Andrzej Bargiela

TRAVEL TIME ESTIMATION THROUGH STATISTICAL CORRELATION OF INDUCTIVE LOOP READINGS

Abstract

Intelligent traffic and travel information systems depend critically on accurate, real-time estimates of urban travel times. The traditional survey-based estimation of travel times is not well suited for this purpose because of the cost and its inherent off-line processing delays. Modern, image-processing based surveys are capable of fulfilling thereal-time requirements but at a cost of some significant investment in the infrastructure. The lower-cost alternative of using lane-occupancy detectors has been tried with some success in the context of freeway travel time estimation but has been found not to be sufficiently accurate in urban traffic situations. This paper proposes a method based on statistical correlation of lane occupancy readings and shows that the method is capable of estimating urban travel times with sufficient accuracy and in real time.

1 INTRODUCTION

Early attempts to estimate link travel time were based on general notions of traffic flow and lane occupancy to estimate speed and consequently, travel time along a link. The relationship used was

$$speed = \frac{flow}{occupancy} * effective\_car\_length \tag{1}$$

In (Hall, etc., 1989) and (Pushkar, etc., 1994) the authors investigated the relationship of various factors in equation (1) and showed that the accuracy of this equation depends on such things like physical location and weather conditions. They also suggested that the relationship is prone to a systematic bias with respect to occupancy. Attempts to quantify such errors have been made by various researchers (Hall, 1989), (Pursula, 1995) and have prompted alternative approaches in which stochastic nature of the measurements was taken explicitly into account (Dailey, 1997, 1999), (Pursula, 1998), (Bargiela, etc., 2004).

The approaches based on Kalman filtering, (Dailey, 1997, 1999), implied the need for identification of several parameters which, in themselves, were not trivial to calculate. The problems associated with this approach seem to have been rooted in the implicit rather than explicit identification of the sources of errors in the fundamental traffic flow relationship (1). Refinements proposed by Coifman (Coifman, 2001) have produced some improvement but did not resolve the root problem.

In a more recent work (Petty, etc., 1998) a more general model of traffic flow has been proposed. This model has some similarity to the Dailey's crross-correlation approach described in (Dailey, 1993). The model is using stochastic variables and suggests practical procedures that can potentially lead to accurate travel time estimates. The underpinning assumption in this model is that travel times can be considered to be drawn from the same probability distribution. This allows estimation of probability distribution from the cumulative upstream and downstream arrival processes. However, there are several difficulties associated with the above methodology and these are widely recognised by the transportation researchers. First, the choice of the time windows, in which the upstream and downstream processes are considered, is not trivial. The results of empirical study, intended to inform such a choice, indicate that adaptive choice of parameters (depending on the traffic conditions) is necessary. Second, it is difficult to find "stationary" periods in traffic behaviour from which to draw measurement data. It is clear that the larger is such a period the better is the precision of the proposed model. However, there does not seem to be an easy way of identifying such a period other than monitoring transitions in traffic regime through some measurements of traffic density. Third, the choice of the appropriate level of data aggregation represents a significant challenge. Although it is possible to assess the effects of different levels of aggregation retrospectively, what is needed is the *a-priori* information about the right aggregation level.

This paper proposes a new methodology that is designed to overcome the shortcommings of current travel time estimation techniques that are based on stochastic models. Section 2 specifies the urban link model which serves as a basic reference for the description of travel time estimation methodology in Section 3. Some experimental results are provided in Section 4.

2 THE URBAN LINK MODEL

Most of the travel time estimation methods have been developed for freeway traffic. Unfortunately, urban links have several distinguishing features that do not allow direct application of the freeway methods. Firstly, the distance between consecutive junctions is relatively short compared to freeway links. On such a link, vehicle's travel time can be as little as 15 seconds, which makes the methods relying on aggregated traffic counters inherently inaccurate. Secondly, urban links typically have unobserved side-roads, which allow vehicles to enter and leave the link unnoticed. This makes the methods that rely on comparison of the entering and exiting distributions not applicable in urban traffic situations.

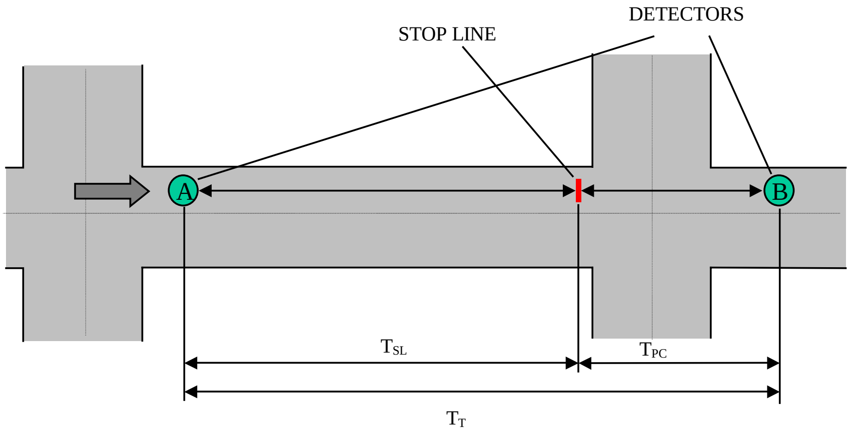

Figure 1. Urban link model and travel time components.

A typical urban link is shown in Figure 1. The main distinguishing feature for urban road links is the presence of a controlled junction in-between the detectors.
It is clear from the figure that total travel time (T_T) on an urban link can be split into two components, which are: T_{SL} – travel time from the upstream measurement point up to the stop line (SL) and T_{PC} – travel time from the stop line to the downstream measurement point. The purpose of such division is that the total time may often include the delay induced by the traffic control. It is worth noting that some UTMC systems (for instance, SCOOT) use T_{SL} in order to optimise traffic and produce optimal control. The division of total travel time into two components can also be used for estimation of other important characteristics of traffic such as turning movements (Bargiela, et all, 2005). In this paper we consider only the combined travel time T_T.

3 THE METHODOLOGY

Using notation from Figure 1, we assume that measurements that come from detectors *A* and *B* represent time of arrival of a vehicle at that particular detector. These measurements form a time series, or a process, of arrival times. Let us denote by $\{A(t), t \geq 0\}$ the series of upstream arrival times and by $\{B(t), t \geq 0\}$ the series of downstream arrival times.

It is further assumed, that traffic flows freely from *A* to *B*; i.e. we consider a subset of arrival times for which traffic does not stop at the stop line. Series of arrival times $A(t)$ and $B(t)$ are discrete and countable. Let us enumerate and denote individual arrival times of the above time series by $\{a_i\}$ and $\{b_j\}$. It is intuitively appealing that if a_i is the arrival time of a vehicle at the upstream detector and b_j is the arrival time of the same vehicle at the downstream detector, then subject to the free traffic flow condition, a_i and b_j are related through a relationship of the form

$$b_j = a_i + \varepsilon \tag{2}$$

where ε_k is a stochastic component that represents the delay between arrival of a vehicle at upstream and downstream points and takes into account fluctuation of speed with time. All ε_k, k=1, 2, …, can be considered as realisations of the same random variable ε. It is justifiable therefore that for short links, as is the case in urban networks, variable ε is considered to have a symmetrical bell-shaped distribution, such as normal. Then ε can be characterised by its mean and standard deviation: $\varepsilon = \varepsilon\,(T_T,\ \sigma)$. With these assumptions in mind the problem of travel time estimation essentially reduces to the problem of estimation of parameters of the random variable $\varepsilon\,(T_T,\ \sigma)$.

The methodology is as follows. Arrival times of downstream process are shifted into the past by a time τ; that is the following linear transformation is performed

$$b^*_j = b_j - \tau \tag{3}$$

Then the upstream time series is combined with the modified downstream time series in a way that keeps their relative order. Figure 2 illustrates this idea

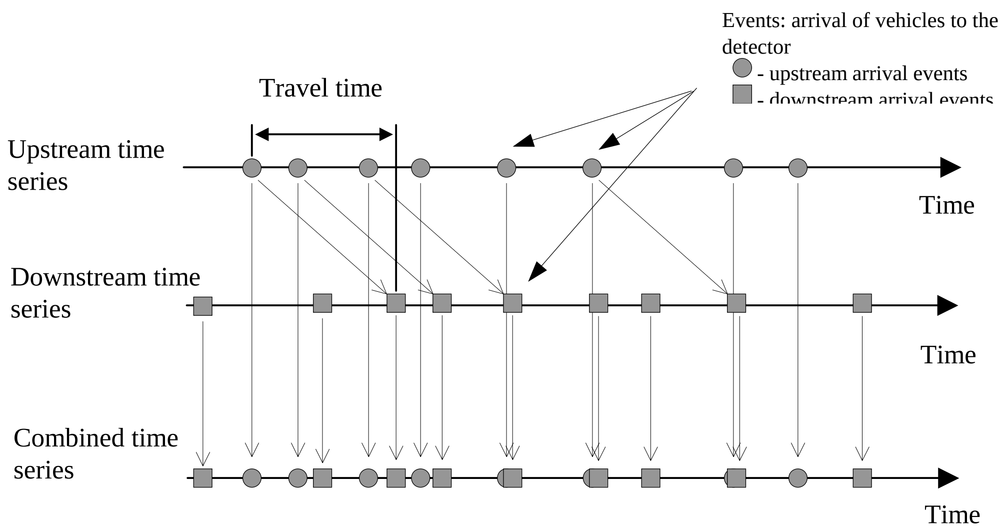

Figure 2. Merging upstream and shifted downstream time series

The dashed arrows point to the related events, i.e. link the upstream arrival events with the corresponding downstream arrival events. Using the combined process we extract pairs of neighbouring arrival times from upstream and downstream sensors. That is upstream arrival time is associated with the closest downstream arrival. The following Figure 3 illustrates the pair extraction procedure

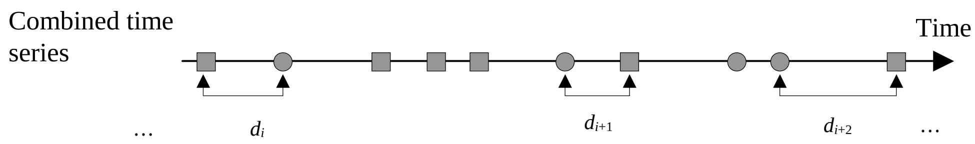

Figure 3. Pairs of neighbouring events that come from different time series

Although the optimal selection of pairs of events P_k={a_i, b_j} would require minimisation of the sum of corresponding time differences d_k over all possible choices of pairs (which would imply computational complexity O(N^2), where N is the number of elements in the sample of combined time series) it can be shown that local optimisation of pairing of events produces near-optimal results and has computational complexity of O(N). At every step in the local optimisation of pairing we assume that current position of an event in the resulting time series is k. There are six possible configurations of events following the event at k as given in the below Figure 4.

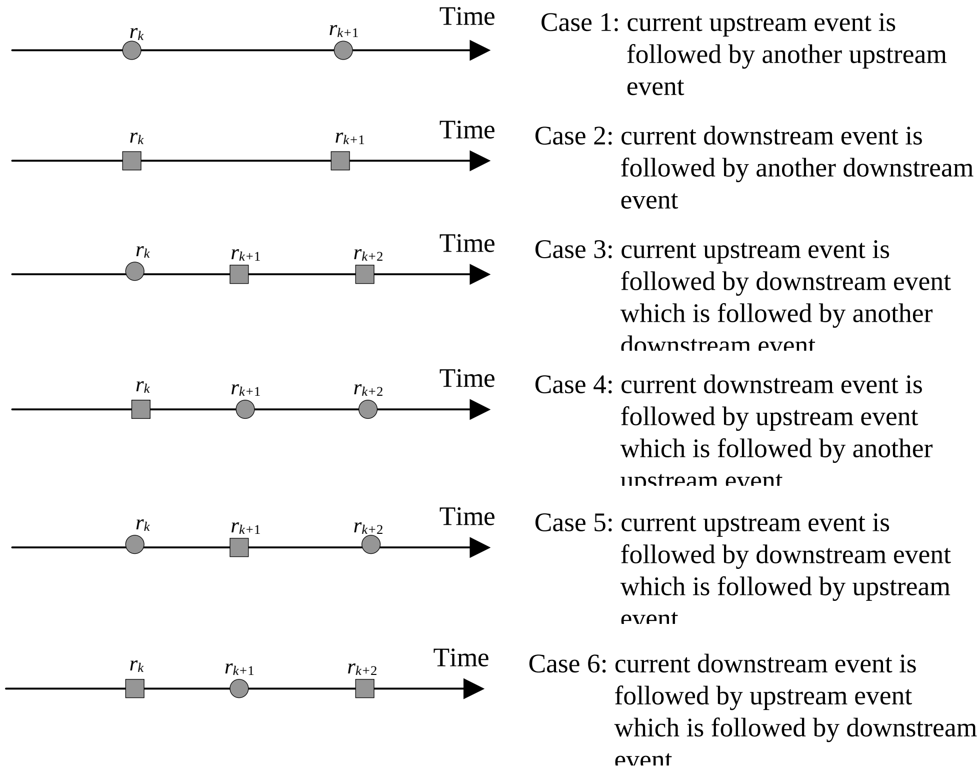

Figure 4. Combinations of events ahead of the current even at k

In the cases 1 and 2 above, the algorithm does not produce pairs since the current event at k cannot be associated with the following event since they have the same origin. If case 1 or case 2 is encountered at step k, the algorithm makes event k+1 its current and repeats the procedure. In cases 3 and 4 there are clear pairs formed by events at k and k+1, since event at k+1 cannot form a pair with event at k+2 because of their same origin. Thus, if case 3 or 4 is encountered, a pair of events at k and k+1 is formed and the algorithm makes the event at k+2 its current event for the next cycle. In case 5 and 6 it is not clear what pair is to be taken, the one formed by events at k and k+1 or events at k+1 and k+2. In this case, the algorithm forms a pair of such events, whose time difference is smaller. That is if $r_{k+1}-r_k < r_{k+2}-r_{k+1}$ then algorithm produces a pair formed by events k and k+1, and advances to k+2 as its current event for the next cycle otherwise it produces a pair formed by k+1 and k+2 and advances to k+3 as its current event for the next cycle.

The above algorithm ensures that if an event is associated with a pair, it is associated with one pair only. It performs a local minimisation of the time differences associated with pairs (as shown above for cases 5 and 6) and its computational complexity is O(N) where N is the number of events in the resulting time series.

Once a set of pairs P_k has been obtained, an average of time differences between the paired events is calculated. The calculation uses current time shift τ and consequently the result is a function of τ. Let us denote the arrival time of an upstream event that forms part of pair k by a^k and the arrival time of a downstream event that forms part of the same pair by b^k. Then, consider the following *cost function.*

$$F(\tau) = \frac{1}{N_P} \sum_{k=1}^{N_P} \left| a^k - b^k \right| \tag{4}$$

where N_p is the number of pairs.

All pairs selected by the algorithm can be split into two classes – the ones that are formed by independent upstream and downstream events and those formed by the events that are related through equation (2). Assume that the set of independent pairs is denoted by Ω_i and the set of dependent pairs is denoted by Ω_d. The cardinalities $N_i=|\Omega_i|$ and $N_d=|\Omega_d|$ satisfy $N_p = N_i + N_d$ and consequently (4) can then be written as the following sum.

$$F(\tau) = \frac{1}{N_P} \Big(\sum_{k:P_k \in \Omega_i}^{N_i} \left| a^k - b^k \right| + \sum_{l:P_l \in \Omega_d}^{N_d} \left| a^l - b^l \right| \Big) =$$

$$\frac{N_i}{N_P} \left(\frac{1}{N_i} \sum_{k:P_k \in \Omega_i}^{N_i} \left| a^k - b^k \right| \right) + \frac{N_d}{N_P} \left(\frac{1}{N_d} \sum_{l:P_l \in \Omega_d}^{N_d} \left| a^l - b^l \right| \right) =$$

$$k\overline{X_i}(\tau) + (1-k)\overline{X_d}(\tau); \quad 0 \le k \le 1 \tag{5}$$

where $\overline{X_i}(\tau)$ is the average of time differences $\{d_i\}$ for independent pairs, $\overline{X_d}(\tau)$ is the average of time differences $\{d_j\}$ for dependent pairs.

Although it is difficult to express analytically the relationship between differences $\{d_i\}$ produced by independent pairs (with events shifted by time τ), it can be anticipated that such a relationship will be weaker (or more irregular) than that of the differences produced by dependent events. Using (2) and (3), the relationship between dependent difference d_j and τ can be expressed as

$$d_j(\tau) = b_j^* - a_i = b_j - \tau - a_i = \varepsilon(T_T, \sigma) - \tau \qquad (6)$$

It can be shown that if τ_m is defined as

$$\tau_m = \arg\min E\left|\varepsilon(T_T, \sigma) - \tau\right| \qquad (7)$$

then τ_m is the median of the random variable $\varepsilon(T_T, \sigma)$. In the case when $\varepsilon(T_T, \sigma)$ has a bell-shaped symmetrical distribution and its first two moments exist and are finite, then the median equals the mean and equals the mode of $\varepsilon(T_T, \sigma)$. Therefore, if the mean of independent differences does not change significantly with τ, then it is reasonable to expect (5) to reach its minimum in the point $\tau = T_T$.

The above arguments can be considered to be a heuristic, as they cannot be proven formally. Nevertheless, the rationale of this calculation is strongly appealing and a number of experiments with real traffic data have shown it to be correct. Another argument supporting the correctness of this approach is that when τ approaches T_T, coefficient ($1-q(\tau)$) in (5) tends to grow while $q(\tau)$ itself decreases provided that σ of $\varepsilon(T_T, \sigma)$ is smaller than the average time gap between consecutive arrival events. This, in turn, makes the second term of the sum (5) having stronger influence on the whole sum and thus minimum in the point $\tau=T_T$ is being emphasised more strongly.

Several modifications of the presented algorithms can be suggested so as to extract other useful information from the time series of arrival events. The methodology demonstrates several advantages over previously developed approaches. These can be summarised as follows.

1. The methodology operates at the resolution of data and no aggregation is required thus yielding the most accurate estimates possible.
2. The method is robust with respect to undercounting and overcounting of vehicles as well as uncounted vehicles that are leaving and joining the link. Such events will be related to the independent pairs or excluded from consideration automatically
3. The presented algorithm is efficient. The complexity of the algorithm is *kN*, where *k* is the number of steps (depends on the method used for generating the steps), and *N* is the number of events in both the upstream and downstream series of arrival times. It is worth noticing that the correlation-based algorithm (Dailey, 1993) is also *MN*, where *M* is the number of time

lags and N is the number of elements of the aggregated arrival time series (both the downstream and the upstream), but M is normally bigger than k and N does not depend on the number of actual events and in the case of small number of arrival events, the proposed methodology will have better performance.

4. The set of time differences produced by the pair extraction algorithm can be used to estimate other useful parameters, such as variance, of the travel time as a random variable.

Several methods of choosing the step $\Delta\tau$ of the algorithm can be proposed. The simplest way is to let $\Delta\tau$ be a constant equal to the smallest unit of the time resolution of the data. In this case the optimum solution is guaranteed to be found.

4 EXPERIMENTAL RESULTS

The proposed methodology has been used in an empirical study of a SCOOT controlled region of the town of Mansfield, UK. The following figures present the plots of the cost functions calculated for one of the links of the region.

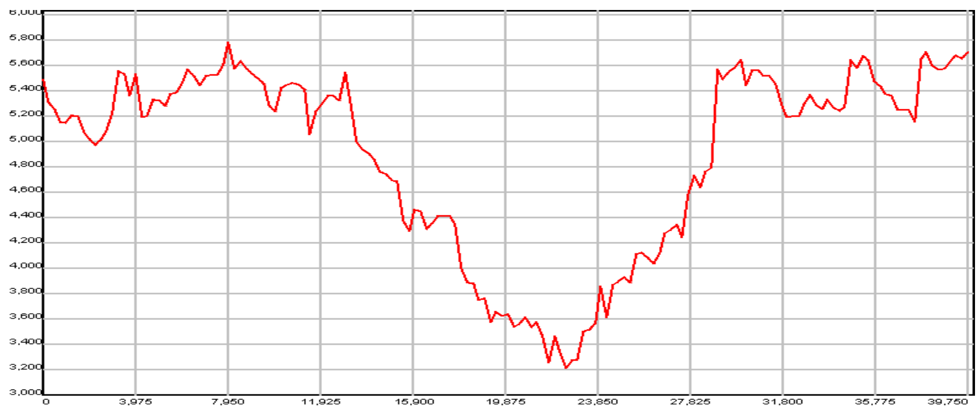

Figure 5. Cost function calculated over one hour period.

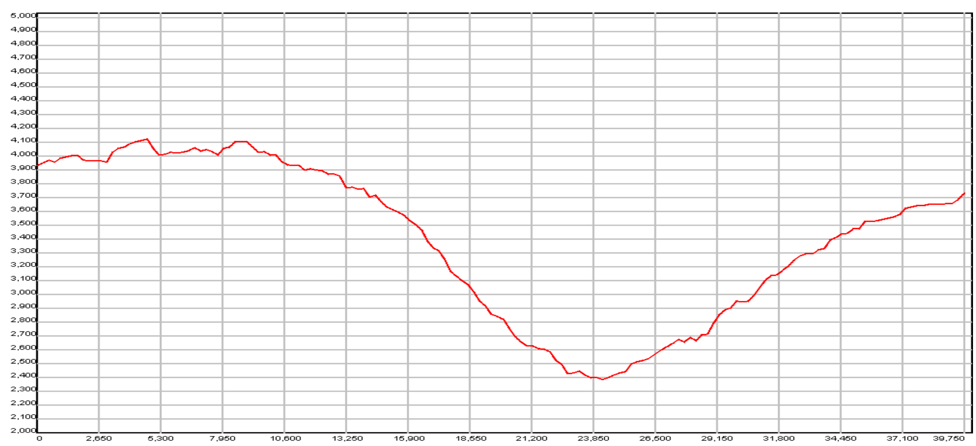

Figure 6. Cost function calculated over whole day

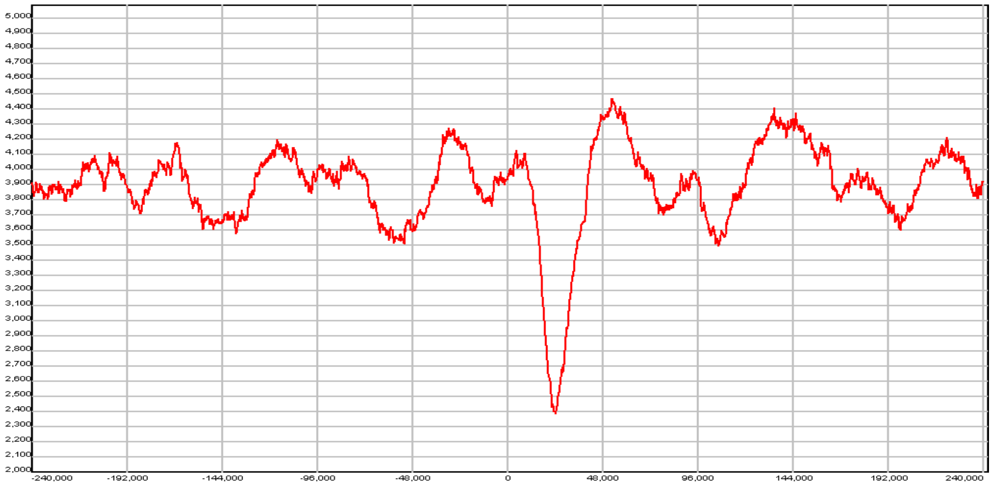

Figure 7. The cost function calculated over a broader range of time shifts.

Figure 7 illustrates the cost function calculated over a range of time shifts (from –4 min to 4 min). It is clear from the plot that there is only one point of the "best match" between the upstream and downstream series which is located at the point of 24.25 seconds. The cost function behaves as expected: its bigger values correspond to the areas of time shift τ where most of the events are uncorrelated and smaller values correspond to the neighbourhood of 24.25 seconds where many events are correlated.

5 CONCLUSIONS

The proposed approach to travel time estimation, based on statistical correlation of inductive loop readings, has been justified as a strong meta-heuristic and has proven itself in the analysis of real-life data. Although the methodology has been derived in the context of urban travel time estimation, it is quite general and can be applied in other situations where the problem can be reduced to finding a delay between two stochastic point processes.

The proposed method of statistical correlation represents significant improvement over previously developed approaches to travel time estimation from inductive loop readings.

LIST OF REFERENCES

Bargiela A, Kosonen I, Pursula M, Peytchev E. (2004). *Granular analysis of traffic data for turning movements estimation,* International Journal of Enterprise Information System, vol. 2, 2, 2006, 13-27.

Coifman, B. (2001), *Improved Velocity Estimation Using Single Loop Detectors.*

Transportation Research: Part A, vol 35, no 10, pp. 863-880

Dailey, D.J. (1993), *Travel time estimation using cross-correlation techniques*. Transportation Research, Part B, 27B (2), 97-107

Dailey, D.J. (1997), *Travel time estimates using a series of single loop volume and occupancy measurements*. Transportation Research Board, 76th Annual Meeting, Washington

Dailey, D.J. (1999), *A statistical algorithm for estimating speed from single loop volume and occupancy measurements*. Transportation Research B. Vol. 33, pp 313-322

Hall, F.L., Persaud, B.N. (1989), *Evaluation of speed estimates made with single-detector data from freeways traffic managemtnt systems*. Transportation Research Record, (1232):9-16

Petty, K., Bickel, P., Jiang, J., Ostland, M., Rice, J., Ritov, Y., Schoenberg, F. (1998), *Accurate estimation of travel times from single-loop detectors*. Transportation Research, Part B: Methodological, 32(1):1-17

Pursula M. (1995). *The approximation of percentage time delay with local measurements.* Transportation Research Record 1484. Washington, D.C. p.58-65.

Pursula M. (1998). *Two dimensional estimation of speed flow relationships with backpropagation neural networks*. In: Himanen V., Nijkamp P., Reggiani A., Raitio J. (Eds.). Neural Networks in Transport Applications. Ashgate, Aldershot, England. p. 263-294..

Pushkar, A., Hall, F.L., Acha-Daza, J.A. (1994), *Estimation of speeds from single-loop freeway travel times in real time from flow measurements*. Journal of Transportation Engineering 122, pp. 185-191