The Netherlands
Research School for
Transport, Infrastructure
and Logistics

TRAIL

11th TRAIL Congress
November 2010

ESTIMATING URBAN TRAVEL TIME DISTRIBUTION USING PROBE VEHICLE DATA

Fangfang Zheng MSc[1,2],Prof. dr. Henk van Zuylen[1,3]
[1]Faculty of Civil Engineering and Geosciences, Department of Transport and Planning,
Delft University of Technology, the Netherlands
[2]College of Traffic and Transportation, Southwest Jiaotong University, Chengdu, P.R. China
[3]Hunan University, Changsha, P.R. China

ABSTRACT

In recent years, more and more research has been devoted to urban travel time estimation and prediction. However, very few promising results can be found in literatures. One reason is that urban travel times are intrinsically uncertain due to stochastic properties of traffic flow and stochastic arrivals and departures at intersections. As a result, given known traffic states and traffic control schemes, a range of travel times (a certain travel time distribution) can be observed. In this paper, a delay (travel time) distribution model for an isolated intersection as derived from an analytical model under different circumstances was calibrated using probe vehicle data. The Least squares and maximum likelihood methods were applied to estimate the parameters in the delay distribution model. Based on the estimated parameters, delay distribution is reconstructed.

KEYWORDS

Urban travel time, delay distribution, probe vehicle data

INTRODUCTION

Travel time prediction on motorways has been investigated by several researchers (van Hinsbergen et al., 2008; Wei et al., 2007; You et al., 2000; Yu et al., 2008)and from the analysis of the regularities in travel times, travel time prediction methods have been developed. Both model based predictions and heuristic methods have been developed and both methods have shown satisfactory results.

The idea behind both approaches is that travel times are determined by the traffic states along a route. These traffic states are estimated in the model based method, while heuristic methods just look for relationships between certain parameters (e.g., speeds and flows) and the (future) travel times without physical models behind.

For urban travel times this approach is less successful until now (see (Liu, 2008)). The mechanism on urban trips is significantly different from that on expressways.
The travel time is mainly determined by four mechanisms:

1. Distance and free flow speed;
2. Parking movements, (un)loading vehicles and buses at stops, crossing pedestrians and cyclists;
3. Queuing process before signalized intersections;
4. Traffic control at signalized intersections

The free flow speed on urban roads is mainly determined by the speed limit. It is slightly influenced by the vehicle composition, the density of traffic on the roads, and the spacing between two intersections. The mid-link delay is mainly caused by the movements of e.g. buses at bus stops, vehicles at parking places along the road and pedestrians crossing the road. The time spent in queues is determined by the queue lengths and the effective capacity of the bottleneck in front of the queue. Since this is in most cases a signalized intersection, this part of the travel time has a strong relation with traffic signals. The delay at a signalized intersection is determined by the regular process of the green and red status of the signal and by the queuing process. However, some irregular behavior, e.g., visiting a shop during the trip along the roadside, is difficult to be modeled. This gives outliers for deriving travel times.

The result of these mechanisms we see in urban travel time is that given the same traffic condition, the travel time (delay) is not a single value but a range of travel times (delays) can be observed. An analytical delay distribution model has been developed by authors (Zheng et al., 2010). The applications of the delay (travel time) distribution are two folds. First of all, given the delay or travel time distribution, the variability of travel time or delay can be investigated using statistical measures, e.g., percentile or standard deviation. Secondly, it helps to determine the prediction interval when dealing with travel time prediction. However, the difficulty in applying this model remains the question how to estimate parameters in the model, especially the overflow queue distribution. In (Zheng et al., 2010), the overflow queue distribution is estimated in an analytical way by applying a Markov chain model with the assumption of a certain arrival distribution (e.g. Poisson distribution) within a certain time period. However, it is not feasible when it comes to the oversaturated condition when the overflow queue is time dependent and growing. The difficulties in applying this model for real time estimation or prediction inspire us to think about adopting heuristic methods for parameter estimation instead of analytical methods. The Least Squares estimation and Maximum Likelihood estimation are adopted to perform parameter estimation in this paper. The Genetic Algorithm is adopted to find the optimal parameter set both to minimize the Square Error function and maximize the likelihood function. Based on the estimated parameters, the delay distribution is reconstructed.

METHODOLOGY

Delay distribution

As discussed in (Zheng et al., 2010), travel time has a large variability and this distribution depends on the moment of arrival at intersections and on the length of queue at that moment. Given that vehicles arrive with constant time headway, in the undersaturated condition, when a vehicle arrives at thebeginning of the red time, delay equals to the red time plus the time to release the initial overflow queue and decreases linearly until zero. While in the oversaturated condition, an arriving vehicle needs to wait for another cycle or more cycles to depart due to the large overflow queue in front of it. The analytical delay distribution functions of a fixed time controlled intersection as derived by authors (Zheng et al., 2010) are given by the following equations:

$$P(w|n_i) = \alpha(n_i)\delta(w) + \sum_{k=0}^{N} \beta B(w, w_{2k+1}(n_i), w_{2k+2}(n_i)) \quad (1)$$

$$P(w) = \sum_{n_i}^{\infty} P(w|n_i)\, p(n_i) \quad (2)$$

where $P(w|n_i)$ is the delay distribution given a fixed overflow queue n_i; k is the number of extra red times that an arriving vehicle needs to wait; N is the maximum number of red times that the arriving vehicle needs to wait for, given the overflow queue n_i; $\delta(W)$ is the Dirac delta function and $B(w, w_{2k+1}, w_{2k+2})$ is a box function with the property:

$$B(w, w_{2k+1}, w_{2k+2}) = \begin{cases} 1 & w_{2k+1} < w < w_{2k+2} \\ 0 & otherwise \end{cases}$$

The different parameters, e.g., follow from the traffic state (e.g. the flow, overflow queue, the red phase and cycle time). Figure 1 shows a characteristic shape of the delay probability distribution for the undersaturated condition.

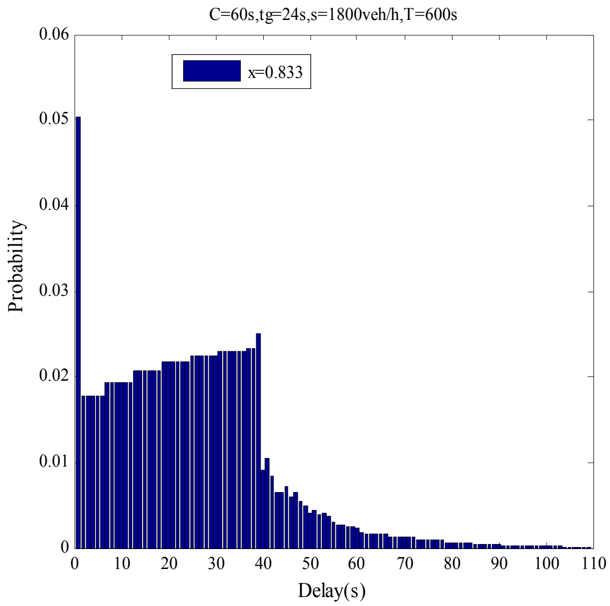

Figure 1: Probability distribution of the delay at a moderately loaded (x=0.833) signalized intersection

Parameter estimation

As can be seen from Eq. (1), the traffic state parameters in this function include α and β, and the delay boundaries in the box function w_{2k+1} and w_{2k+2}. However, α, w_{2k+1} and w_{2k+2} also depend on the overflow queue n_i. The question is whether it is possible to recognize the traffic state which in this case is the overflow queue distribution from the travel time (delay) distribution. In order to estimate the parameters in the delay distribution functions, we assume a maximum overflow queue. The overflow queue distribution is estimated based on the measurements, e.g., the measured delays, flows and cycle time.

Least Squares (LS) estimation

The objective of the least square method is to adjust the parameters of a model function to best fit a set of data and to characterize the statistical properties of estimates. Here, the model function is the delay probability function $P(w)$ with parameters $\alpha, \beta, p_0, p_1, ..., p_{n_0}$ and the data set is the measured delays. Therefore, the objective function can be formulated as:

$$\min f(\alpha, \beta, p_0, p_1, ..., p_{n0}) = \sum (P(w) - P_{measurements})^2$$
$$= \sum_{i=1}^{m} (\sum_{j=0}^{n_0} (\alpha(j)\delta(w) + \sum_{k=0}^{N} \beta B(w_i, w_{2k+1}(j), w_{2k+2}(j)))p_j - P_{measurements})^2 \quad (3)$$
$$\text{St. } 0 \le \alpha \le 1, 0 \le \beta \le 1, \quad \sum_{j=0}^{n_0} p_j = 1$$

where i is the number of vehicles in the overflow queue and n_0 is the maximum overflow queue we assume. It can also be approximated based on the maximum delay. $P_{measurements}$ is the measured delay distribution which in this case is the PVD.

Maximum Likelihood (ML) Estimation

Let $P(w_1, w_2, ..., w_m | \alpha, \beta, p_0, p_1, ..., p_{n_0})$ denotes the probability density function of measured delays $w_1, w_2, ..., w_m$ given parameters $\alpha, \beta, p_0, p_1, ..., p_{n_0}$, where $w_1, w_2, ..., w_m$ are independent. For a single delay, the probability can be calculated by the following function:

$$P(w_i | \alpha, \beta, p_0, p_1, ..., p_{n_0}) = \sum_{j=0}^{n_0} [\alpha(j)\delta(w_i) + \sum_{k=0}^{N} \beta B(w_i, w_{2k+1}(j), w_{2k+2}(j))] p_j \quad (4)$$

Then for independent delays $w_1, w_2, ..., w_m$, the probability is formulated as:

$$P(w_1, w_2, ..., w_m | \alpha, \beta, p_0, p_1, ..., p_{n_0}) = P(w_1 | \alpha, \beta, p_0, p_1, ..., p_{n_0}) P(w_2 | \alpha, \beta, p_0, p_1, ..., p_{n_0}) ... P(w_m | \alpha, \beta, p_0, p_1, ..., p_{n_0})$$
$$= \prod_{i=1}^{m} \left\{ \sum_{j=0}^{n_0} [\alpha(j)\delta(w_i) + \sum_{k=0}^{N} \beta B(w_i, w_{2k+1}(j), w_{2k+2}(j))] p_j \right\} \quad (5)$$

The likelihood function can be derived as:

$$L(\alpha, \beta, p_0, p_1, ..., p_{n_0} | w_1, w_2, ..., w_m) = P(w_1, w_2, ..., w_m | \alpha, \beta, p_0, p_1, ..., p_{n_0})$$
$$= \prod_{i=1}^{m} \left\{ \sum_{j=0}^{n_0} [\alpha(j)\delta(w_i) + \sum_{k=0}^{N} \beta B(w_i, w_{2k+1}(j), w_{2k+2}(j))] p_j \right\} \quad (6)$$

In practice, it is more convenient to work with scaled logarithm of the likelihood function which is calculated as:

$$\ln L = \sum_{i=1}^{m} \ln \left[\sum_{j=0}^{n_0} [\alpha(j)\delta(w_i) + \sum_{k=0}^{N} \beta B(w_i, w_{2k+1}(j), w_{2k+2}(j))] p_j \right] \tag{7}$$

As can be seen from Eqs.(3) and (7), finding the optimal solution for the parameter estimation in an analytical way is not applicable since both objective functions are very complicated and highly nonlinear. The Genetic Algorithm is applied to find the optimal solution to minimize the square error for the objective function.

PRELIMINARY RESULTS

In order see whether parameters in the delay distribution model can be estimated using PVD, an experiment was set up in VISSIM. Individual delays were recorded in VISSIM and 10% of the total measurements were used for parameter estimation. Figure 2 compares the delay distributions from VISSIM simulation data with reconstructed delay distributions. The parameters are estimated based on LS method and ML method. The optimal solutions for both methods are obtained using Generic Algorithm. The results show that the delay distribution estimated based on both LS method and ML method under the highly undersaturated condition (left figure with degree of saturation=0.68) can well represent the simulation data. However, when the degree of saturation (0.9) increases, ML method performs better than LS method.

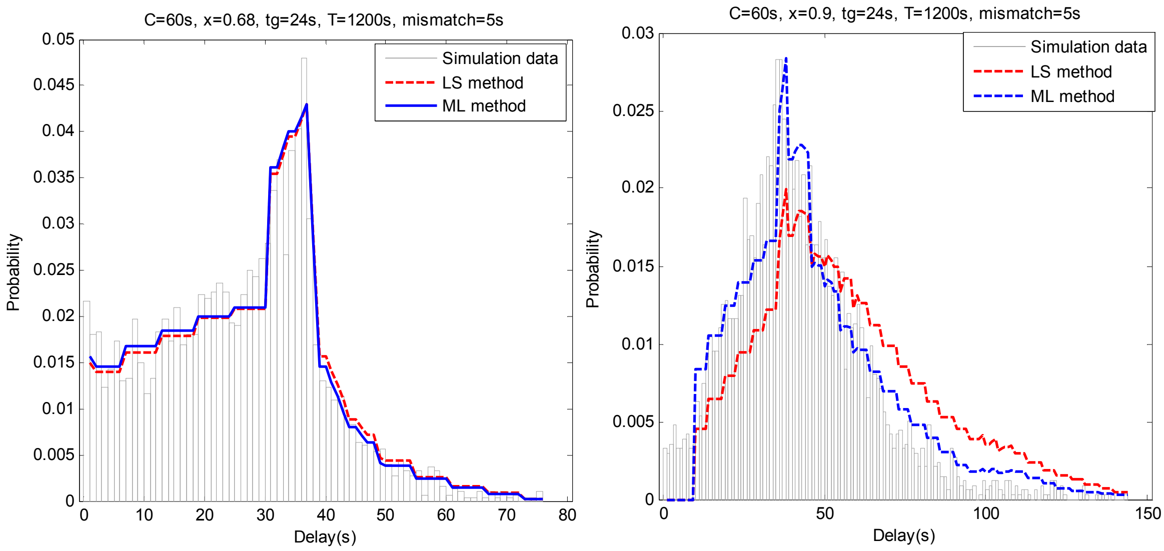

Figure 2: Comparison of delay distribution based on the simulation and GA optimized parameters

REFERENCES

Liu, H. (2008) *Travel time predictions for urban networks*, Thesis TRAIL, Delft, the Netherlands.

Zheng F., H.J. van Zuylen (2010) An investigation of urban link travel time estimation based on field sparse probe vehicle data, *89th TRB annual meeting*, Washington DC, USA.

Zuylen, H.J. van, F. Zheng, Y. Chen (2009) Using Probe vehicle data for traffic state estimation in signalized urban networks, in: *Traffic Data Collection and its Standardization,* editors Jaume Barceló and Masao Kuwahara, Springer, New York.

Zuylen, H.J. van, F. Viti, F (2007) A probabilistic model for queues, delays and waiting time at controlled intersections, *TRB 86th Annual Meeting*, Washington DC, USA.

van Hinsbergen, C.P.I., J.W.C. van Lint (2008) Bayesian combination of travel time prediction models, *87th annual meeting of the Transportation Research Board*,Washington D.C.

Wei, C.H., Y. Lee (2007). Development of freeway travel time forecasting models by integrating different sources of traffic data, in: *IEEE Trans. Veh. Tech.*, 56(6), pp. 3682-3694.

You, J., T.J. Kim (2000) Development and evaluation of a hybrid travel time forecasting models, in: *Transport Research Part C: Emerging Technologies*, 8(1-6), pp. 231-256.

Yu, J.,G. Chang, H.W. Ho, Y. Liu (2008) Variation Based Online Travel Time Prediction Using Clustered Neural Networks, *11th International IEEE Conference on Intelligent Transportation Systems*, Beijing.

Zheng, F., H.J. Van Zuylen (2010) Uncertainty and Predictability of Urban Link Travel Time: A Delay Distribution Based Analysis, in: *Transportation Research Record*.