Technical Memorandum

Estimating the Cost of Lane Closures Due to Incidents and Planned Work Zones Literature Review

Prepared for
TXDOT Austin IAC

Prepared by
The Center for Transportation Research, University of Texas Austin

DRAFT

Introduction

Work zone road user cost (WZRUC) is the additional cost borne by motorists and local communities due to work zone activity on roadways (Mallela and Sadavisam 2011). A comprehensive evaluation of WRZUC considers user delay costs, vehicle operating costs (VOC), crash costs, and emission costs, as well as harder-to-measure effects like local community inconvenience and noise (Mallela and Sadavisam 2011).

The first section of this report, focused on WZRUC, outlines the factors influencing user delay costs, VOC, crash costs, and emission costs. Then, focusing on the measurable aspects of WZRUC—particularly user delay costs—we outline the types of mobility impact analyses and studies that describe cost estimation practices in Texas, Illinois, and Ohio. Available software tools used to estimate work zone road user cost are summarized. Finally we review recent research on work zone capacity and delay estimation.

As with WZRUC, incident-induced delay (IID) also costs users. Incidents are defined as “any event that temporarily reduces roadway capacity” (Jiang 2001). According to the 2010 *Traffic Incident Management Handbook* (IMH), 50 to 60 percent of total roadway delay in metropolitan areas is caused by incidents. Eighty percent of incidents include vehicle disablements and 10 percent are due to accidents. Of the total number of incidents, most occur on the roadway shoulder, and approximately 30 percent block at least one lane. Moreover, blocked lanes due to accidents typically cause forty-five to ninety-minute delays, which result in 1,200 to 5,000 vehicle loss hours (VLH) (Jenkins and McAvoy 2015). Section 2 outlines recent IID research in greater depth.

1. Work Zone Road User Cost Estimation

User delay costs are often the primary metric considered by planners and managers when investigating work zone related costs, given the availability of data and methodologies to support their calculation. Meng and Weng (2013) define a work zone as “a segment of road in which maintenance or construction operations impinge on one or more lanes available to traffic, or affect the operational characteristics of traffic flow through the segment.” A work zone is comprised of an “advance warning area,” a “transition area,” an “activity area,” and a “termination area” (Weng and Meng 2013).

There are four methodologies used to estimate user delay costs, as well as work zone capacity, which is often a required input in the delay estimation process.

Emission costs assess the additional vehicle emissions resulting from reduced speeds and queueing, as well as air pollutant emissions and greenhouse gases. Vehicle emissions are influenced by roadway characteristics, traffic characteristics (e.g., vehicle composition) as well as driver, vehicle, and weather characteristics. Emission models are categorized into static emission factor models or dynamic instantaneous models. Static emission factor models calculate emissions based on average operating conditions but do not capture precise driving conditions such as acceleration and idling. Dynamic instantaneous emission models are capable of capturing these detailed conditions but require extensive data and computation. Assigning a monetary value to emissions informs the economic analysis of health effects caused by pollutants and greenhouse gases but there has been no consensus on how to compute emission costs (Mallela and Sadavisam 2011).

Crash costs are evaluated from the expected change in crash rates due to the work zones. This aspect of WRZUC considers crash rate at the work zones, crash severity rating, and the unit cost of crashes, along with either human capital costs or comprehensive costs. Human capital costs include the monetary value directly associated with the crash such as property damage, medical expenses, and others. Comprehensive costs include tangible, non-monetary losses in addition to human capital costs such as diminished quality of life (Mallela and Sadavisam 2011).

Vehicle Operating Cost (VOC), defined as the operating costs that vary with the degree of vehicle use, are time dependent so that extra time in the roadway due to work zones incurs higher VOC for road users. Three models are often implemented to estimate VOC: a) the National Cooperative Highway Research Program (NCHRP) Report 133; b) the Texas Research and Development Foundation (TRDF) method; and, c) the HERST-ST model. The HERST-ST method provides a more modern framework in which to compute VOC from vehicle types, roadway conditions, and traffic characteristics from constant speed operating conditions. It also calculated excess resource consumption due to speed-change cycles and roadway geometry (Mallela and Sadavisam 2011).

User delay costs include passenger car travel time delays, truck travel time delays, , and freight inventory delay. User delay costs distinguish between the dollar estimates of time for personal local, personal intercity, business local, and business intercity travel and convert delay time into a dollar value. Work zone capacity and queue formation estimation methods—essentially supply and demand analyses—estimate extra travel time due to work zones and the dollar value associated with the delays (Mallela and Sadavisam 2011). Critical datasets used in the process include hourly traffic demand, traffic composition, travel speed, work zone configuration, and maintenance of traffic (MOT) strategy (Mallela and Sadavisam 2011). The delays are converted into monetary values using a value of time (VOT) factor. The VOT recommended by TxDOT for the year 2017 is $22.4/hr for passenger cars and $32.7/hr for trucks (TxDOT 2017).

1.1 User Delay Estimation

User delay estimation methodologies typically rely on the computation of a representative value for the work zone capacity. Determinants of work zone capacity include work zone configuration, roadway conditions, work activity characteristics, weather, proportion of heavy vehicles, and driver composition (Weng and Meng 2013). User delay computation can be conducted using analytical models based on deterministic queuing theory or by using simulation. The following sections outline typical approaches for work zone capacity estimation and user delay calculation.

An additional aspect of the estimation of work zone related delays is the modeling of network diversion effects. Quan and Zhang (2012) compare full and partial closure on Interstate 5 in Sacramento, California, and argue that the optimal of number of lanes to close depends on demand reduction due to closure and the availability of diversion routes. Zheng et al. (2014) optimize the schedule for work zones by reducing network delay as an optimization objective. They model network delay by rerouting travelers on k-shortest paths between an origin-destination pair when a closure is scheduled on one of the paths.

Simulation-based models, like dynamic traffic assignment, also exist to capture network diversion; however, such models are difficult to calibrate and validate.

1.1.1 Work Zone Capacity Estimation

Work zone capacity is a critical component of the user delay estimation, since most methodologies use a supply/demand type of analysis to determine queues and corresponding delays.

In parametric work zone capacity estimation approaches, predictors take an established form and coefficients are estimated through field data. Non-parametric approaches, on the other hand, do not assume the structure of a model as fixed because nonlinear relationships and high-order interactions may affect capacity. Although parametric methods may have lower accuracy, non-parametric models grow in size with more complex data, leading to more complicated models. Simulation methods allow for detailed modeling of work zone conditions (varying distance in warning signs and different lane distribution, for example), but due to the high levels of computational resources as well the extensive calibration requirement, simulation approaches be inefficient when compared to parametric and non-parametric methods (Weng and Meng 2013).

In other recent studies evaluating work zone capacity estimation, capacity is modeled as a stochastic variable with a derived distribution. Weng and Yan (2015) use a right-truncated lognormal distribution to model work zone capacity using maximum likelihood estimation optimized through an iterative Newton-Raphson algorithm. The distribution parameters depend on work zone traffic and other characteristics like urban vs. rural road type, “number of closed lanes, number of opened lanes, lane closure location, work duration, work intensity, work time, heavy-vehicle percentage, and capacity measurement method.” The capacity measurement method is a categorical variable differentiating between definitions of work zone capacity: as hourly traffic, mean queue-discharge rate, and full hour average volumes counted at lane closures with upstream-queue. The parameters are calibrated and tested by aggregating data from eighteen previous work zone publications (a total of 237 datasets).

The authors argue that for work zone capacity to be approximated by a probability distribution instead of a single value because, according to Edara et al. (2012), “the measured work zone capacity may not be close to the true value.” Approximating work zone capacity by a probability distribution enables prediction of travel time reliability in work zones. Findings conclude that, as expected, urban roads are strongly associated with higher mean work zone capacity as is longer-term work zone duration, on average. Additionally, mean work zone capacity is negatively associated with heavy-vehicle proportion so that a 1 percent increase in heavy vehicles through a work zone results in a mean work zone capacity decrease of 0.281 percent, on average. Other significant findings describe the effect of capacity measurement method (on average, queue-discharge method significantly produces lower capacity estimations) and the significant difference in lane closure location variation, with the log work zone capacity standard deviation being higher than left lane closures. Furthermore, various distributions are compared so as to validate the choice of a lognormal probability distribution. Lastly, Weng and Yan (2015 and 2011)contrasted methods of accuracy evaluation, determining the former as a better fit due to the consideration of increased work zone capacity determinants, such as work zone activity intensity. Similarly, Calvert et al. (2016) model roadway capacity as a Weibull distribution and calibrate it using field data; however, their model applies to a general freeway and does not consider special characteristics of a work zone.

1.1.2 User Delay Estimation

A macroscopic analysis of work zone **traffic delay** typically entails deterministic queueing methods because vehicle queues form upstream of a work zone when traffic flow exceeds work zone capacity. Macroscopic models commonly require demand volume, freeway capacity, work zone capacity, and work zone duration as inputs. Although widely implemented, **deterministic queue methods** often do not account for deceleration and acceleration before and after the work zone, which would result in an underestimation of delay. Jiang (2001) proposes using mechanics equations to account for additional delay due to acceleration and deceleration in a work zone, and combines it with deterministic queueing theory to estimate queue lengths. The proposed methodology does not take into account road characteristics such as curvature. Furthermore, the equations require proper estimation of work zone capacity, traffic flow rate, and speed to compute traffic delays.

Simulation models may also be used to estimate delays due to work zones. Weng and Meng (2013) distinguish between macroscopic analytical, macroscopic simulation, and microscopic simulation methods for estimating traffic delay due to work zones. Existing macroscopic simulation methods include QUEWZ-92 and QuickZone, while microscopic simulation software includes CORSIM, PARAMICS, and VISSIM. Because microscopic simulation tools model individual vehicle behavior, microscopic simulation may estimate traffic delays with greater accuracy than macroscopic analysis or simulation. Nonetheless, like simulation methods for capacity estimation, microscopic simulation is computationally intricate and time consuming. Mesoscopic simulation provides a solution that addresses some of the limitations of macroscopic models without excessive computational/modeling burden (Daganzo 1994 and Yperman 2007). Mesoscopic models do not consider individual vehicles, but use small time and space discretization to better understand describe the queue formation and dissipation process.

1.2 Software tools

There are several macroscopic simulation tools available for estimating traffic delay in work zones. QUEWZ-92, Highway Capacity Software (HCS+), and QuickZone are the most widely used. Microscopic traffic simulation tools have also been used by researchers to estimate traffic delay in work zones. Among the most widely used microsimulation software are CORSIM and VISSIM.

1.2.1 QUEWZ-92

QUEWZ-92 was developed by the Texas A&M Transportation Institute in the early 1990s to evaluate the impacts of freeway work zone lane closures. It is a computerized version of commonly used manual techniques for estimating queue lengths and user costs resulting from work zone lane closures. It simulates traffic flows through freeway segments both with and without a work zone lane closure in place, and it estimates the changes in traffic flow characteristics and additional road user costs resulting from a lane closure. The road user cost that can be calculated by the QUEWZ-92 software package includes vehicle operating costs and travel time costs. QUEWZ-92 can also identify time schedules without causing excessive queuing under a given number of closed lanes.

QUEWZ-92 requires four types of input data: lane closure configuration, schedule of work activity, traffic volumes approaching the freeway segment, and alternative values to the default model constants. In this software, several default values are used unless the specific values are determined by the user.

QUEWZ-92 requires AADT for traffic flow analysis. It can estimate the directional hourly volume from the AADT using the two sets of adjustment factors (one for urban and the other for rural), which are based on data collected in rural and urban interstates in Texas in 1985. Therefore, the estimated directional hourly volumes from the AADT reflect only the average distribution in Texas based upon 30+year old information.

The cost update factor adjusts the road user costs for the effect of inflation. All the costs computed in QUEWZ-92 are expressed in 1990 dollars. QUEWZ-92 estimates the capacity of the work zone during work activity based on HCM equation.

The output from QUEWZ-92 includes input data summary, summary of user costs, summary of traffic conditions, and the summary of traffic volumes. The lane closure schedule indicates the hours under certain lane closure conditions in which work activity can continue without resulting in excessive queues.

1.2.2 HCS

The McTrans Center at the University of Florida developed the HCS+ Release 5 of the Highway Capacity Software for Windows 98/Me/NT/2000/XP. This version of the HCS implements the procedures defined in the *HCM 2000*. While HCS+ does not provide specific functions to estimate work zone capacity, it estimates the capacity of a general freeway segment with the considerations of effects of lane width, lateral obstructions, and interchanges.

Input requirements include general freeway characteristics and traffic characteristics. Generally, HCS+ can provide a level-of-service (LOS) measure and service volume estimation based on different traffic condition characteristics.

1.2.3 MicroBENCOST

MicroBENCOST was developed in the early 1990s through the National Cooperative Highway Research Program as a comprehensive and convenient framework for doing highway user benefit-cost analysis on personal computers. It allows up to a total of three work zones on any route in order to evaluate the impact of lane closings and reduced capacity at the route segment.

1.2.4 QuizkZone

QuickZone is a software that estimates user delays in work zones developed by the Office of Research, Development, and Technology in the FHWA and Mitretek Systems. This software package can evaluate the traffic impacts for work zone mitigation strategies and estimates the costs, traffic delays, and potential backups associated with these impacts.

The software provides information in spreadsheet form and can accommodate networks with up to 100 nodes and 200 links. The user would enter data on the planned work zone such as location, projected detour routes, anticipated traffic volumes, and construction dates and times. The program then displays the amount of delay in vehicle hours, the maximum length of the project traffic queue, and the costs associated with the work zone activity.

While the software is easy to use, it does require a significant amount of data to accurately model the roadway system. As QuickZone is a network flow model that analyzes individual segments in specified time steps, the software relies on a network composed of links and nodes. Hourly and daily demand patterns are needed to populate the links with traffic flow. If hourly volumes are not available, QuickZone can estimate these values using *HCM* procedures. QuickZone is very flexible in terms of input parameters and the open-source software product allows users to modify and customize to fit local conditions as desired.

1.2.5 VISSIM

VISSIM is a microscopic, behavior-based, multi-modal traffic simulation model developed by PTV. It can model traffic and public transport operations in a network integrating street and freeway systems and it can analyze traffic operations under traffic constraints.

One of the unique features of VISSIM is the Vehicle Actuated Programming (VAP) Language Module. It allows the user to externally control vehicle detection, traffic control, and driver behavior logic. The VAP Language Module makes it possible to model and evaluate the effectiveness of various ITS applications, such as variable message signs, traffic diversion, and various work zone traffic control strategies.

With its Dynamic Assignment model, VISSIM can also answer route choice dependent questions such as the impacts of work zone or incident related congestion. This feature may be useful to study the regional impact of a single or multiple road construction projects on a complex freeway arterial network. VISSIM has an optional vehicle emission module that can be used to estimate vehicle emissions produced as a function of the vehicle's operating mode. The model can predict second-by-second emissions for HC, CO, NOx, and CO2 as well as fuel consumption for a range of vehicle categories. Thus, VISSIM can be used to estimate the environmental impact of work zones under various traffic conditions.

VISSIM requires very detailed information on network geometry, traffic demand, vehicle composition, and traffic control. Setup of the simulation network and calibration of the model is a fairly complex task that requires personnel with significant experience of the system. Therefore, this model may be more applicable as a planning tool to estimate the regional impact of some very complex road construction projects on a large freeway-arterial network.

Also, there are several features of VISSIM that limit its ability to faithfully and easily model traffic operations through work zones. For example, it is possible to define different lane widths for each lane within a link, but the narrower lanes do not automatically influence vehicle free-flow speeds. Therefore they do not have an effect on lane capacity.

1.2.6 CA4PRS

One of the more recent innovations developed in California is the Construction Analysis for Pavement Rehabilitation Strategies (CA4PRS) software. This is a schedule and traffic analysis decision-support tool used to help transportation agencies select more efficient and economical rehabilitation strategies. CA4PRS is beneficial when utilized during the planning and design stages to evaluate different scenarios to balance construction scheduling, traffic delay costs, and project costs. While the software was not available for extensive evaluation in this research project, the FHWA endorsed the product in 2008 for nationwide deployment. The most appropriate use of this product, which was developed specifically for rehabilitation projects, begins during the initial stages of project development. CA4PRS can aid agencies in estimating working days and CPM schedules, developing construction staging plans, traffic control plan design, and identifying incentive and cost (A) + time (B) contracts.

1.3 State Practices

We reviewed three papers with recommendations on WZRUC estimation by independent researchers for the Texas Department of Transportation, the Illinois Department of Transportation, and the Ohio Department of Transportation. The Texas and Illinois researchers collected field data and assessed varying models, while the Ohio report outlines an updated methodology. Collectively, the studies present a survey of the broad methodologies implemented for WZRUC assessments.

1.3.1 Texas

A report by Borchardt et al. (2009) provides revised guidelines on expected capacities of freeway work zones and compares different models for assessing traffic conditions and road user costs for the Texas Department of Transportation (TxDOT). Borchardt includes a survey of current practices in Texas districts, work zone field data gathering methods for Texas highways, an overview of various mobility assessment models, and a comparison of road user cost estimations by three different models.

The survey of current practices asked TxDOT districts about the guidelines and software tools implemented for work zone road user cost estimation as well as key determinants of work zone capacity. The thirteen (out of twenty-five) surveys returned revealed that in 2007, nine of the thirteen responding TxDOT districts applied previous field experience and engineering judgement to estimate work zone capacity, three used *HCM* guidelines, and two did not implement any guidelines for work zone capacity estimation. Ten of districts that responded did not utilize any software to estimate work zone capacity and others used QUEWZ, FREW, or microscopic traffic simulation. The three main factors considered crucial to estimate work zone capacity include roadway geometry, lane width and lane treatment (open vs. closed) (Borchardt et. al 2009).

The study collected data on lane closures on freeways in the Houston and San Antonio regions. When available, data from ITS devices and radar from Houston's TransStar system were used to calculate the capacity of each work zone. If not applicable, portable devices or manual counts served as data collection strategies. For each study site, the maximum observed volume throughput of the work zone served to estimate work zone capacity. Data considered valid for capacity estimation was collected in the presence of an observed vehicle queue either within or on the approach to the work zone. Moreover, traffic demand data was collected upstream in selected zones. However, as pointed out in Benekohal et al. (2004), both maximum observed volume approach and queue discharge flow approach do not accurately estimate capacity due to the presence of occasional large headways between vehicles. This alludes to other approaches to measuring work zone capacity which are discussed in the previous section.

Borchardt's overview of multiple models to analyze traffic operations distinguished "straight-line" vs. "network" models. The linear models evaluated include QUEWZ-92, Highway Capacity Software (HCS+), MicroBENCOST, and Kim's (2001) multiple regression model, in addition to QuickZone and

VISSIM network models. Network models are more detailed and computationally complex than linear models. In evaluating the linear models, the models' capacity estimation was compared to the observed maximum volume from the work zone studies (Borchardt et. al 2009). It is difficult to assess the validity of this comparison as there exists drawbacks to field data collection methods (Benekohal et al. 2004), including measurement error.

The road user cost analysis contrasts cost estimates from three models (VISSIM, Jiang (1999), and QUEWZ-92) for four field sites in San Antonio. The estimate from the VISSIM simulation serves as a baseline metric for assessing the other two models. To obtain cost in dollars, hour delay is multiplied by $20.09 per vehicle hour for passenger cars and $36.71 per vehicle hour for trucks in 2007 dollars. Jiang's (1999) estimate approximates VISSIM's closely, while QUEWZ-92 underperforms by almost half. The authors note that the RMS error from Jiang (1999) is less than QUEWZ-92, implying a better model fit.

1.3.2 Illinois

As in Texas, Benekohal et al. (2010) estimate work zone capacity using field data collection results for Illinois Department of Transportation. The authors collected data, which included traffic and work zone characteristics, by analyzing video of the traffic stream obtained from a roadside video camera. Using the upper and lower bounds on capacity from four different methods of capacity estimation, the authors propose a platooning-factor based method to estimate the work zone capacity. The reported capacity values range between 1,200 pcphpl to 1,750 pcphpl based on traffic conditions, time of day, proportion of heavy vehicles, and average speed with or without platoon. The authors calibrated the speed flow curve diagram for each case using the collected field data (Benekohal et al. 2010).

1.3.3 Ohio

Ohio Department of Transportation's (ODOT) queue analysis includes traffic volume data, pre-lane closure conditions, lane closure conditions, and observed queue data as input information, then performs a supply and demand macroscopic analysis to output total delay in hours. It converts hours to dollar amounts using the Consumer Price Index annual average values from the Bureau of Labor Statistics. The CPI annual average value varies by the following regions: "US Urban City Average," "Midwest Urban Average," "Cleveland-Akron," and "Cincinnati-Hamilton." Traffic volume data is composed of hourly or average annual daily traffic (AADT) data and a heavy vehicle adjustment factor is applied. Traffic volume calculations also include a diversion threshold over which a percentage of vehicles are expected to divert to other routes. Volumes over the threshold are not computed as part of the queue length or average hour delay calculation. Pre-lane closure conditions include roadway geometry, number of lanes per direction, terrain type, and free flow speed. Lane closure conditions are the conditions within the work zone, such as number of open and/or closed lanes, work intensity and capacity (from *HCM* 2010), and other considerations like lighting. The supply and demand analysis determines queue length as a function of average minimum headway assumed under jammed conditions, which the authors recognize as an underestimation. From there, the analysis differentiates four traffic conditions: 1) no lane closure and no diversion; 2) no lane closure with diversion; 3) lane closure and no diversion; and, 4) lane closure with diversion. To determine queue length, the number of passenger cars in a queue is calculated as the difference between the cumulative number of cars entering the work zone and the cumulative number exiting per hour. The queue length per hour (Qt) is determined as:

$$Qt = hNQ5280N \qquad (1)$$

where N_Q is the difference between passenger cars entering and exiting the work zone, *h* stands as the average minimum headway of passenger cars in a queue, and *N* is the number of open lanes upstream of the work zone. Total delay is calculated from queue length by summing up Qt for all *t* and subdivided into passenger cars vs. trucks and buses based on the average proportion of trucks and buses. To compute road user cost in dollars, average delay (calculated as the total delay divided by the number of vehicles travelling through the work zone) and the CPI annual average value by region are used (Jenkins and McAvoy 2015). Two drawbacks of ODOT's approach include ignoring the cost to diverting traffic and assuming queue analysis under only jammed conditions.

2. Incident-induced Delays

Habtemichael et al. (2015) describe three methodological strategies to estimate incident-induced delay (IID): capacity reduction estimation for deterministic queue analysis, microscopic simulation, and statistical methods. Capacity is the "maximum hourly rate at which persons or vehicles can reasonably be expected to traverse a point or uniform section of a lane or roadway during a given time period" (Qin and Smith 2001). Table I summarizes the percentage of freeway capacity available as a result of incident characteristics according to the *Highway Capacity Manual* 2000 (*HCM* 2000).

Table I: Freeway capacity available under incident conditions (percent)

Number of freeway lanes in each direction	Shoulder disablement	Shoulder accident	Lanes blocked		
			One	Two	Three
2	0.95	0.81	0.35	0	N/A
3	0.99	0.83	0.49	0.17	0.00
4	0.99	0.85	0.58	0.25	0.13
5	0.99	0.87	0.65	0.40	0.20
6	0.99	0.89	0.71	0.50	0.25
7	0.99	0.91	0.75	0.67	0.36
8	0.99	0.93	0.78	0.63	0.41

From *Highway Capacity Manual* 2000.

Qin and Smith (2001) as well as Knoop et. al (2008) implement modern methods to calculate capacity reduction due to incidents. Nevertheless, Hadi et. al (2012) points to both of these methods (Qin and Smith; Knoop et.al) as being limited in scope because they fail to account for non-linear traffic demand. On the other hand, microscopic simulation relies on theoretical choice models to duplicate the behavior of individual vehicles (Habtemichael et. al 2015). It is capable of simulating a vehicle's speed, acceleration, as well as position (*HCM* 200), and consequently incident delay. As such, microscopic simulation is suited for "complex traffic conditions" (Zhang 2012). However, microscopic simulation requires time-consuming, precise, and accurate calibration under incident and incident-free conditions for both peak and off-peak hours (Habtemichael et. al 2015), making it a costly approach. Zhang et. al (2012) and Hadi et. al (2007) delineate procedures for utilizing and evaluating microscopic simulation. Finally, the statistical methods presented leverage high quality data to estimate IID and differentiate between recurrent and non-recurrent congestion, and can directly estimate vehicle loss hours (VLH). Chung (2011), Snelder et. al (2013), Habtemichael et. al (2015), and Adler et. al (2013) have contributed recent and innovative research in the incident delay literature.

2.2 Capacity Estimation

To estimate capacity, data on traffic volume (vehicles per hour per lane), speed (mph), and density (vehicles per mile per lane) are essential. Accident characteristics such as number of lanes blocked, duration, and time of day also affect accident capacity estimation. To calculate capacity under prevailing conditions, Qin and Smith (2001) used the speed-plot approach developed by Venkatanarayayana (2001), which is a time-invariant measure of capacity for a roadway segment. The authors measured accident capacity in the field as the fifteen-minute over-saturated flow at the upstream of the bottleneck created by the accident, and calculated the difference between prevailing capacity and accident capacity to deduce reduced capacity. The authors argue the incident capacity is best modeled as a Beta distribution random variable due to inherent variance in incident conditions, with distribution parameters depending on the number of available and blocked lanes (Qin and Smith 2001). As an input into queuing analysis, reduction in capacity estimation determines maximum queue length, average queue length, maximum and average individual delay, and total delay of an accident (Qin and Smith 2001). According to Qin and Smith (2001), reduction in capacity (measured in VPHPL) is the difference in capacity under prevailing (i.e., typical) conditions and accident capacity.

Knoop et. al (2008) implement a different approach to estimate reduced capacity. The authors utilize high quality videos of traffic flow around two accidents that resulted in blocked lanes recorded from a helicopter in The Netherlands. The first incident involved a car rolling over a median resulting in one eastbound lane closed for emergency vehicles. The second incident, involving multiple trucks and passenger cars, blocked one lane and the shoulder in one direction. From the videos, vehicle travel times were calculated and aggregated to flow-rates over thirty-second intervals. To estimate capacity, the maximum flow-rate values when driving out of congestion were used. The authors found the capacity of the lanes of the opposite direction lowered by fifty percent. For lanes in the direction of the incident, capacity was reduced by more than half, aligning with previous capacity measurements. Although limited to two incidents, the study provides evidence of delay caused by incident-induced rubbernecking. Because rubbernecking contributes to total incident delay, evaluation of such effects in IID should be considered.

The capacity approach is limited in one key aspect. Chung (2011) states, “because these methods have been focused mainly on undersaturated conditions rather than on over-saturated or congested conditions, they cannot capture the additional congestion impact caused by incidents that occur under congested conditions.” In other words, a time-invariant value for prevailing capacity fails to account for variation in demand, such as in peak and off-peak hours. For example, Snelder et. al (2013) found that incidents which “occur in the hours before the start of the morning and evening peaks have the greatest [delay] impact,” providing evidence towards Chung’s critique.

2.3 Microsimulation modeling

Zhang et. al (2012) developed a three-step procedure to calibrate and validate microscopic simulation when applied to incident delay. The procedure is composed of the following: 1) the establishment of a simulation platform and parameter calibration; 2) the enactment of traffic data for model calibration and validation; and, 3) the evaluation of the model through a case study of six incidents along the Ning-Tong Freeway in China.

The first step in the authors’ methodology determines the calibration parameters, the performance evaluation metrics and goals, and the sample size of the simulation. The parameters considered in the study encompass traffic composition (e.g., number of trucks vs. passenger cars) and driver behavior (e.g., acceleration times, deceleration times, headway, and lane-changing preferences). Performance evaluation metrics include speed, density, traffic volume, and travel time, which the authors do not allow to deviate by more than five percent from their true value. The second step calibrates the simulation using geometric data, such as number of lanes and lane width, as well as traffic volume, truck traffic composition percentage, average speed, and incident duration information. Delay estimates are then validated by comparing to actual incident delay data.

When comparing the simulated incident delays to the actual delay values for six incidents in China, the simulation scheme underperforms. Two-thirds of the simulated delays have an absolute error greater than 20 percent, with an incident in December 2010 having a relative error of approximately 60 percent (Zhang et al 2012). Table II summarizes the incident simulation delay compared to the true delay value. Zhang et al. (2012) present several explanations, including that incident characteristics (such as number of blocked lanes) also affect incident delay.

Table II: Incident delay simulation results

Incident time	Simulation delay (vh)	True delay value (vh)	Relative error
2010-2-13	83370	81973	2%
2010-2-28	19468	26126	-25%
2010-4-20	37847	30775	23%
2010-10-24	212828	172940	23%
2010-10-27	42596	40761	5%
2010-12-30	26175	63490	-59%

From Zhang et. al (2012)

The authors' methodology t is resource intensive and requires careful calibration with a large amount of roadway, traffic, and incident duration information. Nonetheless, Zhang et. al (2012) fail to account for characteristics such as incident severity, resulting in their case study results missing critical incident attributes.

Hadi et. al (2007) examine the impact of incident characteristics on simulation results by adjusting microscopic modeling parameters (lane speed, blocked lane location, traffic demand, and rubbernecking effects) to attain *HCM* 2000 "target-link capacity values for non-incident and incident conditions." In order to attain the targeted outputs, three software tools—CORSIM, AIMSUM, and VISSIM—are evaluated by varying calibration parameters. The authors found CORSIM to best achieve the capacity reduction estimates established by *HCM* 2000 due to the inclusion of rubbernecking effects, which their study recommends for inclusion in any simulation. Knoop et. al's (2008) research provides strong evidence in support of Hadi et. al's findings.

2.4 Statistical Methodologies

In order to account for non-linear demand, Chung (2011) differentiates between recurrent and non-recurrent congestion: recurrent congestion is "the predictable delay caused by the high traffic demand"; non-recurrent congestion is "the unpredictable delay caused by incidents." Other studies implement a "reference day" to approximate congestion under conditions in which the incident did not occur. The reference day approach corrects for recurrent congestion during an incident to estimate its consequent delay.

Snelder et. al (2013) use traffic and incident information from loop detector data and incident databases in The Netherlands to measure incident delay. Their approach measures total incident congestion and differentiates between recurrent and non-recurrent congestion by constructing a weighted median reference day from similar weekdays within a four-week period. The authors implemented a six-step procedure to estimate incidents delay in vehicle loss hours. The first step correlates incident start time and location to traffic counts, speeds, and lane closure status from the loop detector data so to filter and merge relevant data. Then, the congestion caused by the incident (through traffic flow, speed, and lane closure information) is tracked upstream over the network until the congestion effects are no longer detectable.

A parallelogram method, outlined by Olmstead (1999), delineates congestion in vehicle hours through place and time. Actual vehicle hours within the parallelogram are compared to vehicle hours at a hypothetical speed of 100 km/h. This measure estimates vehicle loss hours due to an incident upstream of its location. Spillback effects and rubbernecking effects can be computed using the same method throughout the network and in the opposite direction, respectively.

In the next step, vehicle loss hours for a reference day are computed. A reference day measure is constructed by selecting data from four weeks before and four weeks after the incident for the same day of the week and as closely matched to month and weather characteristics. The authors weigh the data so that data for one week before or after the incident receives a weight of four, two weeks before or after receives a weight of three, three weeks before or after receives a weight of two, and four weeks before or after receives a weight of one. Vehicle hours for the same place/time area as the incident are compared to vehicle hours at a speed of 100 km/h, and the median VLH is taken as the reference. Lastly, to estimate extra travel time, the summation of the VLH due to upstream, spillback, and rubbernecking effects is taken minus the reference VLH.

Among the authors' several findings for their most detailed dataset, they conclude the following: a) upstream effects constitute, on average, 70 percent of total delay caused by all incidents; b) spillback and rubbernecking effects characterize 13 and 17 percent, on average, respectively of total delay; c) incidents occurring within an hour before the morning and evening peaks, on average, have the greatest delay impacts; and, d) on average, Tuesdays, Thursdays, and Fridays constitute the majority of total extra travel time due to incidents. The authors recommend that the method be extended to consider re-routing delays, downstream gains, as well as demand effects. Nevertheless, Habtemichael et. al (2015) assert that using a "day-of-the-week-based" method to construct a reference day "provides a very poor match of traffic patterns" due to unaccounted for within-day traffic pattern variation. Instead, Habtemichael et. al (2015) recommend a K-nearest neighbor method to choose comparable traffic behavior based on volume and travel time patterns.

Adler et. al (2013) evaluate the impact of incident duration on non-recurrent congestion (NRC), implementing the approach presented by Snelder et. al (2013), but widen the reference day window to eight weeks instead of four. Incident duration is the time period necessary for incident-associated traffic measures to be lifted. To examine the effect of duration on non-recurrent congestion, the study uses a log-log regression specification, controlling for recurrent congestion and other factors such as incident severity, car type involved, and weather. Additionally, the regression includes fixed effects for hour of the day, weekend, month, year, and location. The authors differentiate between accidents, which involve incidents with damage or injuries, and non-accidents. They allow regression parameters to differ for each category. The study results showcase that incident duration has a "strong positive, but concave" relationship with non-recurrent congestion, suggesting that lowering incident duration will, on average, reduce NRC. More specifically, a one-minute increase is associated with an increase in non-recurrent congestion of 3.95 VLH, on average, for accidents in their dataset. Assuming a €20 per VLH (the authors state this value is in line with travel time cost literature), a 3.95 VLH average for accidents results in an incident-induced delay cost of €79 (Adler et al 2013). A similar procedure can be extended to evaluate the effect of the number of blocked lanes due to an incident, rather than incident duration, on non-recurrent congestion, thus providing a robust and applicable methodology that leverages novel data collection technologies.

3. Concluding Remarks

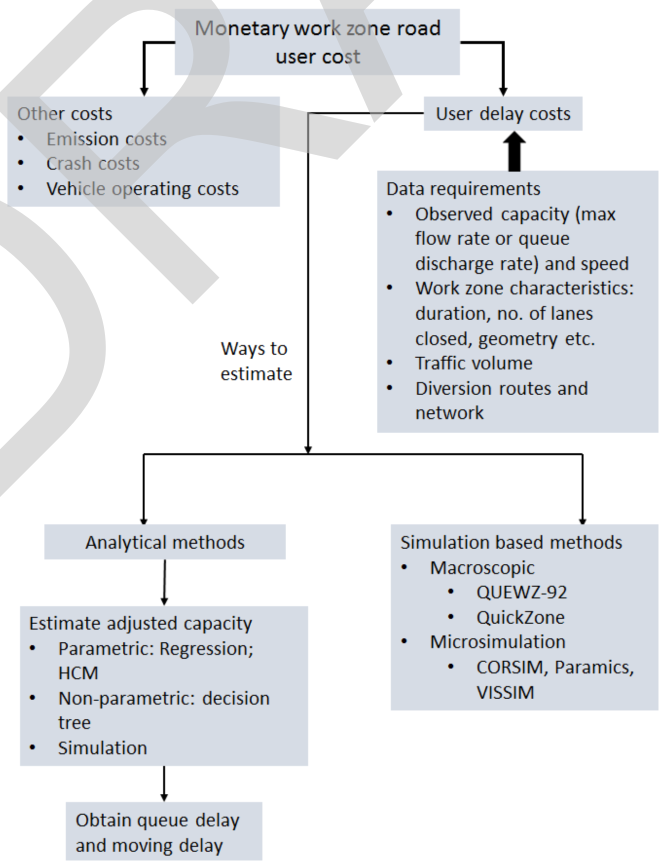
This literature review is an outline of three distinct methodologies used to estimate incident delay: capacity models, microscopic simulation, and empirical strategies. Capacity models account for the reduction in roadway capacity due to blocked lanes. Although established, capacity approaches suffer from an inability to distinguish between non-recurrent and recurrent congestion at the incident site. As such, capacity estimation procedures do not reflect true incident-induced delays.

Microscopic simulation handles complex scenarios, and has the potential to compute incident delays appropriately. However, microscopic analysis is time-consuming and cost-ineffective due to the large amount of detailed data required to calibrate and validate the software for various scenarios.

With novel data collection capabilities, statistical methods provide greater precision and cost-efficiency than the other two techniques. Chung (2011) writes, "the empirical methods require only the fundamental data such as accident time, location, speed, and volume data," so that IID estimation can be applied "as long as such data are available." By exploiting new technology (e.g., sensors and video recording), researchers can reliably estimate incident-induced delays.

The state of the practice for estimating the user delay cost component of WZRUC can be summarized as shown in Figure 1. Common practice includes macroscopic analytical methods with parametric estimation of adjusted capacity where analytical equations are used to determine queue length and total delay. Jenkins (2015) and Benekohal et al. (2010) present examples of this method. Macroscopic analytical models are simple and easily transferable to different locations. However, due to the lack of complex geometry considerations and interactions of the entire network, the results are not considered to be very accurate. Simulation models, on the other hand, represent real field conditions accurately, making sensitivity analyses easier to conduct. However, they require building and calibrating a complex model for specific locations, which can be time consuming.

Figure 1. Monetary Work Zone Road User Cost.



The literature review offers insight into ways the current models for estimating WZRUC may be improved for Texas. Potential directions to enhance existing methodologies include:

1. **Incorporating more data sources for input estimation.** The performance of both analytical and simulation based methods depend on the quality and accuracy of data provided. Even though AADT and average speed can provide an approximate measure for traffic volume and speed, they do not capture the time of day variation, which may lead to inaccurate estimates of traffic delay. Using the rich data provided by the modern location- and probe-based sensors, we can improve the accuracy of the input data incorporated into the user delay cost estimation models. Haseman et al. (2010) provide one such example of updated travel time estimation in work zones using Bluetooth dataset that can be used validate volume and capacity thresholds in which queuing conditions occur. The inputs to emission cost and VOC estimation models can also be improved using the same methods.
2. **Using data-driven models for estimation of capacity and queue delays.** Current models in practice use a deterministic capacity value obtained from field measurements, engineering judgement, or the *Highway Capacity Manual* (*HCM*). However, studies have shown the capacity of a roadway varies over time and shows stochastic behavior. Using the data-driven methods proposed in literature to estimate probability distribution of work zone capacities, we can improve the accuracy of the models. Similarly, the model for queue delays can be updated to include acceleration and deceleration components, as shown by Jiang (2001).
3. **Incorporating diversion of traffic.** Modeling network-wide diversion is an important component for estimating user delay. The diversion is captured using correct estimates of traffic volumes approaching a work zone. Using data-driven methods, we can estimate probability distribution of traffic volume and the rate of diversion to improve the accuracy of models.

Based on the information contained herein, as well as data availability, CTR will further lane user cost research by conducting a test case study in Austin, Texas. The goals of the test case study are to 1) understand how available data may be used to compute user costs, and 2) test the performance of three lane user cost analysis methodologies. Using available data, CTR will identify a test case work zone, and then estimate queue length, duration of queue, and use cost using three methodologies. The three methodologies, as outlined in Figure 1 above, include the following:

1. Data Analysis
2. Analytical method using parametric analytics to estimate adjusted capacity
3. Simple simulation-based method

References

Adler, M. W., J. van Ommeren, and P. Rietveld. 2013. “Road congestion and incident duration.” *Economics of Transportation*, 2(4), 109–118.

Benekohal, R., A. Z. Kaja-Mohideen, and M. Chitturi. 2004. “Methodology for estimating operating speed and capacity in work zones.” *Transportation Research Record: Journal of the Transportation Research Board*, (1883), 103–111.

Benekohal, R. F., H. Ramezani, and K. A. Avrenli. 2010. *Queue and user's costs in highway work zones*. Research Report ICT-10-075

Borchardt, D. W., G. Pesti, D. Sun, and L. Ding. 2009. “Capacity and road user cost analysis of selected freeway work zones in Texas.” *Austin, TX. Texas Department of Transportation, Austin, TX*.

Calvert, S. C., H. Taale, and S. P. Hoogendoorn. 2016. “Quantification of motorway capacity variation: influence of day type specific variation and capacity drop.” *Journal of Advanced Transportation, 50*(4), 570–588.

Chung, Y. 2011. “Quantification of nonrecurrent congestion delay caused by freeway accidents and analysis of causal factors.” *Transportation Research Record: Journal of the Transportation Research Board*, (2229), 8–18.

Daganzo, C. F. 1994. “The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory.” *Transportation Research Part B: Methodological*, *28*(4), 269–287.

Farradyne, P. B. 2000. “Traffic incident management handbook.” *Prepared for Federal Highway Administration, Office of Travel Management.*

Habtemichael, F. G., M. Cetin, and K. A. Anuar. 2015. “Incident-induced delays on freeways: quantification method by grouping similar traffic patterns.” *Transportation Research Record: Journal of the Transportation Research Board,* (2484), 60–69.

Hadi, M., P. Sinha, and A. Wang. 2007. “Modeling reductions in freeway capacity due to incidents in microscopic simulation models.” *Transportation Research Record: Journal of the Transportation Research Board*, (1999), 62–68.

Haseman, R., J. Wasson and D. Bullock. 2010. “Real-time measurement of travel time delay in work zones and evaluation metrics using Bluetooth probe tracking.” *Transportation Research Record: Journal of the Transportation Research Board*, (2169), 40–53.

Highway Capacity Manual (2000).

Jenkins, J. M. and D. McAvoy. 2015. *Evaluation of Traffic Flow Analysis and Road User Cost Tools Applied to Work Zones*. Ohio Department of Transportation, Office of Statewide Planning & Research.

Jiang, Y. 2001. “Estimation of traffic delays and vehicle queues at freeway work zones.” Proceedings of the 80th Transportation Research Board Annual Meeting, Washington, DC.

Mallela, J. and S. Sadavisam. 2011. “Work Zone Road User Costs: Concepts and Applications.” US Department of Transportation, Federal Highway Administration.

Olmstead, T. 1999. “Pitfall to avoid when estimating incident-induced delay by using deterministic queuing models.” *Transportation Research Record*: Journal of the Transportation Research Board, (1683), 38–46.

Qian, Z. and H. Michael Zhang. 2012. “Full closure or partial closure? Evaluation of construction plans for the I-5 closure in downtown Sacramento.” *Journal of Transportation Engineering*, *139*(3), 273–286.

Qin, L. and B. Smith. 2001. “Characterization of accident capacity reduction.” Smart Travel Lab Report 2001-02). University of Virginia Center for Transportation Studies. https://www.researchgate.net/profile/Brian_Smith29/publication/228785399_Characterization_of_accident_capacity_reduction/links/55b7ab6608ae9289a08bfda4/Characterization-of-accident-capacity-reduction.pdf. Last accessed Fe. 21, 2018.

Ramezani, H. and R. Benekohal. 2011. “Analysis of Queue Formation and Dissipation in Work Zones.” *Procedia-Social and Behavioral Sciences*, *16*, 450–459.

Snelder, M., T. Bakri, and B. van Arem. 2013. “Delays Caused by Incidents: Data-Driven Approach.” *Transportation Research Record: Journal of the Transportation Research Board*, (2333), 1–8.

Texas Department of Transportation. 2017. “Road User Costs”. Available online at <http://www.txdot.gov/inside-txdot/division/construction/road-user-costs.html>. Last accessed: November, 2017

Yperman, I. 2007. “The link transmission model for dynamic network loading.” Available at: https://lirias.kuleuven.be/bitstream/1979/946/2/PhD-FINAL-ALL.pdf, last accessed Feb. 28, 2018

Weng, J. and Q. Meng. 2013. “Estimating capacity and traffic delay in work zones: An overview.” *Transportation Research Part C: Emerging Technologies*, *35*, 34–45.

Weng, J., and X. Yan. 2015. “Probability distribution-based model for work zone capacity prediction.” *Journal of Advanced Transportation*, *50*(2), 165–179.

Zhang, H., F. Ni, and S. Yang. 2012. “Prediction of freeway traffic incident delay based on simulation.” *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference*, 167–171.

Zheng, H., E. Nava, and Y. C. Chiu. 2014. “Measuring networkwide traffic delay in schedule optimization for work-zone planning in urban networks.” *IEEE Transactions on Intelligent Transportation Systems*, *15*(6), 2595–2604.