavailable on urban street networks. However, as new ITS data collection techniques, such as cell phone monitoring, improve this will not be an impediment. It is hypothesized that universal function approximator models, such as ANN, would work best for this problem. However, these models require some effort to calibrate and the best model for a given situation would have to be determined on a case by case basis. A process that would in "self calibrating" would be extremely useful for these types of applications.

The ability to provide accurate and timely travel time information would be very useful to transit patrons as well as transit authorities. Because variability in travel time (both waiting and on-board) is extremely important for transit choice it would also be useful to extend the model to provide not only estimates of the travel time but also confidence intervals. For the ANN model this could be accomplished by using a bootstrap technique.

REFERENCES

[1] Z. Wall, and D. J. Dailey, "An algorithm for predicting the arrival time of mass transit vehicles using automatic vehicle location data", *TRB 78th Annual Meeting (CD-ROM)*, Washington DC, January 1999.

[2] S. I. J. Chien, Y. Ding, and C. Wei, "Dynamic bus arrival time prediction with artificial neural networks", *Journal of Transportation Engineering*, September/October 2002, p.p.429-438, 2002.

[3] A. Shalaby, and A. Farhan, "Bus travel time prediction for dynamic operations control and passenger information systems", *Transportation Research Board, the 82nd Annual Meeting (CD-ROM)*, Washington DC, January 2003.

[4] M. Hagan, H. Demuth, and M. Beale, Neural Network Design, PWS, Boston, 1996.