Technical Report 130

Using Collected Data to Improve Dynamic Traffic Assignment Modeling: Volumes 1 and 2

Research Supervisor
Natalia Ruiz Juri

Project Title: Using Collected Data to Improve Dynamic Traffic Assignment Modeling

September 2018

Data-Supported Transportation Operations & Planning Center (D-STOP)

A Tier 1 USDOT University Transportation Center at The University of Texas at Austin

D-STOP is a collaborative initiative by researchers at the Center for Transportation Research and the Wireless Networking and Communications Group at The University of Texas at Austin.

Technical Report Documentation Page

1. Report No. D-STOP/2018/130	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Using Collected Data to Improve Dynamic Traffic Assignment Modeling: Volumes 1 and 2		5. Report Date September 2018 6. Performing Organization Code
7. Author(s) Natalia Ruiz Juri, Stephen D. Boyles, Tengkuo Zhu, Kenneth Perrine, Amber Chen, Yun Li		8. Performing Organization Report No. Report 130
9. Performing Organization Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 3925 W. Braker Lane, 4th Floor Austin, Texas 78759		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTRT13-G-UTC58
12. Sponsoring Agency Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 3925 W. Braker Lane, 4th Floor Austin, Texas 78759		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes
Supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.
Project Title: Using Collected Data to Improve Dynamic Traffic Assignment Modeling

16. Abstract

Volume 1: In the field, queue spillback contributes substantially to urban congestion. Therefore, most dynamic network loading models either explicitly include spillback, or frame a failure to model spillback as an unfortunate consequence of mathematical or computational tractability that should be relaxed in future work. While models with spillback are undeniably more realistic, they are also less robust to errors in input demand. We show that when there is high uncertainty in input demand, excluding spillback can actually reduce error by reducing sensitivity to demand errors. Our demonstrations include a small network that can be solved analytically, and dynamic user equilibrium on two real-world networks (Austin and San Antonio). We conclude that model realism must be carefully balanced against accuracy of the input parameters, and the sensitivity of the model to any such errors in these inputs.

Volume 2: Realistic estimations of queue propagation and traffic delays due to freeway lane closures are critical for planning and managing work zones. The current practical methods include deterministic queuing theory and microsimulation. However, these approaches are either too simple to capture the complex traffic entry/exit patterns or require significant modeling effort. This paper proposes a framework to use kinematic wave theory to estimate queue position and the user's traffic delay. Link transmission model (LTM) is implemented in the analysis of work zone impacts due to its efficiency and is validated in different scenarios. A comparison of LTM method with microsimulation suggests that LTM method produces comparable queue propagation and dissipation patterns, with runtimes of only one second. LTM method estimations were also consistent with field data collected on IH35 through Austin, TX, suggesting that model results are realistic, and could support decision-making given adequate input data and parameter calibration. The method presented can be used by agencies to support planning and operational decisions, such as the placement of variable message signs, for events that happen frequently and cannot be modeled in extensive detail. It can also facilitate the computation of meaningful performance metrics to communicate with stakeholders for strategic planning, and to conduct cost-benefit analyses, among others

17. Key Words Dynamic traffic assignment, network loading, demand uncertainty, sensitivity analysis, queue spillback; freeway lane closure, link transmission model, queue length, user delay cost	18. Distribution Statement No restrictions. This document is available to the public through NTIS (http://www.ntis.gov): National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report) Unclassified	20. Security Classif.(of this page) Unclassified	21. No. of Pages	22. Price

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Acknowledgements

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers.

Volume 1: Queue Spillback and Demand Uncertainty in Dynamic Network Loading

QUEUE SPILLBACK AND DEMAND UNCERTAINTY IN DYNAMIC NETWORK LOADING

Stephen D. Boyles
Associate Professor
Civil, Architectural, and Environmental Engineering
The University of Texas at Austin

Natalia Ruiz Juri
Director
Network Modeling Center
Center for Transportation Research
The University of Texas at Austin

Word Count: 6475 words + 8 figure(s) x 0 + 4 table(s) x 250 = 7475 words

Submission Date: August 1, 2018

ABSTRACT

In the field, queue spillback contributes substantially to urban congestion. Therefore, most dynamic network loading models either explicitly include spillback, or frame a failure to model spillback as an unfortunate consequence of mathematical or computational tractability that should be relaxed in future work. While models with spillback are undeniably more realistic, they are also less robust to errors in input demand. We show that when there is high uncertainty in input demand, excluding spillback can actually reduce error by reducing sensitivity to demand errors. Our demonstrations include a small network that can be solved analytically, and dynamic user equilibrium on two real-world networks representing Austin, TX, and San Antonio, TX. We conclude that model realism must be carefully balanced against accuracy of the input parameters, and the sensitivity of the model to any such errors in these inputs.

Keywords: Dynamic traffic assignment, network loading, demand uncertainty, sensitivity analysis, queue spillback

INTRODUCTION

The natural course of research is to move from simpler, stylized models to more realistic ones by relaxing assumptions. Obviously, more realistic models are *ceteris paribus* preferable to less realistic ones. Realism, however, is not the only criterion for model selection, and realism must be balanced against tractability, availability of data for calibration and validation, mathematical regularity, and robustness to errors in the input data. Practitioners must make modeling choices taking all these factors into account, given their relevance in a specific application.

This paper specifically examines the tradeoffs between model realism and robustness to input errors, in the context of dynamic network loading. The point queue model is perhaps the simplest link model, but neglects queue spillback, which is responsible for considerable delay in congested networks. More sophisticated link models do represent queue spillback, which improves their realism and accuracy — however, this feature also increases sensitivity of the model to the input parameters. Queue spillback can introduce discontinuities into flow models, which suggests that even small errors in the input data could potentially propagate into much larger errors in the output.

In practice, input data are never known with exact precision due to measurement, estimation, and forecasting errors. Therefore, it is not obvious that the more "realistic" model is preferable to one which is more robust to these errors. Rather, one or the other model may be preferable depending on (1) the level of uncertainty in the input data, (2) the relative sensitivities of the two models, (3) and the magnitude of the error introduced by ignoring spillback.

We explore this idea in two settings. The first is a simple freeway interchange, where dynamic network loading can be performed in closed-form. The second is regional dynamic traffic assignment, using real-world networks representing the cities of San Antonio, TX, and Austin, TX. In addition to being larger, the latter scenarios allow for route choice, and the resulting equilibrium may partially counteract errors in demand. In both of these settings, we vary the uncertainty in the demand, compare dynamic network loadings with spillback and no-spillback models, and identify the relative performance of the two models. In particular, if there is high uncertainty in input demand, we find cases where the no-spillback model produces a more accurate answer than the model with spillback.

This investigation contributes to the transportation science literature by highlighting the importance of model robustness to input error, and showing that a more realistic model need not produce more accurate answers. Practitioners should thus make modeling choices based not only on arguments of realism and tractability, but also in awareness of input data quality and model sensitivity. If modeling spillback is critical for a particular application, our findings suggest that input data should be known with high confidence.

The remainder of the paper is structured as follows. We first review literature related to transportation modeling, focusing on dynamic network loading, and on the effects of parameter uncertainty. We then describe the general framework for our tests, and then present results on a single interchange, and on two large networks. The paper concludes by discussing the implications of these findings for practice.

LITERATURE REVIEW

Queue spillback has long been recognized as a major source of congestion, and one that must be managed in a different way from queues that are confined to a single link. For instance, signalized intersections should be controlled differently on a corridor where queues spill back (*1–4*). Ramp

meters must balance restricting freeway inflows with arterial disruptions if the queue exceeds the ramp length, and these disruptions are a common source of complaints (*5–7*). Daganzo (*8*) showed that, paradoxically, increasing the capacity on a bottleneck may actually *decrease* network capacity in the presence of queue spillback. In more severe cases, queue spillback can cause gridlock (*9*), as when a cycle of network links becomes blocked and no vehicles can move.

Therefore, queue spillback is often treated as a *sine qua non* in contemporary dynamic traffic assignment. Models without spillback, such as point queues, are generally used only because of their favorable mathematical properties — some recent examples are Ban et al. (*10*), Friesz et al. (*11*), and Han et al. (*12*) — and the same researchers often follow such work with more sophisticated models where queues can spill back (*13–15*), as a natural progression to their earlier work. Dynamic network loading procedures intended for large-scale simulation, including the cell transmission model (*16, 17*), link transmission model (*18*), and double-queue model (*19*), all allow queues to spill back.

A separate line of research concerns the challenges in obtaining accurate inputs for transportation models, particularly predictions in future years. Demand forecasts are often substantially different than the demand realized in the field; the research of Flyvbjerg et al. (*20*) found that roughly 90% of rail projects forecasted demand in excess of what was actually realized, by an average of 106%. Errors for road projects are smaller, but still significant, with particular implications for toll revenue forecasting. The same authors also noted that forecast accuracy has not improved significantly over the last 30 years, despite substantial advances in travel demand modeling. Other authors have confirmed these findings (*21–23*), but also note that there is no consensus as to how forecasting error should be addressed in practice.

To study the effects of uncertainty, researchers have applied Monte Carlo simulations (*24–27*) or sensitivity analyses (*28*) to model parameters. For instance, Zhao and Kockelman (*29*) tracked how errors propagate through the traditional four-step travel demand model, finding that input errors were amplified through the trip generation, trip distribution, and mode choice steps; but then reduced through route choice (likely because the equilibrium principle acts as a compensating mechanism for demand errors), resulting in output errors of the same relative size as the input errors.

Today, researchers know that uncertain demand has significant implications for project selection (*30, 31*), network design (*32–35*), evacuation planning (*36*), roadway pricing (*37–41*), and other applications of transportation models. However, less attention has been given to the effect of demand uncertainty *on the choice of models themselves*. Two exceptions are Bar-Gera (*42*), who lists "ability to obtain inputs" and "stability" among a list of modeling desiderata, and Daganzo (*43*). In the logistics context, the latter paper shows that replacing continuous costs or demand with step functions (to make a model more "realistic," as orders and truckloads are necessarily discrete) may in fact increase error if the inputs are only known approximately, a result that deserves more attention than it has received. Our manuscript is in the same spirit as Daganzo (*43*), but in the setting of dynamic network loading and spillback, as we investigate the interaction between demand errors and model realism.

DYNAMIC NETWORK LOADING AND SPILLBACK

Dynamic network loading is the process of mapping driver behavior (e.g., route and departure time choice) to observed network traffic conditions (e.g., congestion and delay). Dynamic network loading thus plays a central role in dynamic traffic assignment. Many dynamic network loading

procedures are modular, and can be divided into *link models* and *node models*. A link model represents traffic flow on a single network link, treating the upstream and downstream nodes as boundary conditions or constraints. A node model serves as the "glue" connecting link models, representing priority and vehicle interactions at physical intersections.

In the discrete-time version of this framework, each link reports its *sending flow* $S(t)$ and *receiving flow* $R(t)$ at each timestep.[1] The sending flow $S(t)$ is the number of vehicles which would exit the link during the t-th timestep, if there were no constraints from downstream. The receiving flow $R(t)$ is the number of vehicles which would enter the link during the t-th timestep, if there were an infinite number of vehicles waiting to enter from upstream. The values $S(t)$ and $R(t)$ thus are upper bounds on the number of vehicles that actually leave and enter the link during the t-th step. These values are calculated only from the conditions on the link, without referring to other links or nodes, and can thus be computed in parallel.

A simple link model is the *point queue* (PQ). Consider a link with a uniform capacity q_{max} and an integer free-flow travel time t_0 (both measured with a time unit of one timestep). Denote by $N^{\uparrow}(t)$ and $N^{\downarrow}(t)$ the cumulative number of vehicles which have entered and left the link by the start of the t-th timestep, since the beginning of loading. The PQ sending flow is given by

$$S(t) = \min\left\{N^{\uparrow}(t+1-t_0) - N^{\downarrow}(t), q_{max}\right\}, \tag{1}$$

and the PQ receiving flow by

$$R(t) = q_{max}. \tag{2}$$

This model is called a "point" queue because $R(t)$ does not depend on the link's condition, and vehicles are never blocked from entering the link — it is as if the vehicles on the link occupy no physical space, so the queue will never spill back to upstream links.

The *spatial queue* (SQ) model modifies the point queue model to represent the physical space used by queued vehicles. (*44*) If the link can hold at most $\bar{N}$ vehicles at one instant, the SQ receiving flow is given by

$$R(t) = \min\left\{\bar{N} - (N^{\uparrow}(t) - N^{\downarrow}(t)), q_{max}\right\}. \tag{3}$$

The SQ sending flow is identical to the PQ sending flow, and is given by equation (1). Equation (3) ensures that the number of vehicles on the link will never exceed $\bar{N}$ at any time.

More sophisticated link models exist, such as the cell transmission model (*16, 17*), link transmission model (*18*), and double-queue model (*19*). Compared to the SQ, these models further reflect the time needed for shockwaves to propagate on a link.

Node models capture the interactions which occur between links at a junction, and calculate the *transition flows* $y_{ab}(t)$ representing the actual number of vehicles which leave link a and enter link b during the t-th time step. The simplest node has a single incoming link i, and a single outgoing link o. A typical model for such a node is

$$y_{io}(t) = \min\{S_i(t), R_o(t)\} \tag{4}$$

The number of vehicles which transition between link i and o is thus the smaller of the sending flow from the upstream link and the receiving flow for the downstream link.

These y values are then used to update the cumulative entries and exits to links:

$$N_i^{\uparrow}(t+1) = N_i^{\uparrow}(t) + \sum y_{\cdot i}(t) \tag{5}$$

$$N_i^{\downarrow}(t+1) = N_i^{\downarrow}(t) + \sum y_{i\cdot}(t) \tag{6}$$

[1] Some authors use the terms *demand* and *supply* for the same concepts.

TABLE 1 Errors from Model Choice and Data Quality

	Right model	Wrong model
Right data	Accurate	Modeling errors
Wrong data	Data errors	Modeling AND data errors

where the sums are over all upstream and downstream links, respectively. Node models must respect the sending and receiving flows for links, so that $\sum y_{\cdot i}(t) \leq R_i(t)$ and $\sum y_{i\cdot}(t) \leq S_i(t)$.

These latter conditions permit modeling of the spillback phenomenon: if $\bar{N}$ vehicles are on an SQ link, no additional vehicles can enter, and no vehicles can leave the upstream links, spreading the queue to previous links. If this process continues, the queue can spill over multiple links, and if the queued links form a cycle, no vehicles can move and gridlock results.

At a diverge node with one incoming link i and two outgoing links 1 and 2, let $p_{io}(t)$ be the fraction of the sending flow $S_i(t)$ desiring to exit the node on link $o \in \{1,2\}$. Under the standard assumption that vehicles exit a link in the same order they enter, we have

$$y_{io}(t) = \phi p_{io}(t) S_i(t) \qquad \text{for } o \in \{1,2\} \tag{7}$$

with

$$\phi = \min\left\{ \frac{R_1}{p_{i1}S_i}, \frac{R_2}{p_{i2}S_i}, 1 \right\}. \tag{8}$$

At a merge node with two incoming links 1 and 2, and one outgoing link o, each incoming link is given a nonnegative priority value α_i, with $\alpha_1 + \alpha_2 = 1$. There are three possible cases for the merge; in the first case, the merge is uncongested ($S_1(t) + S_2(t) \leq R(t)$) and flows freely:

$$y_{io}(t) = S_i(t) \qquad \text{for } i \in \{1,2\}. \tag{9}$$

In the second case, the merge is congested ($S_1(t) + S_2(t) > R(t)$) and there is a queue on both links (that is, $S_i(t) > \alpha_i R_o(t)$ for $i \in \{1,2\}$):

$$y_{io}(t) = \alpha_i R_o(t) \qquad \text{for } i \in \{1,2\}. \tag{10}$$

In the last case, the merge is congested, but a queue only exists on one of the links. Without loss of generality, assume the queue is on link 1; then $S_2(t) \leq \alpha_2 R_o(t)$ and we have

$$y_{1o}(t) = R_o(t) - S_2(t) \qquad y_{2o}(t) = S_2(t). \tag{11}$$

More discussion on these node models, as well as alternatives and generalizations for nodes with multiple incoming and outgoing links, can be found in the literature. (*17, 18, 45–47*)

MOTIVATION AND EXPERIMENTAL PHILOSOPHY

The purpose of our experiments is to compare a model with spillback to one without spillback when model parameters are uncertain. Queue spillback and gridlock clearly exist in practice, and have significant impacts on congestion. These are strong arguments for the spillback model, which is clearly more realistic.

We thus assume that the spillback model is the right choice, if the input data are known precisely. With this assumption, using the no-spillback model will introduce modeling errors, even with the correct input data. If the input data are not fully known, then even using the spillback model will introduce errors due to data uncertainty. Using the no-spillback model would seem to add insult to injury, introducing modeling error on top of data errors. (Table 1 illustrates these possibilities schematically.)

However, if the "wrong" model is more robust to errors in the input data, it may in fact

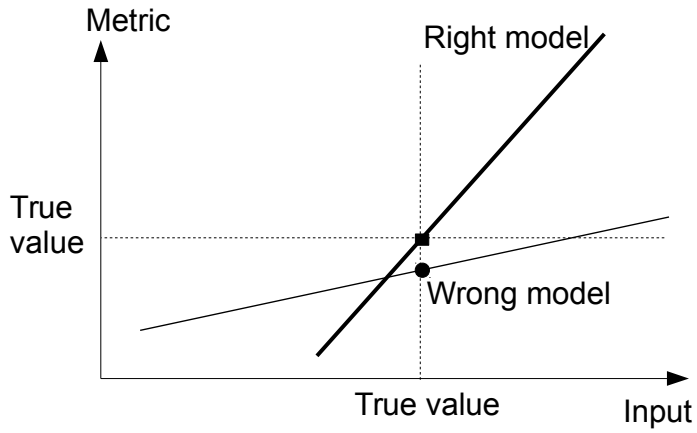


FIGURE 1 A stylized example illustrating right and wrong models.

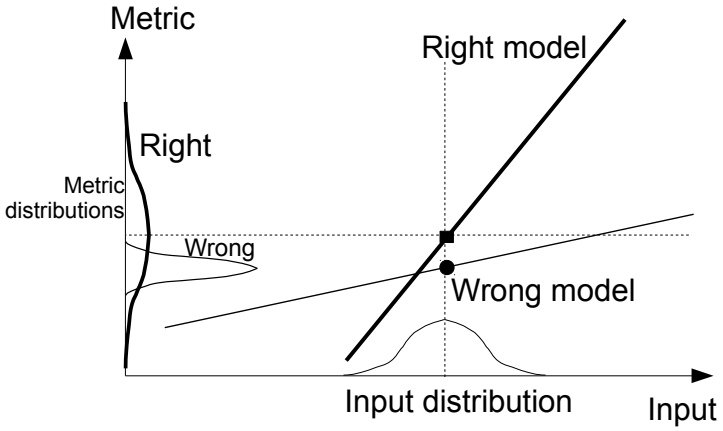


FIGURE 2 Mapping from a distribution on the input to distributions on the metric.

yield a more accurate answer. Figure 1 shows a stylized scenario with a single input parameter, and a single metric of interest. The dashed lines indicate accurate (ground truth) values, while the solid lines indicate the predictions of two models. The right model passes through the intersection of the dashed lines, as it should: given the correct input, the right model gives the correct metric. The wrong model, by contrast, does not produce the correct metric even with the correct input, and should not be used.

This conclusion holds true even if the input data are not known precisely, as long as the error is fairly small; the right model will still produce a metric closer to the horizontal line (true value) than the wrong model. If input errors are large, this need not be true. Figure 2 shows a distribution of errors on the inputs, and the resulting impact on the distribution of output error, on the metric. The metric error distribution for the right model is centered around the correct value — but the greater sensitivity of the right model yields a large spread. The metric error distribution for the wrong model is not centered around the correct value, but it is more tightly distributed, and the expected *absolute* deviation from the horizontal line is smaller for the "wrong" model. The wrong model actually produces the smaller error!

We hypothesize that similar effects exist in dynamic network loading, and manifest when

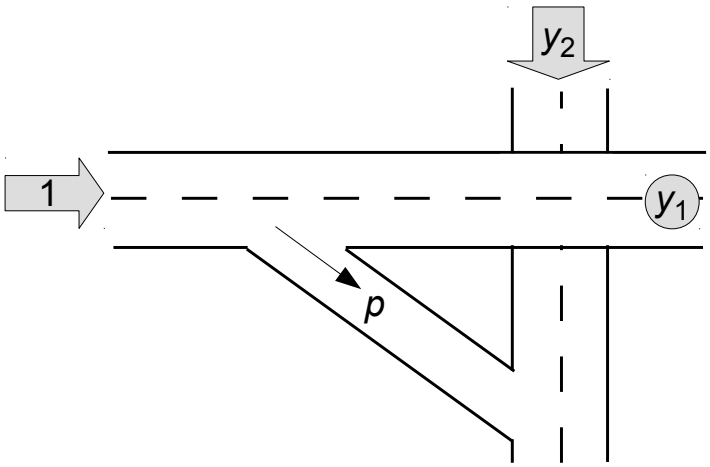


FIGURE 3 Freeway interchange for first experiment; mainlines have capacity 1 and ramp has capacity 1/2. y_2 is a model *parameter* that must be input; y_1 is a metric that must be calculated.)

the error in the input parameters is sufficiently large. The following sections describe our experiments to investigate this hypothesis, by comparing the spillback model (the "right" one) to the no-spillback model (the "wrong" one). We introduce errors into the demand rates (and in one of the experiments, also in the route choice), and compare metrics such as link flows and travel times, using the spillback model results with zero error as the assumed truth.

EXPERIMENT 1: A SINGLE INTERCHANGE

Consider the freeway interchange shown in Figure 3. The two mainlines have equal capacity, and the ramp connecting them has half the capacity of the mainlines. For simplicity, assume that units are chosen such that the capacity of each mainline is equal to one, and that the inflow rate on the horizontal mainline is also one. There are two input parameters which must be estimated: the proportion p of flow on the horizontal link choosing the ramp, and the flow y_2 on the vertical mainline link. In this experiment both are assumed stationary in time. We wish to predict the steady-state flow rate y_1 on the horizontal link downstream of the diverge, located at the circular detector in Figure 3, after any initial transient conditions have subsided.

We use the standard merge and diverge equations in dynamic network loading (described in equations (7)–(11)), giving the ramp a merge priority value of $\alpha_r = 1/3$ and the southbound mainline a priority value of $\alpha_2 = 2/3$. Since the inflow to the ramp is initially p, if $p + y_2 \leq 1$ the merge is uncongested and all flow can move freely. If $p + y_2 > 1$, three cases are possible: if $p < 1/3$, a queue forms on the vertical link but not the ramp; if $y_2 < 2/3$, a queue forms on the ramp but not the vertical link; and if $p \geq 1/3$ and $y_2 \geq 2/3$, queues form on both approaches. Without spillback, the presence of queues at the merge has no bearing on behavior at the diverge, and the total outflow from the diverge is given by $\min\{1, 1/(2p)\}$, with the detector registering

$$y_1^{NS} = (1-p)\min\{1, 1/(2p)\}. \tag{12}$$

In the case of spillback, equation (12) only holds if there is no queue on the ramp. Otherwise, at steady-state, the ramp outflow (and thus its inflow) is given by $1 - y_2$ if the queue is only on the

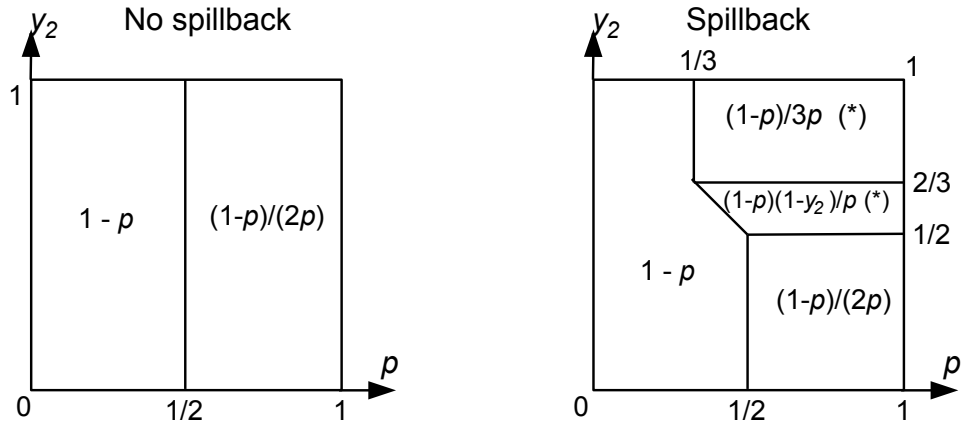


FIGURE 4 Steady-state y_1 values for no-spillback and spillback cases. Asterisks indicate regions where the ramp queue spills back and restricts flow.

ramp, and by $1/3$ if queues exist on both the ramp and vertical link. The corresponding flow at the detector is obtained by multiplying the ramp flow by $(1-p)/p$. These results are summarized in Figure 4, and are denoted by the mappings $y_1^{NS}(y_2,p)$ and $y_1^{S}(y_2,p)$ for the no-spillback and spillback cases, respectively. Full derivations of these results are routine and omitted here for reasons of space.

The possible values of inputs y_2 and p lie within the unit square $[0,1]^2$. Within this range, twenty evenly-spaced values of y_2 and p were combined to produce four hundred scenarios for analysis. For each of these scenarios, the following procedure was performed.

1. Let $\hat{y}_2$ and $\hat{p}$ denote the values corresponding to this scenario. These are assumed to be the true values of these parameters, corresponding to a true flow rate of $y_1^S(\hat{y}_2,\hat{p})$. (That is, the spillback model is presumed completely accurate if given the true y_2 and p values.)
2. Generate n sampled values of y_2 and p, using independent normal distributions with respective means $\hat{y}_2$ and $\hat{p}$, and a provided standard deviation. These samples were truncated to the range $[0,1]$ to respect feasibility.
3. For each sample, the error associated with the no-spillback model is calculated as
$$\varepsilon^{NS} = |y_1^{NS}(y_2,p) - y_1^S(\hat{y}_2,\hat{p})|, \tag{13}$$
while the error associated with the spillback model is
$$\varepsilon^{S} = |y_1^{S}(y_2,p) - y_1^S(\hat{y}_2,\hat{p})|. \tag{14}$$
4. Based on the sampled values of ε^{NS} and ε^S, calculate the additional expected error in the no-spillback model to be
$$\delta = E[\varepsilon^{NS} - \varepsilon^S], \tag{15}$$
along with the standard deviation s of this difference.
5. Calculate the t score $t = \delta/(s/\sqrt{n})$.

If the resulting t score is greater than a specified positive critical value, the error in the no-spillback model is significantly greater than that in the spillback model, and the spillback model is to be preferred. If it is smaller than a negative critical value, the error in the no-spillback model is significantly less than that of the spillback model, and the no-spillback model is preferred. Otherwise, there is no significant difference in the errors produced by the models.

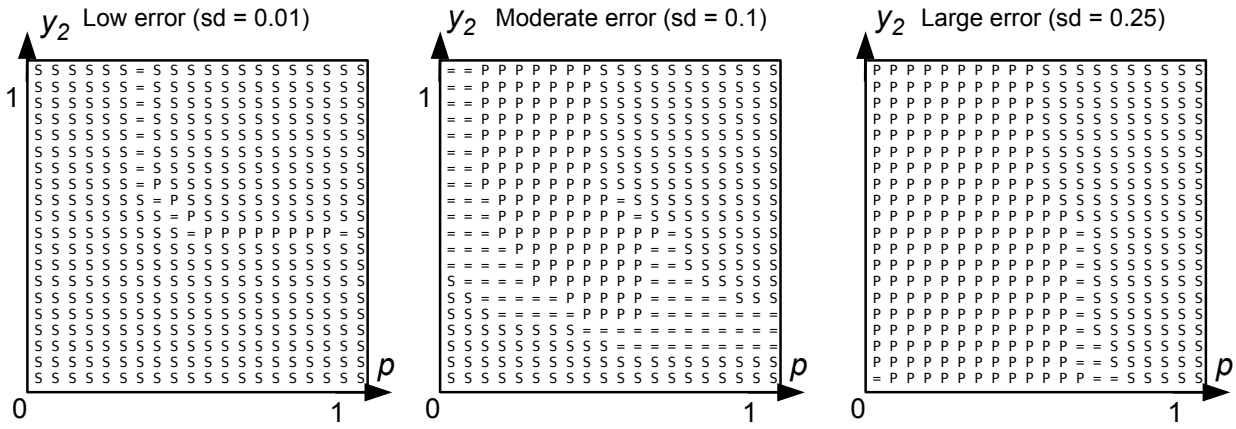


FIGURE 5 Model with less error for each scenario, if significant ($p = 0.05$).

Figure 5 presents the results of these simulations for three cases: when the standard deviation of the sampled Q_2 and p values was small (0.01), moderate (0.1), and large (0.25). A sample size of $n = 2500$ was used for each scenario, and critical values of ± 1.96 were used for the statistical test, corresponding to 5% significance. In this figure, an S denotes that the model with spillback produces less expected error, N denotes that the no-spillback model produces less expected error, and '=' denotes no statistically significant difference in errors. Owing to the sample size, the t scores were typically quite large, averaging $+87.4$ across all scenarios where the spillback model was preferred, and -13.9 across all scenarios where the no-spillback model was preferred.

When input error is small, the model with spillback is almost always preferred, while the no-spillback model often produces less expected error when input errors are large, specifically when the diverge proportion is small. This demonstrates that the no-spillback model may be preferable to the model with spillback under certain circumstances: even though it contains model error, it is more robust to data error.

EXPERIMENT 2: LARGE NETWORKS

The second experiment involves regional dynamic traffic assignment on two calibrated, real-world networks representing portions of San Antonio, TX, and Austin, TX. Table 2 gives descriptive statistics for these networks, and Figure 6 shows their topology. The San Antonio network is centered around US-281 and represents a suburban area, with larger spacing between signalized intersections and higher speed limits. The Austin network represents an urban area, with closely-spaced signals and speed limits generally below 40 mph, with a north-south freeway on the western boundary.

Base data for these networks was provided by the Capital Area Metropolitan Planning Organization and San Antonio Metropolitan Planning Organizations. Information on signal timings was provided by the City of Austin and the City of San Antonio. Both sets of data were then manually refined and adjusted to improve consistency with field observations.

The VISTA software (*48*) was used to solve for dynamic user equilibrium on each network. This software uses the cell transmission model (CTM) as its link model. The primary difference between CTM and the SQ model is that backward-moving shockwaves propagate with finite speed, so link space freed when vehicles depart the downstream end of the link is not immediately avail-

TABLE 2 Descriptive Statistics for Experiment 2 Networks

	San Antonio	Austin
Total demand	341,710	236,940
Time period	4 PM–7 PM	6 AM–9 AM
Links	1258	1247
Nodes	864	587
Centroids	122	166
Signalized intersections	100	137
Average capacity (vph)	1505	1492

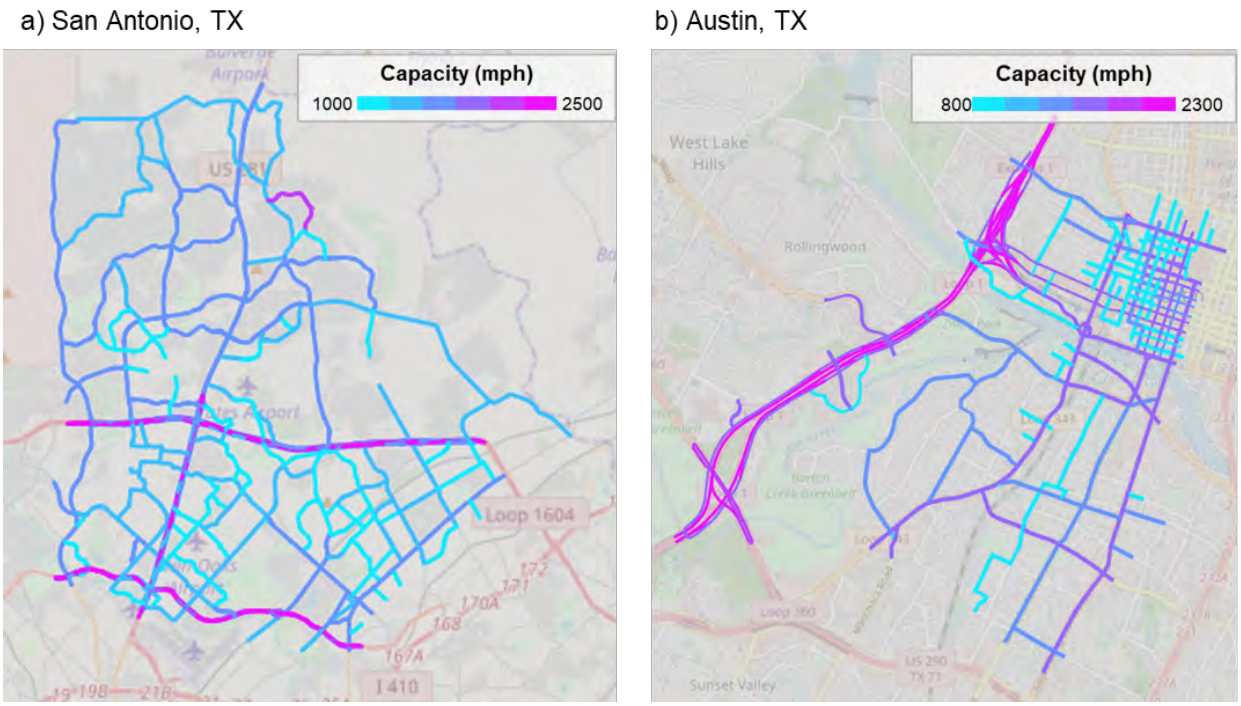


FIGURE 6 Schematic of San Antonio and Austin networks.

TABLE 3 Metrics from Demand Error Scenarios.

	TSTT (veh-hr $\times 10^3$)		Average OD travel time (min)	
Demand error	Spillback	No-spillback	Spillback	No-spillback
San Antonio, TX				
-30%	19	19	5	5
-15%	25	23	6	5
0%	***42***	33	***8***	6
$+15\%$	58	47	10	8
$+30\%$	90	69	14	10
Austin, TX				
-30%	9	6	3	3
-15%	9	8	3	3
0%	***13***	10	***3***	3
$+15\%$	18	13	4	3
$+30\%$	30	20	6	4

able for entering vehicles at the upstream end. CTM models spillback. To create a no-spillback version of CTM, the jam density on links was set to infinity, allowing an unlimited number of vehicles to fit on a link. It is possible to show that the CTM sending and receiving flow coincide with the PQ formulas (1) and (2) with this modification.

For the experiments in this paper, VISTA was applied with a timestep of 6 seconds. The method of successive averages was used to solve for an approximate dynamic user equilibrium. In the dynamic user equilibrium state, for each origin, destination, departure time, all used paths have equal and minimal travel time. It is possible that the equilibrium principle can act as a "restoring force" compensating for errors in the input demand, for instance, by routing vehicles to less-congested routes if demand is too high. We thus want to see if the phenomena seen in Experiment 1 persist in practical networks with equilibrium.

To simulate demand errors, the origin-destination (OD) matrices for the two networks were first inflated or deflated by a uniform value across all scenarios. This value ranged from -30% (underprediction of demand) to $+30\%$ (overprediction). Table 3 shows the five demand scenarios, and the total system travel time (TSTT) and average OD travel times obtained from VISTA. The boldface values indicate the assumed true values: metrics obtained from the spillback model with no error in demand. Figure 7 shows the results for TSTT in Austin, in the manner of the stylized plots in Figures 1 and 2; plots for San Antonio and average OD travel times are similar and omitted for brevity.

Using the results from Table 3, we can now test the performance of the spillback and no-spillback versions of CTM for different levels of uncertainty in demand. We represent demand uncertainty by creating a probability distribution on the demand error, parameterized by a scalar e. When $e = 0$, the forecast is perfect and the error is 0% with probability 1. When $e = 1$, the forecast is uniform among all five possible demand errors, as if the errors in the modeling process combined to produce a random value within 30% of the truth. (Figure 8) For $e \in (0,1)$, the distribution is obtained by a linear interpolation between these two extremes.

For values of e ranging from 0 to 1, we compute the expected absolute TSTT error of the

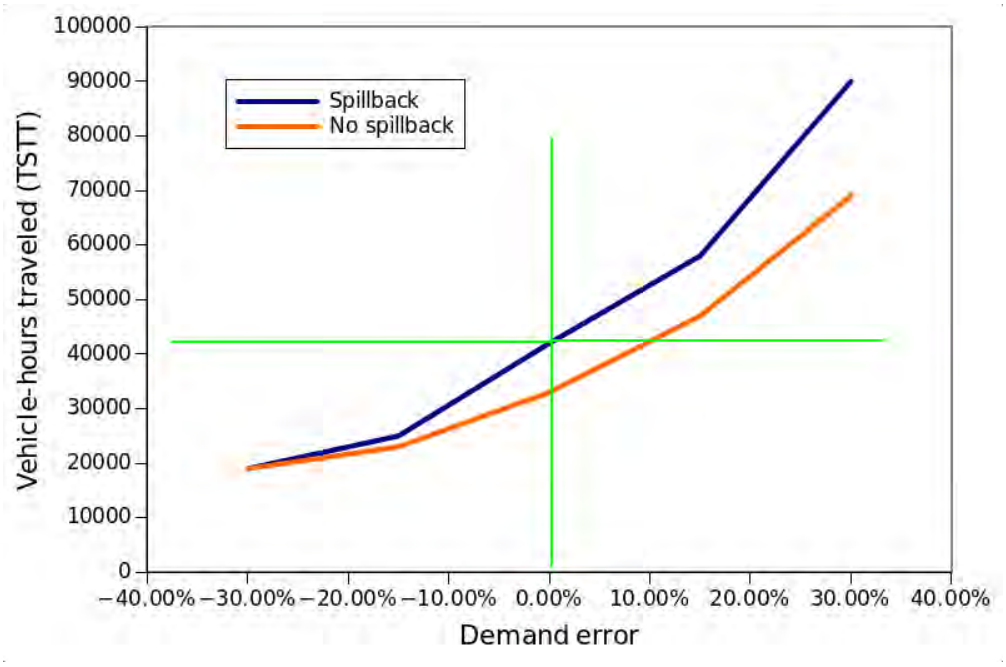


FIGURE 7 TSTT for Austin, TX as demand varies. Horizontal/vertical lines indicate true values.

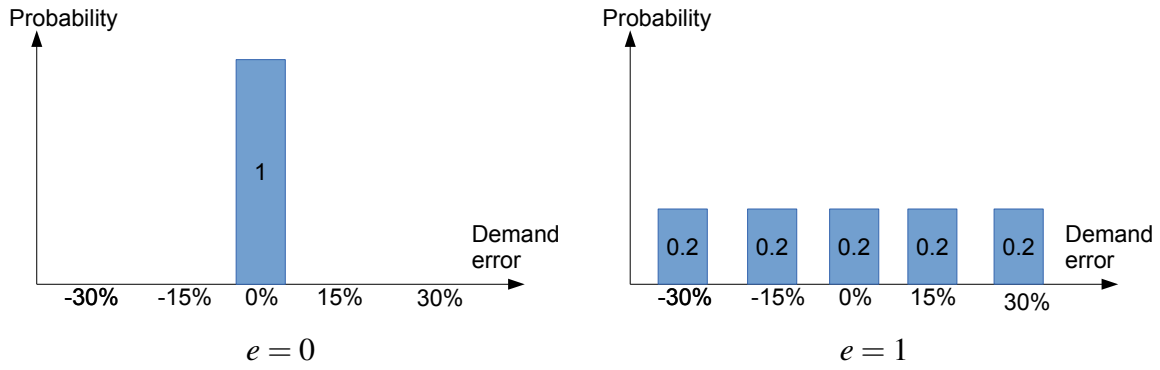


FIGURE 8 Forecasting error distributions; for $e \in (0, 1)$ the distribution is a convex combination of the two shown here.

TABLE 4 Expected Absolute Errors for Demand Uncertainty Levels.

e	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
San Antonio, TX: TSTT (veh-hr $\times 10^3$)											
S	**0**	**2.1**	**4.2**	**6.2**	**8.3**	**10.4**	**12.5**	14.6	16.6	18.7	20.8
NS	9.0	9.8	10.5	11.3	12.0	12.8	13.6	**14.3**	**15.1**	**15.8**	**16.6**
San Antonio, TX: Average OD travel time (min)											
S	**0**	**0.3**	**0.5**	**0.8**	**1.0**	**1.3**	**1.6**	**1.8**	2.1	2.3	2.6
NS	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	**2.0**	**2.0**	**2.0**
Austin, TX: TSTT (veh-hr $\times 10^3$)											
S	**0**	**0.6**	**1.2**	**1.8**	**2.4**	**3.0**	**3.6**	4.2	4.8	5.4	6.0
NS	3.0	3.1	3.3	3.4	3.6	3.7	3.8	**4.0**	**4.1**	**4.3**	**4.4**
Austin, TX: Average OD travel time (min)											
S	**0**	0.08	0.16	0.24	0.32	0.40	0.48	0.56	0.64	0.72	0.80
NS	0.00	**0.02**	**0.04**	**0.06**	**0.08**	**0.10**	**0.12**	**0.14**	**0.16**	**0.18**	**0.20**

spillback and no-spillback models as $E[|TSTT^S - TSTT^*|]$ and $E[|TSTT^{NS} - TSTT^*|]$, respectively, where $TSTT^*$ is the assumed true value (e.g., 42,000 for Austin; see the boldface values in Table 3). Expected error in average OD travel times is calculated analogously.

Table 4 shows these expected absolute errors for both networks and metrics. The row labels S and NS correspond to the spillback and no-spillback models, respectively. The values in boldface indicate the model with lower error.

For both networks and metrics, when the forecast uncertainty e is low, the model with spillback produces lower absolute error, as is expected. When the forecast uncertainty is sufficiently high, however, the model without spillback produces lower error. This happens for the reasons discussed in the Motivation section: although the model with spillback contains modeling error (it does not pass through the intersection of the lines in Figure 7), it is less sensitive to data errors (the slope is less steep than the spillback model). If the demand is sufficiently uncertain, this latter effect dominates the former, and the model without spillback gives an answer closer to the true value, on average.

The threshold uncertainty level where the spillback model produces less error varies by network and the metric chosen. In particular, for the Austin network and average OD travel times, the no-spillback model is almost always more accurate than the spillback model. This is because the no-spillback model gives nearly constant OD travel times in our tests, and this value happens to be very close to the true one — it is essentially immune to demand errors over the range we considered. By contrast, the model with spillback is sensitive to the demand level, and predicts much higher delay if demand is overpredicted.

DISCUSSION AND CONCLUSIONS

We compare dynamic network loading models with and without spillback, with a view of trading off modeling errors with robustness to data errors. Queues do spill back in reality, causing instability in traffic flow; but this instability also makes models more sensitive to having the correct input parameters. In two experimental settings, we show that if demand uncertainty is sufficiently high, the model with spillback produces higher errors than the model with spillback. In such cases,

the fidelity of the spillback model to reality (higher demand would indeed cause spillbacks and increase delay) is in fact a liability, if our ability to predict that reality is limited.

The distributions used in this paper (and the use of the *e* parameter) are a simple way to model demand uncertainty, but we have no evidence that this is an accurate model for demand errors in practice. The question of *how* forecasting and demand errors arise and should best be described remains open. There are many factors which combine to produce errors in demand, including sampling errors, land use and demand model specification errors, uncertainty in future economic or technological scenarios, and so forth. Further research in this area would be highly valuable.

We thus advise against interpreting the significance of the particular *e* thresholds observed in this paper. Nevertheless, the conclusions of the paper still have qualitative value: that models with spillback perform well when the input demand is known with high precision, but less well if there is much uncertainty. In the latter case, neglecting spillback may in fact improve the accuracy of the model. In practice, there may be scenarios where modeling spillback is essential. In such cases, one must be willing to invest the resources needed to produce accurate demand inputs.

Future research should also broaden the scope of these experiments to encompass more modeling choices. For concreteness, we have focused on whether spillback should be included or not, but there are many other decisions which must be made when building a model: to name only a few, the network size, modeling resolution, node models, whether demand should be elastic, whether perception errors are present, whether network conditions are stochastic or deterministic, and even the choice of a microsimulator, dynamic traffic assignment, or static traffic assignment.

In all of these cases, the realism of alternative models should be compared with confidence in the inputs, and the models' robustness to errors in these inputs. A more realistic, but less robust model should only be used if the inputs are known with high precision.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the National Science Foundation (CMMI-1254921) and the Data-Supported Transportation Operations and Planning Center. Thanks are also due to the Capital Area Metropolitan Planning Organization and San Antonio Metropolitan Planning Organization for providing the base network and travel demand data used to generate the DTA models for the second experiment.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: S. D. Boyles and N. Ruiz Juri; data collection: N. Ruiz Juri; analysis and interpretation of results: S. D. Boyles and N. Ruiz Juri; draft manuscript preparation: S. D. Boyles and N. Ruiz Juri. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] Skabardonis, A. and N. Geroliminis, Real-time monitoring and control on signalized arterials. *Journal of Intelligent Transportation Systems*, Vol. 12, 2008, pp. 64–74.

[2] Wu, X., H. X. Liu, and D. Gettman, Identification of oversaturated intersections using high-resolution traffic signal data. *Transportation Research Part C*, Vol. 18, 2010, pp. 626–638.

[3] Liu, Y. and G.-L. Chang, An arterial signal optimization model for intersections experiencing queue spillback and lane blockage. *Transportation Research Part C*, Vol. 19, 2011, pp. 130–144.

[4] Ban, X. J., P. Hao, and Z. Sun, Real time queue length estimation for signalized intersections using travel times from mobile sensors. *Transportation Research Part C*, Vol. 19, 2011, pp. 1133–1156.

[5] Havinovski, G. N., *Ramp queues? Not in my backyard: a survey of queue detector design and operating criteria for metered freeway entrances*, 1991, institute of Transportation Engineers, Compendium of Technical Papers.

[6] Taylor, C. and D. Meldrum, *Evaluation of a Fuzzy Logic Ramp Metering Algorithm: A Comparative Study Among Three Ramp Metering Algorithms Used in the Greater Seattle Area*. Washington State Transportation Center (TRAC), 2000.

[7] Gordon, R., Algorithm for controlling spillback from ramp meters. *Transportation Research Record*, Vol. 1554, 2014, pp. 162–171.

[8] Daganzo, C. F., Queue spillovers in transportation networks with a route choice. *Transportation Science*, Vol. 32, 1998, pp. 3–11.

[9] Daganzo, C. F., Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B*, Vol. 41, 2007, pp. 49–62.

[10] Ban, X., J.-S. Pang, H. X. Liu, and R. Ma, Continuous-time point-queue models in dynamic network loading. *Transportation Research Part B*, Vol. 46, 2012, pp. 360–380.

[11] Friesz, T. L., K. Han, P. A. Neto, and T. Yao, Dynamic user equilibrium based on a hydrodynamic model. *Transportation Research Part B*, Vol. 47, 2013, pp. 102–126.

[12] Han, K., T. L. Friesz, and T. Yao, Existence of simultaneous route and departure choice dynamic user equilibrium. *Transportation Research Part B*, Vol. 53, 2013, pp. 17–30.

[13] Ma, R., X. Ban, and J.-S. Pang, Continuous-time dynamic system optimum for single-destination traffic networks with queue spillbacks. *Transportation Research Part B*, Vol. 68, 2014, pp. 98–122.

[14] Ma, R., X. Ban, and J.-S. Pang, A link-based differential complementarity system formulation for continuous-time dynamic user equilibria with queue spillbacks. *Transportation Science*, Vol. 52, 2017, pp. 564–592.

[15] Han, K., B. Piccoli, and T. L. Friesz, Continuity of the path delay operator for dynamic network loading with spillback. *Transportation Research Part B*, Vol. 92B, 2016, pp. 211–233.

[16] Daganzo, C. F., The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, Vol. 28, No. 4, 1994, pp. 269–287.

[17] Daganzo, C. F., The cell transmission model, part II: network traffic. *Transportation Research Part B*, Vol. 29, No. 2, 1995, pp. 79–93.

[18] Yperman, I., *The Link Transmission Model for Dynamic Newtork Loading*. Ph.D. thesis, Katholieke Universiteit Leuven, Belgium, 2007.

[19] Osorio, C., G. Flötteröd, and M. Bierlaire, Dynamic network loading: a stochastic differentiable model that derives link state distributions. *Transportation Research Part B*, Vol. 45, 2015, pp. 420–431.
[20] Flyvbjerg, B., M. K. S. Holm, and S. L. Buhl, How (in)accurate are demand forecasts in public works projects?: the case of transportation. *Journal of the American Planning Organization*, Vol. 71, 2005, pp. 131–146.
[21] de Jong, G., A. Daly, M. Pieters, S. Miller, R. Plasmeijer, and F. Hofman, Uncertainty in traffic forecasts: literature review and new results for The Netherlands. *Transportation*, Vol. 34, 2007, pp. 375–395.
[22] van Wee, B., Large infrastructure projects: a review of the quality of demand forecasts and cost estimations. *Environment and Planning B*, Vol. 34, 2007, pp. 611–625.
[23] Nicolaisen, M. S. and P. A. Driscoll, *Ex-post* evaluations of demand forecast accuracy: a literature review. *Transport Reviews*, Vol. 34, 2014, pp. 540–557.
[24] Ashley, D. J., Uncertainty in the context of highway appraisal. *Transportation*, Vol. 9, 1980, pp. 249–267.
[25] Leurent, F., *An analysis of modelling error, with application to a traffic assignment model with continuously distributed values of time*, 1996, presented at the European Transport Conference, Uxbridge, United Kingdom.
[26] Boyce, A. M., *Risk analysis for privately funded transport schemes*, 1999, presented at the European Transport Conference, Uxbridge, United Kingdom.
[27] Boyce, A. M. and M. J. Bright, *Reducing or managing the forecasting risk in privately-financed projects*, 2003, presented at the European Transport Conference, Uxbridge, United Kingdom.
[28] Rodier, C. J. and R. A. Johnston, Uncertain socioeconomic projections used in travel demand and emissions models: could plausible errors result in air quality nonconformity? *Transportation Research Part A*, Vol. 36, 2002, pp. 613–631.
[29] Zhao, Y. and K. M. Kockelman, The propagation of uncertainty through travel demand models: an exploratory analysis. *Annals of Regional Science*, Vol. 36, 2002, pp. 145–163.
[30] Waller, S. T., J. Schofer, and A. K. Ziliaskopoulos, Evaluation with traffic assignment under demand uncertainty. *Transportation Research Record*, Vol. 1771, 2001, pp. 69–74.
[31] Duthie, J. C., A. Unnikrishnan, and S. T. Waller, Influence of demand uncertainty and correlations on traffic predictions and decisions. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 26, 2011, pp. 16–29.
[32] Ukkusuri, S. V., T. V. Mathew, and S. T. Waller, Robust transportation network design under demand uncertainty. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 22, 2007, pp. 6–18.
[33] Yin, Y., S. M. Madanat, and X.-Y. Lu, Robust improvement schemes for road networks under demand uncertainty. *European Journal of Operational Research*, Vol. 198, 2009, pp. 470–479.
[34] Ukkusuri, S. V. and G. Patil, Multi-period transportation network design under demand uncertainty. *Transportation Research Part B*, Vol. 43, 2009, pp. 625–642.
[35] Chen, A., Z. Zhou, P. Chootinan, S. Ryu, C. Yang, and S. C. Wong, Transportation network design problem under uncertainty: a review and new developments. *Transport Reviews*, Vol. 31, 2011, pp. 743–768.

[36] Ng, M. and S. T. Waller, Reliable evacuation planning via demand inflation and supply deflation. *Transportation Research Part E*, Vol. 46, 2010, pp. 1086–1094.
[37] Gardner, L. M., A. Unnikrishnan, and S. T. Waller, Robust pricing of transportation networks under uncertain demand. *Transportation Research Record*, Vol. 2085, 2008, pp. 21–30.
[38] Boyles, S. D., K. M. Kockelman, and S. T. Waller, Congestion pricing under operational, supply-side uncertainty. *Transportation Research Part C*, Vol. 18, 2010, pp. 519–535.
[39] Gardner, L. M., A. Unnikrishnan, and S. T. Waller, Solution methods for robust pricing of transportation networks under uncertain demand. *Transportation Research Part C*, Vol. 18, 2010, pp. 656–667.
[40] Gardner, L. M., S. D. Boyles, and S. T. Waller, Quantifying the benefit of responsive pricing and travel information in the stochastic congestion pricing problem. *Transportation Research Part A*, Vol. 45, 2011, pp. 204–218.
[41] Li, Z.-C., W. H. K. Lam, S. C. Wong, and A. Sumalee, Environmentally sustainable toll design for congested road networks with uncertain demand. *International Journal of Sustainable Transportation*, Vol. 6, 2012, pp. 127–155.
[42] Bar-Gera, H., Traffic assignment by paired alternative segments. *Transportation Research Part B*, Vol. 44, 2010, pp. 1022–1046.
[43] Daganzo, C. F., Increasing model precision can reduce accuracy. *Transportation Science*, Vol. 21, 1987, pp. 100–105.
[44] Zhang, H. M. and Y. Nie, *Modeling network flow with and without link interaction: properties and implications*, 2006, presented at the 84th Annual Meeting of the Transportation Research Board, Washington, DC.
[45] Nie, Y. M., J. Ma, and H. M. Zhang, A polymorphic dynamic network loading model. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 23, No. 2, 2008, pp. 86–103.
[46] Tampère, C. M. J., R. Corthout, D. Cattrysse, and L. H. Immers, A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B*, Vol. 45, 2011, pp. 289–309.
[47] Corthout, R., G. Flötteröd, F. Viti, and C. M. J. Tampère, Non-unique flows in macroscopic first-order intersection models. *Transportation Research Part B*, Vol. 46, 2012, pp. 343–359.
[48] Ziliaskopoulos, A. K. and S. T. Waller, An Internet-based geographic information system that integrates data, models and users for transportation applications. *Transportation Research Part C*, Vol. 8, 2000, pp. 427–444.

Volume 2: An Efficient Simulation Framework for Estimating Work-Zone Impacts

AN EFFICIENT SIMULATION FRAMEWORK FOR ESTIMATING WORK-ZONE IMPACTS

Tengkuo Zhu
Researcher
Network Modeling Center
Center for Transportation Research
Cockrell School of Engineering
The University of Texas at Austin
3925 W Braker Ln
Austin, TX 78759
E-Mail: zhutengkuo@utexas.edu

Natalia Ruiz Juri
Director
Network Modeling Center
Center for Transportation Research
Cockrell School of Engineering
The University of Texas at Austin
3925 W Braker Ln
Austin, TX 78759
E-Mail: nruizjuri@mail.utexas.edu
Phone: (512) 232-3099

Stephen D. Boyles
Associate Professor
Department of Civil, Architectural, and Environmental Engineering
Cockrell School of Engineering
The University of Texas at Austin
301 E Dean Keeton St
Austin, TX 78712
E-Mail: sboyles@utexas.edu

Kenneth Perrine
Research Associate
Network Modeling Center
Center for Transportation Research
Cockrell School of Engineering
The University of Texas at Austin
3925 W Braker Ln
Austin, TX 78759
E-Mail: kperrine@utexas.edu
Phone: (512) 232-3123

Amber Chen
Researcher
Network Modeling Center
Center for Transportation Research
Cockrell School of Engineering
The University of Texas at Austin
3925 W Braker Ln
Austin, TX 78759
E-Mail: amber14@utexas.edu
Phone: (512) 232-3075

Yun Li
Jiangsu Key Laboratory of Traffic and Transportation Security
Huaiyin Institute of Technology
No.1 Meicheng Road
Huaian, Jiangsu Province, China 223003
E-mail: liyunxiaobu@gmail.com
Phone: +86-182-523-40440

Word count: 6,511 words text + 3 tables × 250 words (each) = 7261 words

Submission Date: November 13, 2018

ABSTRACT

Realistic estimations of queue propagation and traffic delays due to freeway lane closures are critical for planning and managing work zones. The current practical methods include deterministic queuing theory and microsimulation. However, these approaches are either too simple to capture the complex traffic entry/exit patterns or require significant modeling effort. This paper proposes a framework to use kinematic wave theory to estimate queue position and the user's traffic delay. Link transmission model (LTM) is implemented in the analysis of work zone impacts due to its efficiency and is validated in different scenarios. A comparison of LTM method with microsimulation suggests that LTM method produces comparable queue propagation and dissipation patterns, with runtimes of only one second. LTM method estimations were also consistent with field data collected on IH35 through Austin, TX, suggesting that model results are realistic, and could support decision-making given adequate input data and parameter calibration. The method presented can be used by agencies to support planning and operational decisions, such as the placement of variable message signs, for events that happen frequently and cannot be modeled in extensive detail. It can also facilitate the computation of meaningful performance metrics to communicate with stakeholders for strategic planning, and to conduct cost-benefit analyses, among others.

Keywords: Freeway Lane Closure, Link Transmission Model, Queue Length, User Delay Cost

INTRODUCTION

Freeway lane closures, which include restoration, resurfacing, reconstruction of traffic facilities, accidents, or special events, typically lead to traffic congestion and travel time delays. Although a variety of traffic management techniques, from ramp closure to variable message signs (VMS), are available to mitigate the impacts of temporary closures, their effective implementation typically depends on understanding the expected impacts of a given work zone. A detailed understanding of work-zone impacts is also critical for quantifying their costs, which plays an important role in work-zone planning, communication with stakeholders, and when evaluating the benefits of delay mitigation technologies and strategies. Queue length plays an important role in characterizing traffic in the proximity of a freeway lane closure. An accurate estimation of position and length can be used, for example, to inform the placement and activation of VMS, which are expected to be most effective when their placement allows drivers to take alternative paths before entering the congested portion of the freeway.

Several methods exist to estimate work-zone impacts, including microsimulation models (1), simpler simulation-based or analytical approaches (2), and data-centric artificial intelligence techniques (3). While the latter is promising, the volume and quality of data needed to ensure reliable performance differ in a variety of contexts, and the associated computational effort still presents a significant challenge. Technologies based on traffic flow and/or car-following theory are appealing because, once properly calibrated, these methods could be effective under a range of scenarios. Among these, analytical approaches, which implement macroscopic flow models or queuing theory (4), are most commonly used in practice. According to FHWA's *Work Zone and Traffic Analysis Technical Resources* (5), at least fifteen state department of transportation use this method, which is implemented into a simple spreadsheet, to estimate queue length and user's traffic delay of construction and maintenance activities. However, this technique may not be accurate when analyzing work zones that affect several entry/exit ramps and do not always provide a fair evaluation of user delay. Microsimulation models are also popular, which are expected to provide the most realistic assessment of work-zone impacts, but model preparation effort and the corresponding operation times may be prohibitive for some applications.

This paper aims to propose an approach for realistically simulating the queue propagation and dissipation process resulting from freeway closures and estimating the user's traffic delay. The goal is to obtain results comparable to microsimulation but with significantly lower modeling effort, runtime and budget, so that it can be used for planning and managing short-term/nightly work zones with limited resources. The efficient approach can also enable the computation of more detail metrics for economic analysis and strategic planning, and potentially be extended to be used in real time.

The method is based on Yperman's link transmission model (LTM) (6), which provides a simulation-based approximation to the hydrodynamic theory of traffic flow (7, 8). This approach can account for varying geometric configurations and can be efficiently implemented through spreadsheets or programming. The paper also proposes a heuristic adjustment of model parameters to account for weaving effect in the proximity of entry and exit ramps using *Highway Capacity Manual* (HCM). The issue of model parameter selection and how they affect estimate results are also addressed.

Two types of numerical experiments are used to validate the performance of the LTM method: a set of microsimulation experiments using synthetic data explores model performance under varying origin-destination (OD) demand patterns, and real-world data from a lane closure on IH35 in Austin, Texas, which is used to understand the model's ability to replicate field conditions based on field traffic counts. Results suggest that LTM models provide an accurate approximation of microsimulation results when analyzing queues resulting from freeway closures. Furthermore, the LTM framework can generate acceptable estimates of work-zone impacts when incorporating field data.

The following sections provide additional background on work-zone impacts estimation, LTM implementation, numerical experiments, and corresponding results.

LITERATURE REVIEW

The estimation of work-zone impacts is critical to support strategic planning, communication among relevant stakeholders, and the quantification of user delay costs for cost-benefit analyses. Existing approaches in the literature include analytical and simulation-based methodologies. Among analytical approaches, the deterministic queuing theory has been widely used due to its simplicity (2, 4). According to the FHWA (5), more than fifteen state departments of transportation use tools based on deterministic queuing theory to analyze traffic delays. Research on this topic is typically focused on quantifying adequate capacity value for the work, e.g. (4), which is later used to estimate queue length. While efficient, many such tools are not suitable to study work zones that lead to long queues and affect several entry and exit ramps because they are unable to modeling complex traffic in such scenarios. Microsimulation models have also been successfully used in the context of work-zone impact analysis (e.g., 1), but model development effort and running times are often excessive for some applications, such as short-term construction.

In this context, the mesoscopic model, which combines the advantages of the analytical approach and microsimulation, has drawn increasing attention in recent years. Researchers have studied the use of mesoscopic models, including the cell transmission model (CTM) (10) and link transmission model (5), to analyze congestion propagation and delays during closures. Corthout et al. (11) propose a framework to assess the network-level impact of incidents without re-computing link flows throughout the network using LTM. Osorio et al. (12) develop a dynamic network loading (DNL) model based on finite capacity queuing theory to compute the probability of a queue being present at the downstream end of a link, as well as the probability of a spillback.

Previous work has also considered corridor-level mesoscopic models that leverage data for improved realism. Ruiz et al. (13) utilize CTM to predict travel times on freeway segments based on real-time traffic sensor data. Xia et al. (14) predict delays under incident conditions based on loop detector data and kinematic wave theory, while Liu et al. (15) use similar data and assumptions to study queues at congested urban intersections. Also, in the context of urban intersections, Comert and Cetin (16) propose a different approach that estimates queue lengths based on probe-vehicle data. To the knowledge of the authors, previous research has not discussed some of the practical considerations of sustainable implementing mesoscopic models such as parameter adjustment, or explicitly analyzed their performance given typically available field data.

This research implements and extends an offline version of the framework proposed in Ruiz et al. (13) to the analysis of closure-related queues on freeways using LTM. Discussion about how traffic data available from smart work-zone trailers may be used to inform model development and validation and to refine model parameters is also included in the paper. Additionally, a heuristic method, based on the *HCM* (17), is used to adjust model parameters to account for lane-changing maneuvers.

METHODOLOGICAL APPROACH

This section introduces the LTM model and its corresponding parameters. Additionally, this section describes the procedure to estimate queue length and user's traffic delay and procedures to analyze work zone impact using the proposed framework.

The Link Transmission Model

The link transmission model (5), first introduced by Yperman, is an effective tool that implements Newell's simplified kinematic traffic flow theory (18), in which traffic is modeled as a fluid that is a continuous function of time and space. LTM discretizes space and represents roadway segments as links connected at nodes. Each link is characterized by its length L_a, free-flow speed u_a^f, capacity q_a^{max}, jam density k_a^j and backward wave speed w_a. Parameter values assume a homogeneous, time-independent, and triangular fundamental diagram (Figure 1).

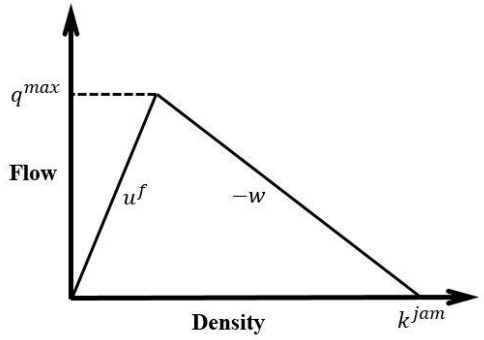


FIGURE 1. Fundamental diagram for LTM

LTM updates the variables representing cumulative counts upstream and downstream end of a link, $N^{\uparrow}(t)$ and $N^{\downarrow}(t)$, at a pre-defined time step based on the flow transmitted between links. Transmitted flow between links 1 and 2 at time step t, $q_{12}(t)$, is a function of link receiving flow $R(t)$ and sending flow $S(t)$, which characterize the maximum flow that link a could receive or send at time t, respectively (Equations [1] and [2] in which Δt denotes the time step size),

$$R_a(t) = \min\{q_a^{max}\Delta t, k_a^j L_a + N_a^{\downarrow}\left(t - \frac{L_a}{w_a} + \Delta t\right) - N_a^{\uparrow}(t)\} \quad [1]$$

$$S_a(t) = \min\{q_a^{max}\Delta t, N_a^{\uparrow}\left(t - \frac{L_a}{u_a^f} + \Delta t\right) - N_a^{\downarrow}(t)\} \quad [2]$$

The calculation of transmitted flows depends on the type of geometry and corresponding node model assumptions. General node models and intersection types are proposed in (19). The following sections describe the approach used in this work to represent links in series, merges, and diverges.

Links in series

For nodes in which there is one incoming and one outgoing link, $q_{12}(t)$ is given by equation [3], where link 1 is the incoming link and link 2 is the outgoing link.

$$q_{12}(t) = \min\{S_1(t), R_2(t)\} \quad [3]$$

Diverge

Diverge scenario is characterized by one incoming link (link 1) and two or more outgoing links (links 2 and 3). Assuming the fraction of flow destined to each outgoing link is given by p_{12} and p_{13}, two possible scenarios for flow calculation may arise. Case I represents a situation in which both downstream links can accommodate the sending flow, making the transmitted flow equal to the appropriate fraction of sending flow (Equations [4] and [5]).

$$p_{12}S_1(t) \le R_2(t),\ p_{13}S_1(t) \le R_3(t) \quad [4]$$

$$q_{12}(t) = p_{12}S_1(t),\ q_{13}(t) = p_{13}S_1(t) \quad [5]$$

In Case II, at least one of the downstream links cannot accommodate the incoming flow, then a moving fraction φ(t) is computed to maximize flow (Equations [6] and [7])

$$\varphi(t) = \min\left\{\frac{R_2(t)}{p_{12}S_1(t)}, \frac{R_3(t)}{p_{13}S_1(t)}\right\} \quad [6]$$

$$q_{12}(t) = \varphi(t)\, p_{12}S_1(t),\quad q_{13}(t) = \varphi(t)\, p_{13}S_1(t) \quad [7]$$

Merge with priority

In a merge scenario, there is one outgoing link (link 1) and two incoming links (link 2 and link 3). The calculation of transmitted flow requires defining the relative priority of the incoming links, ρ, for cases in which the receiving link cannot accommodate the total sending flow.

Three possible scenarios may arise, depending on downstream conditions. In Case I the receiving link can accommodate all incoming flow (Equations [8] and [9]).

$$S_2(t) + S_3(t) \le R_1(t) \quad [8]$$

$$q_{21}(t) = S_2(t),\ q_{31}(t) = S_3(t) \quad [9]$$

Case II represents the situation where flow from both upstream links is restricted by the downstream link, and transmitted flow is given by Equations [10] and [11].

$$S_2(t) > \frac{q_2^{max}}{q_2^{max} + q_3^{max}} R_1(t),\ S_3(t) > \frac{q_3^{max}}{q_2^{max} + q_3^{max}} R_1(t) \quad [10]$$

$$q_{21}(t) = \frac{q_2^{max} \cdot \rho_2}{q_2^{max} + q_3^{max}} R_1(t),\ q_{31}(t) = \frac{q_3^{max} \cdot \rho_3}{q_2^{max} + q_3^{max}} R_1(t) \quad [11]$$

In Case III only one of the upstream links is restricted by the downstream link receiving flow. Equations [12] and [13] illustrate the case where a queue forms in link 2.

$$S_2(t) > \frac{q_2^{max}}{q_2^{max} + q_3^{max}} R_1(t),\ S_3(t) < \frac{q_3^{max}}{q_2^{max} + q_3^{max}} R_1(t) \quad [12]$$

$$q_{21}(t) = \frac{q_2^{max} \cdot \rho_2}{q_2^{max} + q_3^{max}} R_1(t),\ q_{31}(t) = R_1(t) - q_{21}(t) \quad [13]$$

The value of link priority depends on the geometric characteristics of the intersection and incoming link capacities. In the applications studied in this paper, the throughput of entry ramps suggests a much higher priority as a result of the presence of an auxiliary lane, and of generally low speeds during the analyzed period, because

geometric conditions make it easier for entering vehicles to join and merge into the main lanes even at the expense of main lane vehicles.

Queue Length Estimation

This section describes the calculation of queue length and queue tail based on traffic shockwave position.

Equation [14] illustrates the calculation of shockwave speed v, where $q^{\uparrow}(t)$, $q^{\downarrow}(t)$ are upstream and downstream outflow rate, and $k^{\uparrow}(t)$, $k^{\downarrow}(t)$ are upstream and downstream densities, respectively.

$$v = \frac{q^{\uparrow}(t) - q^{\downarrow}(t)}{k^{\uparrow}(t) - k^{\downarrow}(t)} \quad [14]$$

The shockwave movement per time step is $v\Delta t$ and queue length can be estimated by tracking the shockwave in every time step. The computation of densities is shown in Equations [15] and [16].

$$k^{\uparrow}(t) = \begin{cases} \frac{q^{\uparrow}(t)}{u^f} & if\ q^{\uparrow}(t) < R(t) \text{ and } R(t) = q^{max} \\ k^j - \frac{q^{\uparrow}(t)}{w} & \text{Otherwise} \end{cases} \quad [15]$$

$$k^{\downarrow}(t) = \begin{cases} \frac{q^{\downarrow}(t)}{u^f} & \text{if } q^{\downarrow}(t) = S(t) \\ k^j - \frac{q^{\downarrow}(t)}{w} & \text{Otherwise} \end{cases} \quad [16]$$

User's Traffic Delay Calculation

The user's traffic delay is defined as the difference between the total system travel time (TSTT) under construction and normal operating conditions. TSTT could be estimated as illustrated in Figure 2. At time t, $N_a^{\uparrow}(t) - N_a^{\downarrow}(t)$ represents the number of vehicles on link a. So, TSTT of link a during time interval $(t, t + \Delta t)$ is $(N_a^{\uparrow}(t) - N_a^{\downarrow}(t)) * \Delta t$. Denote T as the total number of time steps in the simulation, and A as the set of all the links in the network, the TSTT of the network is given by:

$$TSTT = \sum_{a \in A} \sum_t^T [(N_a^{\uparrow}(t) - N_a^{\downarrow}(t)) * \Delta t] \quad [17]$$

and the user's traffic delay is

$$Delay = TSTT^c - TSTT^0 \quad [18]$$

where $TSTT^c$ and $TSTT^0$ are the TSTT of the network under construction and normal operating conditions.

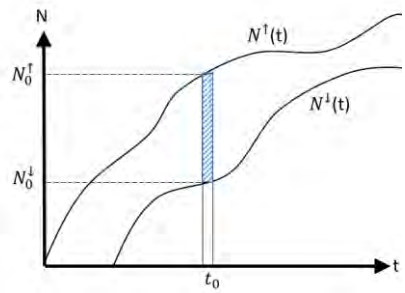


FIGURE 2. Total system travel time between t_0 and $t_0 + \Delta t$

Estimation of capacity through a work zone

The work zone capacity is a critical parameter in any user delay cost estimation approach. This study implements the *HCM* (17) method to estimate work-zone capacity, presented in Equation [19],

$$C_{wz} = \frac{QDR_{wz}}{100 - a_{wz}} \times 100 \quad [19]$$

where C_{wz} is work-zone capacity (passenger car per hour per lane[pcphpl]); a_{wz} is the percentage drop in pre-breakdown capacity at the work zone; QDR_{wz} is the average fifteen-minute queue discharge rate (pcphpl) and can be calculated using Equation [20],

$$QDR_{wz} = 2093 - 154 \times LCSI - 194 \times f_{BT} - 179 \times f_{AT} + 9 \times f_{LAT} - 59 \times f_{DN} \quad [20]$$

in which f_{BT} is an indictor variable for barrier type; f_{AT} is an indictor factor for area type; f_{LAT} is lateral distance from the edge of travel lane to the barriers; f_{DN} is an indicator variable for day or night conditions.

Accounting for Lane-changing Effects

When work zones are present, additional lane-changing activity is typically observed upstream from closed lanes and exit ramps, leading to a reduced capacity. In this paper, this impact is accounted by limiting the receiving flow on links downstream from the "lane-changing influence area" (Figure 3) based on *HCM*. The detailed procedure is as follows.

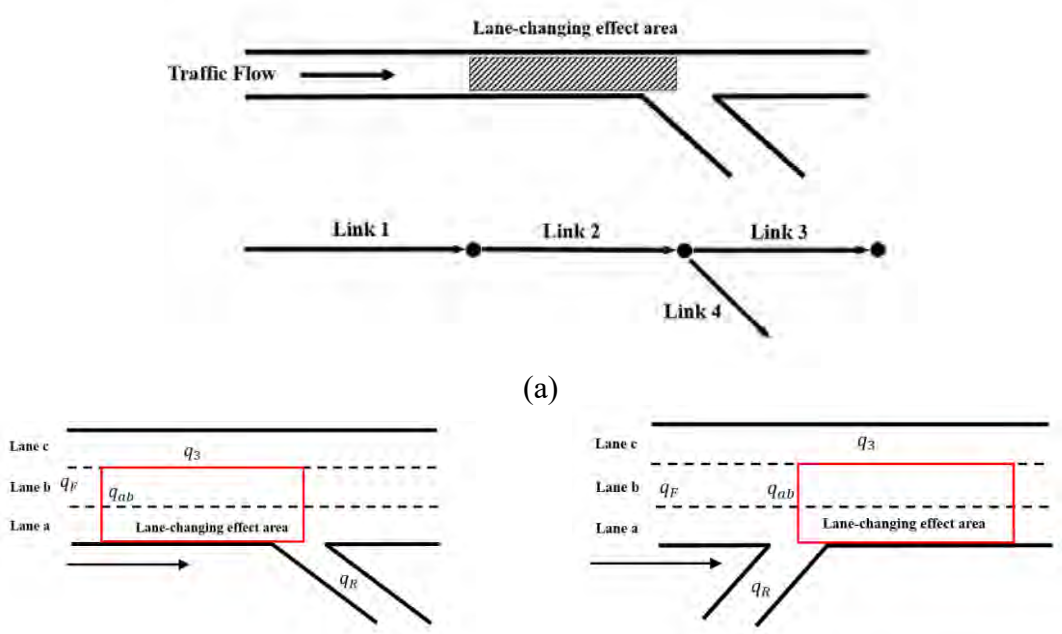


FIGURE 3. Illustration of (a) diverging scenario and (b) lane-changing effect area for diverge and merge.

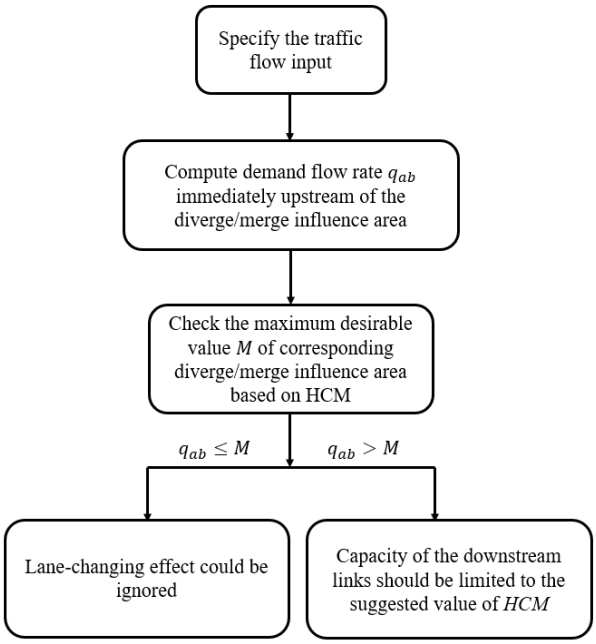


FIGURE 4. Workflow to account for the lane-changing effect.

Imagine a diverging scenario as shown in Figure 3 (a), based on *HCM*, a lane-changing effect area appears upstream of the exit ramp. If the upstream traffic flow is low, the lane-changing effect is marginal and could be ignored. However, as the upstream traffic flow increases, the straight-moving vehicles would be affected by the lane-changing maneuvers of the diverging vehicles and the actual number of straight-moving vehicles is restricted.

In LTM, this scenario could be model as four links as shown in the Figure 3 (b), in which link 2 represents the link to account for the lane-changing effect. If the traffic that flows into the lane-changing effect area exceeds the suggested value of HCM, the capacity of link 3 should be restricted so that the actual flow that diverges and moves straight would decrease. Note that for multiple lane freeways, for example, a three-lane freeway as shown in Figure 3 (b), the link upstream flow q_F is different from the traffic that flows into the lane-changing effect area q_{ab} because the lane-changing effect area only covers the lateral two lanes while the link upstream flow is the traffic that flows into all three lanes. Typically, q_{ab} is a function of q_F and readers could refer to (17) for a detailed calculation of the lane-changing effect flow.

In this regard, the workflow to account for lane-changing effect is shown in Figure 4.

Implementing the LTM Method

When lane closure presents due to construction or maintenance activities, the capacity of the road segment decreases, and the queue begins to accumulate upstream of the work zone segment when the traffic flow exceeds the capacity. The proposed method could simulate this process by assigning a low capacity to the links that represent the work zone area. Besides, the user's delay could be estimated as the difference between the TSTT under construction and normal operating conditions. The analysis procedure is illustrated in Figure 5.

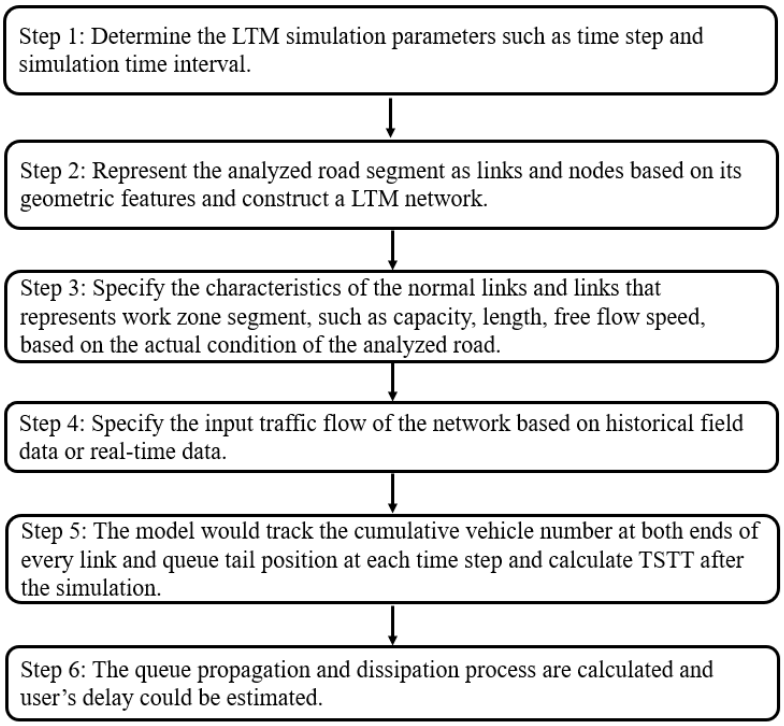


FIGURE 5. Illustration of the work zone impact analysis procedure.

VALIDATION USING MICROSIMULATION

This section presents the comparison of the queue propagation and dissipation process as simulated by LTM with VISSIM microsimulation under a range of travel demand inputs.

The experiments consider a hypothetical freeway (Figure 6(a)) with four continuous main-lane roadway segments. In Figure 6(b) Link 1 is one mile long, and the remaining links are 0.5 miles long. The free-flow speed is assumed to be 60 mph in the main-lane segments and 45 mph in exit ramps. A lane closure in Link 7 is represented by a reduction in the number of lanes from 3 to 2. The experiments consider different scenarios involving various demand patterns (Table 1).

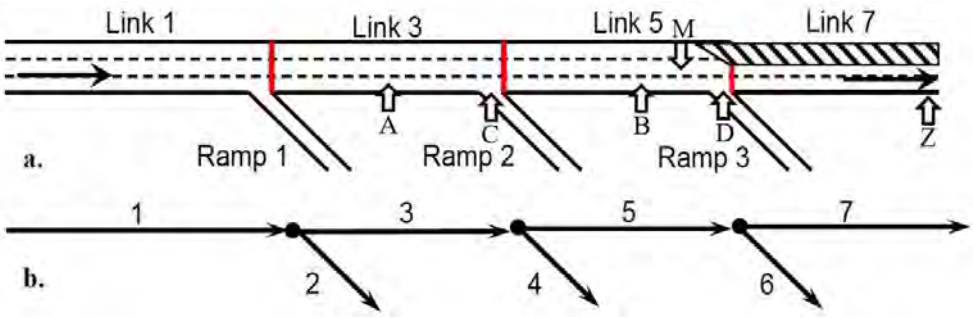


FIGURE 6. (a) Freeway network in which the arrows represent the location of potential bottleneck points and (b) corresponding LTM model

Microsimulation model

Vehicle distributions and driver behavior parameters are chosen to mirror the United States driving patterns. Trucks percentage is chosen as 3 percent. The decision-making distance is 2,000 feet upstream of the respective diverges and lane closure. The simulator is configured to work at 5 times per simulation second.

For each scenario, simulation runs are conducted with different random number seeds. The queue tail position results are then analyzed to find the averages and 95 percent confidence intervals according to the t-distribution with 12 degrees of freedom. It is indirectly obtained by analyzing VISSIM queue counter primitives that are placed throughout the main lanes, which are "active" when vehicles crossing the counter location travel less than 3 mph, with a headway of fewer than 66 feet. Then, queue counters deem the queue dissipated (or "deactivated") when speeds exceed 6 mph.

LTM implementation

For all scenarios, the backward wave speed is assumed to be half of the free-flow speed, and the jam density is 193 vehicles per mile per lane (vpmpl). A time step of 10 seconds is selected. The base capacity of main lanes and ramps are 2300 vehicles per hour per lane (vphpl) and 2000 vphpl respectively.

Comparison scenarios

Eighteen scenarios are considered (Table 1), each of which presents a different demand defined by the total volume entering on Link 1 and the diverge proportion at each exit ramp.

In scenarios with "simple" geometry type, Ramp 3 is closed. The "compound" cases include a 10 percent exit volume on Ramp 3 that causes merge and diverge maneuvers to overlap in the same freeway region.

In both approaches, the demand is maintained for simulation minutes and then reduced to a "cool down" traffic of 2000 vehicles per hour (vph). In some scenarios, the combination of input flow and OD pattern results in travel demands higher than available capacity ($q_a/q_a^{max} > 1$) at one or more locations. The last column identifies these locations, which are also shown in Figure 6(a).

TABLE 1. Demand Scenarios

ID	Geometry Type	Demand	Ramp 1	Ramp 2	Ramp 3	$q_a/q_a^{max} > 1$
U1a	Simple (without merge/diverge)	5500 vph	10%	10%	Closed	M
U1b			10%	25%	Closed	
U1c			10%	40%	Closed	A
U2a		6000 vph	10%	10%	Closed	MZ
U2b			10%	25%	Closed	A M
U2c			10%	40%	Closed	A C
U3a		6500 vph	10%	10%	Closed	A MZ
U3b			10%	25%	Closed	A MA
U3c			10%	40%	Closed	A C
D1a	Compounded (merge/diverge)	5500 vph	Closed	10%	10%	M
D1b			Closed	10%	25%	
D1c			Closed	10%	40%	B

D2a		6000 vph	Closed	10%	10%	MZ
D2b			Closed	10%	25%	B M
D2c			Closed	10%	40%	B D
D3a		6500 vph	Closed	10%	10%	B MZ
D3b			Closed	10%	25%	B M
D3c			Closed	10%	40%	B D

* A: Link 4 diverge influence area, B: Link 6 diverge influence area, C: Link 4 diverge, D: Link 6 diverge, M: the lane closure merge influence area, and Z: the downstream main lanes link.

Results and Analysis

In this section, the evolution of the queue tail position over time is presented for all scenarios. The queue position is measured in feet from the tail of Link 7.

As the demand and ramp exit fractions are varied, the evolution of the queue tail position and corresponding estimation accuracy are largely dependent upon dominant bottleneck locations.

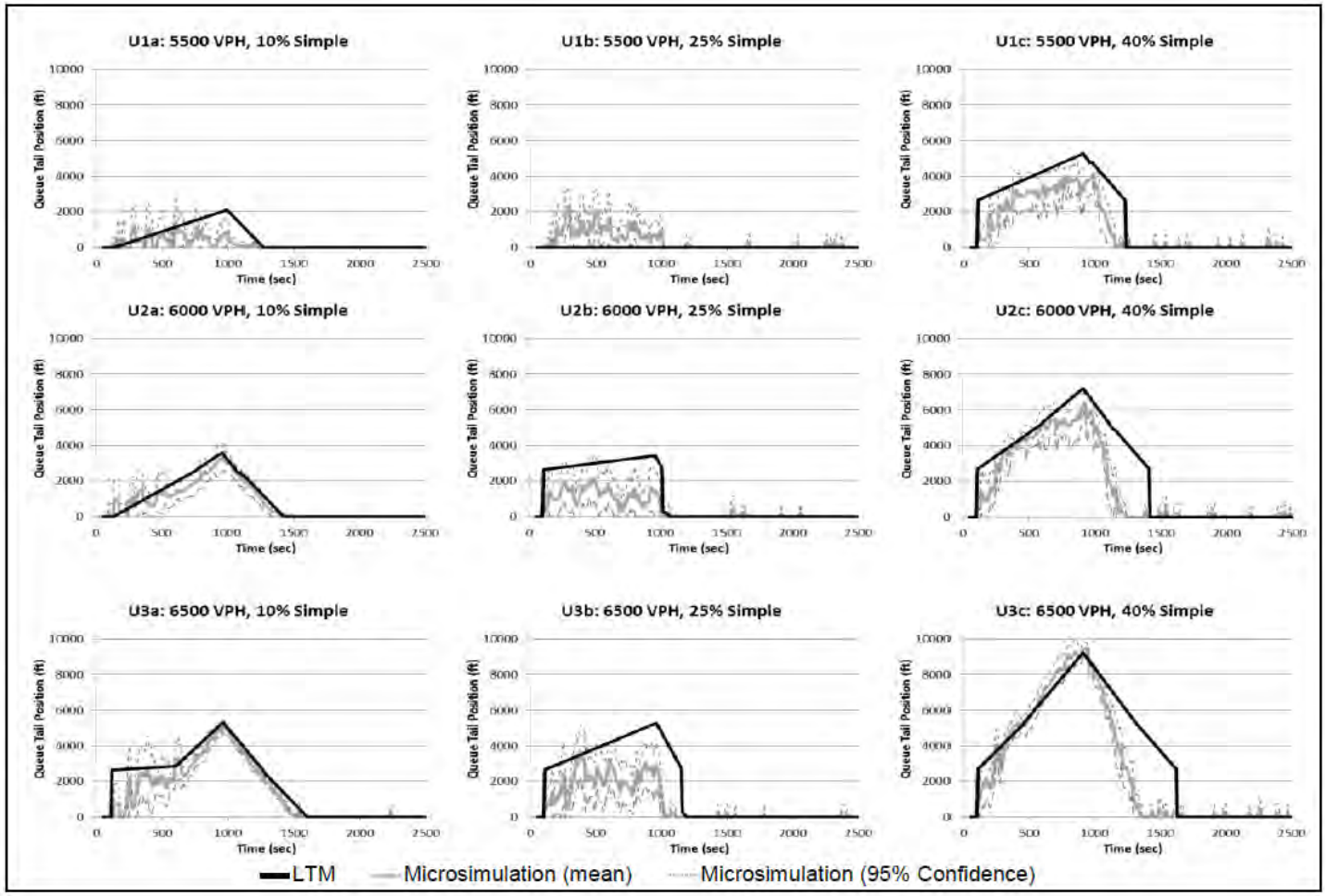
For all the "simple" geometry type scenarios (Figure 7(a)), the bottleneck is located at the merge section leading to the lane closure, and the queue starts forming at the beginning of the simulation. The queue build-up continues until the travel demand is reduced to its cool-down value. In all these cases, both the queue build-up process and maximum queue length are captured fairly accurately. For scenarios D1c, D2c, and D3c, the maximum tail position is slightly underestimated. The former is expected to improve for a better selection of model parameters, particularly jam density (k_a^{jam}). In the more congested scenarios in the compound geometry category (D1a, D2a, and D3a) the LTM-based model tends to underestimate the queue dissipation speed. This is likely the result of implementing a static approach to adjust the capacity of the combined merge/diverge section, which relies on the worst-case demands during the analyzed period. Such an approach may over-penalize the merge section, leading to the observed slower queue dissipation. In reality, the impact of lane changing maneuvers is a function of prevailing traffic volumes that evolve during the simulation. Further research will implement a dynamic analysis considering time-varying demands at the corresponding nodes and links. A second factor that may influence the speed at which the queue tail recedes is the definition of queue in the context of this study.

Figure 7(a) displays the results corresponding to the "simple" geometry-type scenarios. In the least congested cases (scenarios U1and U1b), a single queue forms at the merge location. The evolution of the queue tail position follows a similar pattern to that in the "compound" geometry cases. For scenario U1b, the LTM approach does not predict a queue formation. The corresponding microsimulation results present considerable "noise," produced by very small queues that form and disappear erratically. Such queues, observed across most low-volume scenarios, are likely to be the result of random vehicle interactions which cannot be captured by the mesoscopic model. Their impact in terms of traffic operations is expected to be minimal, so this is not considered a serious limitation of the proposed framework.

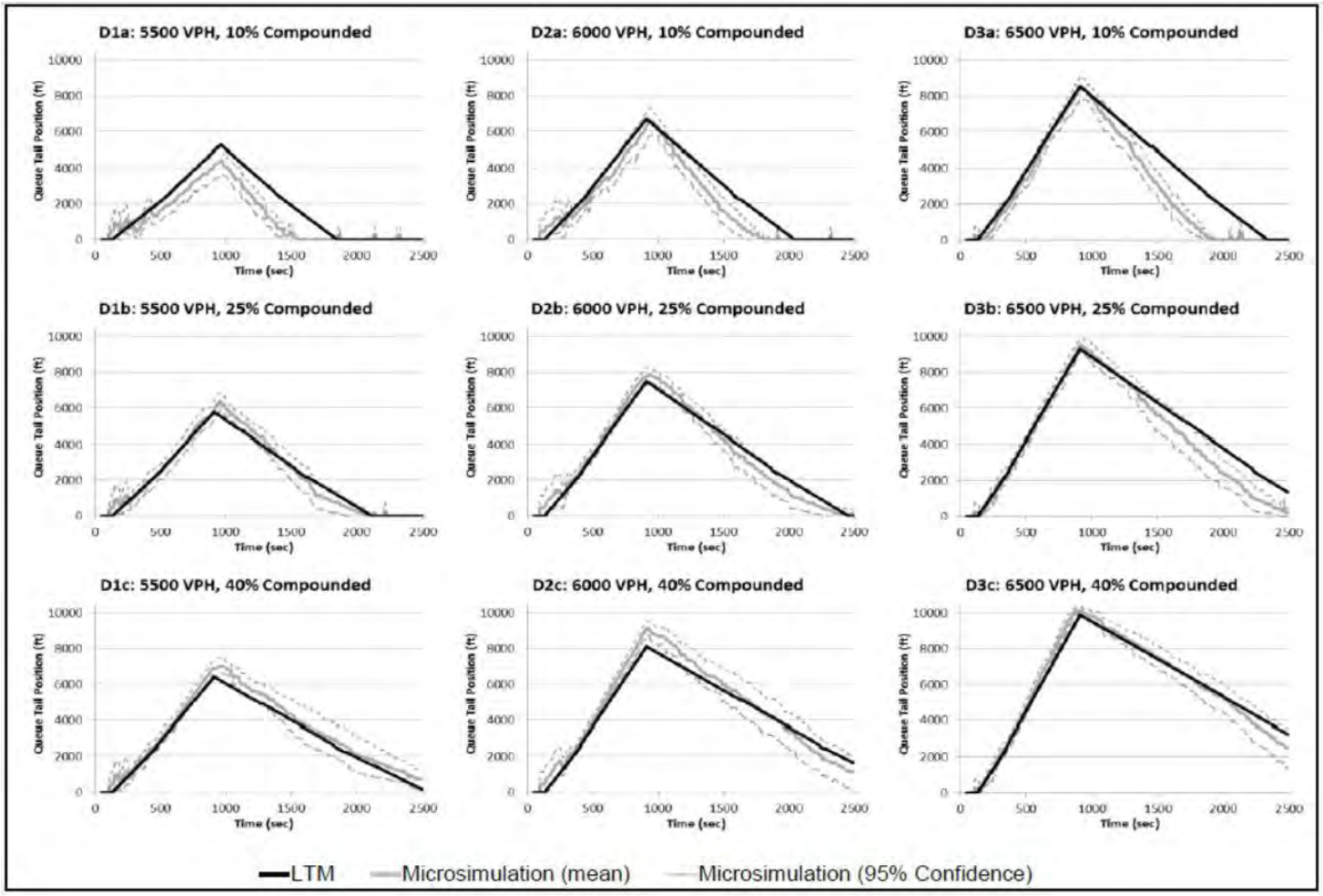
In the remaining scenarios, the queue formation consistently begins at Link 4, given that $q_a/q_a^{max} > 1$ at such location. For scenario U3a, LTM correctly captures the early formation of a queue at Link 4 followed by the arrival of a shockwave originated on

Link 6 due to the higher volume of vehicles that do not exit the freeway. In scenarios U2b, U3b, and U1c, the LTM model overestimates the queue tail position throughout the simulation. These results support the previous conclusion suggesting the need to refine the proposed heuristic method.

The results are very encouraging: the queue formation and dissipation patterns obtained from the LTM-based framework are remarkably similar to the ones observed in the microsimulation experiments. The corresponding root means squared errors values are below 0.1 miles in most cases. Further, the proposed LTM-based method requires a quarter-second to run in a mainstream 3.30 GHz, 4 GB Pentium Core i3 desktop computer. This is a sizeable improvement over the time of a microsimulation cycle, which is about 25 minutes for a set of 12 runs.


U1a: 5500 VPH, 10% Simple
U1b: 5500 VPH, 25% Simple
U1c: 5500 VPH, 40% Simple
U2a: 6000 VPH, 10% Simple
U2b: 6000 VPH, 25% Simple
U2c: 6000 VPH, 40% Simple
U3a: 6500 VPH, 10% Simple
U3b: 6500 VPH, 25% Simple
U3c: 6500 VPH, 40% Simple
Queue Tail Position (ft)
Time (sec)
LTM
Microsimulation (mean)
Microsimulation (95% Confidence)

(a)



(b)

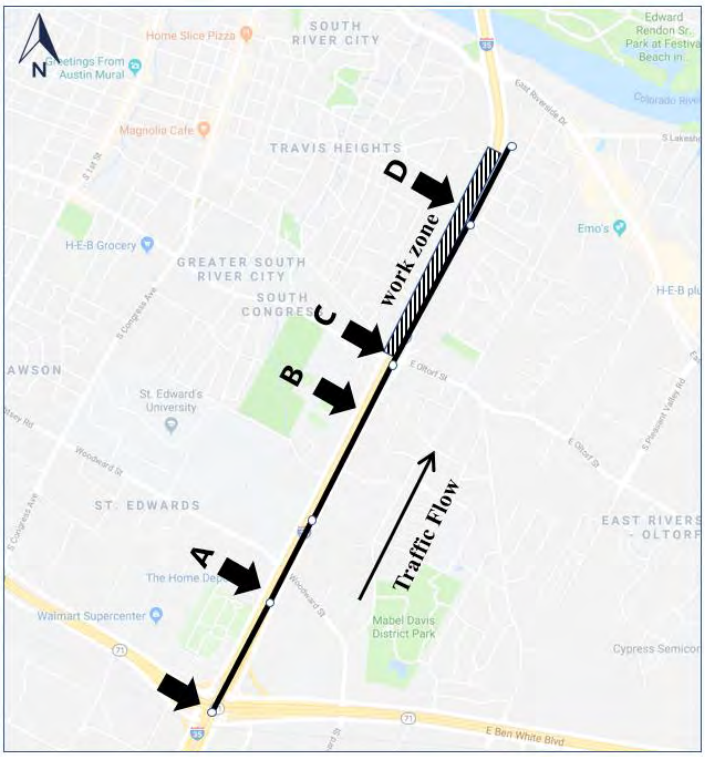
FIGURE 7. Queue tail position for scenarios with (a) "simple" geometry and (b) "compound" geometry

VALIDATION USING FIELD DATA

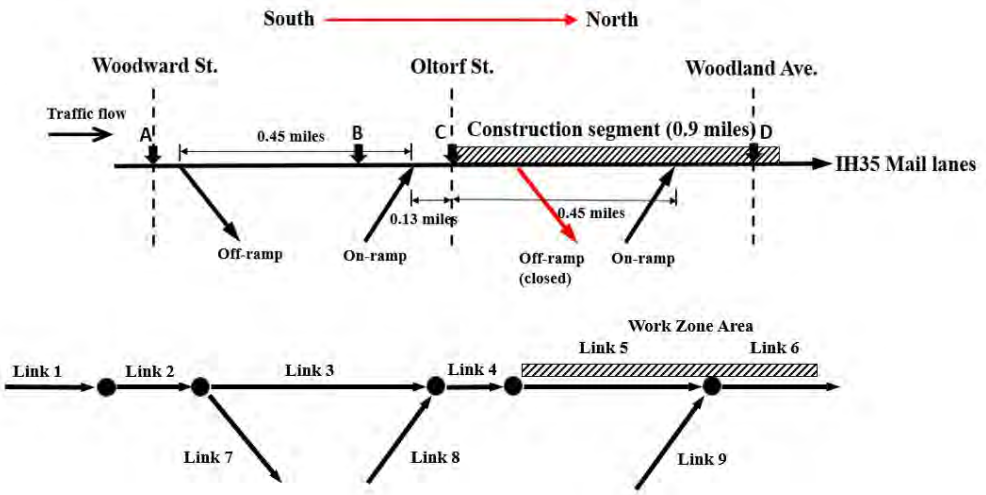
This section describes the implementation of the proposed LTM framework to the analysis of a freeway closure on IH35 in Austin, Texas. The model is developed with traffic volume data collected by the Texas Department of Transportation (TxDOT) using smart work-zone trailers. Probe-based speed data from the National Performance Management Research Data Set (NPMRDS) (20) and point-speed data from work-zone trailers are used to validate estimated queue lengths, duration, and tail position.

Work-zone Description and Simulation Parameters

The analyzed road segment locates on northbound IH35 between Exit 232B and the East Riverside Drive interchange (Figure 8(a)). This road section consists of three lanes, two of which were closed for 7 hours for construction activities beginning on 10/16/2017 at 21:50 pm. There are two on-ramps and two off-ramps in the study area, all of which are single-lane ramps. One off-ramp was closed during the construction period, as illustrated in Figure 8(b). Figure 8(c) presents the LTM representation of the study area. As illustrated, link 1 to link 4 correspond to the IH35 main lanes, with links 5 and 6 representing construction segments and links 7 to 9 representing exit or entry ramps.



(a)



(b)

FIGURE 8. (a) Work-zone location and study area where the arrows represent the trailer location and black lines represent NPMRDS segments and (b) schematic representation and LTM network set-up of the study area

LTM Model Network and Parameters

The total capacity of normal links, work zone links, and ramps are 5,700 vph, 2,300 vph, and 1,350 vph, respectively, based on *HCM*. Jam density is 190 vplpm for normal links. However, the jam density of links 2 and 4 is calculated using only three of the four lanes, to account for observed behavior that auxiliary lanes are not filled by queued vehicles. The free-flow speed is 60 mph for normal links and 45 mph for ramps and work-zone links according to available traffic control plans. The relative priority of the entry ramp is set to 10 based on the previous discussion (see "methodological approach").

In the study, the simulation starts at 21:00 pm to incorporate a "warm-up" period before construction starts (21:50 pm), with a time step of 10 seconds.

Input Traffic Volumes

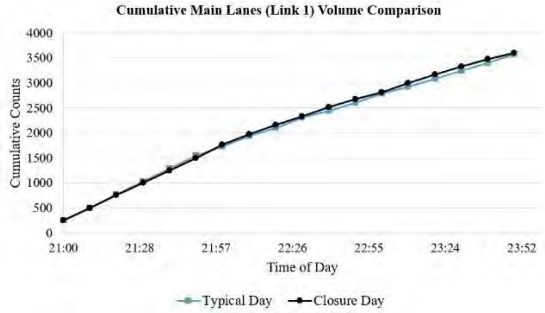
In this study, the input traffic volume is derived from smart work-zone trailers data, which use a radar-based detection system to provide 5-minute traffic counts. The location of work-zone trailers is illustrated in Figure 8(b).

The input traffic volume is estimated as follows: Trailer A provides the main lane traffic volumes, while the difference in volumes between trailer A and trailer B is used to estimate the volumes on the first off-ramp. Similarly, the volume of the remaining ramps is calculated as the difference between its nearby upstream and downstream link volumes.

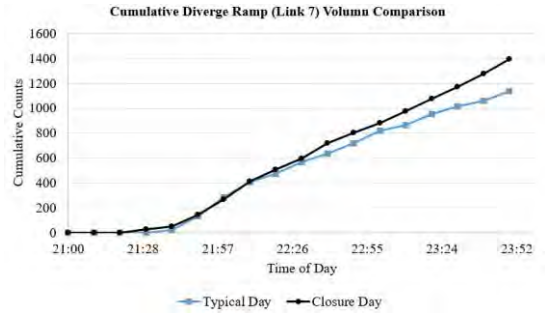
However, when modeling non-recurrent congestion, through traffic of the study area on a closure day may be different from a typical day. A comparison of cumulative traffic counts at three locations (main lanes and first on and off ramps) on a typical day and a closure day (Figure 9(a)) suggests that the main lanes volumes are not significantly affected by construction. However, a larger number of vehicles are observed to exit on link 7 and to enter on link 8 and link 9 on closure day after the work-zone start time

(21:50). The phenomenon suggests that some drivers may be diverted to alternative paths, while others may be queue-jumping: exiting before joining the queue and re-entering closer to the end of the queue. Such behaviors are likely to have an impact on the resulting traffic patterns and should be considered when developing decision support models.

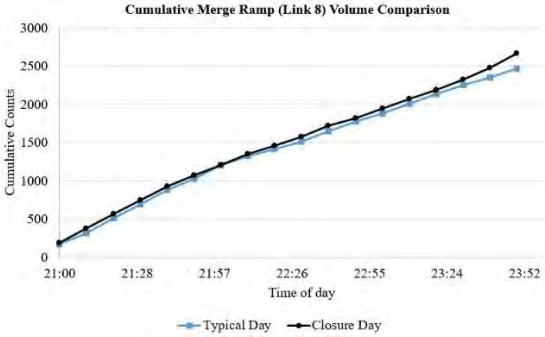
In this context, two sets of results are presented, one that uses typical day traffic input ("typical day inputs"), and one that uses closure day traffic input ("closure day inputs"). While the closure day inputs would not be available on a typical predictive application, the corresponding results illustrate the potential performance of LTM when accurate inputs are available, highlighting the importance of developing adequate techniques to approximate driver behavior during non-recurrent congestion. Table 2 presents the corresponding values.



(a)



(b)



(c)

FIGURE 9. Comparison of cumulative counts at main lanes and on-/off-ramps on a typical day and closure day

TABLE 2. Two input traffic volume profiles

	Closure Day				Typical Day			
Time Interval	Main lanes Traffic Flow (vph)	Exit ramp 1 diverge rate	Entry ramp 1 flow (vph)	Entry ramp 2 flow (vph)	Main lanes Traffic Flow (vph)	Exit ramp 1 diverge rate	Entry ramp 1 flow (vph)	Entry ramp 2 flow (vph)
21:00 – 21:30	1500	0.0%	1128	480	1534	0.0%	1024	482
21:30 – 21:50	1516	10.0%	1100	285	1570	17.0%	1100	280
21:50 – 22:00	1500	38.0%	816	624	1500	82.0%	876	444
22:00 – 22:10	1626	73.0%	870	954	1110	57.0%	1062	606
22:10 – 22:20	1149	50.0%	726	624	1272	42.0%	720	666
22:20 – 22:30	1056	60.0%	690	930	1000	46.0%	564	648
22:30 – 22:40	900	70.0%	700	516	1230	50.0%	582	354
22:40 – 23:00	900	70.0%	700	516	900	40.0%	700	0
23:00 – 5:00	500	60.0%	500	0	500	40.0%	500	0

Validation Data

In this case, the field data is considered as "ground truth". The LTM results are compared with the validation data to examine the accuracy of the proposed framework. The NPMRDS are chosen as validation data, which provides 5-minute average speeds for freeway segments denoted traffic message channel (TMC). Relevant TMCs for this study are shown in Figure 8(a).

Results Analysis

Figure 10 presents the queue length comparison of LTM with NPMRDS. Grey scales are used to indicate average speeds: the darker color indicate the lower speed. Dark areas roughly indicate the position of the queue over time. This is an approximate location, considering that TMC segments may be long and that identifying the position of the queue within a segment requires further processing.

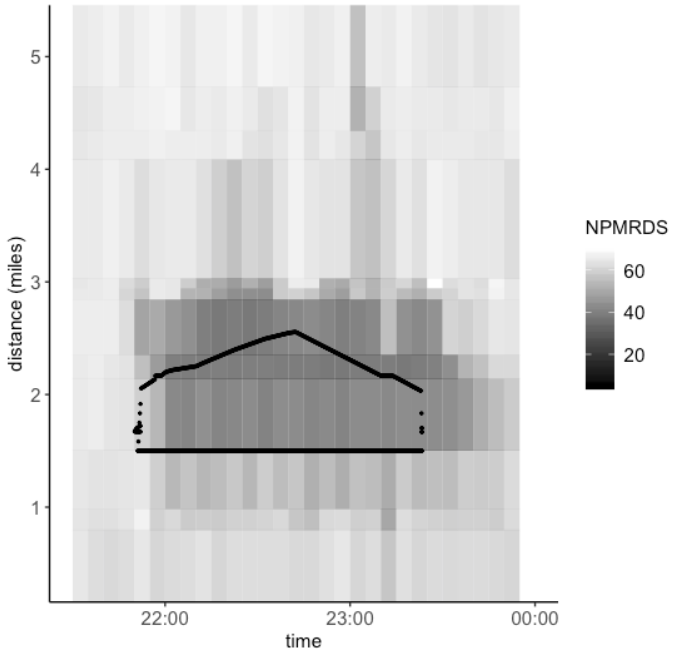


FIGURE 10. Illustration of queue propagation and dissipation process using NPMRDS and LTM

The two curves in Figure 10 represent the position of the tail of the queue over time, as estimated by the LTM model using typical day and closure day traffic volume input. Table 3 compares maximum queue length, queue duration, and estimated total traffic delay.

According to Figure 10, the closure day curve reaches its maximum length (1.1 miles) around 22:25, which coincides with the lowest speeds reported by NPMRDS. The typical day curves peaks at almost the same time, with maximum length (0.9 mi.) shorter than the closure day. This is probably a result of the increased input volume on the entry ramp due to queue jumping and suggests that preventing such behavior could mitigate the impacts of construction. For both traffic input, the queue begins to dissipate around 23:00 and fully dissipated by 23:50, when the NPMRDS speed increases to 60 mph.

According to Table 3, the total traffic delay is 505 hours when using closure day inputs, and 348 hours for typical day data, which corresponds to 10 and 7 minutes per vehicle, respectively. Based on the value of time of $17.91 per hour for passenger cars (21), the total user delay cost is $9,044 and $6,233, respectively. These values may be used to estimate the economic impact of construction activity on the road system and served as a reference for DoTs to charge to the related construction agencies.

The result suggests that LTM adequately captures the queue duration, and roughly captures the propagation and dissipation process. Although the maximum queue length of the closure day is lower than the NPMRDS, note that LTM is a first order model that do not consider the vehicle acceleration and deceleration process, which means that there might be vehicles approaching the queue with a decreased speed. As a result, the actual queue length should be longer when taking these decelerating vehicles into account. Moreover, the actual queue length is probably lower than the length indicated by the figure, when considering the possibilities that actual queue occupies part of the TMC segments and decrease the average speed of the whole segment. For these reasons, the difference between the LTM and NPMRDS are not as significant as indicated by the figure.

TABLE 3. Comparison of field data with LTM results

	NPMRDS Data	Closure Day Input	Typical Day Input
Maximum Queue Length (miles)	1.4	1.1	0.9
Queue Start and End Time	21:50 – 23:40	21:50 – 23:45	21:50- 23:25
Total Traffic Delay (hours)	Unknown	398.8	378
User's Delay Cost (dollars)	Unknown	7142.5	6770

CONCLUSIONS

This work develops and implements an LTM-based model for rapid calculation of queue length, tail position, and total traffic delay in freeway segments. One of the salient characteristics of the proposed method is its ability to easily accommodate any desired geometric configurations. It is also computationally efficient, producing results comparable to those of a microsimulator but with less set-up and run-time efforts. The approach also accounts for the impact of lane-changing in the proximity of lane closures and exit ramps through a heuristic procedure based on *HCM* recommendations.

Model results are validated using microsimulation on a hypothetical freeway segment to explore model performance under a variety of traffic flow inputs and network configurations. Results are promising, suggesting that the LTM-based framework can replicate the queue formation and dissipation patterns produced by a microsimulator with remarkable accuracy. The analyzed cases include situations with more than one queue in the considered segment. The LTM approach also captured successfully the effect of queues that form one or more ramps upstream from the lane closure. Such queues, which result from an increased utilization of exit ramps, can be inadvertently introduced by using VMS to suggest detours. The former exemplifies the value of implementing simulation-based approaches in the design and implementation of traffic management plans.

This work also discusses the use of field data, increasingly available from fixed and mobile sensors, in the development and validation of simulation models. Traffic count and speed data from work-zone trailers are combined with segment-level speed data NPMRDS to test the LTM performance during a real-world closure. A qualitative analysis of the results suggests that the proposed approach can recreate the queue duration, and reasonable estimate the formation and dissipation process and estimate the maximum queue length. Our proposed use of probe-based and sensor data to validate model results qualitatively may be extended to the analysis of other models.

Model results are sensitive to input data and model parameters. In this effort, two traffic volume inputs, closure day and typical day are presented. Aside from illustrating the sensitivity of the model to these inputs, the analysis suggests that queue-jumping prevention may improve traffic flow in the analyzed location, thus demonstrating the value of the proposed framework.

This work made relatively simple assumptions on the values of other parameters such as work-zone capacity and merge-ramp priority. These parameters have a significant impact on model performance, and understanding how to inform their estimation based on a systematic analysis of field data and model results will be the subject of further research. Developing methods to better understand the position of the queue using field data can support quantitative validation efforts, which is the subject of ongoing research by the authors. Additionally, the framework may also be extended to incorporate real-time information and predict expected queues within given planning horizons, providing a simple yet accurate modeling framework to support traffic operations in real time.

The results presented in this paper suggest that the proposed simulation framework can provide results comparable to those of microsimulation, and consistent with field observations, with only one-second running time. This characteristic makes it suitable to support traffic management decisions, including the location of variable message signs in the traffic planning stage. The data workflows may be adjusted to use the tool in real time to support quick-response decisions for unexpected variations in traffic. The relatively low effort and negligible running time make this tool appropriate for applications for which budget and time constraints do not allow for detailed modeling, but which can benefit for more precise results than those from simpler analytical methods.

ACKNOWLEDGMENT

The work presented in this paper was conducted with support from the Texas Department of Transportation. We would like to thank Brandy Savarese for her editorial assistance.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm the contribution to the paper as follows: study conception and design: Natalia Ruiz Juri; data collection: Amber Chen; microsimulation test: Kenneth Perrine; Yun Li; analysis and interpretation of results: Tengkuo Zhu, Natalia Ruiz Juri, Kenneth Perrine; draft manuscript preparation: Tengkuo Zhu, Stephen D. Boyles, Natalia Ruiz Juri. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Corthout, R., C. Tampère, L. Immers. Marginal Incident Computation - An Efficient Algorithm to Determine Congestion Spillback due to Incidents. *Transportation Research Record*. No. 2099, 2009, pp. 22-29.
2. Jiang, X., and H. Adeli. Object Oriented Model for Freeway Work Zone Capacity and Queue Delay Estimation. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 19, 2004, pp. 144-156.
3. Ghosh-Dastidar, S. and H. Adeli. Neural Network-Wavelet Microsimulation Model for Delay and Queue Length Estimation at Freeway Work Zones. *Journal of Transportation Engineering*, Vol. 132, No. 4, 2006, pp. 331-341.
4. J. Weng, Q. Meng, Estimating capacity and traffic delay in work zones: An overview, *Transportation Research Part C: Emerging Technologies*, Volume 35, 2013, pp. 34-45,
5. *Work Zone and Traffic Analysis Technical Resources*. U.S. Department of Transportation Federal Highway Administration, Work Zone Management Program. https://ops.fhwa.dot.gov/wz/traffic_analysis/techresources.htm Accessed July 10th, 2018,
6. Yperman, Isaak. The Link Transmission Model for Dynamic Network Loading. Ph.D. Thesis, 2007, Katholieke Universiteit Leuven, Belgium.
7. Lighthill, M. J., and J. B. Whitham. (1955). On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society*, London, Series A, Vol. 229, No.1178, 1955, pp.281-345.
8. Richards, P.I. Shockwaves on the Highway. *Operations Research*, Vol. 4, 1956, pp. 42-51.
9. Benekohal, R.F., Kaja-Mohideen, A.Z., Chitturi, M., 2004. Methodology for estimating operating speed and capacity in work zone. *Transportation Research Record* 1833, 103–111.
10. Daganzo, C. The Cell Transmission Model: a Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research Part B*, Vol. 28, No. 4, 1994, pp. 269–287.
11. Corthout, R., C. Tampère, L. Immers. Marginal Incident Computation - An Efficient Algorithm to Determine Congestion Spillback due to Incidents. *Transportation Research Record*. No. 2099, 2009, pp. 22-29.
12. Osorio, C., G. Flötteröd, and M. Bierlaire. Dynamic Network Loading: A Stochastic Differentiable Model that Derives Link State Distributions. *Transportation Research Part B: Methodological*, Vol. 45, No. 9, 2011, pp. 1410-1423.
13. Ruiz-Juri, N., A. Unnikrishnan, and S. T. Waller. Integrated Traffic Simulation-Statistical Analysis Framework for Online Prediction of Freeway Travel Time. *Transportation Research Record*, No. 2039, 2007, pp 24-31
14. Xia, J., M. Chen, and Z. Qian. Predicting Freeway Travel Time under Incident Conditions. *Transportation Research Record.* No. 2178, 2010 pp. 58-66.
15. Liu, H. X., X. Wu, W. Ma, and H. Hu. Real-time Queue Length Estimation for Congested Signalized Intersections. *Transportation Research C,* Vol. 17, No. 4, 2009, pp. 412–427.

16. Comert, G., and M. Cetin. Queue Length Estimation from Probe Vehicle Location and the Impacts of Sample Size. *European Journal of Operational Research*, Vol. 197, No. 1, 2009, pp. 196-202.
17. Transportation Research Board. "Uninterrupted Flow." Highway Capacity Manual, 2016.
18. Newell, G. F. A Simplified Theory of Kinematic Waves in Highway Traffic, Part I: General Theory. *Transportation Research Part B: Methodological*, Vol. 27, No. 4, 1993, pp. 281-287.
19. Tampère, C., Corthout, R., Cattrysse, D., Immers H., Lambertus. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. Transportation Research Part B: Methodological. 45. 289-309. 10.1016/j.trb.2010.06.004.
20. *National Performance Management Research Data Set*. https://npmrds.ritis.org/analytics/ Accessed July 10th, 2018,
21. Ellis, D. (2017). Value of Delay Time for Use in Mobility Monitoring Efforts. Texas A&M Transportation Institute.