

# On the estimation of arterial route travel time distribution with Markov chains

Mohsen Ramezani, Nikolas Geroliminis\*

*School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

## ARTICLE INFO

### Article history:

Received 24 August 2011

Received in revised form 22 August 2012

Accepted 22 August 2012

### Keywords:

Grid clustering

Markov chain

Probe vehicle

Travel time distribution

Travel time variability

## ABSTRACT

Recent advances in the probe vehicle deployment offer an innovative prospect for research in arterial travel time estimation. Specifically, we focus on the estimation of probability distribution of arterial route travel time, which contains more information regarding arterial performance measurements and travel time reliability. One of the fundamental contributions of this work is the integration of travel time correlation of route's successive links within the methodology. In the proposed technique, given probe vehicles travel times of the traversing links, a two-dimensional (2D) diagram is established with data points representing travel times of a probe vehicle crossing two consecutive links. A heuristic grid clustering method is developed to cluster each 2D diagram to rectangular sub spaces (states) with regard to travel time homogeneity. By applying a Markov chain procedure, we integrate the correlation between states of 2D diagrams for successive links. We then compute the transition probabilities and link partial travel time distributions to obtain the arterial route travel time distribution. The procedure with various probe vehicle sample sizes is tested on two study sites with time dependent conditions, with field measurements and simulated data. The results are very close to the Markov chain procedure and more accurate once compared to the convolution of links travel time distributions for different levels of congestion, even for small penetration rates of probe vehicles.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Introduction

Nowadays, traffic congestion is a widespread time-consuming phenomenon in urban areas and the primary step for improving conditions is traffic observation and data collection. Hence, network monitoring is a crucial component in management of transportation systems for traffic control and guidance purposes. The introduction of Intelligent Transportation System (ITS) technologies and new sensing hardware promise significant progress in reducing the congestion level in cities. The integration of Global Positioning System (GPS) technology within the ITS framework introduces a new paradigm in traffic surveillance: probe vehicles. Compared to fixed traffic sensors (e.g. inductive loop detectors), probe vehicles offer further data like vehicle trajectory in a more convenient manner. In principle, a steadily incremental public deployment rate, low maintenance cost, and inherent distributed characteristics lead to tackling GPS-equipped vehicle challenges in traffic monitoring research. Nevertheless, the estimation methodology should not be constrained to GPS information from cars, but may be applied to any type of Automatic Vehicle Location (AVL) mobile sensors found in abundance in the form of commercial fleets like UPS, FedEx, taxis and transit vehicles, given the broad existence of filters which distinguish between actual congestion and a stopping delay (see for example Bertini and Tantiyanugulchai, 2004, for transit vehicles as probes or Geroliminis and Daganzo, 2008; Herrera et al., 2010 for other vehicles). The chief struggle with utilizing probe data is that

\* Corresponding author.

E-mail addresses: [mohsen.ramezani@epfl.ch](mailto:mohsen.ramezani@epfl.ch) (M. Ramezani), [nikolas.geroliminis@epfl.ch](mailto:nikolas.geroliminis@epfl.ch) (N. Geroliminis).

travel time of a probe vehicle is principally a sample of a random variable, i.e. travel time. This raises inquiries about the probe vehicles penetration rate required to have a proper sample set for travel time estimation.

Travel time is a crucial index in assessing the operational efficiency of traffic networks. It establishes a common perception among all the perspectives of individual travelers and practitioners. In addition, it can be an indicator of congestion level of transport network once compared to the free flow travel time. With respect to monitoring, reliable and efficient estimation of travel time and other performance measures is still not a wide spread accomplishment on arterials, since it requires extensive sensor infrastructure, normally found only on freeway networks. The issue is not only that the existing monitoring infrastructure in arterials is less dense than in freeways but also that arterial network traffic dynamics are inherently different than these of freeways and fixed sensors cannot always provide the required level of data. The main reasons are randomness in supply and demand of the dynamic urban network (see for example [Du et al., 2012](#)), the signaling effect (alternation of green and red phases in short time intervals), and the characteristics of route choice (vehicles in arterials can randomly turn at intersections and either begin or finish their trips along the street itself, which is not the case in freeways). Meanwhile, speed of vehicles at a given time in the network is not a deterministic quantity over space because of drivers' behaviors (conservative vs. aggressive drivers), the spatial effect of signals (near the stop line vs. further upstream) and temporal-spatial pockets, where average speed is temporarily different than the widespread average, e.g. point bottleneck in a freeway system. To simplify matters, if we track a vehicle on a freeway for 15 s we can estimate with high level of confidence the mobility level in the time-space proximity of the vehicle. Instead, this is not the case in arterials as there are variations in travel times even for vehicles traveling in the same link during the same cycle length.

Reduction in travel time variability is at least as desirable as reduction in mean travel time ([Jenelius, 2012](#)), since it decreases commuting stress and uncertainty of mode- and route-choice decision making. Travel time variability designates the variation of various trip travel times over a specific path. Travel time variability can be investigated from several point of views ([Noland and Polak, 2002](#)): vehicle-to-vehicle variability which corresponds to different vehicles traveling the same route at the same time, period-to-period variability corresponding to vehicles traveling the same route at different periods within a day, and day-to-day variability addressing the travel time variations of vehicles crossing the same route at the same period of time on different days. Different indexes of travel time stochasticity-reliability are presented in ([Kaparias et al., 2008](#)). In this paper, we model the vehicle-to-vehicle variability and analyze the probability distribution of travel time for arterial routes (expressed as series of links). The input to the proposed model is probe vehicles travel times of all links that are traversed and belong to the route; and the output is the Travel Time Distribution (TTD) over the study time horizon. The developed model is based on Markov chain to address both the traffic progression and correlation between links.

Note that the urban TTD should not be estimated for periods less than one cycle, because this might over- or under-estimate travel times depending on the period of observation (red or green period). Also the analysis of the travel time characteristics for a short period, e.g. two or three times the cycle length, can significantly be influenced by the size of study period. For example, choosing the start or end of the study period at the beginning or end of the cycle could cause drastic changes in the results ([Zheng and Van Zuylen, 2010](#)). However, by analyzing longer study periods; we can smooth out the traffic variations caused by time-dependent signal capacity. Nevertheless, we stress that the period of analysis should not be too large, as in this case the spatial correlations between links may be influenced by variant traffic regimes. In the latter case, classifying the period of analysis to traffic regimes is required.

The remainder of this paper is organized as follows. Section 2 provides a literature review of travel time estimation. In Section 3, we briefly discuss the Markov chain procedure, its application in traffic engineering, and we illustrate our motivation via an example. Then, we introduce our proposed methodology in Section 4. The study site, data, and simulation details are presented in Section 5, while results and discussion are described in Section 6. Finally in Section 7, conclusions are drawn and future work is summarized.

## 2. Travel time estimation literature review

There is a vast literature addressing different travel time estimation approaches for diverse applications and terms. Initial approaches for travel time estimation on signalized links include point delay models ([Webster, 1958](#); [Allsop, 1972](#)), speed vs. volume to capacity ratio relations, and procedures based on the Highway Capacity Manual ([HCM, 2000](#)). The latter calculates average travel time as the sum of the running time and the intersection delay, based on a deterministic point delay model plus some stochastic components. Such approaches are not well-suited for real-time applications, especially for time-dependent congested conditions. The results from the application of several conventional approaches on two arterial sites indicate that these methods produce large differences against field measured and predicted travel times ([Tsekeris and Skabardonis, 2004](#)). The same reference presents a detailed review of arterial travel time models.

Recently, an analytical model was developed to estimate the travel times on arterial streets based on 15–30 s flow and occupancy data provided by loop detectors and the signal settings at each traffic signal ([Skabardonis and Geroliminis, 2005](#)). The model considers the spatial and temporal queuing characteristics at the traffic signals and the signal coordination to estimate travel time in arterials. Several extensions and enhancements to the analytical model were developed and implemented by the same authors which explicitly address the issues of long queues and spillovers that frequently occur on arterials in urban areas ([Skabardonis and Geroliminis, 2008](#); [Geroliminis and Skabardonis, 2011](#)). The model has also been integrated into a pilot arterial performance measurement system in California (APemS). A similar model was also developed by [Liu and Ma \(2009\)](#), which utilizes the exact times that the vehicles cross the upstream detector (from individual vehicle detector actuations).

The aforementioned approaches require precise signal phase times and instrumentation at each intersection, which might be expensive or non-existing. Also, these studies focused on estimation of *average travel time* of vehicles traveling in one cycle time. Although statistical scalar indexes (e.g. mean, variance, percentiles, etc.) might characterize distributions to some extent, they are not fully enlightening about travel time variability, once compared to TTD. An alternative to characterize the TTD could be to introduce an appropriate number of percentiles and approximate the TTD, by choosing these percentiles when sharp changes occur in the slope of cumulative probability distribution. In this paper, we utilize the full TTD approach.

To address the uncertainty issue, a new trend is seen in recent travel time research. Kwong et al. (2009) developed a methodology for vehicle re-identification using wireless magnetic sensors in arterial routes, without the need of signal settings. The high accuracy of matching vehicle signatures between different locations provided accurate estimates of empirical travel times (mean and distribution) for different locations. In (Guo et al., 2010), a multistate model is employed to fit a mixture of Gaussian distributions into travel time observations of an expressway corridor. Each normal distribution is associated with an underlying traffic state providing quantitative uncertainty evaluation. The multistate mixture model results in better fitting, revealing that TTD usually has more than one mode which is entirely dependent on time horizon of study, demand, topology, etc. A recent work of the same group (Park et al., 2011) tries to quantify the impact of traffic incidents on TTD using multi-state (3 states) models demonstrating that incidents increase the travel time variability. The results indicate that the state corresponding to the congested regime become more dominant, yet there is no need to add a new component to the multi-state model. A similar approach is investigated, utilizing mixtures of normal distributions to estimate mean travel times for arterial routes with Next Generation Simulation (NGSIM) data (Feng et al., 2011).

Uno et al. (2009) discuss route travel time variability using bus probe data. In pre-processing stage of their method, a map matching procedure and a data filtering for dwell time elimination are performed. Afterwards, they decompose a route to sections with lognormal TTD to estimate the bus route TTD. In (Ferman et al., 2005) a statistical evaluation is investigated, which tries to assess the feasibility of probe vehicles employment for collecting traffic information. The authors provide analytical solutions about data sampling, reporting rates and probe vehicles penetration level in a single road link without any validation. They assume a binomial distribution for number of probe vehicles and a Poisson distribution for reporting rate, deriving formulas for mean and variance of reports number and speed estimation, and confident reporting intervals. Delay at signalized intersections is the main source of uncertainty in urban TTD which is tackled in (Zheng and Van Zuylen, 2010). The authors propose an analytical method for estimation of an urban link delay distribution. The results indicate a correlation between arrival time and link travel time under different degree of congestion. Their model demonstrates evolutions of delay distribution as well, so that both average and variance of delay increase cycle by cycle. Non-recurrent and peculiar events, such as incidents, lane closures, and sport events can also cause significant deviations from recurrent conditions (Kwon et al., 2011).

Sparse probe vehicles is discussed in Herring et al. (2010) for arterial traffic estimation and short term prediction of travel time. The authors propose a statistical modeling framework that captures the evolution of traffic flow as a Coupled Hidden Markov Model (CHMM). The authors assume that each link in the network has one state evolving over time based on a time-invariant state transition matrix. Given the states of the spatial neighbors of the link, independence of state transitions from all other current and past link states is assumed. They also consider a time-invariant travel time distribution (Gaussian) for each state of each link, representing independence of link travel time from other traffic variables, given the link state. This work also decomposes path travel times from probe data to individual link travel times. The evaluation is done using dataset from a taxis fleet in San Francisco, CA, as a part of *The Mobile Millennium Project*. To model the correlation and dependency in transport networks, Rakha et al. (2006) have tried to estimate the variance of freeway route travel time by modeling correlation between segment travel time variances. Geroliminis and Skabardonis (2006) also estimated the variance of urban route travel time by assuming linear correlations between successive links travel times. To the best of our knowledge, correlation between travel times in urban links has not been explicitly addressed in the recent literature.

### 3. Motivation

Given all individual link TTDs, the simplest model for route TTD estimation is to aggregate those independently. Assume a route consisted of  $K$  links with signalized intersections. The route TTD is then computed according to:

$$TTD_K = TTD_1 * TTD_2 * \dots * TTD_k, \quad (1)$$

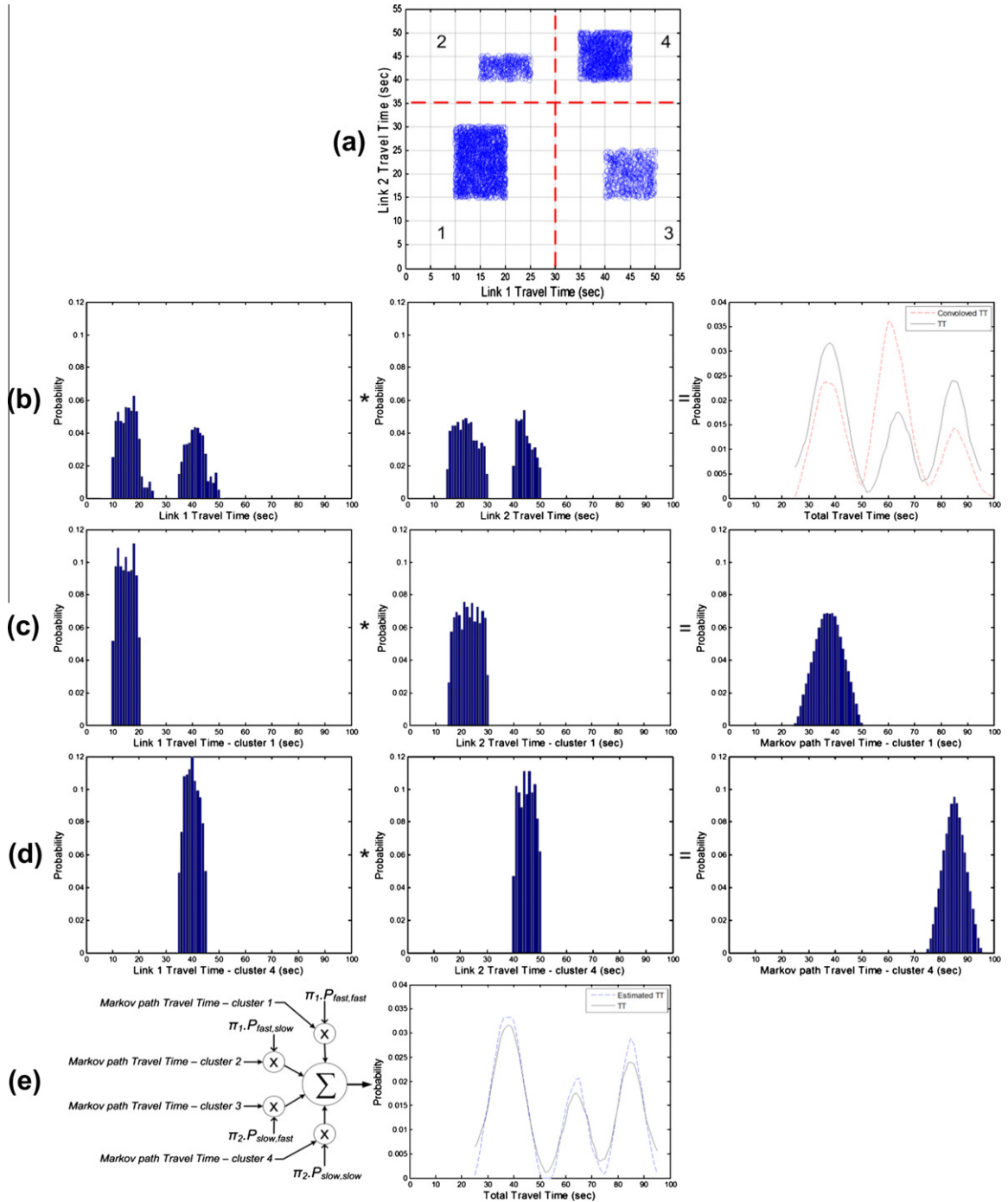
$$(TTD_i * TTD_j)(t) \triangleq \int_{-\infty}^{\infty} TTD_i(\tau) TTD_j(t - \tau) d\tau, \quad (2)$$

where the (\*) mathematical operator expresses convolution (Killmann and Collani, 2001) and the left term in (2) is the probability density over time  $t$  for two links ( $i, j = 1, 2, \dots, k$ ) given both TTDs a priori. Evidently, the above method considers independence between link TTDs, and consequently, any spatiotemporal correlation information is neglected.

To show the impact of correlation in estimation of TTD, we demonstrate a hypothetical ‘toy example’ in which there is a strong correlation between travel time data. Imagine a route, which consists of two serial signalized links. Fig. 1a depicts a 2D diagram in order that each point denotes links 1 and 2 travel times of one probe vehicle. In this case, we can infer there

are four types of vehicles; fast in both links (left-lower part), fast in link 1 and slow in link 2 (left-upper part), slow in link 1 and fast in link 2 (right-lower part), and slow in both links (right-upper part). Given the link TTDs in Fig. 1b, the outcome of the convolution method (convolved route TTD) is far from the real TTD.

To capture correlation patterns between link travel times, one can cluster travel times of one link to different states, e.g. in this example 2 states (slow and fast) are defined for each link. The initial probability,  $\pi$ , of each of the states in link 1 and the transition probabilities,  $P$ , between states of link 1 and states of link 2 as traffic progresses are (numbers represent the toy example)



**Fig. 1.** Explanatory example about motivation of Markov chain integration into TTD estimation; (a) 2D diagram representing four groups of vehicles, (b) convolution of mapping of all vehicles link 1 travel time and mapping of all vehicles link 2 travel time distributions, (c) convolution of fast vehicles link 1 and 2 travel time distributions, (d) convolution of slow vehicles link 1 and 2 travel time distributions, (e) mixture of TTD of all groups of vehicles.

$$\pi = \begin{bmatrix} \pi_{fast} \\ \pi_{slow} \end{bmatrix} = \begin{bmatrix} \frac{N(1)+N(2)}{N(1)+N(2)+N(3)+N(4)} \\ \frac{N(3)+N(4)}{N(1)+N(2)+N(3)+N(4)} \end{bmatrix} = \begin{bmatrix} 0.58 \\ 0.42 \end{bmatrix}, \tag{3}$$

$$P = \begin{bmatrix} P_{fast,fast} & P_{fast,slow} \\ P_{slow,fast} & P_{slow,slow} \end{bmatrix} = \begin{bmatrix} \frac{N(1)}{N(1)+N(2)} & \frac{N(2)}{N(1)+N(2)} \\ \frac{N(3)}{N(3)+N(4)} & \frac{N(4)}{N(3)+N(4)} \end{bmatrix} = \begin{bmatrix} 0.84 & 0.16 \\ 0.29 & 0.71 \end{bmatrix}, \tag{4}$$

where  $N(i), i = 1, \dots, 4$  is the number of points inside each cluster (as shown in Fig. 1a),  $\pi_a, a \in \{slow, fast\}$  is the initial probability of vehicles being slow or fast in link 1, and  $P_{a,b}, a, b \in \{slow, fast\}$  is the corresponding probability of being slow and/or fast in links 1 and 2, respectively. Given 2 states for each link, there are 4 combinations of states, hereafter named as Markov path. Every one of 4 Markov paths has a probability of occurrence and a corresponding TTD. The TTD of each Markov path refers to the convolution of partial TTD of link states where the TTD of a Markov path is conditioned on the states of the links in the corresponding Markov path. The partial TTD of link states is the TTD of a link conditioned on the state of that link (The complete definition of partial TTD is given in Section 4.1). For the sake of brevity, only the TTDs for Markov paths of fast–fast and slow–slow states are depicted in Fig. 1c and d, respectively. At the end of procedure, the TTD of a Markov path is multiplied by the path initial and occurrence probabilities and the mixture is calculated to find the route TTD. The final result of estimation method is closely matched with the real TTD as it is apparent in Fig. 1e.

Note that in the above example, the structure of correlation is known a priori which makes the state definition straightforward. Nevertheless, improper state definition (erroneous clusters) will degrade the significance of the results. Further, travel times of consecutive links in reality might not be well-ordered as in Fig. 1a. Instead, they are more similar to Fig. 2 which makes the state boundaries less clear and the privilege effect of partitioning less straightforward. This contamination of data leads to a grid clustering problem, yet the whole concept remains the same. The above example can be applied to several links using Markov chain procedure which is briefly reviewed in next subsection. The complete version of methodology is presented in Section 4.

### 3.1. Markov chain procedure

Markov chain is a technique for statistical modeling of a random process in which the state of system changes through progression. A Markov chain is entirely demonstrated with the set of state definition, initial probabilities and transition probabilities. The transition probabilities are associated with the manner of state progression during the system evolution. A system which has the Markov property satisfies the following: the conditional probability of the system being at the next state,  $s_{t+1}$ , given the current state,  $s_t$ , depends only on the current state and not on the previous states of the system. Mathematically speaking:

$$\Pr\{s_{t+1} = s' | s_t, s_{t-1}, \dots, s_1\} = \Pr\{s_{t+1} = s' | s_t\}. \tag{5}$$

The Markov property empowers Markov chain to capture both probabilistic nature of travel time and the fundamental correlated feature of successive links travel times. In other words, *traffic spatial progression* in arterials is similar to a Markov chain where the current link travel time of a vehicle depends only on the travel time of immediate upstream link, equivalent to the right term in (5), which is well-matched with physics of traffic. Note that, the traffic spatial progression designates

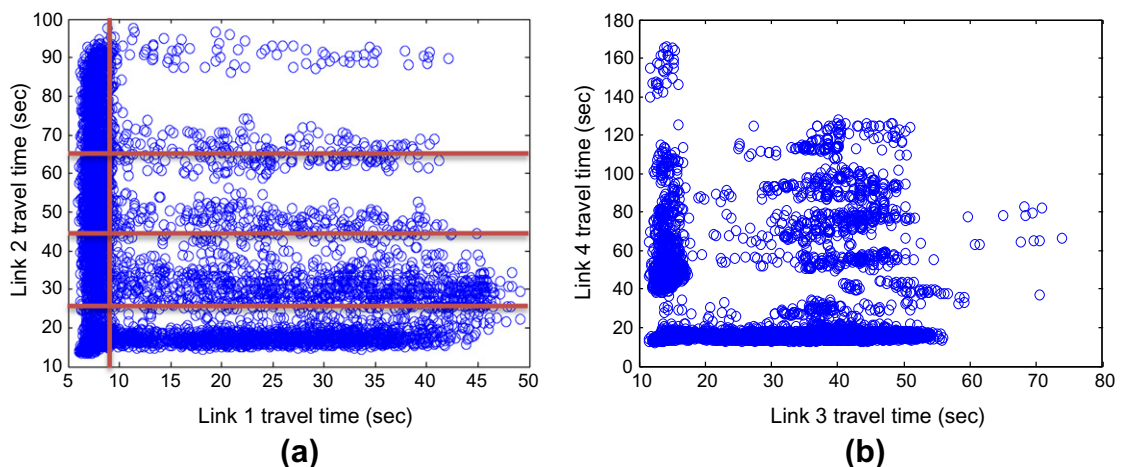


Fig. 2. 2D diagrams showing joint distributions of successive link travel times; (a) links 1–2, (b) links 3–4.



how the moving vehicles in the arterials encounter travel time states while traveling along links; and this definition is different from the evolution of link congestion levels (congestion propagation).

Markov chain is utilized in vast fields of transportation research. Discrete time Markov chain for estimation of expected freeway travel time is investigated in (Yeon et al., 2008) where states correspond to congestion level of links. The authors find the average travel time of each link both in non-congested and congested states using field data and with consideration of transition probabilities between different states, the route mean travel time is estimated. In Dong and Mahmassani (2009) a Markov chain is developed to model the effect of freeway flow breakdown and recovery in travel time reliability. Geroliminis and Skabardonis (2005) also proposed an analytical model using Markov chain for prediction of platoon arrival profiles and queue length considering platoon dispersion in arterials. The CHMM have been also applied in (Herring et al., 2010) for traffic estimation and prediction.

#### 4. Methodology

In our proposed model, the raw measurements are experienced individual link travel times traversed by a set of probe vehicles. A probe vehicle path may consist of one to as many links as the study route has, which makes the number of link travel times reported by every probe vehicle different. With high resolution GPS data (position and time stamp), finding the trajectory of a moving vehicle and link travel times is not a complicated task. Note that errors in trajectories, map matching, or low resolution data are not addressed in this work. Afterwards, travel times of all probe vehicles crossing two successive links during data collection period are used to construct a 2D diagram. The 2D diagram is a graphical representation of vehicles travel times joint distributions. Given a route consisting of  $K$  links,  $K - 1$  2D diagrams are established in order to identify the Markov chain, i.e. defining its states and determining the initial and transition probabilities. Fig. 2a and b illustrate such a diagram for links 1–2 and links 3–4 of a study site that will be described later. Each dot represents the travel time of one vehicle in each of the two links. Fig. 2a depicts a significant fraction of vehicles crosses link 1 or 2 without any delay, but also a large variation of travel times. Note that there is some correlation between travel times in successive links, which cannot be expressed in a linear way (linear correlation is almost zero).

##### 4.1. Markov chain identification

As 2D diagrams are constructed, states, transition probabilities and initial state probabilities of Markov chain should be identified. We define a state of a link as travel times inside a certain interval between two values. Let  $\mathcal{X}_l = \{x_1^l, \dots, x_{m_l}^l\}$  and  $\mathcal{Y}_l = \{y_1^l, \dots, y_{n_l}^l\}$  denote sets of boundaries in 2D diagram  $l$  (correspond to links  $l$  and  $l + 1$ ) producing  $m_l$  and  $n_l$  states for link  $l$  and  $l + 1$ , respectively. In this manner, the first state of link  $l$  indicates travel times in  $[\min \tau_l, x_1^l)$  and the last state represents travel times within  $[x_{m_l}^l, \max \tau_l]$ , where  $\tau_l$  denotes the set of probe vehicles travel time measurements in link  $l$ . This kind of state definition yields to rectangular clusters in 2D diagrams which is used to define initial state and transition probabilities. A possible clustering of a 2D diagram is depicted in Fig. 2a where there are 2 states in link 1 and 4 states in link 2. The clustering procedure will be more elaborated in Section 4.2.2.

Initial state probabilities in Markov chain are the probabilities of link 1 states which are as follows:

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_{m_1} \end{bmatrix} = \begin{bmatrix} \frac{N(1)}{\sum_{i=1}^{m_1} N(i)} \\ \vdots \\ \frac{N(m_1)}{\sum_{i=1}^{m_1} N(i)} \end{bmatrix}, \tag{6}$$

where  $N(i)$  denotes number of data in state  $i$  of link 1 (i.e. link 1 travel times in  $[x_{i-1}^1, x_i^1)$ ). To identify transition matrix between two successive links, transition probabilities should be defined. The generic transition matrix between each pair of successive links is as following:

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,n_1} \\ \vdots & & \vdots \\ p_{m_l,1} & \cdots & p_{m_l,n_1} \end{bmatrix} = \begin{bmatrix} \frac{N(1,1)}{\sum_{i=1}^{m_l} N(1,i)} & \cdots & \frac{N(1,n_1)}{\sum_{i=1}^{m_l} N(1,i)} \\ \vdots & & \vdots \\ \frac{N(m_l,1)}{\sum_{i=1}^{m_l} N(m_l,i)} & \cdots & \frac{N(m_l,n_1)}{\sum_{i=1}^{m_l} N(m_l,i)} \end{bmatrix}, \tag{7a}$$

$$p_{i,j} = \Pr\{S_{l+1} = j | S_l = i\}, \tag{7b}$$

where  $S_l$  indicates the state of travel time in link  $l$ , and  $N(i,j)$  is the number of data points within each rectangular cluster confined by state  $i$  in link  $l$  and state  $j$  in link  $l + 1$ .

By definition of 2D diagrams, travel time observation of each link (except the first and the last links) are used in two 2D diagrams, one as upstream link in  $x$ -axis and the other as downstream link in  $y$ -axis. If the states for link  $l$  are identical in both 2D diagrams ( $l - 1$ ) and  $l$ , i.e.  $\mathcal{Y}_{l-1} = \mathcal{X}_l$  and  $m_l = n_l, l = 1, \dots, k$ , there will be  $Q = \prod_{l=1}^k m_l$  state combinations from origin

to the destination of the route and each one of them is named as a Markov path (violation of this condition is discussed in Section 4.2.2). For a given Markov path, all of the transition probabilities between states of all links are multiplied to compute Markov path occurrence probability (note that Markov property makes joint probability of each cluster of 2D diagrams independent),

$$\Pr\{S_1 = i_1, S_2 = i_2, \dots, S_k = i_k\} = \pi_{i_1} p_{i_1, i_2} p_{i_2, i_3} \dots p_{i_{(k-1)}, i_k}. \quad (8)$$

Then we obtain a Markov path TTD using convolution of partial TTD of each link state, assuming a conditional independence between partial TTDs (hence state definition is of great importance to gain the conditional independence),

$$\text{TTD}\{S_1 = i_1, S_2 = i_2, \dots, S_k = i_k\} = \text{TTD}(i_1) * \text{TTD}(i_2) * \dots * \text{TTD}(i_k). \quad (9)$$

$\text{TTD}(i_l)$ , partial TTD of link  $l$ , is the empirical distribution (histogram) of data points conditioned on the state of links  $l - 1$ ,  $l$ , and  $l + 1$ , i.e.  $i_{l-1}$ ,  $i_l$ , and  $i_{l+1}$ . In the same way, for the first and last links this conditioning is only valid for the following and preceding links, respectively. Finally, the route TTD is computed as a mixture of distributions such that each primary distribution component is the TTD of a Markov path (Eq. (9)) and the corresponding weight of each distribution component is the path occurrence probability (Eq. (8)),

$$\text{TTD} = \sum_{q=1}^Q \Pr(\text{Markov path}_q) \cdot \text{TTD}(\text{Markov path}_q). \quad (10)$$

## 4.2. Clustering

The chief challenge in our methodology is to identify rectangular clusters properly (i.e.  $\mathcal{X}_i$  and  $\mathcal{Y}_i$ ) so states exhibit homogenous travel time characteristics. In other words, any type of correlation or non-random pattern within a cluster is undesirable. Intuitive thumb rules like travel time of free flow, near capacity, and oversaturated conditions may be utilized to address the state identification problem. Furthermore, a heuristic grid clustering method is introduced to determine boundaries based on the structure of 2D diagrams in order to have more homogenous states. Although the rectangular clusters do not have the maximum flexibility to capture the entire correlation patterns in a 2D diagram, their introduction is a matter of traffic physics such that link states are expressed as time intervals. For example with this approach, the probability of a driver to be fast in the next link, given that its current link speed is slow can be directly estimated. This approach makes the proposed method to be direct and intuitively applicable, even for transportation practitioners. Note that, utilizing alternative methods to define the states of the Markov chain or different models for traffic progression, e.g. Hidden Markov Models (see Morris and Trivedi, 2008; Herring et al., 2010) can also be considered.

### 4.2.1. Traffic engineering guidelines

Intuitive traffic engineering rules to estimate travel time for conditions with different level of congestion are employed to define state boundaries. For instance in uncongested regime, we consider free flow travel time boundary ( $TT_{ff}$ ) as:

$$TT_{ff} = \left( \frac{\text{link length}}{V_{ff}} \right) (1 + \varepsilon), \quad (11)$$

where  $V_{ff}$  denotes free flow speed and  $\varepsilon$  is a coefficient, to count for the variability of drivers' behavior (e.g.  $\varepsilon = 0.1$ ). Furthermore, the second boundary which estimates the maximum travel time in near capacity regime is  $R + TT_{ff}$ , where  $R$  is the red interval for through movement. The next boundary for oversaturated conditions where vehicles are delayed for more than one cycle time can be set as  $C + TT_{ff}$ , where  $C$  is the cycle time (Skabardonis and Geroliminis, 2008). The major drawback of these guidelines is that they are practical when all of the route intersections are controlled by traffic signals. They also ignore the effect of offsets and the fact that within these boundaries some correlation might be present. It might also be difficult to be estimated in case of actuated traffic signals. Offsets and turning movements create sharp functions of arrival profiles and travel time characteristics and thus lead to different boundaries in the 2D diagrams. Note that in this case, the numbers of states are the same for all the links which makes the number of Markov paths ( $Q$ ) equal to  $\prod_{l=1}^k m_l$ ,  $m_l = 2, 3, 4$ .

### 4.2.2. Grid clustering

Due to the implication of travel time states as time intervals, grid clustering seems a reasonable clustering approach. The chief trait of grid clustering method is that it uses a multi-dimensional imaginary grid structure which partitions the data space to hyperrectangles (rectangles in 2D space) in order to find hidden patterns in data. Then as the next step, the hyperrectangles are grouped with respect to a topological-neighboring criterion and the problem attributes to produce clusters. The grid clustering has shown to be very effective for analyzing enormous datasets and demonstrates superior performance over fuzzy  $k$ -means and Radial Basis Function (RBF) methods (Yue et al., 2008).

In the proposed model, we discretize the time space and we draw a  $1 \text{ s} \times 1 \text{ s}$  2D Cartesian grid over each 2D diagram and our goal is to find a subset of boundaries in both axes to have homogenous rectangular clusters. In other words, our objective is to identify clusters with negligible correlation between successive links travel times, where convolution is a proper

estimator of sum of distributions. We introduce a metric to measure the discrepancy between data of two consecutive columns (or rows) of Cartesian grid as follows (link index is omitted for simplicity):

$$\alpha_c = \sum_{u=1}^{n_r} \frac{|N(c, u) - N(c + 1, u)|}{N(c, u) + N(c + 1, u)}, \tag{12}$$

where  $n_r$  (or  $n_c$ ) is the number of rows (or columns) of the Cartesian grid over 2D diagram  $l$ , which represents the range of travel time (maximum minus minimum value). Metric  $\alpha_c$  is the measured discrepancy between columns  $c$  and  $c + 1$ ;  $N(c, u)$  and  $N(c + 1, u)$  are the number of data points in small squares in column  $c$  and  $c + 1$ . A similar procedure should be done for the rows of 2D diagram ( $\alpha_r$  instead of  $\alpha_c$ ). The discrepancy metric  $\alpha_c$  represents how data in two successive columns of grid (travel time of  $c$  and  $c + 1$  for a link) are similar. The less  $\alpha_c$  means that the distributions of travel times at columns  $c$  and  $c + 1$  of the 2D diagram projected on the other axis (next link) is more similar. Thus, points with high  $\alpha_c$  value are more likely to represent a drastic change in the distribution of travel time, which can be used as an identification mark for a state boundary. Thus, according to discrepancy values ( $\alpha_r$  and  $\alpha_c$ ) we should choose a few high peak value points as state boundaries. Fig. 3 illustrates the discrepancy values for columns and rows of 2D diagram of Fig. 2a.

The objective of the heuristic grid clustering algorithm is to generate (i) a few numbers of states in order to have reasonable computational complexity, (ii) large enough states that can characterize a traffic condition (uncongested, semi-congested, etc.) and (iii) negligible correlation of any type within each cluster. We introduce two mechanisms to control these settings. First, two threshold levels are defined so that the selected points (boundaries) should have a discrepancy value greater than that; one static and one dynamic. The former is a multiplication of a static factor ( $\mu > 1$ ) by the average discrepancy value of columns (rows), i.e.  $\bar{\alpha} = \sum_{u=1}^{n_c} \alpha_u / n_c$ . This static threshold ensures that the selected point is a peak that is greater than a factor of the average value,  $\bar{\alpha}$ . The latter is the product of an increasing parameter ( $\lambda < 1$ ) by the latest selected  $\alpha$  value, where  $\lambda$  is initiated near zero and increases at each iteration. This dynamic threshold guarantees that the boundary point candidate has a sensible discrepancy difference with previous selected boundary point. Hence using both thresholds, a large number of boundaries is avoided. For example, point 90 s in Fig. 3b is rejected because its  $\alpha$  value is less than  $\mu\bar{\alpha}$ . Second, the discrepancy values of  $\beta$  data points within the neighborhood area are forced to zero to prevent a small size state whenever a new boundary is selected. For example, point 21 s in Fig. 3b is rejected since it is in the vicinity of already chosen point 25 s. In addition, a few initial and final data points should be forced to zero to avoid small states at the beginning and end of travel time interval. For instance, point 15 s in Fig. 3b is rejected since it would create a small size state in the beginning. Note that in Fig. 3a, only one boundary is selected, as the discrepancy values do not have any further peak greater than  $\mu\bar{\alpha}$  after point 10 s. In this manner the procedure achieves the predefined three objectives. We have also noticed that the effectiveness of two aforementioned mechanisms is robust to random fluctuations of  $\alpha$  and not very sensitive to parameter calibration. Values of the parameters after a fine-tuning procedure are selected as  $\mu = 1.6$  and  $\beta = 5$  s. The proposed heuristic algorithm for grid clustering performs iteratively and independently for columns and rows to identify the state boundaries set. The pseudo code of the heuristic algorithm is as follows:

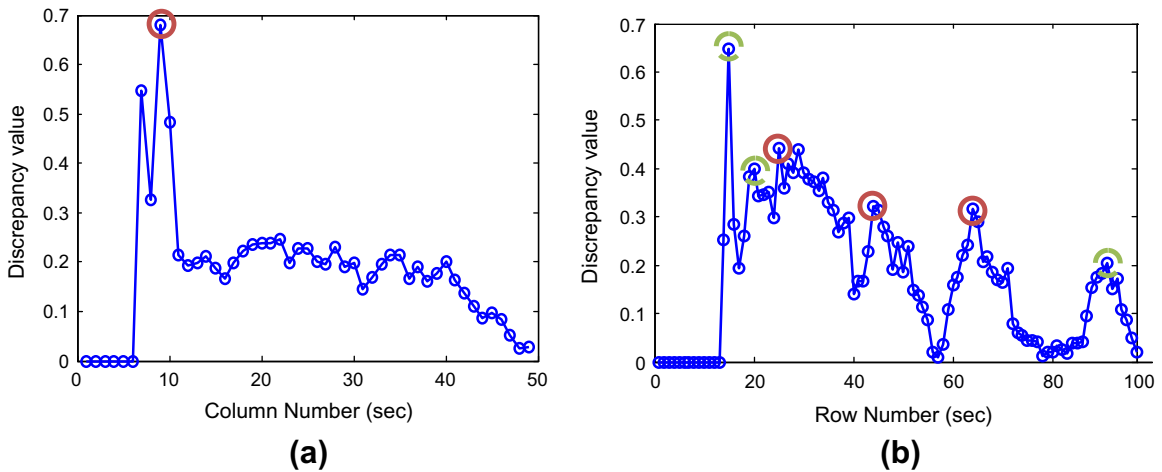


Fig. 3. Discrepancy values for; (a) columns, (b) rows. The red solid circles show the selected data points as boundaries and the green dashed circles show peaks with high discrepancy values that are rejected as boundaries (see explanation in the text). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



1. Compute the discrepancy values ( $\alpha$ )
2. Loop for  $n = 1, 2, \dots$ 
  3. Select the highest  $\alpha$ , i.e.  $\alpha_{max}^n$
  4. If ( $\alpha_{max}^n > \alpha_{max}^{n-1} \times \lambda$ ) and ( $\alpha_{max}^n > \mu\bar{\alpha}$ ) then
    - 4.1. Append the corresponding number of  $\alpha_{max}^n$  to the boundaries set
    - 4.2. Make  $\alpha$  of  $\beta$  neighbors of  $\alpha_{max}^n$  equal to zero
    - 4.3. Increase  $\lambda$
  5. else End
6. Return

Since grid clustering algorithm operates based on each 2D diagram structural data, it is apparent that the states of link  $l$  in the 2D diagrams  $l - 1$  and  $l$  might not be consistent, i.e.  $\mathcal{Y}_{l-1} \neq \mathcal{X}_l$ . Therefore some modifications of the Markov chain formulations, i.e. (6-10), are needed. We consider a new Markov chain with intermediate stages such that at the beginning of Markov chain there are link 1 states (from 2D diagram 1) then link  $l$  states ( $l = 2, 3, \dots, k - 1$ ) (as downstream link from 2D diagram of links  $l - 1$  and  $l$ ), and again link  $l$  states (as upstream link from 2D diagram of links  $l$  and  $l + 1$ ), and finally link  $k$  states (from 2D diagram  $k - 1$ ). In this manner, computing the initial state probabilities and the transition probabilities between link  $l$  and  $l + 1$  remain the same as in (6) and (7). Note that there is no concern about the first and the last links since they only belong in one 2D diagram. The concern is about the intermediate links (2nd to  $(k - 1)$ th), which have two different sets of states. Let us denote them  $\mathcal{Y}_{l-1}$  and  $\mathcal{X}_l$  for link  $l$ , where  $\mathcal{Y}_{l-1} = \{y_{i-1}^{l-1}, \dots, y_{n_{l-1}-1}^{l-1}\}$ ,  $\mathcal{X}_l = \{x_1^l, \dots, x_{m_l-1}^l\}$  and  $\tau_l$  is the set of all link  $l$  reported travel times. Fig. 4 illustrates such intermediate stage.

Computing the intermediate transition probabilities for link  $l$  is modified as: (similar to interlink transition probabilities (7))

$$\hat{p}_l = \begin{bmatrix} \hat{p}_{1,1} & \dots & \hat{p}_{1,m_l} \\ \vdots & & \vdots \\ \hat{p}_{n_{l-1},1} & \dots & \hat{p}_{n_{l-1},m_l} \end{bmatrix}. \tag{13a}$$

$$\hat{p}_{ij} = \Pr\{tt_l \in [x_{j-1}^l, x_j^l] | tt_l \in [y_{i-1}^{l-1}, y_i^{l-1}]\} = \frac{N(\{tt_l \in [x_{j-1}^l, x_j^l] \cap [y_{i-1}^{l-1}, y_i^{l-1}]\})}{N(\{tt_l \in [y_{i-1}^{l-1}, y_i^{l-1}]\})}, \tag{13b}$$

$i = 1, \dots, n_{l-1}; j = 1, \dots, m_l; l = 2, \dots, k - 1.$

Consequently, there will be  $Q = \prod_{l=1}^{k-1} m_l \cdot n_l$  Markov paths. For a given Markov path, the path occurrence probability is computed similarly to (8):

$$\Pr\{S_1 = i_1, S_2 = i_2, \hat{S}_2 = \hat{i}_2, S_3 = i_3, \dots, \hat{S}_{k-1} = \hat{i}_{k-1}, S_k = i_k\} = \pi_{i_1} p_{i_1, \hat{i}_2} \hat{p}_{\hat{i}_2, i_2} p_{i_2, i_3} \dots p_{i_{(k-1)}, i_k}. \tag{14}$$

And the Markov path TTD is computed similarly to (9):

$$\text{TTD}\{S_1 = i_1, S_2 = i_2, \hat{S}_2 = \hat{i}_2, \dots, \hat{S}_{k-1} = \hat{i}_{k-1}, S_k = i_k\} = \text{TTD}(i_1) * \text{TTD}(i_2 \cap \hat{i}_2) * \dots * \text{TTD}(i_k). \tag{15}$$

Finally, Eq. (10) remains valid to compute the route TTD. In the following sections, the study sites are presented and the results are elaborated.

### 5. Study sites

#### 5.1. Peachtree Street

The first selected site is Peachtree Street, an arterial in Atlanta, approximately 640 m in length, with five intersections, six links and two to three through lanes in each direction. The intersections are signalized except the most north one. The speed

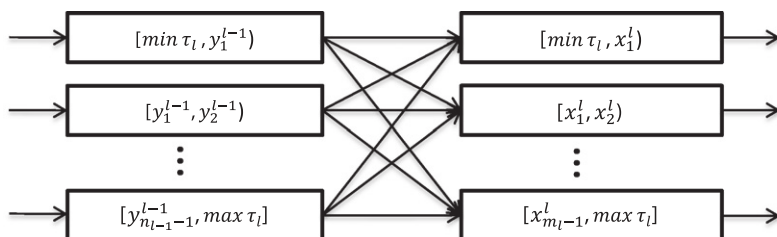


Fig. 4. Schematic of the Markov chain intermediate stage for link  $l$ .

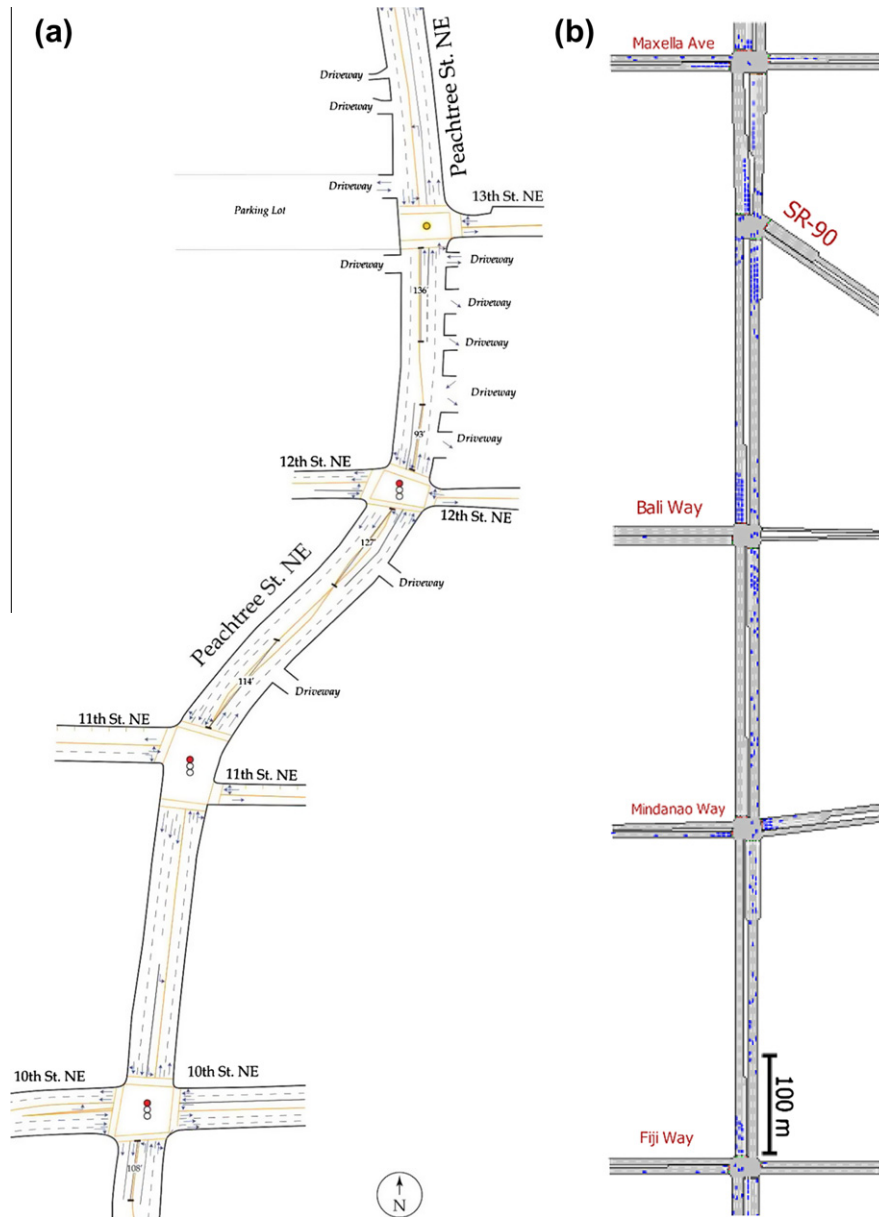


Fig. 5. Study sites; (a) Peachtree Street schematic, (b) Lincoln Blvd. snapshot.

limit on this north–south arterial is 35 mph and we only study the southbound direction of traffic, which is the most congested. The Peachtree Street dataset is a part of NGSIM program (NGSIM, 2006) which managed by Federal Highway Administration, intended to provide a dataset of arterial vehicle trajectories for traffic behavioral analyses. Detailed traffic data (vehicle trajectories) were collected using video camera between 12:45 p.m. and 1:00 p.m. at a resolution of 10 frames per second on November 8, 2006. The dataset contained comprehensive individual vehicle trajectories with time, link and direction stamps, from which the link travel times of vehicles are calculated. Fig. 5a shows the study area schematic.

## 5.2. Lincoln Boulevard

This study site is 1.1 km long stretch of a major urban arterial with speed limit of 35 mph, north of the Los Angeles International Airport, between Fiji Way and Maxella Ave. in the cities of Los Angeles and Santa Monica. The study section includes five signalized intersections with link lengths varying from 150 to 300 m. The number of lanes for through traffic per link is three and additional lanes for turning movements are provided at intersection approaches. Traffic signals are all multiphase operating as coordinated under traffic responsive control as part of the Los Angeles central traffic control system. Loop detec-

tors are located on each lane approximately 90 m upstream of the intersection stop line. Detectors are also placed on the major cross street approaches.

A field study was undertaken to obtain a comprehensive database (loop detector data, manual turning movement counts, probe vehicles, etc.) of operating conditions. The study period enabled us to attain data for a wide range of traffic conditions: from low volume off-peak conditions, peak period conditions and post-peak mid-day flow conditions. Traffic demand is high especially during the peak hour and heavily directional with the higher through and turning volumes in the northbound direction. For a more detailed analysis of the study network the reader can refer to (Skabardonis and Geroliminis, 2008). The vehicle data (demand and turning movements every 15 min) and signal timing data were incorporated into the AIMSUN microscopic simulator. The proposed model was then applied to estimate the TTD on northbound travel direction in a 4 h simulation. The data sampling rate from probe vehicles is every one second and the travel time of each vehicle at each link is recorded and treated as GPS data. The simulation output was first compared with field data (delays and travel times) to verify that the model reasonably replicates field conditions at the test sites. Fig. 5b shows the snapshot of the study site.

## 6. Results and discussion

We evaluate the proposed methodology for the through movement of both study sites. Results for Peachtree Street are shown in Fig. 6. It is evident that the results of proposed method are promising and can capture the fundamental and the most details of TTD profile. Note that since there is no traffic signal at the first intersection, it is not possible to utilize traffic guidelines as per Section 4.2.1. For more comprehensive comparison, the Mean Absolute Error (MAE) between the ground truth TTD and the estimated TTD is calculated:

$$MAE = \frac{\sum_{u=\min(TTD)}^{\max(TTD)} |estimated_u - real_u|}{Range}, \quad (16)$$

where *Range* is the longest TTDs time range (for ground truth and all estimated TTDs, max minus min value). (It should be noted that, in this article all the calculations are discretized with the accuracy of 1 s. Thus, division by *Range* means to normalize the absolute error between the True TTD probability mass function and the different estimated TTD probability mass functions with various range of travel time. With different accuracy of discretization, different definition of *Range* is needed. Also note that, the largest possible value of MAE is  $2/Range$ .) The MAE of grid clustering TTD estimation is  $13.59 \times 10^{-4}$  whereas the MAE of convolution method is  $29.32 \times 10^{-4}$  which shows a significant (54%) improvement. Note that the real TTD is too spiky and the results (only of the real TTD) have been smoothed in the graphs with a moving average of 9 s every 1 s. Another issue about TTD being spiky is that convolution operator has the disadvantage that generally tends to produce smooth output (Folland, 1999) contrary to our method which can create sharp results.

The ground truth TTD for Lincoln Blvd. case study, convolved estimation and result of our methodology with grid clustering approach are depicted in Fig. 7a. The outputs of the proposed method using traffic guidelines with 2 and 3 states in each link are also illustrated in Fig. 7b. It is apparent that TTD is very close to the outcomes of proposed methodology and more accurate once compared to the convolution estimation. It is worth mentioning that the period of analysis should not be too large as it might contain variant traffic conditions. In this case, the spatial correlations between links may be diminished by different characteristics of traffic regimes, i.e. the grid clustering might result in different boundaries (states) for congested and uncongested conditions. This can be inferred by comparing Figs. 6 and 7. The superiority of grid clustering over convolution is more significant in Peachtree Street than Lincoln Blvd., because Peachtree Street data covers only 15 min of mostly invariant traffic regime while the 4 h data of Lincoln Blvd. includes more variations.

Furthermore, in reality not all vehicles are equipped with GPS devices and only a subset of information might be available. To evaluate the robustness of the methodology, different deployment rates of probe vehicles are studied and the results are given in Table 1a. The values in parentheses in the rightmost column show the average number of states in grid clustering method. Hence, grid clustering results to a better estimation than traffic guidelines even with less number of states (less computation burden and complexity). Note that grid clustering algorithm detects the states deterministically based on the spatial structure of 2D diagrams. So for different sample sizes, different number of states is selected for every link (usually less than 3 states). The effect of probe vehicles penetration rates on evolution of TTD profile is shown in Fig. 8a. The TTD estimation with 2% data are also depicted in Fig. 8b demonstrating that our proposed algorithm still performs very well under sparse probe vehicles condition and low number of states.

The same experiments are also done with less demand (75% and 60%) to investigate the effect of demand level on outcome of TTD estimation. In all cases, the proposed methodology utilizing only 5% of the data, gives a better estimation than convolution (neglecting the correlation between links) with 100% data. By comparing results of 2- and 3-states Markov chain in Tables 1b and c, we notice less good outcomes for 3-states compare to 2-states contrary to results in Table 1a. The reason is that less demand makes less congestion and thus introducing the third state regard to near capacity condition is not sensible and causes difficulties for the procedure. In addition being in the free flow regime, the third state boundary may be higher than the maximum travel time, which makes the procedure to fail (see last rows of Table 1c). The results in Table 1 also reveal that, with very low penetration rates the performances of proposed Grid clustering and convolution are declining. It can be justified such that with very low penetration rate the sample size is too small such that to identify the correlation within the data is not sensible and assuming the total independence

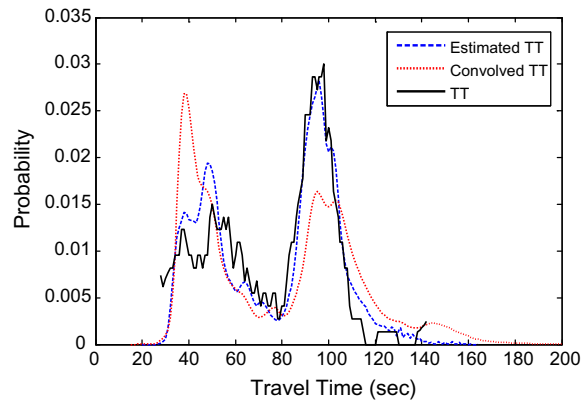


Fig. 6. Estimation and ground truth of TTD for Peachtree Street.

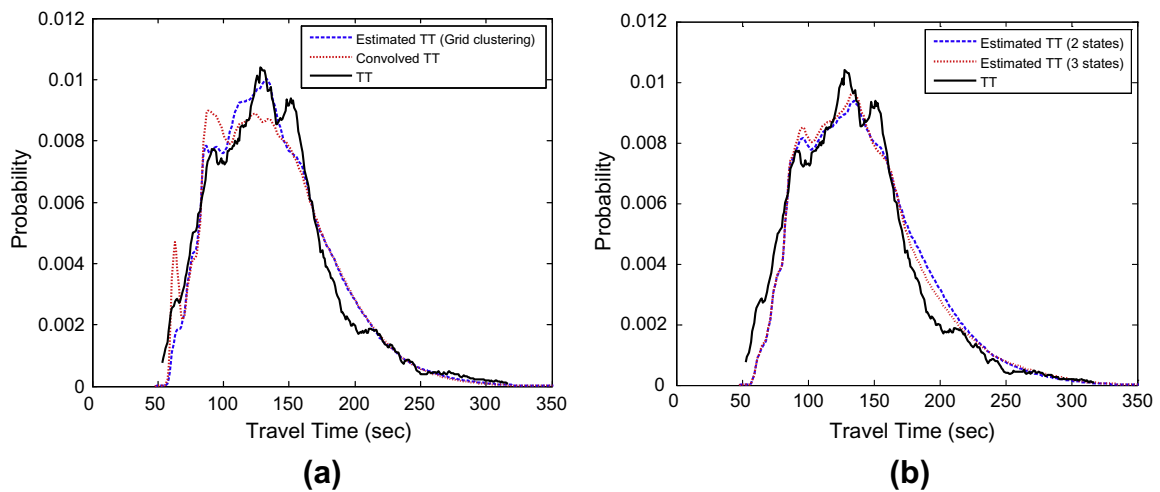


Fig. 7. Estimation of TTD; (a) convolution and grid clustering method, (b) traffic guidelines with 2 and 3 states.

(convolution method) seems reasonable. In cases with so low penetration rate, a proper (Bayesian) fusion of a priori knowledge based on the historical data and online information based on the real-time data is a crucial future work. The ground truth, convoluted, and estimated TTD of 75% and 60% demands are depicted in Fig. 9a and b, respectively. It is evident that convolution method always overestimates the probability of fast and very slow moving vehicles by producing a high peak at the beginning and a long tail.

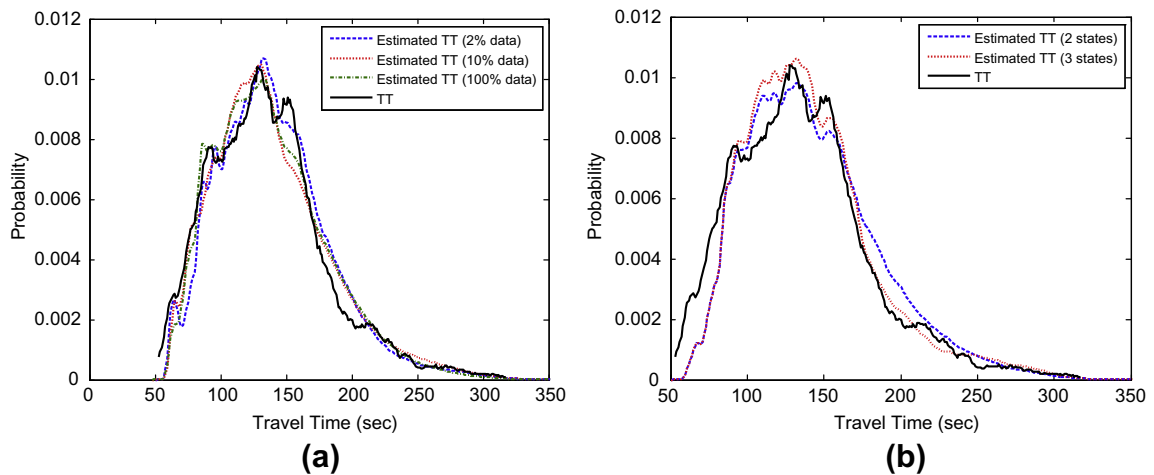
To further investigate the proposed methodology and given literature attempts to approximate the TTD with common single-state and multi-state continuous probability distributions, the TTD of both case studies are fitted with Normal, Gamma, Log-normal, and Gaussian Mixture Models (GMM) (with 2 components; 5 degrees of freedom) and the MAEs of these approximations are given in Table 2. The advantage of proposed grid clustering estimation over the fitting methods seems promising, given that the grid clustering utilizes the link-level data to estimate the unknown route-level information, while the fitting methods try to approximate the route-level TTD, which is considered known. (In addition, the results verify the visual inspections that Lincoln Blvd. TTD resembles more the unimodal distributions whereas Peachtree TTD is a multi-modal distribution.)

To provide more comprehensive comparisons, we do additional statistical investigation such that the link travel time data set is split to two parts, one for estimation methods and the other one for computing the ground truth TTD and subsequently the error metric. In this manner, the data used for the estimation is entirely different from the base ground truth. The Lincoln case study with 100% demand is selected and for various fractions of data, we run the grid clustering and convolution estimation methods and compare those with the TTD computed from remaining part of data set. The results for 5 runs are presented in Table 3 validating the statistical superiority of the proposed grid clustering. Considering all the facts, incorporation of Markov chain into TTD estimation in a synergy produces promising results which can capture the fundamental characteristic of field TTD.

**Table 1**

MAE ( $\times 10^{-4}$ ) of described methods for different levels of demand (value in parenthesis of grid clustering indicates the average number of clusters), (a) 100% demand – range: 440 s; (b) 75% demand – range: 310 s; (c) 60% demand – range: 305 s.

Probe vehicles sample size (%)	Convolution	2 States Markov chain	3 States Markov chain	Grid clustering
100	3.27	3.34	3.09	2.67 (2.875)
50	3.32	3.45	3.18	3.00 (2.75)
20	3.48	3.75	3.31	3.12 (2.625)
10	3.53	3.72	3.40	2.98 (2.625)
5	3.75	3.95	3.47	3.01 (2.875)
2	4.09	4.89	4.44	3.63 (2.625)
1	4.49	4.49	–	3.95 (3.125)
<hr/>				
100	6.90	6.48	6.47	5.23 (2.5)
50	7.06	6.55	6.60	5.22 (2.25)
20	7.11	6.76	6.76	5.30 (2.375)
10	6.90	6.82	6.78	5.88 (2.125)
5	7.32	6.82	6.75	5.65 (2.75)
2	7.89	7.35	–	8.05 (3)
1	9.65	8.98	–	8.61 (2.875)
<hr/>				
100	9.36	7.61	7.35	6.09 (2.5)
50	9.36	7.84	7.60	7.22 (2.25)
20	10.24	7.68	–	7.00 (2.625)
10	9.39	8.06	–	7.88 (2.75)
5	10.54	8.30	–	7.52 (2.625)
2	10.55	10.78	–	9.54 (2.375)
1	11.78	12.64	–	11.48 (3.5)



**Fig. 8.** TTD estimation with; (a) different probe vehicles penetration rates, (b) 2% probe vehicles penetration rate.

## 7. Conclusion and future works

In this article, we introduce a sound approach to address traffic progression and correlation in arterials for travel time distribution estimation. In this method, probe vehicles provide travel time of links of the arterial route. For each pair of consecutive links, a 2D diagram is established to graphically represent the joint distributions of successive link travel times. Then, using these 2D diagrams, we incorporate a Markov chain procedure into the model and identify its initial and transition probabilities from the observed data. For Markov chain state identification, a heuristic grid clustering algorithm is also developed. The procedure is tested with various deployment rates of probe vehicles to tackle the problem of probe vehicles

**Table 2**

MAE ( $\times 10^{-4}$ ) of estimation and fitting methods for both case studies.

Study site	Convolution	Grid clustering	Gamma	Log normal	Normal	GMM
Lincoln	3.27	2.67	2.79	3.11	4.15	2.88
Peachtree	29.32	13.59	32.04	45.32	29.57	14.14



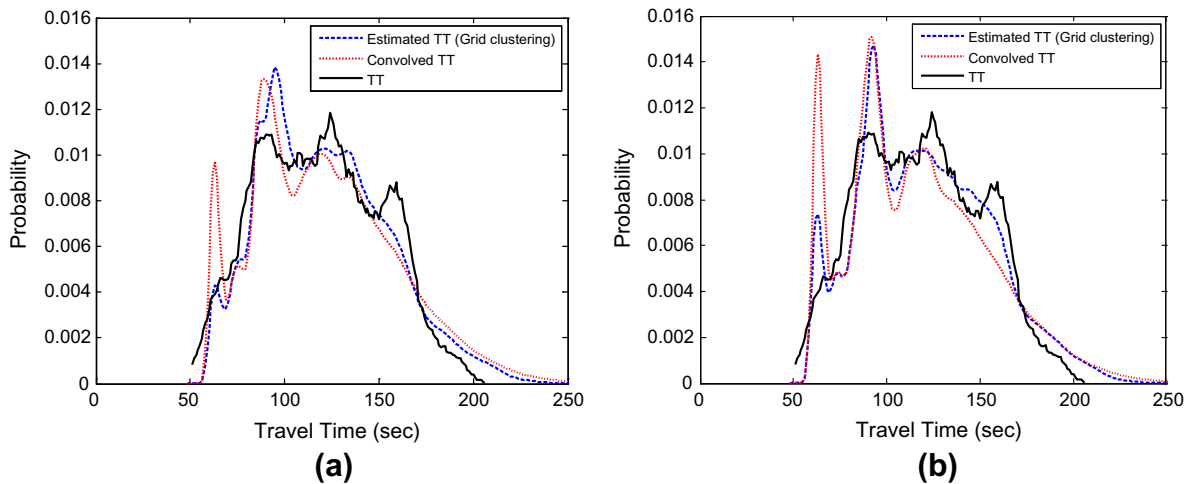


Fig. 9. Estimation of TTD with; (a) 75% demand, (b) 60% demand.

**Table 3**  
MAE ( $\times 10^{-4}$ ) of statistical tests for Lincoln case study.

Fraction of data for estimation (%)	Convolution	Grid clustering
50	3.52	3.08
20	3.56	3.29
10	3.47	3.17
5	3.79	3.29
2	4.35	3.67

sample size. Our proposed methodology shows a coherent performance capturing the fundamental characteristics of field measurements even under condition of sparse probe vehicles.

The inputs to the proposed procedure are link travel times and since they are not directly reported by probe vehicles, the estimation of link travel times from GPS data is of interest to integrate with our methodology. This problem is named as travel time allocation or decomposition in the literature and discussed more in Hellinga et al. (2008), Hofleitner and Bayen (2011). The next crucial step for future studies is to investigate under what traffic conditions and route structures our method significantly outperforms convolution estimation, i.e. when correlation plays a main role in TTD estimation. For instance, in case of very long links because of platoon dispersion effects and different driver characteristics, correlation might be negligible. More empirical experiments with real data are needed to examine the influence of different traffic regimes. Moreover, this approach would be helpful to inspect how the signal timing, offset setting, site topology, etc., affect the travel time variability. Another major aspect of this research is, as GPS data become available in real sites, to determine travel time reliability on as many routes as possible, even when data are not available for some routes or parts of a route, i.e. providing adequate area coverage and integration of historical and real-time data.

The proposed methodology can be integrated in a real-time implementation and estimation of TTDs. In this case, by utilizing historical data and partitioning the data set for different times of days and level of congestion, one can apply the proposed methodology for each temporal partition and estimate a priori TTD offline. In the online part of the real-time implementation, as more data become available, a learning procedure should be developed (e.g. based on Bayesian update), which will update the parameters of the model (states and initial and transition probabilities) and improve the estimation.

## Acknowledgement

This research was supported by Swiss National Science Foundation, Grant #200021-130165.

## References

- Allsop, R., 1972. Delay at fixed time traffic signals – I: Theoretical analysis. *Transportation Science* 6 (3), 280–285.
- Bertini, R., Tantyanugulchai, S., 2004. Transit buses as traffic probes: use of geolocation data for empirical evaluation. *Transportation Research Record* 1870, 35–45.
- Dong, J., Mahmassani, H.S., 2009. Flow breakdown and travel time reliability. *Transportation Research Record* 2124, 203–212.
- Du, L., Peeta, S., Kim, Y.H., 2012. An adaptive information fusion model to predict the short-term link travel time distribution in dynamic traffic networks. *Transportation Research Part B* 46 (1), 235–252.

- Feng, Y., Davis, G., Hourdos, J., 2011. Arterial travel time characterization and real-time traffic condition identification using GPS-equipped probe vehicles. In: *Transportation Research Board 90th Annual Meeting*, Washington, DC.
- Ferman, M.A., Blumenfeld, D.E., Dai, X., 2005. An analytical evaluation of a real-time traffic information system using probe vehicles. *Journal of Intelligent Transportation Systems* 9 (1), 23–34.
- Folland, G.B., 1999. *Real Analysis: Modern Techniques and Their Applications*, second ed. John Wiley & Sons, p. 241 (ISBN 0-471-31716-0).
- Geroliminis, N., Daganzo, C., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transportation Research Part B* 42 (9), 759–770.
- Geroliminis, N., Skabardonis, A., 2005. Prediction of arrival profiles and queue lengths along signalized arterials by using a Markov decision process. *Transportation Research Record* 1934, 116–124.
- Geroliminis, N., Skabardonis, A., 2006. Real time vehicle reidentification and performance measures on signalized arterials. In: *9th International IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada.
- Geroliminis, N., Skabardonis, A., 2011. Identification and analysis of queue spillovers in city street networks. *IEEE Transactions on Intelligent Transportation Systems* 12 (4), 1107–1115.
- Guo, F., Rakha, H., Park, S., 2010. Multistate model for travel time reliability. *Transportation Research Record* 2128, 46–54.
- Hellinga, B., Izzadpanah, P., Takada, H., Fu, L., 2008. Decomposing travel time measured by probe-based traffic monitoring systems to individual road segments. *Transportation Research Part C* 16 (6), 768–782.
- Herrera, J.C., Work, D.B., Herring, R., Ban, X., Jacobson, Q., Bayen, A., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment. *Transportation Research Part C* 18 (4), 568–583.
- Herring, R., Hofleitner, A., Abbeel, P., Bayen, A., 2010. Estimating arterial traffic conditions using sparse probe data. In: *13th International IEEE Conference on Intelligent Transportation Systems Madeira, Portugal*.
- Hofleitner, A., Bayen, A., 2011. Optimal decomposition of travel times measured by probe vehicles using a statistical traffic flow model. In: *14th International IEEE Conference on Intelligent Transportation Systems*, Washington, DC.
- Jenelius, E., 2012. The value of travel time variability with trip chains, flexible scheduling and correlated travel times. *Transportation Research Part B* 46 (6), 762–780.
- Kaparias, I., Bell, M., Belzner, H., 2008. A new measure of travel time reliability for in-vehicle navigation systems. *Journal of Intelligent Transportation Systems* 12 (4), 202–211.
- Killmann, F., Collani, E.v., 2001. A note on the convolution of the uniform and related distributions and their use in quality control. *Economic Quality Control* 1 (16), 17–41.
- Kwon, J., Barkley, T., Hranac, R., Petty, K., Compin, N., 2011. Decomposition of travel time reliability into various sources: incidents, weather, work zones, special events, and base capacity. *Transportation Research Record* 2229, 28–33.
- Kwong, K., Kavalier, R., Rajagopal, R., Varaiya, P., 2009. Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors. *Transportation Research Part C* 17 (6), 586–606.
- Liu, H., Ma, W., 2009. A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. *Transportation Research Part C* 17 (1), 11–26.
- Morris, B., Trivedi, M., 2008. Learning, modeling, and classification of vehicle track patterns from live video. *IEEE Transactions on Intelligent Transportation Systems* 9 (3), 425–437.
- National Research Council, *Transportation Research Board*, 2000. *Highway Capacity Manual*. Washington, DC.
- NGSIM, 2006. Next Generation Simulation. <<http://ngsim.fhwa.dot.gov/>>.
- Noland, R.B., Polak, J.W., 2002. Travel time variability: a review of theoretical and empirical issues. *Transport Reviews* 22 (1), 39–54.
- Park, S., Rakha, H., Guo, F., 2011. Multi-state travel time reliability model: impact of incidents on travel time reliability. In: *14th International IEEE Annual Conference on Intelligent Transportation Systems*, Washington, DC.
- Rakha, H., El-Shawarby, I., Arafeh, M., Dion, F., 2006. Estimating path travel-time reliability. In: *9th International IEEE Annual Conference on Intelligent Transportation Systems*, Toronto, Canada.
- Skabardonis, A., Geroliminis, N., 2005. Real-time estimation of travel times on signalized arterials. In: *16th International Symposium on Transportation and Traffic Theory*, University of Maryland.
- Skabardonis, A., Geroliminis, N., 2008. Real-time monitoring and control on signalized arterials. *Journal of Intelligent Transportation Systems* 12 (2), 64–74.
- The Mobile Millennium Project. <<http://traffic.berkeley.edu>>.
- Tsekeris, T., Skabardonis, A., 2004. On-line performance measurement models for urban arterial networks. In: *83rd Annual Meeting of Transportation Research Board*, Washington, DC.
- Uno, N., Kurauchi, F., Tamura, H., 2009. Using bus probe data for analysis of travel time variability. *Journal of Intelligent Transportation Systems* 13 (1), 2–15.
- Webster, F., 1958. *Traffic Signal Settings*. Road Research Technical Paper No. 39, Road Research Laboratory, England.
- Yeon, J., Elefteriadou, L., Lawphongpanich, S., 2008. Travel time estimation on a freeway using discrete time Markov chain. *Transportation Research Part B* 42 (4), 325–338.
- Yue, S., Wei, M., Wang, J., Wang, H., 2008. A general grid-clustering approach. *Pattern Recognition Letters* 29 (9), 1372–1384.
- Zheng, F., Van Zuylen, H., 2010. Uncertainty and predictability of urban link travel time: a delay distribution based analysis. In: *Transportation Research Board 89th Annual Meeting*, Washington, DC.