Dynamic Travel Time Prediction Models for Buses Using Only GPS Data

Wei Fan[1] **Zegeye Gurmu**[2]

1 Associate Professor, Department of Civil and Environmental Engineering, The University of North Carolina at Charlotte, EPIC Building, Room 3261, 9201 University City Boulevard, Charlotte, NC 28223. Email: wfan7@uncc.edu

2 Traffic Engineer - Teague Nall & Perkins, Inc., 1100 Macon Street, Fort Worth, Texas 76102. Email: zgurmu@tnpinc.com

Abstract

Providing real-time and accurate travel time information of transit vehicles can be very helpful as it assists passengers in planning their trips to minimize waiting times. The purpose of this research is to develop and compare dynamic travel time prediction models which can provide accurate prediction of bus travel time in order to give real-time information at a given downstream bus stop using only global positioning system (GPS) data. Historical Average (HA), Kalman Filtering (KF) and Artificial Neural Network (ANN) models are considered and developed in this paper. A case has been studied by making use of the three models. Promising results are obtained from the case study, indicating that the models can be used to implement an Advanced Public Transport System. The implementation of this system could assist transit operators in improving the reliability of bus services, thus attracting more travelers to transit vehicles and helping relieve congestion. The performances of the three models were assessed and compared with each other under two criteria: overall prediction accuracy and robustness. It was shown that the ANN outperformed the other two models in both aspects. In conclusion, it is shown that bus travel time information can be reasonably provided using only arrival and departure time information at stops even in the absence of traffic-stream data.

1. INTRODUCTION

Traffic congestion is increasing with alarming rapidity and continually posing threat to the quality of life of people in many countries all over the world over the past few decades. It increases travel time, air pollution and fuel use, and decreases accessibility and mobility. In order to alleviate congestion problem, different techniques have been suggested during the past years, including demand-side (congestion pricing, traffic management, etc.) and supply-side (constructing more roads, adding lanes, etc.) or their integration. Expanding public transportation service has been reportedly considered as one of the many promisingly effective improvement techniques on the supply-side [1, 2]. For example, travelers can be provided with reliable travel information through the help of Advanced Public Transport System (APTS), which is one component of Intelligent Transportation Systems (ITS) [1, 3]. Since travel time information is one of the most preferred information by travelers, its timely and accurate provision is vital in order to attract more people to public transport and increase the contentment of transit users [4]. However, real-time travel time information cannot be easily measured and made available directly. Therefore, mathematical models that can predict travel time with reasonable accuracy are strongly needed. A variety of models for predicting traffic states such as travel time and traffic flow have been developed over the years. The most widely used bus arrival time prediction models can be broadly classified into four categories and are discussed below.

1) *Historical Average Models* [4, 5, 6]. These models give the current and future travel time from observed historical bus travel time data of previous journeys by assuming that the current

traffic condition is to remain stationary. Historical average models could be suitable for real-time dynamic travel time information as the algorithms are usually simple and require relatively small computation time. However, the performance of the models are weak unless the traffic pattern in the area of interest is relatively stable over time or where congestion is minimal, e.g., rural areas.

2) *Regression Models* [4, 6, 7, 8, 9, 10]. The regression models require a linear mathematical function to explain a dependent variable with a set of independent variables. Unlike the previous models, these are able to work satisfactorily even if traffic conditions are not stable. They usually measure the simultaneous impact of various factors, which are independent between one and another, affecting the dependent variable. Regression models have been used by many authors in bus travel time prediction. For example, Patnaik et al. [7] developed a set of multiple linear regression models to estimate bus arrival times using distance, number of stops, dwell times, boarding and alighting passengers and weather descriptors as independent variables. Their study showed that the models could be used to estimate bus arrival time at downstream stops. Jeong and Rilett [4] and Ramakrishna et al. [6] also developed multiple linear regression models using different sets of independent variables. In both studies, the regression models were outperformed by other type of models. However, these models have a relative advantage in revealing which independent variables are less or more important for predicting travel times. For example, Patnaik et al. [7] discussed that weather was not an important input in their model. Ramakrishna et al. [6] also found out that two variables, i.e. bus stop dwell times from the origin of the route to the current bus stop in minutes and intersection delays from the origin of the route to the current bus stop in minutes, are less important in predicting bus travel time. Because variables in transportation systems are inter-correlated, the applicability of the regression models is in general limited [8].

3) *Kalman Filtering Model* [8, 10]. Kalman filtering models could be used to predict the future state of the dependent variable. They have elegant mathematical representations (e.g. linear state-space equation) which can adequately accommodate traffic fluctuations with their time-dependent parameters (e.g. Kalman gain) [8]. These models have also been used by many authors in bus travel time prediction [3, 8, 9, 10]. The basic function here is to provide estimates of the current state of the system from previous time steps. They can also serve as the basis for predicting future values or improving estimates of variables at earlier times because of their capacity to filter noise [11].

4) *Machine Learning Models* such as Artificial Neural Networks (ANN) [4, 9]. These models are able to deal with complex and noise data and are suitable to find nonlinear relationships between dependent variable and independent variables. They can be used for prediction purpose, without explicitly specifying the (physical) traffic processes. ANN methods fall under this category. ANNs, inspired by biological neural networks, are constructed with multiple layers of processing units, known as artificial neurons. The neurons contain activation functions which are highly interconnected with one another by synaptic weights. Through learning process, the synaptic weights are adjusted to map the input-output relationship for the analyzed system automatically [12]. ANNs have recently gained popularity in predicting bus arrival time because of their ability to solve complex non-linear relationships as have been seen in many research efforts [4, 6, 8, 9]. However, their disadvantage is that results obtained using these models for one location may not be transferrable to the next because of location-specific circumstances (geometry, traffic control, etc.).

Most research efforts directed towards bus travel time prediction make use of data on the stated traffic-stream variables. For example, ANNs developed by different researchers in predicting bus travel time so far used explanatory variables such as flow, speed, weather, distance etc. as inputs. The same is true with other prediction models such as Regression and Kalman Filtering. However, with limited resources and budget, there is a strong need to make predictions in the absence of traffic stream information. Probably most relevant to this paper, Gurmu and Fan [13] used ANN to predict bust travel time dynamically using only GPS data. However, Kalman Filtering models were not considered in their paper. The purpose of this paper is therefore to explicitly consider arrival and departure time information at bus stops collected only via GPS technology to predict bus travel time dynamically in the absence of the aforementioned variables. In this paper, ANNs and Kalman Filtering are explored as potential models to give dynamic information on travel times for buses. Historical average is also considered as a baseline prediction model and used for comparison purpose.

The reminder of this paper is organized as follows: Section 2 discusses the proposed prediction models. A brief description of the study area is provided in Section 3. Section 4 presents comprehensive numerical results and model comparisons. Finally, section 5 concludes this paper with a summary and discussion of future research directions.

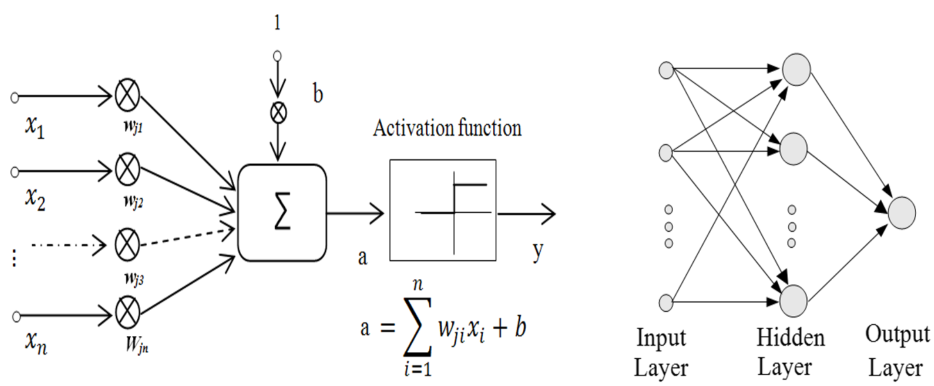
2. THE PREDICTION MODELS

2.1. ANN

ANNs learn from examples and capture subtle functional relationships among the data even if the underlying relationships are unknown or hard to explain. Hence, this unique characteristic can make ANN model an effective application in this paper because non-linear correlations between travel times can be captured to predict bus travel time at the downstream bus stops. ANN model is trained offline and yet can be used to provide real-time travel time information. To achieve this objective, care needs to be given in selecting a unique set input-output combination for the prediction purpose. However, in order to gain the maximum benefit from neural network, there should be enough data or observations [14].

2.1.1. The Network Architecture

A connected multilayer perceptron (MLP), the most popular neural network architecture, is chosen in this study. It has the potential to approximate almost any function if there are enough neurons in the hidden layers, i.e. it has a very good capability of arbitrary input-output matching (Houghton et al. 2009). Not to mention that it is also easy to implement. This ANN architecture typically consists of units arranged in layers. Each layer is composed of nodes and if the fully connected network is considered, each node connects to every node in subsequent layer. Each MLP comprises a minimum of three layers: an input layer, one or more hidden layer, and an output layer. The input layer receives external information and distributes inputs to subsequent layers. The output layer is where problem solution is obtained. The actual processing in the MLP network occurs in the nodes of the hidden and output layers. The data vector is first fed into the network at the input layer, and it then feeds into the hidden layer which in turn feeds into the last layer, i.e. output layer. The connections are typically formed by connecting each of the nodes in a given layer to all of the neurons in the layers after. It is in the hidden layer where the weight of these connections and bias parameter are generated during the training process. In this paper, a single hidden layer is used as it has been proved to be sufficient for ANNs to approximate any nonlinear functions [14]. A suitable number of nodes in the hidden layer are usually determined by experiment during the network learning process. A fully connected MLP with a single hidden layer is presented on Figure 1.



1.a. Neuron 1.b. FeedForward Network

Figure 1. Multilayer Perceptron Neural Network Architecture

2.1.2. ANN Model Development

The basic training procedure of ANNs is normally the same and the implementation of the algorithm is uncomplicated. The accuracy of the result is, however, greatly dependent upon the selection of a specific input/output combinations. The input variables need to be presented in such a way that the function signal appearing at the output of neuron j is computed as:

$$Y_j = \psi(X_1, X_2, X_3, ..., X_n) \tag{1}$$

This can be written formally as:

$$Y_j = \psi_j(\sum_{i=1}^{m} w_{ji} X_i + b_j) \tag{2}$$

Where

m is number of inputs applied to neuron j
$\{X_i\}$ is set of input variables of neuron j
Y_j is output of the jth neuron
w_{ji} is the synaptic weight connecting the ith input to the jth neuron,
b_j is error term and
$\psi_j(\cdot)$ is an activation function.

The activation function $\psi_j(\cdot)$ introduces a nonlinear relationship into the network between inputs and outputs of a node. The sigmoidal functions such as logistic and hyperbolic tangent functions (tanh) are the most common choices because they normalize inputs. From practical point of view, functions such as tanh or arctan that produce both positive and negative values tend to yield faster training than functions that produce only positive values such as logistic [15, 16]. In this study, therefore, tanh function is used to scale inputs and targets to (-1, 1). This function is given by:

$$\psi(X) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

The training procedure chosen was the most commonly used back-propagation algorithm, which is arguably the most popular algorithm for transportation use [4, 8, 9, 17]. The objective of the training process is to improve weights w_{ji} that minimize the mean squared error (MSE) [17].

2.1.3. Training, Testing and Validation

Network validation is important to verify that the training accuracy is sufficient. The whole data is divided into two sets. One set is used to construct the model (training set) and the other set is used to validate the model (testing set). Validation set is also utilized to minimize the over-training problem or to determine the stopping point of the training process. In practice, it is common to use one test set for both validation and testing purposes. During ANN development, the proportion of the data to be set aside for the training and test sets needs to be determined. Although there is no general procedure to do so, several factors such as the data type and the size of available data should be taken into consideration in such division process. Once a data set is sorted, the first 70% of the data can be taken as a training set whereas the next 30% of the data set as a testing set. This division has been used by most researchers [18, 19]. Out of the testing set, 20% of the data set is taken as a validation test.

2.1.4. The ANN Prediction Algorithm

Consider a hypothetical bus route shown on Figure 2. Suppose a journey for a bus *k*, equipped with GPS, is initiated from stop *'0'* at a certain time of day interval *'t'* and the bus is currently at location *'c'* which may or may not be a bus stop after passing stop *'i'*. Travel time information is to be provided for a person at stop *'j'*. Therefore, the time taken for a bus from the current location *'c'* to stop *'j'* can be calculated as:

$$TT^k{}_{cj} = TT^k{}_{ij} - TT^k{}_{ic} \tag{4}$$

where $TT^k{}_{cj}$ is predicted bus travel time from location c to stop j for bus k,
$TT^k{}_{ij}$ is predicted bus travel time between stop ' i' and 'j' for bus k and

$TT^k{}_{ic}$ is travel time to point c after passing stop 'i' for bus k.

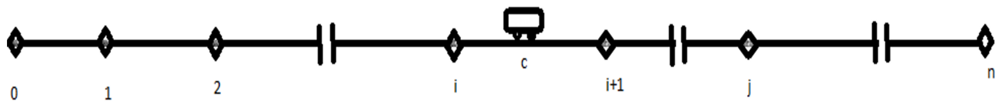


Figure 2. A Hypothetical Bus Route

$TT^k{}_{ic}$ value can be determined by subtracting departure time at stop 'i' from the current time when travel time information is requested. $TT^k{}_{ij}$ is determined from the ANN.

As has been discussed earlier, unique set of input/output variables, which can be obtained from GPS, are considered. The input variables chosen are:

X_1 = the time of day interval 't',

X_2=coded id number of bus station 'i',

X_3= coded id number of bus station 'j', and

X_4= the travel time taken from stop '0' to 'i'.

The output 'Y' will be predicted bus travel time to reach stop 'j' from stop 'i' which is equivalent to $TT^k{}_{ij}$ in equation (4).

During training process, weights and biases are adjusted automatically with the help of training and learning functions [20]. They are applied to all and individual weights and biases of a network respectively. The objective of the training process is to find weight values that will cause the output to match the actual target values as closely as possible. In this study, MATLAB has been used to train a network using the aforementioned set of inputs and a target variable which is the observed travel time between stops *'i'* and *'j'*. The function '*sim*' simulates a network [20]. It takes the network input vector $\{X_n\}$, and the network object NEURAL_NET, and returns the network output Y. That is:

$$TT^k{}_{ij} = Y = sim(NEURAL\_NET, \{X_n\}) \tag{5}$$

Normalizing the input/target data sets before training is important. This is because the contribution of one input will depend heavily on its variability relative to other inputs. For example, if one input has a range of 0 to 1, while another input has a range of 0 to 2,000,000, then the contribution of the first input will be dominated by the second input. Therefore, rescaling is important to fix such problem.

As has been stated earlier, the choice of input variables makes the algorithm possible to predict dynamically even when the neural network is trained off-line. The first input variable, for example, represents time of the day which accounts for the variability of travel time between different hours of the day. The last input variable also takes the current travel time information from the origin up to the recently visited bus stop. In particular, it has been noted that both under-training and over-training could negatively affect the ANN model quality [13] and it is out of this consideration, special care has been given to the ANN model development in this paper to ensure its quality.

2.2. Kalman Filtering

The Kalman Filter-based algorithms estimate state from the previous time steps. The current measurements are also required to make prediction for the next time step. Like the ANN model, Kalman filter can also be applied to data with no traffic parameter other than travel time.

Let $x_{(t|t-1)}$ represents the travel time at time interval *t* that is to be predicted, φ_t is the transition parameter between states at time interval *t* which describes time dependent relationship, and $w_{(t|t-1)}$ denote a Gaussian noise term with zero mean and a variance of Q_t . Assume the prediction model can be written as:

$$x_{t|t-1} = \varphi_{t-1|t-1}\, x_{t-1|t-1} + w_{t-1|t-1} \tag{6}$$

Subscripts are as follows: *t*|*t* current time period, *t-1*|*t-1* previous time period, and *t*|*t-1* are intermediate steps. The travel time observation on time interval *t* denoted by $Z_{(t|t)}$ is also assumed to be:

$$z_{t|t} = x_{t|t} + u_{t|t} \tag{7}$$

Where $u_{(t|t)}$ is the measurement error at time interval t that has a Gaussian distribution with zero mean and a variance of $R_{(t|t)}$. Also, {$w_{(t-1|t-1)}$} and {$u_{(t|t)}$} are not correlated (i.e., $E[w_i u_j] = 0$ for all i and j). The value of $z_{(t|t)}$ is usually obtained from averaging the travel times reported by the buses equipped with GPS at time interval t.

Assume that for all i , j , E[w(i)u(j)]= 0 , and let $P_{t/t}$ represents the covariance of the estimation error at time interval t , then the following simple Kalman filtering equations can be used to make the prediction.

Prediction:

State estimate extrapolation: $$x_{t|t-1} = \varphi_{t-1|t-1} x_{t-1|t-1} \quad (8)$$

Error covariance extrapolation: $$P_{t|t-1} = \varphi_{t-1} P_{t-1|t-1} \varphi^T_{t-1} + Q_{t-1} \quad (9)$$

Kalman Gain Calculation:

Kalman gain: $$K_t = P_{t|t-1}(P_{t|t-1} + R_t)^{-1} \quad (10)$$

Update:

State estimate update: $$x_{t|t} = x_{t|t-1} + K_t(z_{t|t} - x_{t|t-1}) \quad (11)$$

Error covariance update: $$P_{t|t} = (1 - K_t)P_{t|t-1} \quad (12)$$

Detailed derivations of Kalman filtering equations can be found elsewhere [11].

3. Study Bus Route

GPS data were available from November 2008 to May 2009 for buses in Macae, Brazil. Bus line LT11 had been chosen for the case study because it had the largest number of data sets as compared to the other bus lines (over a million records). This route is shown on Figure 3 below, and it has 35 stops numbered 0 to 34.

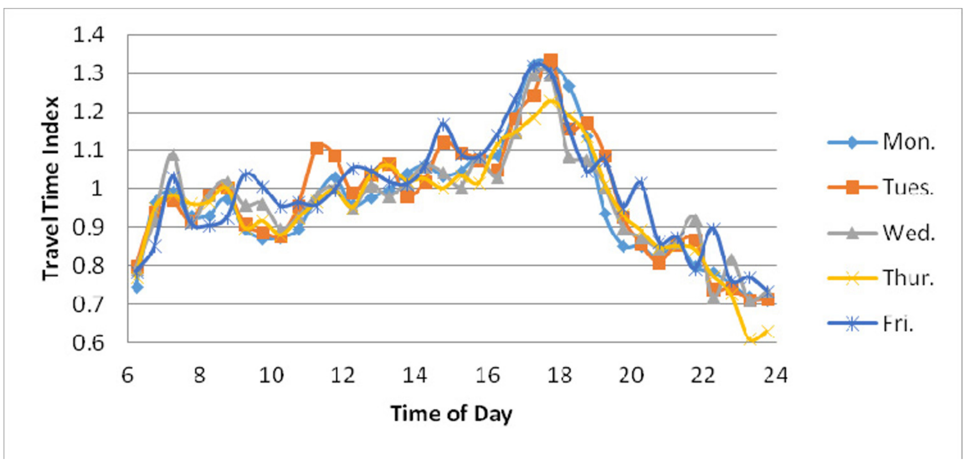
Figure 3. LT11 Bus Route, Macae, Brazil

The data were collected using Automatic Vehicle Location (AVL) systems. In these systems, GPS receivers are usually interfaced with a Global System for Mobile Communications (GSM) modems and placed in the buses. The systems basically record point locations in latitude-longitude pairs, speeds of the buses, date and time. Arrival and departure time information, from the GPS data, at each bus stop are considered for our prediction model development.

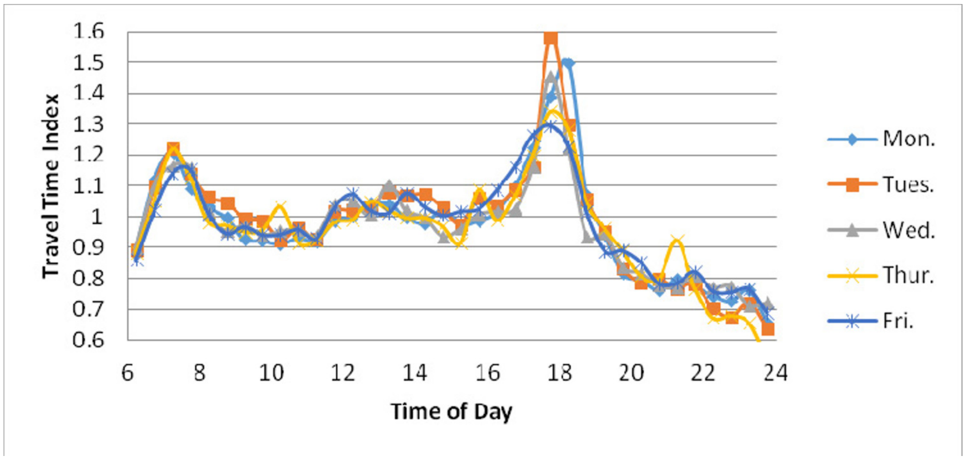
4. NUMERICAL RESULTS

4.1. Preliminary analysis

As part of the preliminary analysis, the arrival and departure travel time information were first translated to travel times. Once the travel times between stops were computed, they were pooled together in 30 minute intervals (e.g.; 6:00-6:30am, 6:30-7:00am etc.). Then, the average values of travel times in each interval were determined. Clustering the travel times by time period is important because travel times between stops usually may vary over time of the day, resulting in different travel patterns. In this paper, the average travel time from historical data is considered as a baseline prediction model which will later on be used for comparison purpose. For each time period, the travel time index, i.e. the ratio of the average travel time per weekday and the average travel time over all days, was calculated. Figure 4 presents the travel time index vs. time of day in northbound and southbound directions, respectively. The different lines (with different colors) correspond with different workdays, i.e. Monday to Friday. The upper panel demonstrates the line LT11 in the Northbound direction (form Stop 0 to Stop 34) and the lower panel shows the opposite direction. From Figure 4, it can be seen that there is a significant variation in (average) travel time over different hours of the day. In the evening rush hour (17:00-18:00), the travel times are about 30% higher than the average (over all time periods and all days) for the northbound direction. During the morning rush hour (around 7:00 – 7:30am), travel times are about 20% and 10% higher than average for the Southbound and northbound directions respectively. Thus, these travel time variations over different hours of the day are significant and should be taken into account in the model development. This also explains why we need to predict travel time dynamically.



4.a. LT11 North Bound Direction



4.b. LT11 South Bound Direction

Figure 4. Travel Time Variation over Time of Day

Another observation is that travel time distributions over different days of the week seem to be nearly the same. To investigate this, Duncan's multiple range test was performed and it was found that there was no significant differences in the travel time distributions for both directions. Thus, these variations were not further considered in the model development.

Historic average prediction models could be improved by considering errors or variations and correlations between different values of the variable under consideration [21]. For example, if the bus takes longer time during the first section of the trip, it is likely that the bus will also be take longer time on the second section of the trip. This means that travel times of successive trip sections for a bus may be correlated. If strong correlations are found between the travel times, the information about the travel time of a previous trip section can be used to update the travel time prediction of the next trip section. In order to investigate this, correlation analysis was conducted between observed travel times of successive sections. For simplicity, the whole trajectory was divided in to four approximately equal sections each with average travel time being close to 20 minutes. The result of the correlation analysis is presented in Figure 5 below.

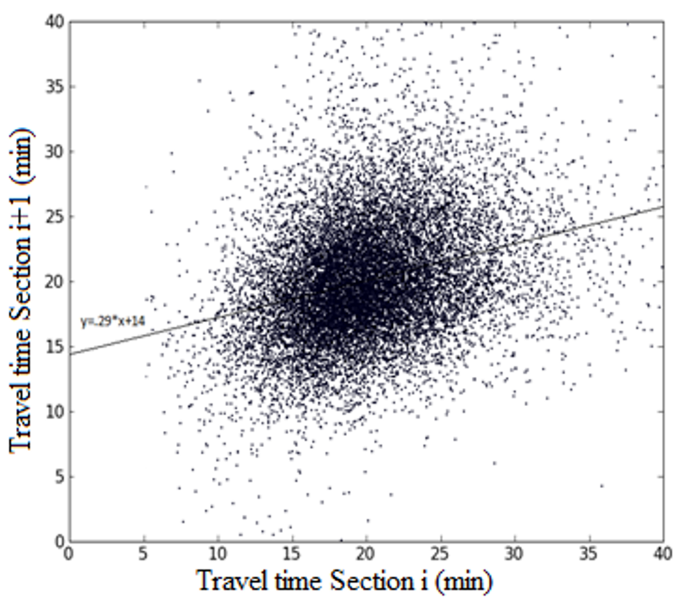


Figure 5. Correlation between Travel Times of Successive Sections

As can be seen from the above Figure 5, it was not possible, however, to get satisfactory linear relationships ($R^2=0.09$) to improve the baseline prediction based on correlation. This indicates a strong need for one look for estimation models that can capture variations between travel times of successive time intervals such as Kalman Filtering to deal with noises in the data and ANNs to capture these highly nonlinear relationships and predict bus travel time at the downstream bus stops.

4.2. Model Performance and Comparisons

After the prediction models have been developed, it is necessary to evaluate their performance in terms of prediction accuracy. Since regression prediction models are not good options in the absence of traffic data, comparisons are made between the ANN model, Kalman Filtering and historical average travel time model in this study. The Mean Absolute Percentage Error (MAPE) was used as the measure of model performance. It represents the average percentage difference between the observed value (in this case observed arrival times at a bus stop) and the predicted value (in this case predicted arrival times at a bus stop).

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t^P - Y_t^{Obs}}{Y_t^{Obs}}\right| \tag{13}$$

Where, Y_t^P is the predicted bus travel time from recent bus stop to target bus stop

Y_t^{Obs} is the observed bus travel time

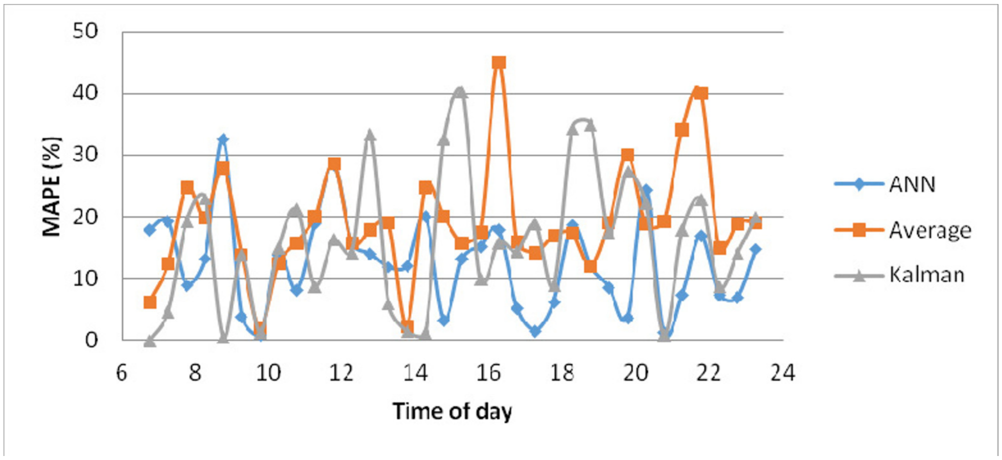
n is the number of test sets

In this paper, three different sections were considered to show the comparison between the two models. The sections are defined for one direction, i.e., north bound in the following Table 1.

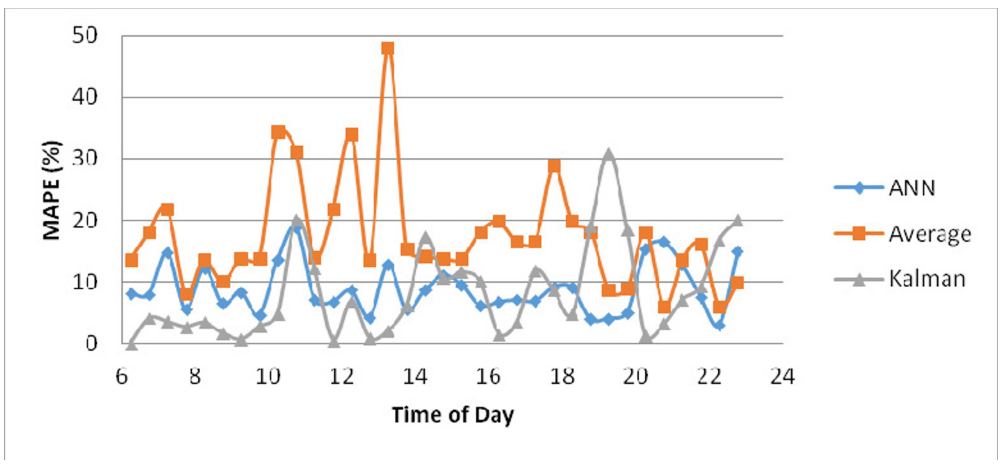
Table 1. Sections

Trajectory	**Stop ID's**	**Remark**
Section 1	Stop 8 to Stop16	Short section
Section 2	Stop 8 to Stop 26	Medium section
Section 3	Stop 0 to Stop 34	Long (whole section)

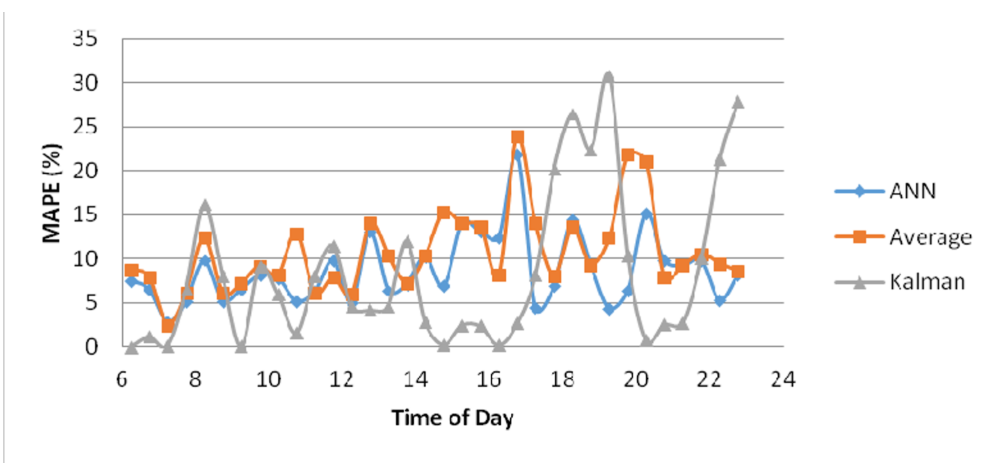
Figure 6 presents the MAPE values of the prediction models over different hours of the day for the three sections. It can be seen from the Figure 6 that there is no much difference in the distribution of MAPE between the three models for the short section. However, it can be clearly seen that ANN and Kalman filter models outperformed the historical average model approach on more than 70% of the time intervals of the day for medium sections. For the long section, Kalman filtering model seems to give a better prediction in most of the time intervals as compared to the other two models. The reason might be that Kalman filtering model has a relative advantage of using the current measurement to predict for the next time step. However, it is still vulnerable to give reasonable estimations during peak hours, specially evening peak. This is because Kalman filter usually fails to handle abrupt change between consecutive travel times as it tries to filter the noise smoothly. Therefore, it may be best used when there is no huge difference in travel times between two consecutive time steps. ANNs on the other hand can handle such differences in travel time and it gives relatively stable prediction during peak and off-peak hour for both medium and large sections.



6.a. Short Section (i.e., Section 1)



6.b. Medium Section (i.e., Section 2)

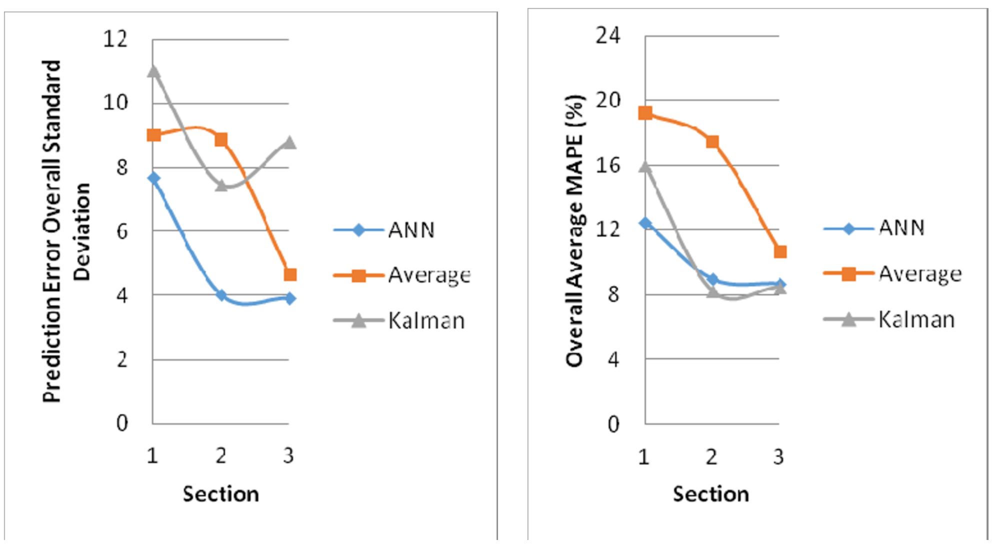


6.c. Long Section (i.e., Section 3)

Figure 6. MAPE vs. Time of Day

It is commonly accepted that usually a low MAPE value is desirable for an algorithm. However, an algorithm with a low value of MAPE may occasionally yield a prediction with a large deviation. This is undesirable since it may divert passengers away from the bus stop and eventually cause them to miss the bus. Therefore, it is important to define a second measure that will be used to detect this behavior. It examines the robustness of an algorithm such that its maximum deviation is within a certain range. Here, the robustness measure Ro is defined as: Ro = max {MAPE} of a section at a certain time interval of the day. As can be seen from Figure 6, the maximum MAPE values of the historical average and Kalman filtering models are greater than that of the corresponding MAPE of the ANN in all the three sections under consideration. Hence, ANN is more robust than the other two approaches.

In order to reinforce the above discussion, further analysis on overall average MAPE and standard deviation of prediction errors over different hour of the day from the three models can be done for the three sections under consideration. These values were computed for each section and are presented in Figure 7.



7.a. Overall Standard Deviation 7.b. Overall MAPE

Figure 7. Overall Standard Deviation and MAPE for each Section

As can be seen from the above Figure 7, the ANN models have the lowest value for the prediction error overall standard deviation for all the three sections on hand. The historical average approach, on other hand, has the highest overall MAPE. It can be seen that Kalman Filtering and ANN models have comparable values of overall MAPE for sections 2 & 3 and the ANN still has a lower value for the short section, i.e. section 1. Therefore, following the above discussions, we can conclude that the ANN model outperforms the other two models in terms of both overall prediction accuracy and robustness.

The model performances were considered and compared for the three sections in above sections. Based on these analysis results, the preferred ANN model performance is further investigated by applying it to any section between two bus stops. Figure 8 below presents the average MAPE values (computed using equation (13)) for a range of observed travel times between stops.

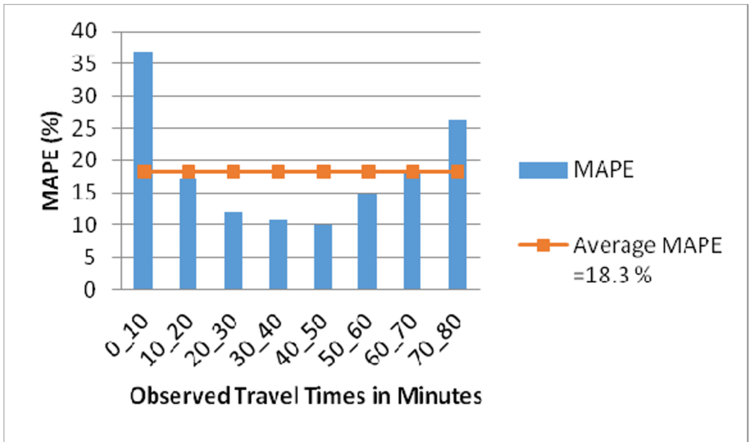


Figure 8. MAPE for a Range of Observed Travel Times

It is depicted in the above Figure 8 that bus travel time information could be provided for the study route and direction with an overall average MAPE value of 18.3% using the preferred ANN model. The extreme high and low trip travel times that accounted for 25% of the observed travel times contributed to their MAPEs higher than the total average MAPE (i.e. 18.3%). This indicates that the model is less effective in predicting bus travel times for very short and/or long trips. For short trips, high variability is expected due to a number of factors. For example, a bus may stop for a minute at a traffic light, stuck in traffic, or happen to run through a series of perfectly coordinated signal lights without any delay, etc. When the observed travel time is more than 50 minutes, the prediction error starts increasing again. This could be due to the boarding/deboarding activities at every stop when there is too high or too low travel demand and therefore more uncertainties (i.e., less forecasting accuracy) may be resulted. However, when the observed travel times are in a range between 20 to 50 minutes, the prediction model is able to provide real-time travel information with less than 10% MAPE. This range of travel times is observed when the person requiring the information is standing at least 5 stops away from the current bus location. Therefore, according to this case study, a person that is 20 to 50 minutes away from the current bus location receives the optimal travel time information. In other words, a person standing approximately 5 to 12 stops away receives travel information with 2 to 5 minutes prediction error correspondingly.

5. SUMMARY AND FUTURE RESEARCH

This paper presents bus travel time prediction between current bus location and any downstream bus stop under consideration using GPS data. Three predictions models, i.e. Historical average, Kalman Filter and ANN were considered in this paper. The performance of the models also tested and compared with each other. Historical average approach, where the predicted travel time is taken to be the average of the travel times of previous buses that traveled between any two stops, was considered as a baseline prediction. Prediction accuracy and robustness were used as performance measures. The overall precision measure determines the average deviation of the predicted travel time from the observed travel time. The robustness measure determines if an algorithm will occasionally give a prediction that is far off the actual arrival time. The ANN outperformed the other two approaches in both performance measures. The ANN model, which is trained offline, enables us to provide real-time travel time information at downstream stations with minimal error. The results obtained from the overall study are promising and the ANN model can be used to implement an Advanced Public Transport System to predict the arrival time at bus stops in area where even there is undisciplined traffic flow. The implementation of this system will improve the reliability of the public transport system, thus attracting more travelers to buses and helping relieve congestion. Kalman filtering model also can give a good prediction with reasonable accuracy but it is vulnerable if there is a huge difference in travel time values between two consecutive time steps. In conclusion, it is highly recommended that the ANN models be used as the statistical modeling approach for the dynamic travel time prediction models for buses using only GPS data.

The study investigated the prediction of travel time to provide a dynamic information to transit users using departure and arrival time information at stops collected via GPS technology. Apart from prediction, creating an optimal bus schedule is also very important. Further study effort may be directed towards incorporating schedule adherence information as an additional independent variable to improve the prediction models. A combined use of the three models may potentially improve the bus travel time prediction quality and therefore can also be another future research direction. A user-interactive system may also need to be developed to provide the travel time information.

ACKNOWLEDGEMENTS

The authors would like to thank Rápido Macaense and APB Prodata for the supply of travel time data of bus lines LT11, used for the case study, in Macae. The data was provided during thesis study in 2010 by the secondary author.

REFERENCES

[1] J. Houghton, J. Reiners, and C. Lim. *Intelligent transport system: how cities can improve mobility*. IBM Institute for Business Value, U.S.A. 2009. Accessed on July 26, 2014. ftp://ftp.software.ibm.com/common/ssi/pm/xb/n/gbe03232usen/GBE03232USEN.PDF.

[2] K. Dewan, and I. Ahmad. Carpooling: a step to reduce congestion (A Case Study of Delhi). *Engineering Letters*, 14:1, International Association of Engineers, 2007. Accessed on September 01, 2014. http://www.engineeringletters.com/issues_v14/issue_1/EL_14_1_12.pdf.

[3] L. Vanajakshi, S.C. Subramanian, R. Sivanandan. Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses. *IET Intelligent Transport System*, 3, 1-9, 2009.

[4] R. Jeong, and L. Rilett. *Bus arrival time prediction using artificial neural network model*. IEEE Intelligent Transportation Systems Conference, Washington, D.C., USA, 988-993, 2004.

[5] A. Farhan, A. Shalaby, and T. Sayed. *Bus travel time prediction using GPS and APC*. ASCE 7th International Conference on Applications of Advanced Technology in Transportation, Cambridge, Massachusetts, 2002.

[6] Y. Ramakrishna, P. Ramakrishna, V. Lakshmanan, and R. Sivanandan. *Bus travel time prediction using GPS data*. Proceedings Map India, 2006. Accessed on September 01, 2014. http://www.gisdevelopment.net/proceedings/mapindia/2006/student%20oral/mi06stu_84.htm

[7] J. Patnaik, S. Chien, and A. Bladihas. Estimation of bus arrival times using APC data. *Journal of Public Transportation*, 7(1), 1–20, 2004.

[8] S. Chien, Y. Ding, and C. Wei. Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering*, Volume 128, Number 5, 429-438, 2002.

[9] M. Chen, X. Liu, J. Xia, and S. Chien. A dynamic bus arrival time prediction model based on APC data, *Journal of Computer-Aided Civil and Infrastructure Engineering* 19 (5), 364-376, 2004.

[10] A. Shalaby, and A. Farhan. *Bus travel time prediction model for dynamic operations control and passenger information systems*. CD-ROM, the 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.

[11] R. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82 (Series D), 35–45, 1960.

[12] M. Hagan, H. Demuth, and M. Beale. *Neural network design*, PWS, Boston, 1996.

[13] Z. Gurmu, and W. Fan. *Artificial neural network travel time prediction model for buses using only GPS data*, Proceedings of the 92nd Annual Meeting of the Transportation Research Board, January 13-17, National Research Council, Washington, D.C., 2013.

[14] G. Zhang, B. Patuwo, and M. Hu. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14 (1), 35–62, 1998.

[15] S. Haykin. *Neural networks: a comprehensive foundation*. 2nd Edition. Prentice-Hall, Upper Saddle River, N.J., 1999.

[16] B. Karlik, and A. Olgac. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, Vol. 1, Issue 4, 111–122, 2010.

[17] D. Rumelhart, G. Hilton, and R. Williams. *Learning representations of back-propagation errors*, Nature (London), Vol. 323, 533-536, 1986.

[18] L. Yu, S. Wang, and K. Lai. An integrated data preparation scheme for neural network data analysis. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No. 2, 217-230, 2006.

[19] C. Patricia, and C. Robert. Method for optimal division of data sets for use in neural networks. *Knowledge-Based Intelligent Information and Engineering Systems*, Vol. 3684, 906, 2005.

[20] The Mathworks Inc. *MATLAB version 6.1 software*, Massachusetts, USA, 2011.

[21] T. Thomas, W. Weijermars, and E. Van Berkum. Predictions of urban volumes in single time series. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 11, No 1, 71-80, 2010.