DEVOPS-AN OPEN ACCESS JOURNAL
Open Journal

DevOps-An Open Access Journal

Review

Characterizing Corridor-Level Travel Time Distributions Based on Stochastic Flows and Segment Capacities

Hao Lei[1*]; Xuesong Zhou[2]; George F. List[3]; Jeffrey Taylor[1]

[1]*Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT 84112, USA*
[2]*School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA*
[3]*Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC 27695-7908, USA*

[*]***Corresponding author***
Hao Lei
School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA; E-mail: hlei@eng.utah.edu

Article information
Received: September 23rd, 2022; **Revised:** December 3rd, 2022; **Accepted:** December 23rd, 2022; **Published:** January 6th, 2023

Cite this article
Lei H, Zhou X, List GF, Taylor J. Characterizing corridor-level travel time distributions based on stochastic flows and segment capacities. 2023; 2(1).
doi: https://doi.org/10.70705/ppp.doaj.2023.v02.i01.pp1-9

ABSTRACT

Trip travel time reliability is an important measure of transportation system performance and a key factor affecting travelers' choices. This paper explores a method for estimating travel time distributions for corridors that contain multiple bottlenecks. A set of analytical equations are used to calculate the number of queued vehicles ahead of a probe vehicle and further capture many important factors affecting travel times: the prevailing congestion level, queue discharge rates at the bottlenecks, and flow rates associated with merges and diverges. Based on multiple random scenarios and a vector of arrival times, the lane-by-lane delay at each bottleneck along the corridor is recursively estimated to produce a route-level travel time distribution. The model incorporates stochastic variations of bottleneck capacity and demand and explains the travel time correlations between sequential links. Its data needs are the entering and exiting flow rates and a sense of the lane-by-lane distribution of traffic at each bottleneck. A detailed vehicle trajectory data-set from the Next Generation SIMulation (NGSIM) project has been used to verify that the estimated distributions are valid, and the sources of estimation error are examined.

Keywords
Travel time reliability; Stochastic capacity; Stochastic demand; Queue model.

INTRODUCTION

Travel time has long been regarded as one of the most important performance measures in transportation systems. Recently, significant attention has been devoted to evaluating and quantifying the reliability of travel times due to the influence of travel time variability on route, departure time, and mode choices. Operating agencies have also increased their efforts to monitor and improve the reliability of their systems through probe-based data collection, integrated corridor management, and advanced traveler information systems. Corridor management strategies have been designed to balance the performance of freeway and arterial networks in response to congestion and non-recurring events. Although noteworthy progress has been made in quantifying the variability in travel times, a number of challenges still need to be addressed. One is how to estimate distributions of individual vehicle travel times for both recurring and non-recurring congestion conditions, especially since both the demand and the capacity is stochastic. A framework for doing this is vital for both travelers and operating agencies (e.g. traffic management team) if they are to make informed decisions about actions that improve reliability.

Within the subject of analytical dynamic traffic network analysis, the "whole-link" model is widely adopted to describe link travel time evolution due to its simple description of traffic flow propagation through an analytical form. The link travel time function introduced by Friesz, Bernstein, Smith, Tobin, and Wie (1993) defines the travel time r(t) on a single link at a time t as a linear function of the number of vehicles x(t) on the link at time t:

$$r(t) = a + bx(t)$$

where a and b are constants in the above general linear form. A non-decreasing and continuous function is defined to calculate the cumulative number of vehicles on the link based on the time- dependent inflow and outflow rates, u(t) and v(t), at time t:

$$x(t)=x(0)+\int((u(s)-v(s))ds \quad \text{----------(2)}$$

Meanwhile, some more general non-linear travel time functions have been proposed as:

$$r(t)=f\ (x(t), u(t),v(t)) \quad \text{----------(3)}$$

A special case of this form, introduced by Ran, Boyce, and LeBlanc (1993), decomposes the link travel time as two different functions: g1 accounts for flow-independent travel time and g2 accounts for the queuing delay. A detailed mathematical representation is shown below.

r(t) = g1 [x(t), u(t)] + g2 [x(t),v(t)] ----------(4)

They later showed that, by assuming g1 and g2 are separable, i.e.

g1 = g1a[x(t)] + g1b[u(t)] and

g2 = g2a[x(t)] + g2b[v(t)], and Equation 4 can be rewritten as

r(t) = α + f (u(t)) + g(v(t)) + h(x(t)) ----------(5)

where is the free flow travel time, and f (□), g(□) and h(□) correspond to the functions of link inflow rate, link outflow rate, and the number of vehicles on the link, respectively.

Daganzo (1995) draws attention to problems with the general form in Equation 3, indicating that either a rapid decline in the inflows u(t) or a rapid increase in outflow v(t)would lead to unrealistic travel time. Thus, he recommended omitting u(t) and v(t) from Equation 3, reducing the link travel time to a function of the number of vehicles on the link, that is, r(t)=f (x(t)). Although the link travel time function models provide some degree of simplification on travel time analysis, there is one significant drawback. Traffic congestion usually occurs at some bottleneck, and queues are produced and often grow beyond the bottleneck, which is difficult for any travel time function to capture (Zhang & Nie, 2005).

In dynamic traffic assignment and other applications, the vertical queue or point-queue model (Daganzo, 1995) was widely adopted to describe bottleneck traffic dynamics (Zhang & Nie, 2005). In a queuing-based travel time model, it is important to capture the variations of queue discharge flow rates and incoming demand to a bottleneck.

Conventionally, freeway capacity is viewed as a constant value—the maximum discharge flow rate before failure (HCM, 2000). However, the capacities vary according to different external factors in real life situations. Over the past decades, many researchers have developed a number of headway models to describe its distribution. Representatives of these models include the exponential- distribution by Cowan (1975), and normal distribution, gamma-distribution, and lognormal- distribution models by Greenberg (1966).

Incidents are one of the major contributing factors in capacity reductions, and the magnitude and duration of capacity reductions are directly related to the severity and duration of incidents (Giuliano, 1989; Kripalani & Scherer, 2007). In quantifying capacity reduction, the HCM (2000) provides guidance for estimating the remaining freeway capacity during incident conditions. Using over two years of data collected on freeways in the greater Los Angeles area, Golob, Recker, and Leonard (1987) found that accident duration fit a lognormal distribution. By extending the research of Golob et al. (1987) and Giuliano (1989) applied a lognormal distribution when analyzing incident duration for 512 incidents in Los Angeles.

It is commonly observed that travel demand fluctuates significantly within a day. During the morning and evening peak hours, surging demand may overwhelm a roadway's physical capacity and results in delays (Federal Highway Administration [FHWA], 2009). Waller and Ziliaskopoulos (2001), Chen, Skabardonis, and Varaiya (2003) and Lam, Shao, and Sumalee (2008) have used the normal distribution for modeling travel demand variation. Other researchers have modeled travel demand using the Poisson distribution (Clark & Watling, 2005; Hazelton, 2001) and the uniform distribution (Unnikrishnan, Ukkusuri, & Waller 2005).

Substantial efforts have been devoted to travel time variability estimation over the last decade. Statistical approaches (Oh & Chung, 2006; Richardson, 2003) have been widely adopted to quantify travel time variability from archived sensor data. In recent studies investigating the different sources of travel time variability, Kwon, Barkley, Hranac, Petty, and Compin (2011) proposed a quantile regression model to quantify the 95th percentile travel time based on the congestion source variables, such as incidents and weather. In their multi-state travel time reliability modeling framework, Guo, Rakha, and Park (2010) and Park, Rakha, and Guo (2011) provided connections between the travel time distributions and the uncertainty associated with the traffic states, e.g. with incidents versus without incidents.

A second approach uses numerical methods to characterize travel time distributions as a result of stochastic capacity and stochastic demand. Given a stochastic capacity probability distribution function (PDF), a Mellin transforms-based method was adopted by Lo and Tung (2003) to estimate the mean and variance of travel time distributions. Using a sensitivity analysis framework, Clark and Watling (2005) developed a computational procedure to construct a link travel time PDF under stochastic demand conditions. Ng and Waller (2010) introduced a fast Fourier transformation approach to approximate the travel time PDF from underlying stochastic capacity distributions. Although it can quantify the impacts of demand and capacity variation on the travel times, the steady-state travel time function-based approach is still unable to address the underlying time- dependent traffic dynamics.

In order to account for the inherent time-dependent traffic dynamics, some researchers (e.g. Zhou, Rouphail, & Li, 2011) have incorporated point-queue models into travel time variability estimation techniques. Using a dynamic traffic assignment simulator, Alibabai (2011) developed an algorithmic framework to investigate the properties of the path travel time function with respect to various path flow variables. While realistic simulation results require significant efforts in simulation/ assignment model calibration, this approach is particularly suited for studying the effects of various uncertainty sources and assessing the benefits of traffic management strategies and traffic information systems.

The objective of this paper is to describe a model that can be used to estimate individual vehicle travel time distributions at the route

level based on relatively simple information about the configuration of the corridor and its flows and capacities. Potential applications are investigated, and illustrated examples are presented. In addition, a highly detailed Next Generation SIMulation (NGSIM) vehicle trajectory data-set is used to check the validity of the travel time distributions under different traffic conditions.

The rest of the paper is organized as follows. Section 2 describes the point-queue-based travel time estimation framework with deterministic inputs. Monte Carlo simulation is then applied in Section 3 to compute the travel time distribution with stochastic inflow, outflow, and discharge rates. Highly detailed vehicle trajectory data from the NGSIM project is utilized to validate our methods in Section 4.

2. Computing route-level travel times

2.1. Problem statement

Consider a corridor with M bottlenecks, where each node in the node-link structure represents a bottleneck, and the road segments between consecutive bottlenecks are links with homogeneous capacity. Assume that node 0 is the starting point of the corridor, node m corresponds to bottleneck m, and each link between bottlenecks is denoted as link (m − 1, m), for 1 m M. Link (m − 1, m) is the same as link m. Figure 1 illustrates a node-link representation for a corridor with M bottlenecks. Possible merge or diverge nodes are connected to bottleneck m and are denoted as m' or m", respectively, so that the on-ramp before node 1 is denoted as (1', 1), and the off-ramp before node 2 is denoted as (2, 2"). In other words, the merge and diverge links are directly connected to the bottleneck. If there is more than one inflow or outflow between two bottlenecks, one can further decompose the link between those inflow entrances or outflow exits to several segments and merge/diverge points so as to construct the above node-link representation.

For the purposes of this analysis, the interest lies in how to estimate the travel time distribution for trips from point 0 to point m for a probe vehicle z, departing at time t0 = 0. The aim is to estimate the distribution of the route-level (route) travel time, pz , based on the following: (1) the number of
vehicles xm(t0) on each link m along the path at time t0, (2) the discharge flow rate for each bottleneck
c , and (3) the on-ramp or off-ramp flow rates f net. The route-level travel time is defined to be the difference between the departure time at bottleneck 0 and the departure time at bottleneck m for probe vehicle z. Further, the departure time at bottleneck m is defined to be the time the probe vehicle leaves the queue at bottleneck m, which is the time when the number of vehicles in the queue before probe vehicle z at bottleneck m is 0.

The number of vehicles on each link is assumed to be observable from sensors, such as loop detectors, and the discharge flow rates and net flow rates on the on-ramps and off-ramps are assumed to be estimable from historical flow patterns or estimated based on prevailing traffic conditions (e.g. capacity reduction due to incidents).

The notation for the route-level travel time is described below.

Indices net flow rate at a merge or diverge corresponding to bottleneck m, that is, from an on-ramp to the mainline segment or from the mainline to the off-ramp

2.2. Travel time calculation

In a point-queue model, a link can be considered as two segments—the free-flow segment and the queuing segment. A vehicle travels at free-flow speed on the free-flow segment until reaching the beginning of the queuing segment, where it joins the queue waiting to be discharged. A queue is only formed if the link demand exceeds the bottleneck capacity; that is, the link arrival rate exceeds the link departure rate.

To construct a numerically tractable model for calculating route-level travel times along a corridor with multiple bottlenecks, several important assumptions are made.

(1) A point-queue model is adopted to calculate the delay on each link. On each link, a First-In, First-Out (FIFO) property is assumed to assure that any vehicles that enter the link before time t will exit the link before those entering after time t.

(2) The link traversal time is assumed to comprise a free-flow travel time and a queuing delay. The free-flow travel time is constant and flow-independent. The queuing delay is dependent on
the number of vehicles in the queue when the probe vehicle arrives at the bottleneck ß(t +FFTT)
and the bottleneck queue discharge rate cm. Thus, the link travel time is

where w(t +FFTT) is the queuing delay when vehicle z reaches the vertical queue at the bottleneck at time t +FFTT.

(3) The merge or diverge location is coincident with the position of the vertical queue.

(4) The bottleneck m remains congested across the estimation horizon, which extends from the z current time t0 to the arrival time of the probe vehicle z at the bottleneck m, t .

The corresponding queue discharge rates cm are also assumed to be constant. and net flow rates f net in the estimation horizon

The first two assumptions are widely used in queuing models. The third makes it easy to incorporate the flow rate from a merge/diverge without explicitly considering the driving distance and free-flow travel time from the merge/diverge point to the bottleneck m.

Equation 6 considers the arrival time at the beginning of a link. By considering the arrival time at bottleneck m for vehicle z, tz , the link traversal time can be rewritten as

For a general queue with time-dependent arrival and departure rates, a continuous transition model can be used in Equation 6 to update the number of vehicles in the queue at any given time t.

The number of queued vehicles ßm

z at time tz on bottleneck m can be derived from Equation 8,

as shown in Equation 9.

Since the fourth assumption requires the bottleneck to remain extant for the entire estimation period, the departure rate is equal to the bottleneck capacity of remaining challenge∫ is to estimate the unknown queue length ßm(t0) at time t0, and calculate the

To illustrate these ideas, consider the example in Figure 1, where m = 1. In this case, the number of vehicles x (t) and the net flow rate f net associated with bottleneck 1 are given. For the specific starting time t , a probe vehicle z enters the vertical queue of bottleneck 1 at time tz = t +FFTT , and the number of vehicles in the queue at time tz is

Now consider a simpler case without merge and diverge points, i.e. f net = 0. Thanks to the first-in and first-out property, we can show that the queuing segment before the probe vehicle z. The right-hand side is the actual number of vehicles observed on the physical link. One can use Figure 2 to map or "rotate" some of the vehicles from the physical link (shaded) to the vertical stack queue, and the other vehicles on the physical link (not saded) correspond to the vehicles that will arrive at the vertical queue between time t0 and time FFTT1 (that is, right before the probe vehicle). Notice that the length of the queue segment in the point-queue model is equal to zero and has unlimited storage capacity. Interested readers are referred to the paper by Hurdle and Son (2001) to examine the connection between physical queues and vertical stack queues.

The individual components of Equation 10 can be described visually using the cumulative vehicle count curves shown in Figure 3. Curve A is equivalent to the integral over the arrival rate, $A(t) = \mu 1(t)dt$, and the cumulative arrival curve at the vertical stack queue V is the cumulative arrival curve shifted by
the free-flow travel time, $V(t) = A(t - FFTT)$, and thus $V(t) = \mu 1(t - FF\int TT1)dt$. The cumulative

By further considering the net flow rate from the merge or diverge point connected to the bottle- neck, we now have Continuing to link 2 in Figure 1, the probe vehicle z will arrive at the queue of bottleneck 2 at tz. Again, considering the FIFO assumption, the number of vehicles transferring from the first link to the second link before the probe vehicle z includes two terms, ß1 + c × t , which are the number of queued vehicles ß1 when the probe vehicle arrives at the first bottleneck at time t , and those
vehicles c × tz already entering the second link before time tz. Following the derivation logic for Equation 11, the number of vehicles waiting in the queue ahead of vehicle z when it arrives at the second bottleneck at time tz is

More generally, for bottleneck m:

(1) The number of vehicles waiting at the vertical queue of bottleneck m at time tz can be

expressed as

In summary, given the number of vehicles on each link, the queue discharge rate and the net flow on each bottleneck, the route-level travel time for a vehicle can be calculated by applying Equations 14–16 iteratively for links 1 through m. In each iteration, one first applies Equation 14 to obtain the number of queued vehicles at the bottleneck and then computes the queuing delay and update the route-level travel time up to the bottleneck of interest.

2.3. Illustrative example

To demonstrate how to the model can be used to calculate the route-level travel time and capture the delay propagation along a corridor, a corridor with 3 bottlenecks (Figure 4) can be used. Bottleneck 1 is on the freeway, bottleneck 2 is associated with an on-ramp, and bottleneck 3 is in conjunction with an off-ramp. The bottleneck discharge rates for those bottlenecks are 90, 90, and 60 vehicles/min, respectively. The initial number of vehicles on each link is 750, 600, and 650, respec- tively. The inflow rate for the on-ramp at bottleneck 2 is 20 vehicles/min (vpm), which is equivalent to 1,200 vehicles per hour. The outflow rate for the off-ramp is 18 vehicles/min (vpm). The free-flow travel time over each link is 5, 4, and 4.5 min, respectively.

For the probe vehicle in Figure 4 (starting at time 7:00 am), we now have the following calculation process for its route-level travel time.

(1) Departing at 7:00, it takes 5 min (free-flow travel time) for this probe vehicle to reach the point-queue of bottleneck 1 at 7:05. At this time instance, the number of vehicles waiting in

the queue is ß1 = 750 – (5 min × 90 veh/min) = 300 vehicles. With the discharge rate of
90 vehicles/min, this probe vehicle will spend wz = 3.33 min waiting in the queue. Thus, the total travel time for this vehicle is 8.33 min at the end of this bottleneck.

(2) The probe vehicle enters link 2 at 7:08.33, spends 4 min traveling through the free-flow segment, and arrives at the vertical stack queue at tz = 7:12.33. From 7:00 to 7:12.33, there have been 12.33 min × 20 veh/min = 246.6 vehicles entering this bottleneck from the on-ramp.

z

The number of vehicles waiting in the queue at this time is ß2 = (750 + 600) +
(12.33 × 20) − (12.33 × 90) = 486.67. With the discharge rate of 90 vehicles/min, this vehicle leaves the queue wz = 5.41 min later. The departure time from the second bottleneck is
7:17.74.

(3) Following the same calculation process, the number of vehicles waiting at the queue of

z

bottleneck 3 is ß3 = (750 + 600 + 650) + (12.33 × 20) + (−18 × 22.24) − (60 × 22.24) = 511.89
vehicles and the waiting time in the queue is wz = 8.53 min. This vehicle leaves bottleneck 3 at 7:30.77. The total route-level-travel time

pz is 30.77 min.

2.4. Travel time calculation algorithm with deterministic inputs

The algorithm for calculating the route-level path travel time for vehicle z entering the corridor with

M bottlenecks at time t0 is summarized below.

Input: The specific starting time t0, the number of vehicles on each link xm(t0), the net flow rate on each bottleneck f net, and the bottleneck discharge rate c , at time t .

m m 0

Route-level travel time calculation

For m = 1 to M

(1) Calculate the arrival time at bottleneck m

(2) Use Equation 14 to calculate the number of vehicles ahead of the probe vehicle z in the vertical

stack queue of bottleneck m, ßz z , when the probe vehicle z reaches the beginning of the

queue at time tz .

(3) Use Equation 15 to calculate the delay experienced by the probe vehicle on bottleneck m, wz .

(4) Use Equation 16 to update the route-level travel time over m, pz

End For

Output: The route-level travel time pz from bottleneck 1 to bottleneck M.

2.5. Discussions

To consider complex real-life conditions, the model must further use the following approximation methods for calculating the route-level path travel time along a corridor with multiple bottlenecks.

2.5.1. Approximating the time-dependent flow rates with average flow rates

In Equations 9 and 10, we use the maximum bottleneck discharge rates to approximate the actual discharge rates. In reality, the rates (including the queue discharge flow rates and net flow rates from and to ramps) are highly dynamic and could fluctuate significantly even in a short time interval, as shown in Figure 5. In this situation, one needs to use the average flow rate (i.e. the dashed line in

z

Figure 5) during the interval from t0 to the arrival time t to approximate the time-dependent

volume. Although this approximation ignores traffic dynamics, in Equation 14 it still gives a reasonable estimate about the total number of vehicles leaving or entering the bottleneck before the probe vehicle. Section 4 will examine the possible impact from using this approximation method for representing travel time distributions.

2.5.2. Considering further reduced bottleneck discharge flow rate due to queue spillback

The proposed point-queue-based model captures the effects of queue spillback from a downstream bottleneck. Essentially, when a queue spillback occurs, the discharge capacity from the upstream bottleneck is then constrained by the discharge rates at the downstream bottleneck. The method detects spillback and then uses the reduced queue discharge rate to calculate the waiting time at the bottleneck with queue spillback.

As illustrated in Figure 6, the physical queue for bottleneck m spills back to bottleneck (m − 1) between time t1 and t5 through backward waves [Interested readers are referred to the paper by Newell (1993) to learn more]. Due to the queue spillback from bottleneck m, the actual discharge

rate cf of bottleneck (m − 1) between time t and t is constrained by the discharge rate of

1 5

bottleneck m, cm, rather than the original discharge rate cm−1. For example, at time t2 (where t2 > t1), a probe vehicle arrives at bottleneck (m − 1), if the effect of queue spillback is not taken into account,

this probe vehicle in the model will leave bottleneck (m − 1) at time t3 after waiting in the queue behind bottleneck (m − 1), using the original, unaffected queue discharge rate cm−1. With the reduced discharge rate cm < cm−1at bottleneck (m − 1), the actual waiting time for the probe vehicle will be longer with a departing time of t4 > t3.

2.5.3. Calculating the net flow rate for on-ramps

When estimating the net flow at a merge or diverge location, the flow rates in previous instances are assumed to be known and time-invariant. However, special attention must be paid to conditions where the mainline and the on-ramp are both congested. In such instances, (1) the number of vehicles that can enter the bottleneck from the on-ramp and (2) the number of vehicles that can enter from the upstream segment to the bottleneck are constrained by the mainline bottleneck discharge rate. In this case, the available bottleneck discharge rate should be allocated to the upstream segment and the on-ramp proportionally, according to certain rules (Zhang & Nie, 2005). One simple rule is to split the mainline discharge rates according to the number of lanes associated with each incoming approach.

2.5.4. Considering vehicle overtaking/passing

Lastly, the FIFO property assumed on each link rules out the possibility that a vehicle can overtake and pass another vehicle. Future research will consider the impact of this condition on route-level travel time estimation using this approach.

3. Methods for calculating route-level travel time distributions

3.1. Assumptions

In the previous discussion, input parameters such as the net rates f net at the merge and diverge points, and the bottleneck discharge rates cm are assumed to be deterministic. In this section, we will further consider the variations or uncertainty in the input parameters, especially in the following two applications: (1) day-to-day travel time variability estimation by considering flow variations at the same time

period; and (2) real-time travel time reliability estimation, where the near-future traffic flows are estimated from different sources of data with various degrees of estimation uncertainty. Emphases are placed on how to calculate the route-level travel time distribution based on the stochasticity of the random input parameters.

3.2. Important observations on path travel time

3.2.1. Simple corridor without merge and diverge

Consider a simple two-bottleneck corridor with no on-ramp and off-ramp, that is, f net and f net are
1 2
equal to 0. According to Equations 14–16, the route-level travel time to bottleneck 1 for probe vehi- cle z entering link 1 at time t0 is:

The above equations introduce more complex dependencies for both pz and pz, and no additive
1 2
formula or decomposed elements can be easily constructed to simplify these equations. This observation reinforces many previous research studies which indicate that development of the route-level travel time distribution is extremely challenging.

3.3. Monte Carlo simulation

Monte Carlo simulation is widely used to simulate the behavior of various physical and mathematical systems, especially for those problems with significant uncertainty in inputs. The model presented here uses Monte Carlo simulation to investigate the route-level travel time distribution based on the proposed travel time calculation framework. In each simulation run, a realization of the random input parameters leads to a realization of the random path travel time outputs, which can be regarded as estimates of the true route-level travel time variable. A sufficient number of simulations then provide a good representation of the travel time distributions under various traffic conditions and uncertainties.

The following procedure assumes all random variables are log-normally distributed, and calculates travel time distribution through K simulation runs.

Input:
The specific starting time t0,
The distribution of the number of vehicles on each link xm

The distribution of the bottleneck discharge rate cm on each bottleneck, at time t0 where

Link free-flow travel time FFTTm, assumed to be constant. Number of simulations = K.

1: Based on the underlying distribution parameters (μ and) of the individual inputs, generate a set of random samples for the following key variables: the number of vehicles on the link, the bottleneck discharge rate, and net flow rates.
2: Call the algorithm introduced in Section 2.4 to calculate the estimated route-level travel time for simulation k: pz [k] from this set of random samples.
End For End For
Output: Calculate the histogram, mean, and variance for the route-level travel time from the results over K simulation runs.

3.4. Numerical experiments

3.4.1. Monte Carlo simulation

For the same example corridor in Section 2.3 (with three bottlenecks), Monte Carlo experiments were conducted to calculate the route-level travel time by assuming that the bottleneck discharge rates, inflow/outflow rates on ramps, and existing number of vehicles on the link are all lognormal variables. K = 100 simulation runs were performed with different scenarios of stochastic input parameters.
vehicle z through bottleneck 2 and through bottleneck 3, respectively. Obviously, the mean travel
time based on pz is larger than that of pz. In addition, a clear propagation of randomness can be observed, as pz has higher variance than pz. It should be noted that, by using different input distributions for flow discharge rates and the prevailing number of vehicles on the road, the resulting travel time distributions will vary. This demonstrates the advantage of the proposed model in recognizing the impact of capacity and congestion levels on travel time reliability.

3.4.2. Evaluating the improvement in reliability for trafic management strategies

In the previous section, Monte Carlo simulation was used to demonstrate the application of the proposed travel time estimation framework on route-level travel time distribution quantification. In this section, the calculation framework is further applied to evaluate the effectiveness of advanced traffic management strategies (ATMS).

To demonstrate the use of the proposed calculation framework, a 1-mile long 4-lane freeway corridor with the average bottleneck discharge rate of 2,000 vehicles per hour per lane is investigated. Before implementing ATMS, the probability of incidents on this corridor is 20% and 3 lanes are closed due to an incident if an incident occurs. After implementing ATMS (e.g. rapid incident response teams), the number of lanes closed due to incident is reduced to 2.

Figure 8(a–b) and (c–d), respectively, show the simulated distribution of the bottleneck discharge flow rate and the calculated route-level travel time, before and after the implementation of the ATMS. The first peak of Figures 8(a) and (c) represents the average capacity under incidents and the second peak represents the capacity under normal conditions. By comparing Figure 8(b) with (d), we can observe that the travel time distribution with ATMS has a shorter tail and less fluctuation. Figure 8(e) further reveals the potential benefit of ATMS in improving the travel time reliability: the 95th percentile travel time is improved from 25 min (without ATMS) to 13 min (with the implementa- tion of ATMS), while interestingly the median travel time has not changed significantly.

4. Model validation using NGSIM data

This section uses vehicle trajectory data available from the NGSIM project (FHWA, 2006) as ground- truth data to verify the proposed methodology and examine the sources of estimation error.

4.1. Data descriptions

The NGSIM vehicle trajectory data used in this study come from the I-80 data-set, which were collected by a video camera located at Emeryville, California. This data collection point is located adjacent to I-80, as shown in Figure 9. The site was approximately 1,650 feet in length, with an on-ramp at Powell Street (indicated in Figure 9 by the circle). The freeway segment covered in the data-set includes six lanes, numbered incrementally from the left-most lane (HOV lane). Video data are available for three time intervals: 4:00–4:15 pm, 5:00–5:15 pm, and 5:15–5:30 pm, on 13 April 2005. Complete, transcribed vehicle trajectories are available with a time resolution of 0.1 s.

4.2. Data extraction from NGSIM data-set

(1) To extract vehicle flow counts data from the NGSIM data-set, we first construct a node-link structure to represent the freeway segment in Figure 9. This stretch of freeway is divided into two links, as shown in Figure 10, with the on-ramp connected with node 1.

(2) In order to obtain the flow rate at the node/bottleneck, this study introduces a set of virtual detectors at node 1 and at node 2, respectively. Meanwhile, another virtual detector is placed on the on-ramp link so that inflow vehicles from the ramp are also counted. In addition, video cameras are assumed to be installed on both links to provide link snapshots (for probe vehicle data).

(3) The vehicle trajectory data is divided into 5-min intervals for counting vehicles. An example of one 5-min span of vehicle trajectories on one lane is shown in Figure 11 to illustrate how the vehicle counts are collected. As mentioned before, two sets of virtual detectors A and B are placed at nodes 1 and 2 (shown as triangles in Figure 10), and video cameras C and D are also installed on both link 1 and link 2. Vehicles are counted along the vertical line drawn at the given time t. At time t0 = 0, probe vehicle z = 0 enters link 1. At this time step, two vehicles are observed on link 1 and five vehicles are observed on link 2 by video cameras C and D, that is,

x1(t0) = 2 and x2(t0) = 5. Similarly, probe vehicle z = 5 enters link 1 at time t. At this moment,

x1(t) = 2 and x2(t) = 4.

Probe vehicle z = 8 is worth mentioning, which enters link 1 at time t'. However, at time t", this vehicle changes lanes. This can be seen in Figure 11 because there is an incomplete vehicle trajectory. During this 5-min interval, 12 vehicles are counted by detector A, including two vehicles entering before probe vehicle z = 1, but excluding probe vehicle z = 8. Meanwhile, 13 vehicles are counted by detector B. This count includes seven vehicles before probe vehicle z = 1, but excludes probe vehicles 7–11, which have not yet departed from link 2.

Figure 12 illustrates the relationship between the actual trajectories extracted from the NGSIM data- set and the number of vehicles waiting in the "modeled" vertical queue. For vehicles 1 and 2, we plot the free-flow travel times (dashed lines), the waiting times (arrowed lines) and the experienced travel times (solid lines). Essentially, the experienced travel time = free-flow travel time + waiting time at the vertical queue, where the waiting time is determined by the number of vehicles in the queue and the capacity, as stated by Equation 7.

4.3. Model validation

Two variants on the basic travel time distribution estimation approach presented here are investi- gated in this section. The first is lane-based; the second is link-based.

The examples require some additional notation to represent lane-specific parameters.

Z: number of probe vehicles

n: index identifying a lane

tz: starting time for probe vehicle z

nz: starting lane number for probe vehicle z

xm(t, n): number of vehicles on lane n at time t

z tz , n : lane n specific number of vehicles behind bottleneck m

cm(n): lane n specific discharge rate of bottleneck m

f net(n): net flow rate from or to ramps by lane n

&(n): vehicle distribution rate from on-ramp to lane n

wz (n): waiting time for probe vehicle z on bottleneck m on lane n

pz (n): lane-based route-level travel time for probe vehicle z through lane n

The following procedure is used to calculate the lane-based travel time distribution.

For z = 1 to Z on the link

Obtain arrival time tz and starting lane number nz for each probe vehicle z. Obtain the lane-based number of vehicles xm t0,n ;

Obtain the lane-specific discharge rate cm(n);

Calculate net flow rate f net(n) from the on-ramp by applying &(n)× f net;

4.3.1. Lane-based route-level travel time estimation

The distribution of the estimated route-level travel times for each 5-min interval, calculated over the three time periods with available data (4:00–4:15 pm, 5:00–5:15 pm, and 5:15–5:30 pm) are plotted in Figure 13 with the ground truth route-level travel time obtained directly from the NGSIM data. As can be observed, the distribution of the estimated route-level travel time is very close to that of the ground truth route-level travel time. This demonstrates that our model is able to accurately esti- mate the route-level travel time distribution.

4.3.2. Link-based route-level travel time estimation

One common practice is to use link-based flow rates or density to estimate travel time reliability. We replace the lane-based variables in the previous approach with link-based variables xm(t0) and cm. That is, xm(t0) is the existing number of vehicles on all the lanes on the link and cm is the link dis- charge rate. The distribution of the estimated route-level travel time and true route-level travel time for different time intervals is shown in Figure 14.

As it can be observed, the distribution of the estimated link-based route-level travel times fails to capture the wide-spread distribution in the ground truth travel times. This is explained by the fact that link-based input variables would yield the same estimated route-level travel times for those vehicles entering the link at the same time, regardless of their driving lanes.

In order to understand the extent and sources of the lane-by-lane travel time variation, we use the time period between 4:00 pm and 4:15 pm as an example. Figure 14 shows the lane discharge rate, the existing numbers of vehicles on the lane, and the average true and estimated route-level travel times for each lane for each 5-min interval in the time period. The lane sequence is sorted by the true route-level travel time.

Several observations can be made based on Figure 15. Lane 1 (HOV lane) has the lowest existing number of vehicles on the lane and has the lowest average route-level travel time. The left-most lanes (lanes 1 and 2) usually have the highest discharge rates while lanes 3 and 4 usually have the lowest discharge rates. In most cases, lanes 3 and 4 have the highest average existing numbers of vehicles on the lane, as well as the highest average route-level travel times.

These observations imply that, due to the variation of the discharge rates and the number of vehicles on the lane, the route-level travel times also present strong lane-by-lane variations. As a result, we suggest using lane-based statistics to better quantify the travel time variability.

4.3.3. Estimation error sources

By comparing the estimated results with the NGSIM ground truth data, we can further uncover other possible sources of errors in the proposed travel time estimation model.

(1) Aggregation errors: The link/lane discharge rates cm and on-ramp flow rates used in the calculations are average flow rates over a certain time interval, e.g. 5-min rates, while the existing number of vehicles on the link/lane xm(t0) is an instantaneous value based on the entering time of a probe vehicle.

(2) Measurement errors: The number of vehicles on the lane observed by the video camera at time t0 is assumed to be error-free. In fact, there are always vehicle detection errors in NGSIM vehi- cle trajectory data associated with the underlying video recognition algorithm.

(3) Modeling errors associated with lane changing: Since the queue model incorporates the first- in-first-out principle, lane change behavior is not considered in the calculation. This will introduce two types of errors in the model:

(a) The model may underestimate or overestimate the number of vehicles behind the bottleneck,

z . For example, some vehicles will enter the lane (from the other lanes) before a probe
vehicle reaches the bottleneck, or some vehicles originally counted in xm(t0) on the current lane will leave to one of the adjacent lanes, corresponding to a lower value of xm(t0).

(b) When a probe vehicle changes lane from, for example, nz = 1 to lane n', the discharge rate used in the calculation should be changed to the one associated with lane n'.

5. Conclusions

This paper has presented a model which is capable of estimating route-level travel time distributions. It does this on the basis of relatively simple information about the corridor's geometric configuration, its entering and exiting flows, its capacities, and the lane-by-lane distributions of traffic at each bottleneck. Monte Carlo simulation and mathematical approximation methods are used to calculate travel times lane-by-lane by tracking probe vehicles through the network. Ground-truth vehicle trajectory data from the NGSIM project have been used to validate the model's estimates. Several interesting observations are obtained from the research:

(1) The model offers a theoretically sound method to estimate corridor-level travel time and its distribution under different capacity and demand variations, and with possible off-ramp and off-ramp volume changes.

(2) The variation of lane-specific traffic flow parameters (such as the number of vehicles on the lanes and lane discharge rates) are significant to lane-by-lane travel time diversity.

(3) A lane-level rather than link-level representation of the system is critical in developing accurate route-level travel time distribution estimates.

Our future research will consider the impacts of downstream queue spillback on the upstream travel time. Under queue spillback, the discharge rate of the upstream bottleneck is constrained by the downstream bottleneck discharge rate, and this significantly increases dynamics and complexity in estimating the capacity for a queuing model. Moreover, it is also desirable to examine the influence of lane change frequencies on the estimated number of vehicles waiting in the queue.

Acknowledgment

This paper is based on research work partially supported through the TRB SHRP 2 L02 project titled "Establishing Monitoring Programs for Mobility and Travel Time Reliability." The work presented in this paper remains the sole responsibility of the authors.

Funding

The authors received no direct funding for this research.

REFERENCES

Alibabai, H. (2011). Properties of simulated path travel times (Doctoral dissertation). Northwestern University, Evanston, IL.

Chen, C., Skabardonis, A., & Varaiya, P. (2003). Travel-time reliability as a measure of service. Transportation Research Record, 1855, 74–79. http://dx.doi.org/10.3141/1855-09

Clark, S., & Watling, D. (2005). Modelling network travel time

reliability under stochastic demand. Transportation Research Part B: Methodological, 39, 119–140. http://dx.doi.org/10.1016/j.trb.2003.10.006

Cowan, R. J. (1975). Useful headway models. Transportation Research, 9, 371–375.

http://dx.doi.org/10.1016/0041-1647(75)90008-8

Daganzo, C. F. (1995). Properties of link travel time functions under dynamic loads. Transportation Research Part B: Methodological, 29, 95–98.

Federal Highway Administration. (2006, December). Next generation simulation fact sheet (Technical report, FHWA-HRT-06-135). Washington, DC: Author.

Federal Highway Administration. (2009, June). Recurring trafic bottlenecks: A primer focus on low-cost operational improvements. (Technical report, FHWA-HOP-09-037).

Washington, DC: Author.

Friesz, T. L., Bernstein, D., Smith, T. E., Tobin, R. L., & Wie, B. W. (1993). A variational inequality formulation of the dynamic network user equilibrium problem. Operations Research, 41, 179–191. http://dx.doi.org/10.1287/opre.41.1.179

Giuliano, G. (1989). Incident characteristics, frequency, and duration on a high volume urban freeway. Transportation Research Part A: General, 23, 387–396. http://dx.doi.org/10.1016/0191-2607(89)90086-1

Golob, T. F., Recker, W. W., & Leonard, J. D. (1987). An analysis of the severity and incident duration of truck-involved freeway accidents. Accident Analysis and Prevention, 19, 375–395.

http://dx.doi.org/10.1016/0001-4575(87)90023-6 Greenberg, I. (1966). The log-normal distribution of headways.

Australian Road Research, 2, 14–18.

Guo, F., Rakha, H., & Park, S. (2010). Multistate model for travel time reliability. Transportation Research Record: Journal of the Transportation Research Board, 2188, 46–54. http://dx.doi.org/10.3141/2188-06

Hazelton, M. L. (2001). Inference for origin–destination matrices: Estimation, prediction and reconstruction. Transportation Research Part B: Methodological, 35, 667–676.

http://dx.doi.org/10.1016/S0191-2615(00)00009-6 HCM. (2000). Highway capacity manual. Washington, DC:

Transportation Research Board, National Research Council.

Hurdle, V. F., & Son, B. (2001). Shock wave and cumulative arrival and departure models: Partners without conflict. Transportation Research Record, 1776(1), 159–166.

http://dx.doi.org/10.3141/1776-21

Kripalani, A., & Scherer, W. (2007, July). Estimating incident related congestion on freeways based on incident severity (Research Report No. UVACTS-15-0-113). Charlottesville, VA: Center for ITS Implementation Research, University of Virginia.

Kwon, J., Barkley, T., Hranac, R., Petty, K., & Compin, N. (2011). Decomposition of travel time reliability into various sources. Transportation Research Record: Journal of the Transportation Research Board, 2229, 28–33. http://dx.doi.org/10.3141/2229-04

Lam, W. H. K., Shao, H., & Sumalee, A. (2008). Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. Transportation Research Part B: Methodological, 42, 890–910. http://dx.doi.org/10.1016/j.trb.2008.02.004

Lo, H. K., & Tung, Y. K. (2003). Network with degradable links: Capacity analysis and design. Transportation Research Part B: Methodological, 37, 345–363. http://dx.doi.org/10.1016/S0191-2615(02)00017-6

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks. Transportation Research Part B: Methodological, 27, 289–303.

http://dx.doi.org/10.1016/0191-2615(93)90039-D

Ng, M. W., & Waller, S. T. (2010). A computationally efficient methodology to characterize travel time reliability using the fast Fourier transform. Transportation Research Part B: Methodological, 44, 1202–1219. http://dx.doi.org/10.1016/j.trb.2010.02.008

Oh, J.-S., & Chung, Y. (2006). Calculation of travel time variability from loop detector data. Transportation Research Record, 1945, 12–23. http://dx.doi.org/10.3141/1945-02

Ran, B., Boyce, D. E., & LeBlanc, L. J. (1993). A new class of instantaneous dynamic user-optimal traffic assignment models. Operations Research, 41, 192–202. http://dx.doi.org/10.1287/opre.41.1.192

Park, S., Rakha, H., & Guo, F. (2011, October). Multi-state travel time reliability model: Impact of incidents on travel time reliability. In Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on (pp. 2106–2111). Washington,

DC: IEEE.

Richardson, A. J. (2003, December). Travel time variability on an urban freeway (TUTI Report 25-2003). Presented at the 25th Conference of Australian Institutes of Transport Research (CAITR), University of Adelaide, Adelaide, Australia.

Unnikrishnan, A., Ukkusuri, S., & Waller, S. T. (2005). Sampling methods for evaluating the traffic equilibrium problem under uncertain demand. In Transportation Research Board 84th Annual Meeting Compendium of Papers.

Washington, DC.

Waller, S. T., & Ziliaskopoulos, A. K. (2001). Stochastic dynamic network design problem. Transportation Research Record, 1771, 106–113.

http://dx.doi.org/10.3141/1771-14

Zhang, H. M., & Nie, X. (2005). Some consistency conditions for dynamic traffic assignment problems. Networks and Spatial Economics, 5, 71–87. http://dx.doi.org/10.1007/s11067-005-6662-7

Zhou, X., Rouphail, N. M., & Li, M. (2011). Analytical models for quantifying travel time variability based on stochastic capacity and demand distributions. Presented at Transportation Research Board 90th Annual Meeting, Washington, DC.