Travel Time Prediction Models for Urban Corridor: A Case Study of Delhi

Sourabh Jain
Assistant Professor, Graphic Era University, Dehardun
sourabhjain.iitr@gmail.com

Parul Madan
Assistant Professor, Graphic Era University, Dehardun
parulmadaan1@gmail.com

***Abstract*— This study attempts to make use of traffic behaviour on the aggregate level to estimate congestion on urban arterial and sub-arterial roads of a city exhibiting heterogeneous traffic conditions by breaking the route into independent segments and approximating the origin-destination based traffic flow behaviour of the segments. The expected travel time in making a trip is modelled against sectional traffic characteristics (flow and speed) at origin and destination points of road segments, and roadway and segment traffic characteristics such as diversion routes are also tried in accounting for travel time. Predicted travel time is then used along with free flow time to determine the state of congestion on the segments using a congestion index (CI). Travel time is calculated using regression and ANN techniques and comparison has been made. A development of this kind may help in understanding traffic and congestion behaviour practically using easily accessible inputs, limited only to the nodes, and help in improving road network planning and management.**

Keywords— Congestion; Delay; Origin and Destination; Traffic; Travel Time; ANN

I. INTRODUCTION

Traffic congestion, not limited to but especially prevalent in metropolitan cities, is one of the most conspicuously worsening problems associated with traffic engineering and urban planning, with clear implications on spheres of urban economy, environment and lifestyle. Traffic in cities continues to grow meteorically especially in major cities of developing countries, which are characterized by heavy economic and population growth and assimilation in business and residential districts. This naturally necessitates intense transportation of goods and passengers, increasing demand for personal vehicular ownership that over the last decade has seen exponential growth worldwide. However, the failure of sufficiently rapid infrastructural development required to cater to this burgeoning traffic frequently leads to failures of the urban transportation system, resulting in traffic jams. Quantification of congestion thus becomes essential in checking congestion in order to provide a sustainable transportation system that necessitates a well-functioning well-integrated urban economy.

II. OBJECTIVE

The complexity of traffic systems in several developing countries is exacerbated due to the prevalence of heterogeneous traffic that only furthers the chaotic nature of the study. This study aims to understand the relationship between the traffic conditions of the source and the destination in portions (“segments”) of an arbitrarily chosen trip on an arterial and sub-arterial road in a major metropolitan city of India characterizing extremely diverse traffic conditions, and analyze the viability of promoting the use of O-D based measures of congestion to estimate the severity of the problem in the route. For this purpose, the basic traffic parameters, such as volume, speed, density and capacity are measured or calculated at different nodes of the study route and tried against the aforementioned indicator of congestion: Congestion Index, and a review for the prepared model and the behavior of the variables used is then prepared. To better understand the superiority of one of the travel time prediction models, results of the different models are compared.

III. RESEARCH SURVEY

Congestion has been variously defined as a physical condition in traffic streams involving reduced speeds, restrained movement, extended delays and paralysis of the traffic network. The definition of congestion has been conventionally categorized on the basis of four parameters: capacity, speed, delay/travel time and cost incurred due to congestion. Accurate prediction of travel time is important because it improves the quality of transportation services. The key to accurate predictions of travel time is two-fold: the prediction algorithm or model, and the data that is used as input to the algorithm [10]. The input to existing travel time estimation algorithms is either point-based traffic parameters such as time mean speed, volume, and / or occupancy, or the direct section / path based travel-time measurements by individual probe vehicles. Point-based traffic data are available from inductive loop detectors, and video cameras with use of image processing techniques. Direct travel time measurements can be obtained using Automatic Vehicle Identification (AVI), GPS, electronic license plate matching, electronic distance measuring instruments, and cellular phone tracking technology. A variety of prediction model has been developed that include historical based model [12][9], regression model[14][2], Kalman filter- based model [11], and Artificial Neural Network (ANN) model [7][3]. In this paper we discuss the comparison between Regression and ANN based travel time prediction models.

Various studies on regression models have been analyzed. A hybrid model has been developed for predicting travel time for the road network that is congested. To implement the hybrid model, the core forecasting algorithm with non-parametric regression technique has been integrated in GIS technology [13]. Zhangc and Rice built an efficient and easily-implementable model to predict freeway travel time using linear regression analysis [16]. A MLR (Multi Linear Regression) model has been developed in 2006 to predict the travel time using GPS data and passenger data [8]. Yu et al., developed a model in 2014 using SVM (Support Vector Machine) regression method to predict travel time [15].

The studies related to ANN provides with an insight of working with ANN. Mahmoudabadi built an ANN model for calculating the vehicle speed. Input parameters, such as types of road, time, volume of traffic and heavy vehicles ratio were considered in modeling [6]. In 2013 an ANN model has been proposed to calculate complete link travel time using sparse probe vehicle data [17]. Li and Chen examined the impact of different variables collected by Dual- Loop Vehicle Detector (VD) and Electronic Toll Collection (ETC) for prediction model of travel time for the freeway with non-recurrent congestion in 2013. Johar et al. in 2015 applied artificial neural network (ANN) for development of bus travel time prediction model. The bus travel time prediction model was developed to give real time bus arrival information to the passenger and transit agencies for applying proactive strategies [4].

IV. METHODOLOGY

The first step was to identify a suitable route that includes both arterial and sub-arterial roads and is often wrought with congestion. Subsequently the route was divided into segments and for this purpose, eight nodes were chosen, most of them being major rapid transit bus stops or major intersections. The next step was identification of potential factors. Both roadway as well as traffic parameters were considered, and the congestion parameter to be modelled was fixed (Congestion Index, CI). Once the expected data input was rightly identified, data were collected on site using video camera for recording node based traffic parameters and moving car method for measuring the real travel time. The data were pre-processed and source-destination and segment variables were calculated. Finally, all variables found were tested for statistical relationships with the dependent variable, CI, in several combinations using 75% of the observed data. This model was validated for the remaining 25% observed data with the help of root mean squared error (RMSE). The calibrated models of both ANN and regression model were compared to determine the model which has higher accuracy.

V. STUDY ROUTE

Delhi is a rapidly growing major city of India that, characterized by heterogeneity of traffic composition and suffers from aggravating traffic system [1]. The Delhi route chosen for study comprised two sections: a long eastern part of the Inner Ring Road, an access controlled divided arterial way, and Sri Aurobindo Marg, a divided sub-arterial that takes diversion from the Ring Road south of AIIMS. Each portion consists of three segments separated by a total of eight nodes (points). The total length of the study route is about 27.4 km, excluding a 640 m long stretch between AIIMS North Gate and AIIMS West Gate that was not used for observations. Tables 1 and 2 include the roadway details of the study route.

TABLE I. OBSERVATION NODE DETAILS

Node ID	Name
1	Kashmere gate ISBT (Inter State Bus Terminal)
2	Sarai kale khan bus station
3	Andrew ganj main intersection
4	AIIMS north gate
5	AIIMS west gate
6	Green park main intersection
7	IIT gate bus stop
8	Mehrauli bus terminal

TABLE II. SEGMENT ROADWAY DETAILS

Seg. ID	Source node ID	Destination node ID	Length (Km)	No. of lanes	No. of major intersections
Arterial road: -					
1	1	2	12.100	6	5
2	2	3	6.080	6	3
3	3	4	1.740	6	1
Sub arterial road:-					
4	5	6	0.880	6	0
5	6	7	1.350	4	0
6	7	8	3.430	4	2

VI. DATA COLLECTION

Data collection primarily involved traffic parameter observation on study points (“nodes”) such as categorized vehicular traffic volume and spot speed using manual counting and radar gun respectively in count periods of 15 minutes, and travel time using the moving car method. The traffic data were collected in six motored vehicular categories: standard cars and vans, two wheelers (scooters and motorbikes), three wheelers (auto-rickshaws), LCV (light commercial vehicles), trucks and buses. The data was collected for morning and evening, peak and non-peak hours.

Travel time was observed in Phase I in the four slots everyday with the help of moving car method by repeated car trips along the route. Clearly perceived congestion, signalized intersection and bus dwell time delays for buses ahead of the car were individually noted for reference. In Phase II, two acquainted commuters were chartered with making the GPS logs with timestamp of arrival.

VII. MODELLING

A. *Selection of Variables*

The primary aim of this study being origin-destination based congestion estimation, different node variables were tested for correlation both among themselves as well as with the dependent variable – Travel Time. In the Pearson correlation matrix of the independent variables, all but five correlation coefficients came out to be between -0.5 and 0.5 (other than the definite correlation between node and averaged segment values), leading to the conclusion that most of them are really independent. Also, all but two variables, viz. ‘Number of Lanes’ and ‘Destination Speed’, were found to be reasonably related to the dependent variable, with their correlation coefficients greater than 0.5. The values are tabulated in Table 4.

B. *Regression Model*

To develop the travel time prediction model, a multi-linear regression analysis was performed on the data collected from the selected urban corridor. With the given data, two datasets were created: one with the source-destination values, the other one with segment values averaged from the first set. Analysis of Variance (ANOVA) was carried out on both the models. F test was carried out on the models with a DDF (denominator degrees of freedom) of 383 on a confidence interval of 95%. The results are given in the Table 3.

TABLE III. REGRESSION MODELS

S. No.	**Models Name**	**Data Set**	**Set of Data Records**	**Adjusted R2 Value**	**F value**	**P Value (at α = 0.05)**
1.	Model I	Origin-Destination	832	0.917	52.745	0.000
2.	Model II	Segment	817	0.896	37.012	0.002

TABLE IV. VARIABLES CONSIDERED IN THE MODEL

Independent variable	**Symbol**	**Unit**	**Correlation coefficient**
Roadway parameters: -			
Segment length	L	Km	0.890
Number of intersections	N_1		0.914
Number of Lanes	N_2		0.356
Origin – destination parameters: -			
Origin volume	Q_S	Pc/hr	0.697
Origin speed	V_S	Km/hr	0.600
Origin density	K_S	Pc/km	0.549
Destination volume	Q_D	Pc/hr	0.810
Destination speed	V_D	Km/hr	0.552
Destination density	K_D	Pc/km	0.726
Segment averaged parameters: -			
Segment volume	Q	Pc/hr	0.778
Segment speed	V	Km/hr	0.513
Segment density	K	Pc/km	0.660

The adjusted R^2 value is a statistical parameter that depicts the proportion of variance in Travel Time, the dependent variable, explained by that of the independent variables themselves, and is a good indicator of the credibility of the model. The obtained P value of Model I following the F-test indicates that its null hypothesis can be safely rejected and it may be concluded that the model is better than the one with only intercepts. Thus Model I, which has the highest R^2 value and the best F value, was chosen as the desired model. The following formula of travel time (in seconds) as shown in Eq. 1 was arrived at.

$$T = -94.1 + 15.2*L + 135.6*N_1 + 0.27*Q_s - 0.15*Q_d - 23.9*V_s + 26.0*V_d - 6.16*K_s + 4.89*K_d \quad (1)$$

C. ANN Model

ANN has been applied for a wide variety of transportation problems and is relatively easy to use. Neural network automatically discover the relationship between the variables and naturally the fitting take place. Generally the network architecture is the single place where intuition of researchers comes into play. In ANN modelling based on the problem one can choose the required number of variables, as there is no limit on the number of variable. ANN provides flexibility, massive parallelism, learning and generalization ability, accuracy and some amount of fault tolerance in prediction of travel time. For design of neural network there is no general theory or method. Mostly trial and error approach is used. During design of neural network the complexity arises while modelling of non-linear problem. Architecture of network, number of input variables, choosing of training algorithm and activation function are the basic features which must be considered in the design of neural network. All these are problem dependent quantities.

Whole database has been divided in to three parts training, validation and testing in the ratio 65, 15 and 25 percentage respectively. In this study Model II was trained nine times using same set of training data (614 data records) and different number of neurons. Validation and testing data was used to compare the performance of ANN models. The performance measures MSE (Mean Square Error) and coefficient of correlation (R^2) were used to estimate the prediction results. Table 5 gives the detail of network with different structure for Model II. From Table 5 it was found that network structure (3, 5, 1) produces best prediction.

After the number of hidden neurons that produce best prediction for Model II has been found. Then, Model I was trained using 307 set of training data and five numbers of hidden neurons (that produces best prediction for model II).

The detail of the network structure for both the models has been illustrated in Table 6. Based on the performance measure that is MSE and R^2 it was found that Model I gives best prediction than Model II. Thus, it was found that Model I is better than the one with only intercepts. Thus Model I, which has the highest R^2 value and the best F value, was chosen as the desired model.

TABLE VI. DETAIL OF NETWORK STRUCTURE FOR MODEL I AND II

Model Name	Model I	Model II
No. of Hidden Layers	1	1
No. of Hidden Neurons	5	5
Training Function	Trainlm	Trainlm
Transfer Function	Tansig	Tansig
Number of epochs	1000	1000
MSE(T)	1.85E-02	6.34E-03
MSE(V)	2.29E-02	5.45E-02
MSE(Y)	2.09E-02	7.59E-03
R^2(Y)	0.929	0.9612

D. Validation of Models

Modelling by regression was carried out with 75% of the pre-processed data, while the remaining 25% was used for validation of the same model. This was done by finding out normalized root mean squared error (NRMSE) and mean absolute percentage error (MAPE), values that determine the predictive power of the model. They are given by the following formulae: -

$$NRMSE = \frac{RMSE}{\bar{X}_o} \tag{2}$$

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{N}\left(X_{o_i} - X_{p_i}\right)^2}{N}} \tag{3}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{X_{o_i} - X_{p_i}}{X_{o_i}}\right| \tag{4}$$

TABLE V. DETAIL OF NETWORK STRUCTURE FOR MODEL I AND II

Trial No	**Train 1**	**Train 2**	**Train 3**	**Train 4**	**Train 5**	**Train 6**	**Train 7**	**Train 8**	**Train 9**
No. of hidden layers	1	1	1	1	1	1	1	1	1
No. of hidden neurons	3	4	5	6	7	9	11	12	15
Training function	Trainlm	Trainlm	Trainlm	Trainlm	Trainlm	Trainlm	Trainlm	Trainlm	Trainlm
Transfer function	Tansig	Tansig	Tansig	Tansig	Tansig	Tansig	Tansig	Tansig	Tansig
Number of epochs	1000	1000	1000	1000	1000	1000	1000	1000	1000
MSE (T)	5.30E-2	4.85-2	4.09E-2	4.83E-2	4.96E-2	4.83E-2	5.67E-2	4.95E-2	4.91E-2
MSE(V)	5.83E-2	5.73E-2	4.61E-2	5.84E-2	5.96E-2	6.31E-2	5.37E-2	4.77E-2	5.41E-2
MSE (Y)	5.56E-2	5.48E-2	4.24E-2	5.75E-2	5.38E-2	5.58E-2	5.93E-2	5.66E-2	6.34E-2
R2(Y)	0.8309	0.8480	0.8612	0.8458	0.8219	0.8505	0.8400	0.8521	0.8362

Here X_o is the actual value of the parameter, X_p is its predicted value according to the model, $\overline{X}_o$ is the mean of the observed values and N is the total number of observations in the validation dataset. A low value of these values is desirable; typically, a value of around 0.1 (10%) of NRMSE and MAPE depict a highly accurate model. The obtained values on the validation dataset in this linear source-destination (Model 1) model are given in table 8. The values of NRMSE and MAPE imply a substantially accurate forecasting. From the Predicted Time vs. Observed Time graph in Fig. 1, it may be understood that the applied model works well for the validation dataset.

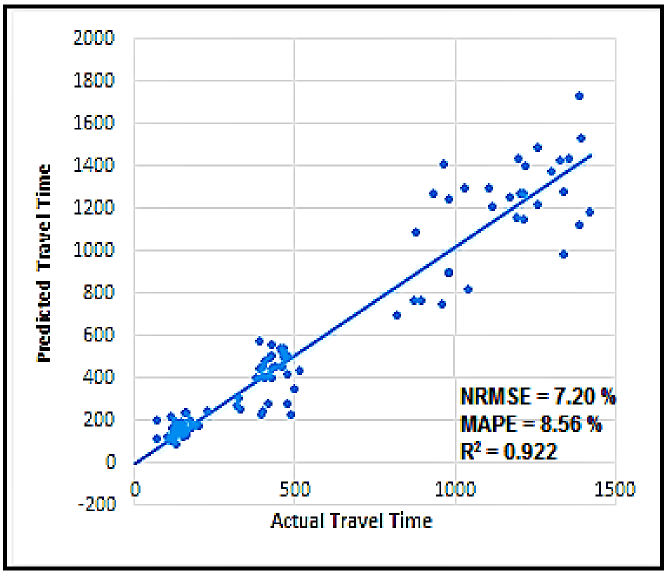

Figure 1. Predicted versus observed travel time results of validation.

TABLE VII. VALIDATION DATASET PARAMETERS

Property	Value
Number of observations, n	96
Mean observed travel time, T_o (sec)	580
Mean predicted travel time, T_p (sec)	572
RMSE (sec)	41.8
NRMSE (%)	7.20
MAPE (%)	8.56

After the development of ANN model it is necessary to estimate the performance in terms of accuracy and robustness. The performance of the developed model was evaluated by applying paired t-test. The paired t-test values of the model at 5% level of significance are shown in the Table 8. Since the calculated t-value is less than the tabulated t-value for the model developed for the selected urban corridor, therefore using null hypothesis it was concluded that here was no major variation among the actual and predicted values. Hence, the model developed for the selected urban corridor is more suitable.

TABLE VIII. SUMMARY OF OUTPUT OF FOUR ANN MODELS FOR TESTING PHASE

Model Name	Model I
R^2	0.929
NRMSE	10.432
MAPE	6.527
Standard Deviation	10.268
t-test $_{calculated}$	-2.263
t-test $_{tabulated}$	1.9788
DOF	127

VIII. COMPARISON OF MODELS

In this study, the result of the developed model was compared to check the accuracy and robustness of the model developed for the selected urban corridor. To compare, three measure of effectiveness that is RMSE, MAPE and R^2 were used. Three measure of effectiveness that is RMSE, MAPE and R^2 for regression and ANN models are shown in the Table 9. Results in the Table 9 shows that ANN model depict greater accuracy and robustness with less error as compared to regression model.

During training process ANN learns from examples and thus weights are adjusted accordingly to use this information during testing and cross validation. As compared to analytical and statistical model, ANN model generally produce better results with minimization of errors. Apart from good results and less error as compared to regression model, still there are some problems related to ANN modelling. In ANN modelling it is not possible to find the effect of each individual variable independently.

TABLE IX. COMPARISON BETWEEN REGRESSION AND ANN MODEL FOR TESTING PHASE

Model Type	**Regression**	**ANN**

Model Name	Model I	Model I
R^2	0.922	**0.929**
NRMSE	7.2	10.432
MAPE	8.56	6.527
Standard Deviation	16.734	10.268

IX. CONCLUSIONS

This paper proposes an idea of determining the model which provides with higher accuracy. For this Regression and ANN models for travel time prediction has been compared.

In this study, the comparison between the models were performed to check the accuracy and robustness of the model developed for the selected urban corridor. Three measure of effectiveness used to compare were RMSE, MAPE and R^2. Results prove that ANN model depict greater accuracy and robustness with less error as compared to regression model.

ANN model generally produce better results with minimization of errors when compared to analytical and statistical model. However, it is not possible to find the effect of each individual variable independently.

It can also be seen that this study, due to the virtue of scope, has some limitations that were duly noted. The use of node data to estimate travel time may help in estimation of travel time, but it falters in providing help for suggesting alternative routes because the node data for alternative routes remain the same notwithstanding anything but roadway parameters such as length and diversions. In order to make this distinction clear, more roadway parameters should be studied for influence on traffic congestion.

REFERENCES

[1] Coscia, V.; Delitala, M.; Frasca, P. 2007. On the mathematical theory of vehicular traffic flow II discrete velocity kinetic models, International Journal of Non-Linear Mechanics, Elsevier, 42(3), pp. 411.

[2] Chang, H., Park, D., Lee, S., Lee, H. and Baek, S. 2010. Dynamic Multi-Interval Bus Travel Time prediction Using Bus Transit Data, Transportmetrica, Vol. 6, No.1, pp. 19-38.

[3] Gurmu, Z. K. and Fan, W. D. 2014. Artificial Neural Network Travel Time Prediction Model for Buses using only GPS Data, Journal of Public Transportation, Vol. 17, No. 2, pp. 45-65.

[4] Johar, A., Jain, S. S. and Garg, P.K. (2015), “Prediction of Bus Travel Time using Artificial Neural Network”, International Journal for Traffic and Transport Engineering (IJTTE), Vol. 5, No. 4, pp. 410-424.

[5] Li, C. S. and Chen, M. C. 2013. Identifying Important Variables for Predicting Travel Time of Freeway with Non-recurrent Congestion with Neural Networks, Neural Computing and Applications, Vol. 23, No. 6, pp. 1611-1629.

[6] Mahmoudabadi, A. 2010. Using Artificial Neural Network to Estimate Average Speed of Vehicles in Rural Roads, International Conference on Intelligent Networking and Computing, Nov. 26, pp. VI-25-VI-30.

[7] Mazloumi, E., Moridpour, S., Currie, G. and Rose, G. 2012. Exploring the Value of Traffic Flow Data in Bus Travel Time Prediction, Journal of Transportation Engineering, Vol. 138, No. 4, pp. 436-446.

[8] Ramakrishna, Y., Ramakrishna, P., Lakshmanan, V. and Sivanandan, R. 2006. Bus Travel Time Prediction using GPS Data, Proceedings, Map India. Avaliable on: http://www.gisdevelopment.net/proceedings/mapindia/2006/, Accessed on 26-05-2015.

[9] Rice, J. and Van Zwet, E. 2004. A Simple and Effective Method for Predicting Travel Times on Freeways, IEEE Transactions on Intelligent Transportation Systems, Vol. 5, No. 3, pp. 200-207.

[10] TRB. 2003a. TCRP Synthesis 48: Real-Time Bus Arrival Information Systems: A Synthesis of Transit Practice, Transit Cooperative Research Program, Federal Transit Administration, Transportation Research Board (TRB), Washington D. C.

[11] Vanajakshi, L., Subramanian, S.C. and Sivanandan, R. 2009. Travel Time Prediction under Heterogeneous Traffic condition using Global Positioning System Data from Buses, IET Intelligent Transport System, Vol. 3, No.1, pp.1-9.

[12] Williams, B. and Hoel, L. 2003. Modeling and Forecasting Vehicle Traffic Flow as a Seasonal Arima Process: Theoretical Basis and Empirical Results, Journal of Transportation Engineering, Vol. 129, No. 6, pp. 664–672.

[13] You, J. and Kim, T. J. 2000. Development and Evaluation of a Hybrid Travel Time Forecasting Model, Transportation Research Part C: Emerging Technologies, Vol. 8, No. 1-6, pp. 231-256.

[14] Yu, B., Lam, W. H. K., and Tam, M. L. 2011. Bus Arrival Time Prediction at Bus Stop with Multiple Routes, Transportation Research Part C: Emerging Technologies, Vol. 19, No. 6, pp. 1157-1170.

[15] Yu, B., Ye, T., Tian, X. M., Ning, G. B., and Zhong, S.Q. 2014. Bus Travel Time Prediction with Forgetting Factor, Journal of Computing in Civil Engineering, ASCE, Vol. 28, No. 3, p. 06014002.

[16] Zhangc, X. and Rice, J. A. 2003. Short-Term Travel Time Prediction using A Time-Varying Coefficient Linear Model, Transportation Research Part C: Emerging Technologies, Vol. 11, pp 187-210.

[17] Zheng, F. and Van Zuylen, H. 2013, “Urban Link Travel Time Estimation Based on Sparse Probe Vehicle Data”, Transportation Research Part C: Emerging Technologies, Vol. 31, pp. 145-157.