Available online at www.sciencedirect.com
ScienceDirect
Transportation Research Part B 42 (2008) 325–338
TRANSPORTATION RESEARCH PART B
www.elsevier.com/locate/trb

# Travel time estimation on a freeway using Discrete Time Markov Chains

Jiyoun Yeon [a,*], Lily Elefteriadou [b], Siriphong Lawphongpanich [c]

[a] Department of Advanced Transportation Technology, Korea Transport Institute, Republic of Korea
[b] Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL 32611, USA
[c] Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA

Received 29 June 2006; received in revised form 23 August 2007; accepted 23 August 2007

**Abstract**

Travel time is widely recognized as an important performance measure for assessing highway operating conditions. There are two methods for obtaining travel time: direct measurement, or estimation. For the latter, previously developed models tend to underestimate travel times under congested conditions because of the difficulties of calculations of vehicle queue formations and dissipations. The purpose of this study is to develop a model that can estimate travel time on a freeway using Discrete Time Markov Chains (DTMC) where the states correspond to whether or not the link is congested. The expected travel time for a given route can be obtained for time periods during which the demand is relatively constant. Estimates from the model are compared to field-measured travel time. Statistical analyses suggest that the estimated travel times do not differ from the measured travel time at the 99% confidence level.
Published by Elsevier Ltd.

*Keywords:* Breakdown; Freeway travel time estimation; Discrete Time Markov Chains; Expected travel time

## 1. Introduction

Travel time is widely recognized as an important performance measure for assessing highway operating conditions. Turner et al. (2004) indicate that travel times are easily understood by practitioners and the public, and are applicable to both the users' and operators' perspective. The California Department of Transportation (Caltrans) suggested that travel time is one of the possible performance measures in the Performance Measurement System (PeMS) which can provide information in various forms to assist managers, traffic engineers, planners, freeway users, and researchers (Choe et al., 2002). A number of studies have attempted to develop algorithms for estimating travel time using advanced surveillance systems: video image processing, automatic vehicle identification, cellular phone tracking, probe vehicles, etc. (Turner, 1996). Most existing infrastructures have loop detectors which provide point detection data, and usually speeds are converted into travel time from

* Corresponding author. Tel.: +82 31 910 3075; fax: +82 31 910 3228.
*E-mail address:* jyyeon@koti.re.kr (J. Yeon).

0191-2615/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.trb.2007.08.005

those data. Oh et al. (2003) indicate that travel time estimates from point detection of speeds under congested conditions are underestimated. They developed a freeway travel time estimation model using traffic density data for a basic freeway segment (i.e., no on-ramps or off-ramps present). Coifman and Ergueta (2003) also illustrated a method to estimate travel time using existing dual-loop-detectors by measuring vehicle signatures.

The objective of this study is to develop a methodology for estimating the expected travel time for a given freeway route consisting of several consecutive segments. The methodology employs the concept of probabilistic breakdown for freeway segments, and is based on Discrete Time Markov Chains (DTMC). The concept of probabilistic breakdown was employed to determine the probability of congestion occurrence at each freeway segment. The scope of the paper includes congestion due to heavy traffic, and not due to weather, accidents, incidents, or work zones. Section 2 briefly reviews the most relevant literature and Section 3 presents the Markov Chain model for travel time estimation. Section 4 summarizes the data obtained for developing and validating the model. Section 5 describes the developed model applied to US 202 in Philadelphia, PA. To evaluate the model, expected travel times are compared to field-measured travel time data, and the results are presented in Section 6. Finally, Section 7 provides conclusions and recommendations.

## 2. Literature review

During the past few years, many researchers have focused on estimating or predicting travel time with various methodologies such as time series models (Shaw, 2002), artificial neural network models (Shaw, 2002), regression models (Zhang and Rice, 2003), and loop detector data (Oh et al., 2003; Coifman and Ergueta, 2003; Chu, 2005). However, travel times are rather sensitive to the given situation such as demands, weather, roadway conditions, traffic conditions, etc. and the models developed do not work well for all conditions. It is particularly complicated to estimate travel time under congested conditions due to difficulties in calculating travel time during queue formations and dissipations. VanLint and van Zuylen (2005) indicated that travel time estimated from input/output flows under non-congested conditions can be regarded as constant due to their small variance, while under congested conditions, travel time increases with increasing demand, and the variance is high. Thus, it is important to identify the impact of congestion in time and space and consider its flows in order to estimate travel time. Stochastic processes are an appropriate method for solving the problem because they can analyze and predict conditions (states) in time and space, subject to probabilistic laws.

Yang et al. (1999) indicate that travel time is easily influenced by high demand which could result in a higher probability of breakdown. Breakdown can be defined as the time period during which speed drops sharply in a time series speed plot, and it can be considered as a transition state which can be used as an indicator to distinguish congested from non-congested conditions (Persaud et al., 2001). Once breakdown occurs, traffic typically becomes congested. However, even at the same location with the same demand levels, breakdown may or may not occur (Elefteriadou et al., 1995). Thus, breakdown has a probabilistic nature, and in order to estimate travel times, it is essential to incorporate the breakdown probability occurrence.

Evans et al. (2001) developed an analytical model for the prediction of flow breakdown based on the zonal merging probabilities. First, the authors determined the arrival distribution of the merging vehicles, and then calculated the transition probabilities from state to state. The probability of breakdown was obtained using Markov Chains and implemented in MATLAB. Del Castillo (2001) developed a model to reflect the propagation of speed drop when the traffic volume reaches capacity. The author considered a speed drop caused by lane changing or merging of other traffic streams as a complex stochastic phenomenon and developed a theoretical model that can provide normalization probability of speed drop behaviors. To apply this disturbance, the proposed model sets a recurrence equation for the speed drop and its duration. Heidemann (2001) developed a model to describe the flow-speed–density relationship using concepts from queuing theory for non-stationary traffic flow, which implies that input and output flow are not in statistical equilibrium. Kharoufeh and Gautam (2004) derived an analytical expression for the cumulative distribution function of travel time for a vehicle traveling on a freeway link. The authors assumed that vehicle speeds are a random environment and can be considered as a finite-state Markov process. The random environment process includes physical factors (e.g., roadway geometry, grades, visibility), traffic factors (e.g., density, presence of heavy vehicles, merging traffic), or environmental factors (e.g., weather conditions, speed limits, etc.). Based on the assumption that the environmental process is a Continuous Time Markov Chain (CTMC), an exact analytical expression is obtained for the Laplace Transformation of the link travel-time cumulative distribution function.

In summary, breakdown is a transition state, signifying the beginning of congestion and travel times are sensitive to the presence of congested conditions, which also affect the upstream and downstream segments. Thus, to estimate travel times, the impact of traffic congestion should be considered, and stochastic processes can be employed to analyze changes of traffic states over time and space and estimate the expected travel time considering the probability of congestion.

## 3. Model development and data collection

A stochastic process is a random function which varies in time and/or space. Its future values can be predicted with a certain amount of probability. This means that the process does not behave in a completely unpredictable manner but it is governed by a random mechanism. The value of the process at time $t$ does not depend on the value at time $t-1$. For a transportation system, if the demands (e.g., approaching flows) do not fluctuate much for a given time period, the system states would not depend on time but could be considered a random mechanism. In these cases, a stochastic process analysis can be applied. A stochastic process $X=\{X(t),\ t\in T\}$ is a collection of random variables. That is, for each $t$ in the index set $T$, $X(t)$ is a random variable. Usually, $t$ is interpreted as time and $X(t)$ is the state of the process at time $t$ (Ross, 1996). If the index set $T$ is countable, the process is called a discrete-time stochastic process, while if $T$ is a non-countable variable, it is called a continuous-time stochastic process. In other words, under the discrete time, $T\in\{0,1,2,\ldots\}$, the change of state occurs at the end of a time unit, while under the continuous time, $T\geqslant 0$, the change of state occurs at any point in time. This research considers how the system states change at every time unit, so a discrete-time stochastic process is employed.

To apply a stochastic process for estimating travel time, relatively uniform demand time periods are used in the analysis. The following five tasks were undertaken: (i) define system states and state variables considering the probability of congestion for a link; (ii) estimate travel time of each link both when the link is non-congested and when it is congested using field data; (iii) determine time periods to consider time of day and daily variation of travel time; (iv) estimate the transition matrices considering how the system changes from one time interval to another for each time period, considering transitions between non-congested and congested flow; (v) estimate route travel time using the link travel time and transition matrix previously developed. This methodology was applied to an 8-mile freeway section along US 202, in Philadelphia, PA. The data collection effort is presented in Section 4, while the application of the developed model using field data is presented in Section 5.

## 4. Data collection

Model development requires the collecting of appropriate data. These can be grouped into two categories: model development data and model validation data. Model development data are needed to build a model to apply the proposed methods, and model validation data are needed to verify statistically the accuracy of the model developed. In this paper, speed and flow data are used for model development, and link travel time or route travel time data are used to validate the developed model. The speed and volume data were obtained by Mobility Technologies and were collected by Remote Traffic Microwave Sensors (RTMS) at each detection location (as seen in Fig. 1) for a 4-month period, from May to August 2004. Specific sensor locations and distances between sensors are shown in Fig. 1. RTMS provided 1 min speed, volume, and occupancy data which were used for further analysis. Travel time data were collected at the Traffic Control Center (TCC) located in Philadelphia, PA. In the TCC, there is a Closed Circuit Television (CCTV) monitoring system where images are displayed from cameras in the field. Each camera can be controlled manually by panning, tilting, and zoom-in, zoom-out from the TCC, and the image viewed can be recorded using a Video Cassette Recorder (VCR). For the validation of this study, travel time data were collected between a pair of cameras: C205 & C213 during different dates and times (Fig. 1). The cameras focused on the adjacent detector locations, for which speed and flow data were obtained from Mobility Technologies. Travel time data were collected on two weekdays: May 25th (Tuesday), and June 16th (Thursday), 2004. Each camera had a time clock displayed

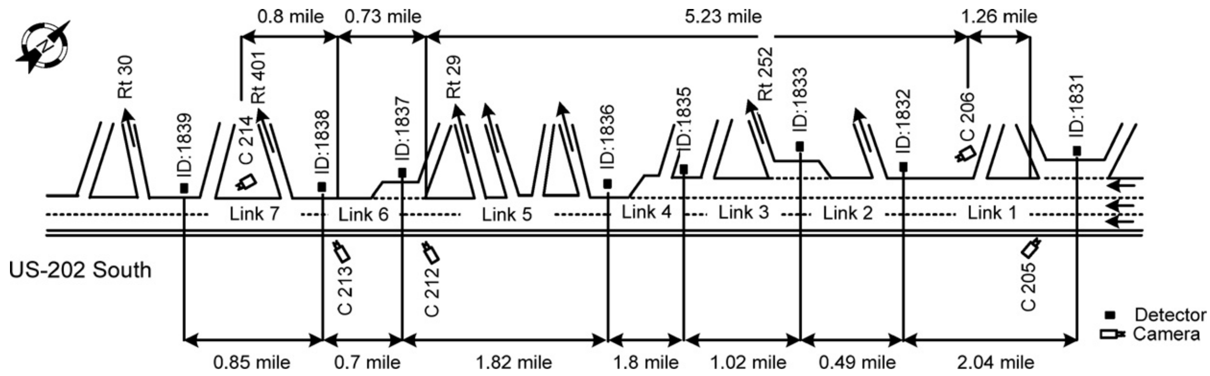


Fig. 1. Camera and detector locations along US-202 SB (not to scale).

on the screen. During the data collection, a vehicle would be identified in the first camera and its arrival time would be recorded for the first upstream camera observed. Then, when that same vehicle was identified in the downstream camera, its departure time would be recorded. Table 1 provides a summary of the travel time data collection.

### 4.1. Definition of states and variables

In stochastic processes, a system has to be defined as a state which is analyzed using a measurable characteristic. For the purposes of this research, the system is defined as a freeway route, and the system state variable is $X(t)$ at time $t$ ($t = 0, 1, 2, \ldots$), where $X(t)$ describes how the states of a given freeway route change every time unit. $X(t)$ can be described as a set of $x_i(t)$, where $x_i(t)$ is a link state variable of link $i$ at time $t$. A link is defined as the segment between detectors on the freeway, and a route is composed of several links. The $x_i(t)$ is a binary variable: if the state of link $i$ at time $t$ is congested $x_i(t)$ is 1, otherwise it is 0. The system state variable and link state variable can be expressed as follows:

$$X(t) : \text{system state variable at time } t \text{ is defined as } \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_k(t) \end{bmatrix}, \tag{1}$$

where $k$ is total number of links

$$x_i(t)\text{: link state variable of link } i \text{ at time } t \begin{bmatrix} 1 & \text{if link } i \text{ at time } t \text{ is congested} \\ 0 & \text{otherwise} \end{bmatrix}.$$

Thus for a freeway route with six links, a system state variable (i.e., $X(t)$) at time $t$ can be $[1, 0, 1, 1, 0, 0]$, where links 1, 3, and 4 are congested. The breakdown occurrence can be used to determine whether the link is congested or not, since after the breakdown, the facility is congested. However, the threshold of breakdown has site-specific characteristics, which will be established based on field data (discussed in Section 5).

Table 1
Route travel time data summary

| Camera O–D pair | Detector ID number | Date and time | Recording hours | Length$^{(1)}$ (mile) | Collected sample size |
|---|---|---|---|---|---|
| C205 & C213 | 1831–1838 | 5/25 17:50–19:00 | 1 h 10 m | 7.87 | 55 |
| | | 6/16 15:30–16:30 | 1 h | | 150 |

*Note:* (1) The length is based on the distance between detectors, not the distance between cameras.

Table 2
Travel time notation by link

| Link information | Link number | | | |
|---|---|---|---|---|
| | 1 | 2 | ··· | $k$ |
| Non-congested travel time | $NT_1$ | $NT_2$ | ··· | $NT_k$ |
| Congested travel time | $CT_1(f)$ | $CT_2(f)$ | ··· | $CT_k(f)$ |

$k$, total number of links; $i$, link number; $NT_i$, non-congested travel time of link $i$ (for $i = 1, 2, \ldots, k$); $CT_i(f)$, congested travel time of link $i$ as a function of flow rates (for $i = 1, 2, \ldots, k$).

Table 2 shows the travel time notation for each link. It is assumed that the segment between detectors (i.e., link) is homogeneous. As shown in Table 2, the travel time under non-congested conditions is considered as constant, because it does not change as a function of flow rates. However, the travel time under congested conditions is estimated as a function of the flow rates. These assumptions were based on previous research (VanLint and van Zuylen, 2005) and were confirmed by the field data collected in this research.

### 4.2. Link travel time estimation

The route travel time is the sum of travel times on each link. To compute this, the link travel time has to be estimated both for non-congested and congested conditions. Fig. 2 presents the general relationships for travel times for both non-congested (i.e., bold solid line) and congested conditions (i.e., bold dashed line). These general relationships – conceptually simplified – are based on previous research (VanLint and van Zuylen, 2005) and were confirmed from the field data collected for this research. As shown in the gray circles of Fig. 2, link travel time has its own distribution at each flow rate level. When flow rates are low or conditions are non-congested, travel time can be considered as constant with small variance, even at high flows. Under congested conditions, travel time decreases exponentially and has large variance. The flows in Fig. 2 correspond to the throughput, which is the number of vehicles passing at each detection location, therefore, as the degree of congestion increases the throughput decreases (i.e., the higher the throughput the lower the travel time). The travel time of link $i$ ($T_i$) is determined by first establishing whether the state of the link is congested or not at time $t$.

### 4.3. Determination of time periods to account for time of day and daily variations

The previous two steps described how link states can be defined and link travel time can be estimated when applying the DTMC, for a given demand. However, demand varies daily, weekly, and monthly. Thus, the model has to distinguish between various demand time periods, such as a.m. peak, off-peak, and p.m. peak. It is assumed that the beginning and end of the peak periods for both a.m. and p.m. can be defined to be when the approaching flows at the upstream end of the study area are greater than 2/3 of the maximum flow.

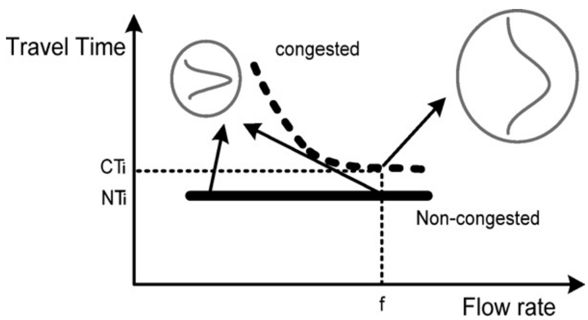


Fig. 2. Conceptual diagram for estimating link travel time.

### 4.4. Transition matrix

This section discusses the development of the transition matrix, which considers whether each link is congested or not. Congestion occurs after the breakdown, and it may propagate to the upstream or downstream segments. Thus, the effects of breakdown at the upstream or downstream segments should be considered in the development of the transition matrix. The transition is a change of state and the one-step transition matrix (e.g., $n = 1$, where $n$ is the number of steps) shows the changing rate from state $i$ to state $j$ as shown below:

$$P^1 = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1m} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2m} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & p_{m3} & \cdots & p_{mm} \end{pmatrix}, \tag{2}$$

where a state is as defined in Section 4.1, and $p_{ij}$ represents the transition rate from state $i$ to state $j$, which can be expressed in conditional probability as follows:

$$p_{ij} = P\{X(t+1) = j | X(t) = i\} \quad \text{for } i,j = 1,2,\ldots,m, \text{ and } t = 1,2,3,\ldots, \tag{3}$$

where $m$ is the total number of possible states.

For example, $p_{12}$ in the transition matrix is computed as the total number of transitions from state 1 to state 2 divided by the total number of transitions from state 1 to all other states including state 1.

The system state variable at time $t$, $X(t)$, includes the link state variable $x_i(t)$, thus, each state (i.e., $i,\ldots,j,\ldots,m$) of the transition matrix can be denoted as $1,2,\ldots,m$. Thus, the number of total possible states ($m$) is equal to 2 to the power of the number of links ($k$) of the system. As the number of steps ($n$) increases, the system becomes more stable. The probabilities when $n$ goes to infinity are called steady-state probabilities, and are the probabilities that a system will eventually arrive at state $j$ whatever the initial state is. Usually, as $n$ increases, the initial state becomes less and less relevant, but this is not always the case. The Markov Chain is *irreducible* (i.e., all states communicate with each other) and *ergodic* (i.e., a process will finally return to the starting state within a certain time period), where there exists a unique steady-state probability for all $j$. The steady-state probability for state $j$ ($\Pi_j$) is defined as follows:

$$\Pi_j = \lim_{n\to\infty} P\{X_{(t=n)} = j / X_{(t=0)} = i\} = \lim_{n\to\infty} P\{X_{(t=n)} = j\} \quad \text{for } i = 1,\ldots,m. \tag{4}$$

These steady-state probabilities show the probabilities that the system eventually will be at each defined state, and they can be used to calculate the expected travel time of a system for the time period being analyzed.

### 4.5. Route travel time estimation

The final task is to estimate route travel time using the output from the previous three tasks. Since the steady-state probabilities for all $j$ can be obtained as described above, the travel time of each link under congested and non-congested conditions can be estimated next. Then, the expected route travel time can be estimated as follows:

$$\overline{T} = \sum_{j=1}^{m} \sum_{i=1}^{k} \Pi_j \{(1 - x_i(t)) \times \mathrm{NT}_i + x_i(t) \times \mathrm{CT}_i(f)\}, \tag{5}$$

where

$\Pi_j$ is the steady-state probability for state $j$.
$x_i(t)$ is the state variable of link $i$ at time $t$.
$\mathrm{NT}_i$ is the non-congested travel time of link $i$.
$\mathrm{CT}_i(f)$ is the congested travel time of link $i$ as a function of flow rates.
$m$ is the total number of possible states.
$k$ is the total number of links.

For each steady-state probability (calculated in Section 4.4), multiply the steady-state probability for a given state by the travel times when the system is in that state. For instance, if the system is in a state where all the links contain no congestion, then the probability is multiplied the state of a link by the sum of the non-congested travel times for each link. Likewise, if the system is in a state that contains one link with congestion, the probability is multiplied the state of a link by the sum of the non-congested travel times for all the non-congested links plus the congested travel time (which includes the sum of the probability of the link being fully-congested, times the fully-congested travel time plus the probability of being semi-congested times the semi-congested travel time). This is the expected travel time for the study route and for the given time period.

## 5. Results from US 202 in Philadelphia, PA

This section describes the case study which implements the above methodology for estimating the route travel time during the p.m. peak (15:00–19:00) period for links 1–6 (Fig. 1). This case study is intended to provide a better understanding how the proposed model can estimate the expected travel time using speed and flow data. In this analysis, breakdown is defined to occur when the speed drops below 50 mph for at least 5 min, and it is assumed that the congestion ends when the speed goes up to 50 mph again (Lorenz and Elefteriadou, 2000). Fig. 3 shows an example of the breakdown occurrence at detector ID 1835 on July 12th. The threshold of breakdown was established based on the time series speed plots for several days at each detection location, which showed the time and magnitude of dramatic speed drops. Breakdown occurrences are usually considered at a specific detection location. However, this study focuses on estimating travel time on a link, and therefore, breakdown is defined as a function of link speed which is the average between two adjacent detection locations. Consequently, in this particular study, breakdown is determined based on link speeds.

- *Step 1: Definition of states and variables*
  Fig. 4 shows how the system state variables are defined at each time interval. It is assumed that all time intervals shown in Fig. 4 are equal (5 min), and the gray bars represent the duration of congestion at each link. There is no congestion at links 1, 2, and 3 during the p.m. peak period, while there are three congestion periods at link 4, and heavy congestion at link 5. For example, the system state variable at time 17:40, $X(t = 17{:}40)$, can be described as $[0, 0, 0, 1, 1, 1]$, and the system state variable at time 18:20, $X(t = 18{:}20)$, can be described as $[0, 0, 0, 0, 1, 0]$. Originally, the defined states

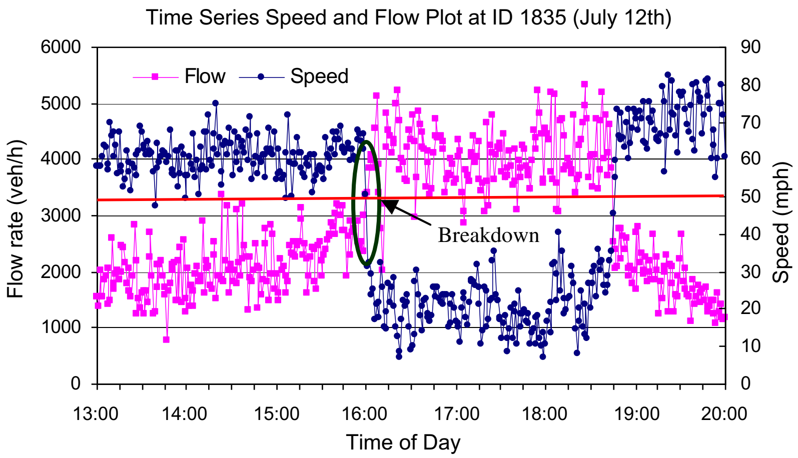


Fig. 3. Time series speed and flow plot at detector ID 1835 on July 12th.

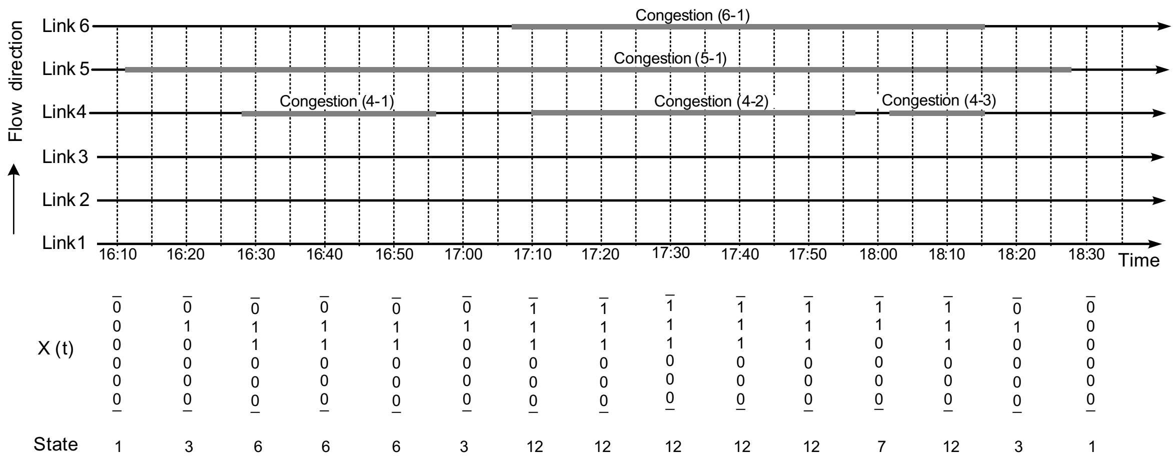


Fig. 4. Defined system state variables on May 25th.

Table 3
Defined states during the p.m. peak period

| $X(t)$ | 1 | 2 | 3 | 4 | 6 | 7 | 12 | 18 | 22 | 28 | 39 | 44 | 46 | 48 | 49 | 50 | 51 | 52 | 54 | 55 | 64 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_6$(t) | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| $x_5$(t) | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| $x_4$(t) | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| $x_3$(t) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| $x_2$(t) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $x_1$(t) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

*Note:* $X(t)$ is composed of link state variable of link $i$ at time $t$, $x_i(t)$, which is 0 (non-congested conditions) or 1 (congested conditions).

were a total of $2^6 = 64$ (i.e., $m = 1, 2, \ldots, 64$) but during the p.m. peak, only the states 1, 2, 3, 4, 6, 7, 12, 18, 22, 28, 39, 44, 46, 48, 49, 50, 51, 52, 54, 55, and 64 occurred. Table 3 shows the defined states during the p.m. peak period.

- *Step 2: Link travel time estimation*
  Fig. 5 is a travel time scatter plot, with travel times obtained from link speed data of each link for four months. These are categorized into non-congested travel times (left) and congested travel times (right), using the threshold of breakdown. As shown in this figure, non-congested travel time at each link, $\mathrm{NT}_i$, is not sensitive to flow rates. For congested conditions however, travel times at each link except links 1 and 2, increase as flow rates increase, while some points appear to have constant travel time (circled in Fig. 5 of link 5). These points represent the following conditions: (i) a small breakdown occurs before a typical breakdown or after recovery from a typical breakdown, (ii) speed drops below the threshold but it stays between the value of the breakdown threshold and the speed of a typical breakdown, (iii) the speed is dropping or recovering through the link, and it is high at one detector and low at the other.The congested travel time is divided into two groups: semi-congested conditions and fully congested conditions. The travel time under congested conditions decreases as flow rates increase, while it decreases exponentially or is constant for semi-congested conditions. The line that separates the data for the two conditions can be determined by two known data points (i.e., the separating start point and right most point) and the shape of the function. Using this separating line, the link travel time function for fully congested conditions is estimated as shown in Table 4. It is assumed that the shape of the function follows an exponential distribution, $y = a \times e^{bx}$ of which the parameters are estimated using linear regression methods.

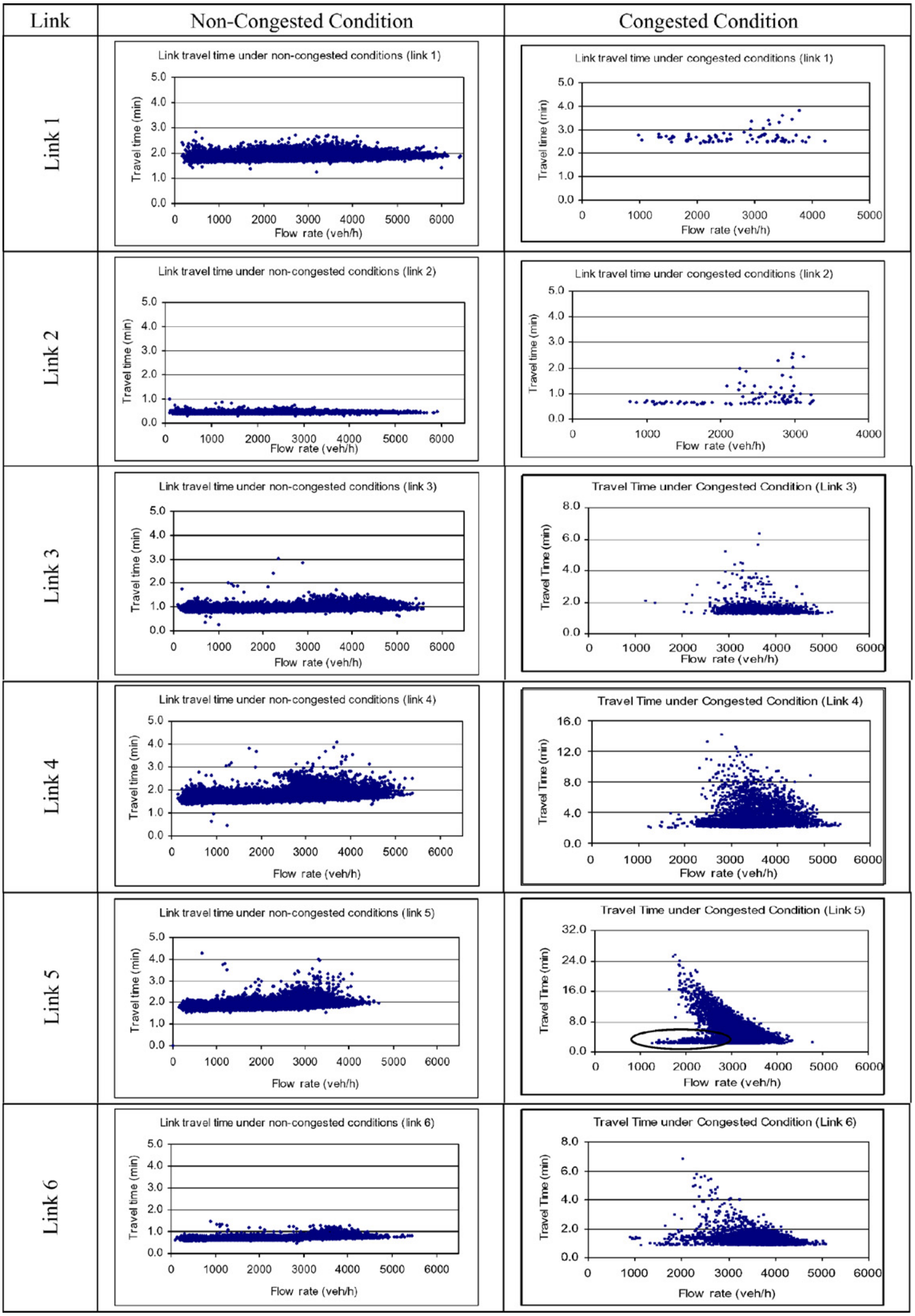


Fig. 5. Travel time scatter plot of each link.

For semi-congested conditions, especially links 3, 4, and 5, travel times are scattered in wide ranges of flow rates with small variance (but somewhat larger than the variance for non-congested conditions), while, the travel time for link 6 decreases slowly as flow rates increase. Thus, the travel time for links 3, 4, and 5 for semi congested conditions is determined to be constant, and for link 6, it is estimated as a linear function of flows. For links 1 and 2, congestion occurrence is rare and its variance is small compared to other links. Thus, in these cases, congested travel time at link 1 and link 2 is considered as the average of all the congested travel times.

- *Step 3: Determination of time periods to account for time of day and daily variations*
  Fig. 6 presents an illustration of daily variations of demand during August 2004 at detector ID 1831 which is the very beginning location of detection locations (total of 14 weekdays not including days with incidents or accidents occurring in either direction), and peak time periods based on traffic volume data. The peak periods are determined to occur when the flow rates exceed 4000 veh/h, which is about 2/3 of the maximum flows at this location. These time periods are distinguished as follows; peak time periods from 6:30 to 9:30 for morning peak, from 15:00 to 19:00 for evening peak, and all other times for off-peak.
- *Step 4: Transition matrix*
  Based on the defined states and congestion occurrences in a 1-min time span, the one-step transition matrices are calculated as shown in Table 5. The process mostly remains in its present state (i.e., $p_{ii}$ has higher probabilities than $p_{ij}$), and it satisfies the *irreducible ergordic* condition. Thus, steady-state probabilities can be obtained by the equations $\Pi = \Pi \cdot P$ (where $\Pi$ is the matrix for the steady-state probabilities and $P$ is the one-step transition matrix) and $\sum_{j=1}^{m} \Pi_j = 1$. The steady-state probabilities, $\Pi_j = [\,\Pi_1 \quad \Pi_2 \quad \Pi_3 \quad \Pi_4 \quad \cdots \quad \Pi_{55} \quad \Pi_{64}\,]$, are calculated using MATLAB, and are $[\Pi_1 = 0.45,\ \Pi_2 = 0.00,\ \Pi_3 = 0.01,\ \Pi_4 = 0.02,\ \ldots,\ \Pi_{55} = 0.00,\ \Pi_{64} = 0.00]$. This implies that most of the time (about 45%) the process is in the non-congested condition (i.e., state 1) whatever the initial state is, and link 6 is congested (i.e., state 4 occurs rarely, about 2%).
- *Step 5: Route travel time estimation*
  Using the estimated travel time of each link and transition matrices of each origin–destination pair-in the previous step, expected route travel times can be estimated by the following equation:

Table 4
Link travel time estimation (min)

| Links | | Non-congested conditions ($NT_i$) | Semi-congested conditions ($SCT_i$) | Fully congested conditions ($FCT_i$) |
|---|---|---|---|---|
| Link 1 | $TT_1$ | 1.88 [65.11] | 2.712 [45.13] | |
| | $\sigma_1$ | 0.101 | 0.285 | |
| Link 2 | $TT_2$ | 0.45 [65.33] | 0.829 [35.46] | |
| | $\sigma_2$ | 0.023 | 0.411 | |
| Link 3 | $TT_3$ | 0.94 [65.11] | 1.431 [42.77] | $T(f) = 0.9982 \times e^{-0.0004 \times f}$ |
| | $\sigma_3$ | 0.059 | 0.152 | $R^2 = 0.4429$ |
| Link 4 | $TT_4$ | 1.71 [63.16] | 2.541 [42.50] | $T(f) = 20.082 \times e^{-0.0004 \times f}$ |
| | $\sigma_4$ | 0.122 | 0.283 | $R^2 = 0.4207$ |
| Link 5 | $TT_5$ | 1.86 [58.71] | 2.569 [42.06] | $T(f) = 68.631 \times e^{-0.0008 \times f}$ |
| | $\sigma_5$ | 0.096 | 0.280 | $R^2 = 0.5654$ |
| Link 6 | $TT_6$ | 0.7 [60] | $T(f) = -0.0002 \times f + 1.7682$ | $T(f) = 12.351 \times e^{-0.0005 \times f}$ |
| | $\sigma_6$ | 0.038 | $R^2 = 0.1889$ | $R^2 = 0.7563$ |

*Note:* [ ] represents equivalent speed in mph $p$-values for semi-congested conditions of links 6 and 7 and fully congested conditions resulting from the $t$-test for each parameter are almost zero. Thus, the null hypothesis (the parameters of each link are greater than or equal to zero) can be rejected.

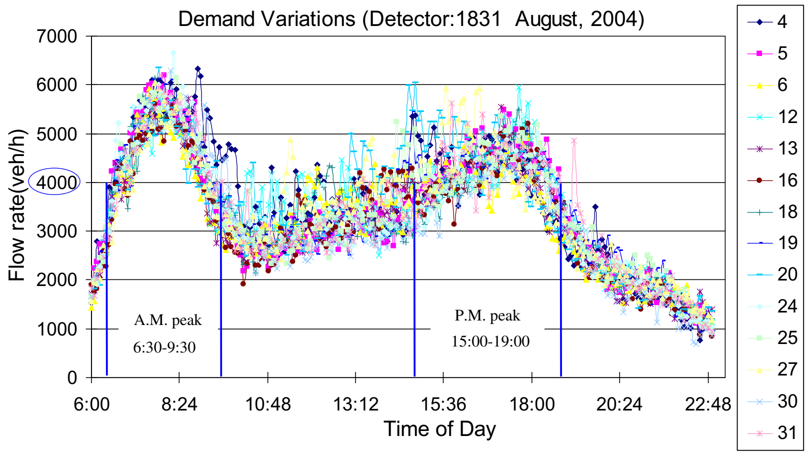


Fig. 6. Demand variations at detector 1831 during August 2004.

Table 5
One-step transition matrix during the p.m. peak period for links 1–6

| $X(t)$ | 1 | 2 | 3 | 4 | 6 | 7 | 12 | 18 | 22 | 28 | 39 | 44 | 46 | 48 | 49 | 50 | 51 | 52 | 54 | 55 | 64 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.99 | 0 | 0 | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0.2 | 0.8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0.07 | 0 | 0.84 | 0 | 0.03 | 0.06 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0.19 | 0 | 0 | 0.67 | 0 | 0.14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0.01 | 0 | 0.04 | 0 | 0.91 | 0.01 | 0.03 | 0 | 0 | 0 | 0 | 0 | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0.01 | 0 | 0.95 | 0.03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0.03 | 0.97 | 0 | 0 | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.95 | 0 | 0 | 0 | 0 | 0 | 0.05 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0 | 0 | 0 | 0 | 0.08 | 0 | 0 | 0 | 0.83 | 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 28 | 0 | 0 | 0 | 0 | 0 | 0 | 0.01 | 0 | 0 | 0.98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 39 | 0 | 0 | 0 | 0 | 0 | 0.5 | 0 | 0 | 0 | 0 | 0.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 44 | 0 | 0 | 0 | 0 | 0 | 0 | 0.33 | 0 | 0 | 0 | 0 | 0.33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.33 | 0 |
| 46 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 48 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.88 | 0.13 | 0 | 0 | 0 | 0 | 0 | 0 |
| 49 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.14 | 0 | 0 | 0 | 0 | 0 | 0.71 | 0.14 | 0 | 0 | 0 | 0 | 0 |
| 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.05 | 0 | 0 | 0 | 0 | 0 | 0.93 | 0 | 0 | 0.03 | 0 | 0 |
| 51 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 52 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.8 | 0 | 0 | 0.2 |
| 54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.05 | 0 | 0 | 0.95 | 0 | 0 |
| 55 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.83 | 0 |
| 64 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.17 | 0 | 0 | 0 | 0.83 |
| $\Pi_j$ | 0.45 | 0 | 0.01 | 0.02 | 0.01 | 0.2 | 0.2 | 0 | 0 | 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

$$
\begin{aligned}
\overline{T} = {} & \Pi_1 \times (\mathrm{NT}_1 + \mathrm{NT}_2 + \mathrm{NT}_3 + \mathrm{NT}_4 + \mathrm{NT}_5 + \mathrm{NT}_6) \\
& + \Pi_2 \times (\mathrm{NT}_1 + \mathrm{NT}_2 + \mathrm{NT}_3 + \underline{p_4 \times \mathrm{SCT}_4 + (1 - p_4) \times \mathrm{FCT}_4(f)} + \mathrm{NT}_5 + \mathrm{NT}_6) \\
& + \Pi_3 \times (\mathrm{NT}_1 + \mathrm{NT}_2 + \mathrm{NT}_3 + \mathrm{NT}_4 + \mathrm{NT}_5 + \underline{p_6 \times \mathrm{SCT}_6 + (1 - p_6) \times \mathrm{FCT}_6(f)}) \\
& \vdots \\
& + \cdots + \Pi_{64} \times (\underline{\mathrm{SCT}_1} + \underline{\mathrm{SCT}_2} + \underline{p_3 \times \mathrm{SCT}_3 + (1 - p_3) \times \mathrm{FCT}_3(f)} \\
& + \underline{p_4 \times \mathrm{SCT}_4 + (1 - p_4) \times \mathrm{FCT}_4(f)} + \underline{p_5 \times \mathrm{SCT}_5 + (1 - p_5) \times \mathrm{FCT}_5(f)} \\
& + \underline{p_6 \times \mathrm{SCT}_6 + (1 - p_6) \times \mathrm{FCT}_6(f)}),
\end{aligned} \tag{6}
$$

Table 6
The $p_i$ of link $i$

| Links | Number of intervals in semi-congested conditions ($A$) | Number of intervals in fully-congested conditions ($B$) | Total ($C = A + B$) | $p_i$ ($=A/C$) |
|---|---|---|---|---|
| Link 1 | – | – | – | – |
| Link 2 | – | – | – | – |
| Link 3 | 1213 | 160 | 1373 | 0.883 |
| Link 4 | 2919 | 1422 | 4341 | 0.672 |
| Link 5 | 2319 | 4881 | 7200 | 0.322 |
| Link 6 | 5817 | 748 | 6565 | 0.886 |

where $p_i$ is the proportion of data points in semi-congested conditions to the total number of points under congested conditions (both semi-congested and fully congested conditions) for link $i$. These values are provided in Table 6.

In Eq. (6), the sum of $\mathrm{NT}_i$ in the first line represents the travel time when the system is in state 1 (i.e., all links are not congested). The underlined portions represent the travel time when the system includes congested conditions. Each line in the equation represents the estimated travel time corresponding to the system status, and the steady-state probability of each state is the probability that the system is in that state. The expected travel time can be estimated by summation of the steady-state probability of each state times the travel time when the system is in that state:

$$\begin{aligned}\overline{T} = {} & 0.45 \times (1.88 + 0.45 + 0.94 + 1.71 + 1.86 + 0.7) \\ & + 0.00 \times (1.88 + 0.45 + 0.9 + \underline{0.672 \times 2.541 + (1 - 0.672) \times 9.02} + 1.86 + 0.7) \\ & + 0.01 \times (1.88 + 0.45 + 0.94 + 1.71 + 1.86 + \underline{0.886 \times 1.37 + (1 - 0.886) \times 4.54}) \\ & \vdots \\ & + \cdots + 0.00 \times (\underline{2.712} + \underline{0.829} + \underline{0.883 \times 1.431 + (1 - 0.883) \times 1.44} \\ & + \underline{0.672 \times 2.541 + (1 - 0.672) \times 9.02)} + \underline{0.322 \times 2.569 + (1 - 0.322) \times 13.856} \\ & + \underline{0.886 \times 1.37 + ((1 - 0.886) \times 4.54)} = 13.52 \text{ min.}\end{aligned} \tag{7}$$

The expected route travel time for links 1–6 is calculated using the above equation for a flow rate of 2000 veh/h and using data from Tables 4–6. For example, $\Pi_1 = 0.453$ (steady-state probability for state 1), $\mathrm{NT}_1 = 1.88$ (Table 4 for link 1), $p_4 = 0.672$ (Table 6 for link 4), $\mathrm{SCT}_4 = 2.54$ (Table 6 for link 4), $\mathrm{FCT}_4$ ($f = 2000$) $= 9.02$ (Table 4 for link 4), etc. The expected route travel time for links 1–6 during the a.m. peak period is estimated to be 13.52 min.

## 6. Model validation

After the model development, the field-measured travel time and the estimated travel time are compared to validate the model. Among the collected travel time data, two sets of route travel time data were used; C205 & C213 (e.g., link 1–link 6) for 17:50–19:00 on May 25th and for 15:30–16:30 on June 16th, 2004. The estimated travel time for 17:50–19:00 is 7.94 min, while the average measured travel time is 8.13 min and its variance is 0.993 min. In addition, the estimated travel time for 15:30–16:30 is 8.28 min, while the average measured travel time is 8.05 min and its variance is 1.129 min. The ratio of estimated travel time to average measured travel time for 17:50–19:00 is 0.977, and for 15:30–16:30 is 1.028. To compare the expected travel time to measured travel time, the paired $t$-test for the mean difference is conducted (i.e., $\mathrm{H}_0$: $\mu_{\text{Estimated TT}} - \mu_{\text{Field TT}} = 0$ and $\mathrm{H}_1$: $\mu_{\text{Estimated TT}} - \mu_{\text{Field TT}} \neq 0$). Table 7 shows the statistical test results, which show there is no statistical evidence that the estimated travel time differs from the measured travel time at the 99% confidence level for both time periods.

Table 7
Statistical results of the model validation ($\alpha = 0.01$)

| Paired $T$ for estimated TT − field TT | $N$ | Mean | St. Dev. | SE mean |
|---|---|---|---|---|
| *1. May 25th* | | | | |
| Estimated TT | 55 | 7.94645 | 0.79060 | 0.10660 |
| Field TT | 55 | 8.13127 | 0.99378 | 0.13400 |
| Difference | 55 | −0.184818 | 0.738757 | 0.099614 |
| 95% CI for mean difference: (−0.384532, 0.014896) | | | | |
| $T$-test of mean difference = 0 (vs not = 0): $T$-value = −1.86, $P$-value = 0.069 | | | | |
| *2. June 16th* | | | | |
| Estimated TT | 150 | 8.28220 | 0.98595 | 0.08050 |
| Field TT | 150 | 8.05187 | 1.13354 | 0.09255 |
| Difference | 150 | 0.230333 | 0.561637 | 0.045857 |
| 95% CI for mean difference: (0.139718, 0.320948) | | | | |
| $T$-test of mean difference = 0 (vs not = 0): $T$-value = 2.37, $P$-value = 0.019 | | | | |

## 7. Conclusions and recommendations

This study focuses on estimating travel time on a freeway using DTMC, in which the system state changes every minute. The model considers the probability of breakdown along each freeway link, and it subsequently estimates the expected travel time for the entire route as a function of those probabilities of breakdown. The model developed was found to match field travel time estimates very well, with deviations of less than 3%. According to statistical testing, there is no evidence that the estimated travel time differs from the measured travel time at the 99% confidence level.

The model developed provides expected route travel time for a given facility. The same methodology could be applied to other freeway segments if appropriate speed and flow data are available. Traffic Management Centers (TMCs) can relatively easily use the model developed and calibrate it to their needs to obtain travel time estimates for various time periods. Also, the model can be used to forecast travel times during a given time period if appropriate demand data could be forecasted. Potential extensions to this model include travel time estimation considering incidents, work zones, and adverse weather conditions.

## Disclaimer

The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views of NSF.

## Acknowledgements

The authors wish to acknowledge Mobility Technologies for providing speed and flow data, and the Pennsylvania Department of Transportation for providing access to the Philadelphia Traffic Management Center. This work was sponsored by the National Science Foundation (NSF) under 03-556.

## References

Choe, T., Skabardonis, A., Varaiya, P., 2002. Freeway Performance Measurement System (PeMS): an operational analysis tool. In: Transportation Research Board 81st Annual Meeting, Washington DC.

Chu, L., 2005. Adaptive Kalman filter based freeway travel time estimation. In: Transportation Research Board 84th Annual Meeting, Washington DC.

Coifman, B., Ergueta, E., 2003. Improved vehicle reidentification and travel time measurement on congested freeways. Journal of Transportation Engineering 129 (5), 475–483.

Del Castillo, J.M., 2001. Propagation perturbation in dense traffic flow: a model and its implications. Transportation Research Part B 35 (4), 367–389.

Elefteriadou, L., Roess, R.P., Mcshane, W.R., 1995. Probabilistic nature of breakdown at freeway merge junction. Transportation Research Record 1484, 80–89.

Evans, J., Elefteriadou, L., Natarajan, G., 2001. Determination of the probability of breakdown on a freeway based on zonal merging probabilities. Transportation Research Part B 35 (3), 237–254.

Heidemann, D., 2001. A queueing theory model of nonstationary traffic flow. Transportation Science 35 (4), 405–412.

Kharoufeh, J.P., Gautam, N., 2004. Deriving link travel time distributions via stochastic speed processes. Transportation Science 38 (1), 97–106.

Lorenz, M., Elefteriadou, L., 2000. A probabilistic approach to defining freeway capacity and breakdown. In: Fourth International Symposium on Highway Capacity, Hawaii.

Oh, J., Jayakrishnan, R., Recker, W., 2003. Section travel time estimation from point detection data. In: Transportation Research Board 81st Annual Meeting, Washington DC.

Persaud, B., Yagar, S., Tsui, D., Look, H., 2001. Breakdown-related capacity for freeway with ramp metering. Transportation Research Record 1748, 110–115.

Ross, S., 1996. Stochastic Processes, second ed. John Wiley and Sons.

Shaw, T., 2002. Reliability Measures for Highway Systems and Segments. TRANSYSTEMS Corporation.

Turner, S.M., 1996. Advanced techniques for travel time data collection. In: Transportation Research Board 75th Annual Meeting, Washington DC.

Turner, S., Margiotta, R., Lomax, T., 2004. Lessons learned: monitoring highway congestion and reliability using archived traffic detector data, FHWA-OP-05-003.

VanLint, J.W.C., van Zuylen, H.J., 2005. Monitoring and predicting freeway travel time reliability using width and skew of the day-to-day travel time distribution. In: Transportation Research Board 84th Annual Meeting, Washington DC.

Yang, H., Lo, H.K., Tang, W.H., 1999. Travel time versus capacity reliability of a road network. In: Bell, M.G.H., Cassir, C. (Eds.), Reliability of Transport Networks. Research Studies Press, Baldock, pp. 119–138.

Zhang, X., Rice, J., 2003. Short-term travel time prediction. Transportation Research Part C 11 (3–4), 187–210.