Transportation Research Part C 35 (2013) 34–45

Contents lists available at SciVerse ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Overview Paper

Estimating capacity and traffic delay in work zones: An overview

Jinxian Weng [a,*], Qiang Meng [b]

[a] MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China
[b] Department of Civil & Environmental Engineering, National University of Singapore, Singapore 117576, Singapore

A R T I C L E I N F O

Article history:
Received 18 January 2012
Received in revised form 14 June 2013
Accepted 15 June 2013

Keywords:
Work zone capacity
Traffic delay
Estimation
Approach
Review

A B S T R A C T

This paper presents a review of the approaches to estimating the work zone operational issues: capacity and traffic delay in work zones. It first explores the factors affecting work zone capacity and then critically reviews three types of approaches including parametric, non-parametric and simulation approaches to estimating work zone capacity. Subsequently, a detailed critical review of the three types of approaches for traffic delay estimation in work zones is presented. Finally, it provides some directions and recommendations for the future research.

1. Introduction

For any country, the road system is an important part of its infrastructure which can affect economic development. An efficient road system can promote economic development because it enhances the performance of local transportation. A good level of service for a road system requires implementation of work zone projects to maintain it. Therefore, various work zone activities, such as pothole patching, crack sealing and pavement resurfacing, are regularly carried out by land transport authorities. Hereafter, a work zone is referred to 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. In general, a work zone comprises four components including the advance warning area, the transition area, the activity area and the termination area, shown in Fig. 1.

However, work zone activities could cause several problems because work zones usually close one or more of the lanes available for traffic, as shown in Fig. 1. The lane closure in work zone might be owing to the following two reasons. First, it is impossible to implement some types of work zone activities such as resurfacing pavements without interrupting traffic. Second, the lane closure could protect workers' safety. However, lane reductions could cause a disturbance to normal traffic flow and speed reductions, further resulting in a reduction of road capacity and an increase of traffic delay. Since vehicles in the closed lanes have to merge into the adjacent available lanes, it may increase the number of traffic conflicts and cause severe traffic safety problems.

In order to increase capacity and mitigate traffic delay in work zones, a few advanced traffic control systems, such as dynamic lane merge traffic control systems, have been proposed in the past. For example, Tarko and Venugopal (2001) devel-

* Corresponding author. Tel.: +86 18612036889.
E-mail addresses: jxweng@bjtu.edu.cn, wjx206@gmail.com (J. Weng).

0968-090X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trc.2013.06.005

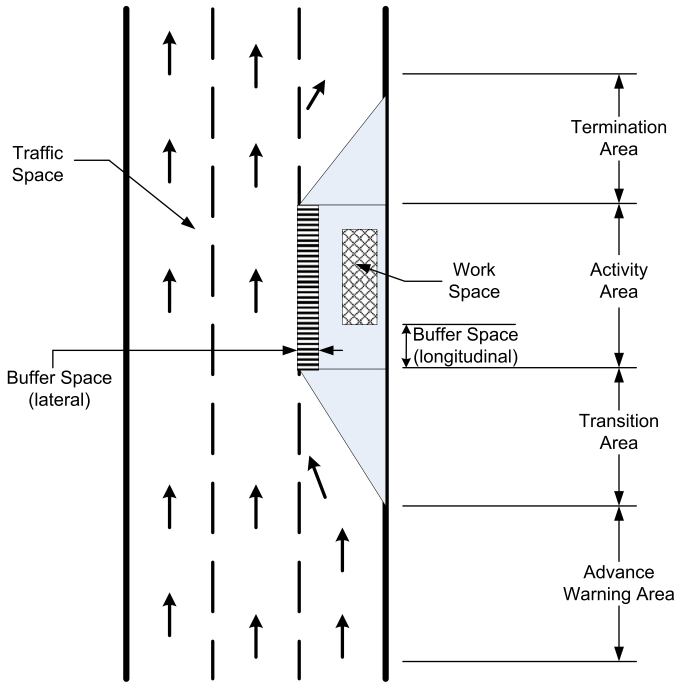

Fig. 1. Component parts of a work zone (MUTCD, 2003).

oped the Indiana Lane Merge System, comprising a series of dynamic "Do Not Pass/When Flashing" signs to smooth the merge operations. Schattler and Datta (2003) implemented the dynamic early merge traffic control system (EMTCS) in work zones which cause lane closures in Michigan. The EMTCS was found to significantly reduce forced lane merges and aggressive driver behavior. To increase the traffic throughput in work zones, Grillo et al. (2008) used the dynamic late lane merge system (DLLMS) to specify a definite merge point in freeway construction work zones. The DLLMS was found to greatly reduce queue lengths in travel lanes. Recently, Shaaban et al. (2011) proposed two Simplified Dynamic Lane Merging Systems (SDLMS) (early merge and late merge) for 3–2 lane short-term work zones in Florida.

It should be noted that work zone capacity estimate is an essential component in planning work zone traffic control systems. In addition, work zone capacity and traffic delay are the two critical indices to determine whether the effectiveness of traffic control systems is acceptable or not. Hence, the accurate estimation of capacity and traffic delay in work zones is of utmost importance and thus there is a need to review the estimation approaches. However, the existing literature on the overview of approaches to estimating work zone capacity and traffic delay is rather limited and incomplete (e.g., Kayani and Bham, 2010). Therefore, this paper aims to provide a thoroughly critical review on the approaches to estimating work zone capacity and traffic delay in work zones. The potential gaps and limitations of these approaches will be identified in this paper. It will also provide some directions and recommendations for future research.

This literature review was conducted using a computerized literature search method. With an access to the digital library of the Beijing Jiaotong University, which includes hundreds of journals as well as conference proceedings, the search is able to cover most published research works worldwide. First, the databases of Scopus, the ScienceDirect®, and the ISI Web of Knowledge were searched using the following key words: "work zone", "capacity", "traffic delay" and "methodology". Beside of these databases, we also used the Google Scholar as a second channel and browsed the personal websites of researchers who are active within traffic operations. Furthermore, we retrieved studies by tracking the references cited in papers we had already found. Finally, we identified 42 papers which relate to the methods and techniques for estimating work zone capacity and traffic delay.

This paper is organized as follows. Section 2 reviews the existing approaches for estimating work zone capacity. The strength and drawbacks of these approaches are also illustrated. Section 3 discusses the available approaches to estimating traffic delay. Future research recommendations and concluding marks are given in Section 4.

2. Work zone capacity

2.1. Factors affecting work zone capacity

There are many factors that could affect work zone capacity. Weng and Meng (2012) concluded 16 important factors which could significantly affect work zone capacity, shown in Fig. 2. These factors are grouped into five categorizes and in detail elaborated below.

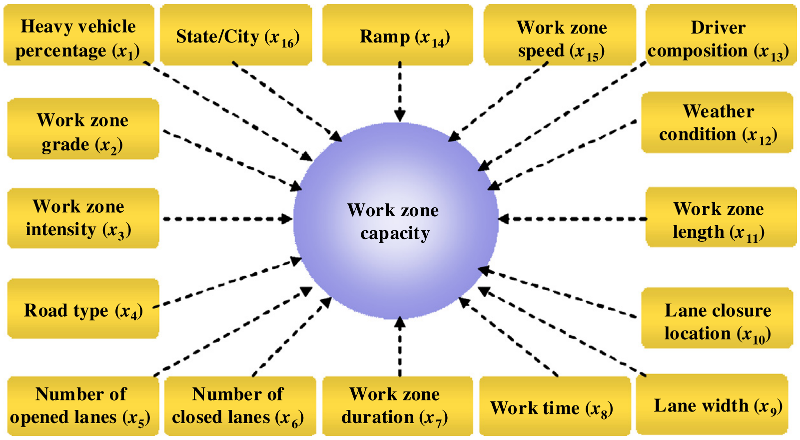

Fig. 2. Sixteen factors affecting work zone capacity.

2.1.1. Work zone configurations

Measurements made in freeway work zones in Texas (Dudek and Richards, 1981; Krammes and Lopez, 1994) and North Carolina freeways (Dixon et al., 1996) showed clearly that work zone capacity varies significantly with the number of freeway lanes as well as the number of lane closures. In addition, lane closure location shows the impacts on work zone capacity. Work time may also exhibit its impacts on work zone capacity. However, from the previous studies, whether work zone length affects work zone capacity or not is a controversial topic. Kim et al. (2001) claimed that longer length could reduce work zone capacity, while Heaslip et al. (2009) found that work zone capacity could not significantly vary with work zone length. Kim et al. (2001) found that the presence of grades may exacerbate the flow constriction in work zones particularly in the presence of heavy vehicles. Adeli and Jiang (2003) reported that a lower work zone speed could decrease the work zone capacity, while improving the safety. In addition, a few of other work zone configuration factors such as lateral clearances, lane taper characteristics and merge control provisions (e.g., signs, cones), could also affect work zone capacity.

2.1.2. Roadway conditions

Dixon et al. (1996) found that work zone capacity on an urban road is usually 20–30% higher than that on a rural road. According to the HCM (2010), ramps' proximity to the work zone, especially the entrance ramps inside the work zone activity area, can create traffic turbulence, resulting in a reduction of the work zone capacity. In addition, the width of the lanes and the distance to lateral obstructions will both affect capacity. The HCM (2010) suggests a reduction factor of up to 14% to account for the effect of lane width on work zone capacity.

2.1.3. Work activity characteristics

The first work activity characteristic is the work intensity. Work zone capacity may decrease as the work intensity increases from the lightest (e.g., guardrail installation) to the heaviest (e.g., bridge repair). Work zone intensity has been classified into three levels (low, medium, or high) by Karim and Adeli (2003). The HCM (2010) suggests a modification of the base capacity value of the work zone to account for the intensity of work activities, without actually providing any modification factors or guidelines. The work time (daytime or nighttime), as another work activity characteristic, also affects work zone capacity. According to the work of Al-Kaisy and Hall (2001), night construction or maintenance could decrease work zone capacity because of the reduced traveler's attention. Work zone duration (short-term/long-term) is the third feature of work activity. In general, the average capacity in long-term work zones is greater than that in short-term work zones because commuters and frequent travelers become familiar with the configuration of the long-term work zone. In addition, means of longitudinal separation from work activities (e.g., cones) and enforcement activities (e.g. regarding speeds) could also influence work zone capacity.

2.1.4. Environmental conditions

Weather (e.g., snowy, rainy, sunny) usually has a significant impact on work zone capacity. The HCM (2010) suggests 10–20% capacity reductions due to bad weather conditions, without providing any specific guidelines.

2.1.5. Others

Since heavy vehicles such as trucks occupy more space on the roadway and move more slowly than passenger cars, a high percentage of heavy vehicles tends to reduce work zone capacity. Krammes and Lopez (1994) carried out a work zone capacity study and found that work zone capacity is affected by the percentage of heavy vehicles. Moreover, Al-Kaisy and Hall (2003) found that the effect of heavy vehicles is greater in queue discharge (capacity) flow than during free-flow operation.

Another factor affecting work zone capacity is the driver composition (commuters/non-commuters). In all situations, commuters and regular travelers are more familiar with work zone configurations and traffic control plans than non-commuters (e.g., tourists). Therefore, the presence of non-commuters could reduce work zone capacity.

2.2. Work zone capacity estimation approaches

As mentioned above, many factors are found to affect work zone capacity and work zone capacity estimation is complex. So far, a number of approaches have been proposed to estimate work zone capacity in the literature. These approaches can be generally categorized into three groups: parametric approaches, non-parametric approaches and simulation approaches, which are tabulated in Table 1. In addition to these approaches, a few researchers (e.g., Kim and Son, 2003; Kim, 2008) also made attempts to estimate work zone capacity using the gap acceptance theory on a basis of lane-changing/merging concept.

2.2.1. Parametric approaches

Many work zone capacity studies focused on estimating work zone capacity by using parametric approaches. Here, the parametric approach is an approach in which the predictor takes a predetermined form. Based on the work zone capacity data collected from the field sites, the coefficients of these predictors can be determined.

Krammes and Lopez (1994) developed the following model to estimate the capacity of short-term freeway work zones using the data collected in 33 work zones in Texas between 1987 and 1991 for the short-term work zones:

$$c = (1600\,\text{pcphpl} + I - R) \times f_{HV} \tag{1}$$

where c is the estimated work zone capacity (vphpl); I is the adjustment value of the work intensity, which is suggested that 160 for low intensity, 0 for medium intensity and -160 for heavy intensity (Karim and Adeli, 2003); R is the adjustment value for the presence of ramps ($R = 0$ if no ramp is present and $R = 160$ if entrance ramp is present); f_{HV} is the heavy vehicle adjustment factor. However, it can be seen from Eq. (1) that quite a few factors affecting work zone capacity were not taken into account. Therefore, taking into account more capacity influencing factors, Kim et al. (2001) developed a model using a multiple regression approach to estimate the short-term work zone capacity using traffic data fore 12 work zone sites with lane closures on four normal lanes in one direction, shown below:

$$c = 1857 - 168.1NUMCL - 37.0LOCCL - 9.0HV + 92.7LD - 34.3WL - 106.1WI - 2.3WG \times HV \tag{2}$$

where $NUMCL$ is the number of closed lanes, $LOCCL$ is the lane closure location (right = 1, left = 0), HV is the heavy vehicle percentage, LD is the lateral distance to the lane closure, WL is the work zone length, WI is the work intensity, and WG is the work zone grade.

For long-term work zones, Al-Kaisy et al. (2000) investigated capacity in work zones involving long-term lane closures in terms of the variations in capacity under the effects of important control and extraneous variables in Ontario, Canada. The variables mainly include temporal variations, grade, the day of week, and weather conditions. The results showed significant variations in long-term work zone capacity. These variables were found to exhibit significant effects on the capacity in long-term work zones. Considering that a multiplicative format is easy for users to understand, Al-Kaisy and Hall (2003) devel-

Table 1
Analysis of literature on approaches to estimating work zone capacity.

Approaches	References	Detailed approaches/tools
Parametric approaches	Al-Kaisy et al. (2000)	Multi-regression approach
	Al-Kaisy and Hall (2003)	A generic multiplicative approach
	Avrenli et al. (2011)	Derived from speed-flow relationship
	Benekohal et al. (2004)	A step-by-step approach based on the speed-flow relationship
	HCM (2010)	Two distinct guidelines
	Kim et al. (2001)	A multiple regression approach
	Krammes and Lopez (1994)	A multiple regression approach
	Racha et al. (2008)	Derived from speed-flow relationship
	Sarasua et al. (2006)	Derived from speed-flow relationship
Non-parametric approaches	Adeli and Jiang (2003)	A neural-fuzzy logic approach
	Karim and Adeli (2003)	Radial basis function neural network
	Weng and Meng (2011)	A decision tree approach
	Weng and Meng (2012)	An ensemble tree approach
	Zheng et al. (2011)	A neural-fuzzy logic approach
Simulation approaches	Arguea (2006)	Improved CORSIM
	Chatterjee et al. (2009)	VISSIM
	Heaslip et al. (2009)	A hybrid approach integrating the simulation and mathematical methods
	Heaslip et al. (2011)	CORSIM
	Ping and Zhu (2006)	CORSIM

oped a generic multiplicative model to estimate the long-term work zone capacity using the data from six reconstruction sites in Ontario:

$$c = C_b \times f_{HV} \times f_d \times f_w \times f_s \times f_r \times f_l \times f_i \tag{3}$$

where C_b = based work zone capacity (i.e., 2000 pcphpl), f_{HV} = adjustment factor for heavy vehicles, f_d = adjustment factor for driver population, f_w = adjustment factor for work activity, f_s = adjustment factor for side of lane closure, f_r = adjustment factor for rain, f_l = adjustment factor for light condition, f_i = adjustment factor for non-additive interactive effects.

However, the above mathematical models were formulated on basis of the measured work zone capacity data. In other words, their model accuracy may greatly rely on the accuracy of measured data from field work zone sites. In general, two different approaches have been used to measure capacity in work zones: one is the queue discharge flow approach (Kianfar et al., 2012) and the other is the maximum flow rate approach. The queue discharge approach counts the actual number of vehicles processed under queuing conditions, while the maximum flow rate approach uses the flow rate before a significant speed drop or after a significant speed gain. Both approaches are based on counting and grouping vehicles. The queue discharge approach gives lower capacity values than the maximum flow rate approach. Both approaches depend on vehicle counts aggregated data over time. According to Benekohal et al. (2004), neither the queue discharge approach nor the maximum flow approach could provide accurate measurements of work zone capacity from the field sites. Therefore, we should avoid the method of measuring capacity first from field sites and then using it to formulate a parametric model.

One alternative way is first to determine the speed-flow relationship and then derive the capacity from speed-flow curves. Some researchers already adopted this method to derive work zone capacity estimation model. Benekohal et al. (2004) gave a step-by-step approach based on the speed-flow information to estimate work zone capacity for a two-to-one lane closure configuration. They established the functional relationship between work zone capacity, flow and operating speed using the extensive data collected from 11 work zones in Illinois. In order to examine the impacts of external factors on traffic speed, the operating speed was further modeled as a function of work intensity, lane width, lateral clearance and other factors. Sarasua et al. (2006) depicted the speed-flow curves for two-to-one, three-to-two and three-to-one lane closure configurations of interstate highway work zones. In addition, The base capacity was considered as a variable which depends on the lane closure configuration. The passenger car equivalent (PCE) value was expressed as a function of traffic speed. Racha et al. (2008) first modeled the traffic flow as a parabolic function of traffic speed while traffic speed was modeled as a non-linear hyperbolic function of traffic density. Based on the parabolic flow function and the hyperbolic speed function, a work zone capacity model was then derived by Racha et al. (2008) to estimate work zone capacity. Avrenli et al. (2011) developed two non-linear speed-flow models for work zones with no lane closure under uncongested and congested conditions, respectively. Based on these two nonlinear models, a work zone capacity model was derived to estimate work zone capacity. They found that the estimation errors from the derived work zone capacity model are acceptable.

Literature search shows that each parametric capacity estimation model is only available for a specific work zone type. In terms of work zone duration, each parametric model is either applicable for long-term work zones or short-term work zones. However, long-term work zones generally have higher capacities than short-term work zones. In this situation, these parametric models for short-term work zones are not applicable for estimating long-term work zone capacity because they may greatly underestimate the capacity of long-term work zones. Likewise, it is inappropriate to estimate the capacity of long-term work zones using those models which are developed using short-term work zone data.

Therefore, it is inadequate to estimate freeway work zone capacity by using a simple model. The current highway capacity manual (HCM, 2010) provides the following two distinct work zone capacity estimation guidelines for the short-term work zones and long-term work zones, respectively:

(i) For short-term work zones, the HCM provides the following equation to estimate capacity in freeway short-term work zones:

$$C = \{[(1600 + I) \times f_{HV}] \times N\} - R \tag{4}$$

where C is the estimated work zone capacity (vph); I is the manual adjustment value of the work intensity; R is the manual adjustment value for on-ramps; f_{HV} is the heavy vehicle adjustment factor and N is the number of lanes open through the work zone.

(ii) For long-term work zones, the HCM gives an average capacity of 1750 vphpl for a two-to-one lane closure and 1860 vphpl for a three-to-two lane closure.

Compared with the existing parametric models, the HCM is applicable for estimating capacity in both work zone types (i.e., short-term work zones and long-term work zones). However, one shortcoming of HCM is that it requires users to decide the adjustment factors themselves, which may lead to significant estimation errors due to possibly subjective judgments being made.

2.2.2. Non-parametric approaches

In work zone capacity analysis, there might exist nonlinear relationships and high-order interactions between influencing factors. Some influencing factors may even hide effects on work zone capacity. Therefore, a conventional parametric approach usually provides low estimation accuracy because it cannot fully describe the complicated effects of influencing factors due to the interaction effects and nonlinearities. Since parametric approaches may not provide high prediction accuracy, many non-parametric approaches were also employed for the estimation of work zone capacity in the past. Here, the non-parametric approach is referred as to a technique that does not assume that the structure of a model is fixed (e.g, neural networks, decision tree). Typically, the corresponding model grows in size to accommodate the complexity of the data.

Adeli and Jiang (2003) was the first to apply a non-parametric approach, namely, a neural-fuzzy logic approach to estimating work zone capacity. This approach introduces a neural network of interacting variables that are quantified using a fuzzy inference method. Conditional rules are generated using a fuzzy logic algorithm and further applied to work zone capacity estimation. The capacity from a conditional rule can be calculated by:

$$c_i = \sum_{j=1}^{M} \exp\left(-\frac{(q_j - c_{ij})^2}{2\sigma_{ij}}\right) q_j \tag{5}$$

where c_i is the work zone capacity from the conditional rule i, q_j is the input value for the variable j, c_{ij} and σ_{ij} are the parameters of membership function for variable j in rule i. Therefore, the work zone capacity is finally calculated as the weighted sum of the work zone capacity from all rules. That is,

$$c = \sum_{i=1}^{N} c_i \frac{w_i}{\sum_{i=1}^{N} w_i} \tag{6}$$

where c is the estimated work zone capacity, w_i is the weighted value for the condition rule i.

In general, this neural-fuzzy logic approach could provide higher estimation accuracy than parametric approaches (Zheng et al., 2011). However, it is so complex that it has poor applicability for the users. In order to provide more accurate estimates, Weng and Meng (2011) employed a decision tree approach to estimating work zone capacity. A decision tree-based model was developed using this approach, shown in Fig. 3. Compared with the neural-fuzzy approach, it has higher estimation accuracy. More importantly, the decision tree approach has a high ease of use. Using the developed tree-based model, users can easily estimate work zone capacity by tracing a path down the tree to a terminal node according to the characteristics of work zone. However, one big deficiency of decision tree approach is that the tree structure is very "unstable" because it might alter dramatically if the training data and testing data are slightly changed.

To compensate for the weak points of the decision tree approach, Weng and Meng (2012) applied an ensemble tree approach to estimating work zone capacity. According to the subjective assessment, an ensemble tree comprising 105 individual trees was selected. A combination of multiple single decision trees could reduce the risk of choosing a bad tree but increase the chance of choosing a good tree. Therefore, the ensemble tree has higher estimation accuracy and stability than the decision tree. Nevertheless, unlike the decision tree, the ensemble tree lacks of a graphical display of results so that it is difficult for users to understand the detailed relationship between work zone capacity and its influencing factors. In addition, it should be noted that both the decision tree and ensemble tree discretize the continuous factors (e.g., heavy vehicle per-

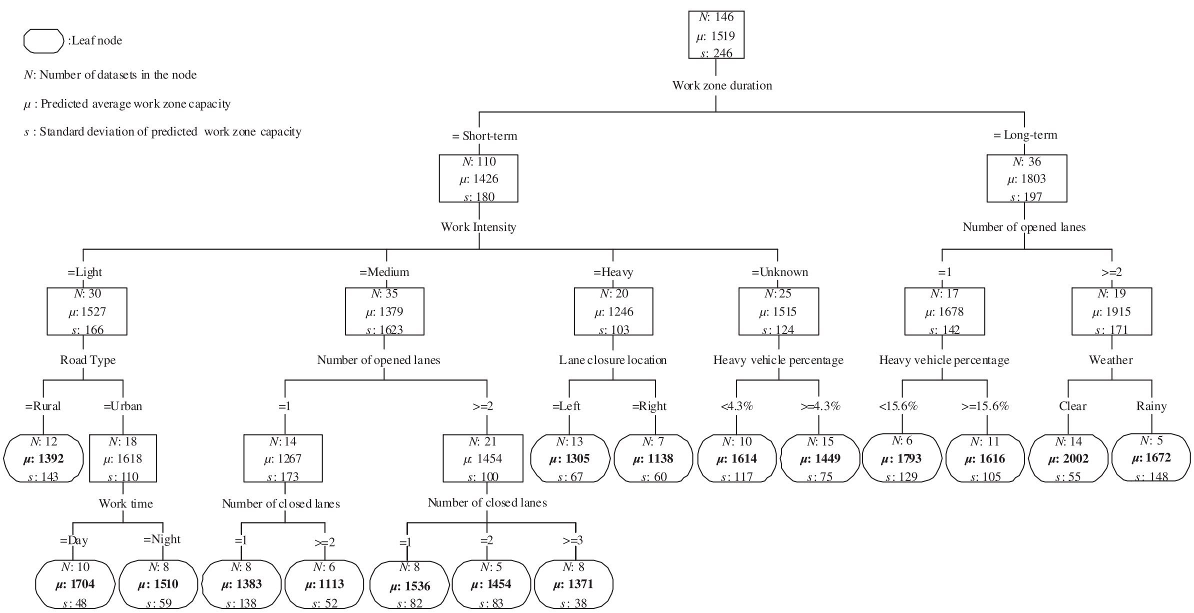

Fig. 3. A decision tree based model developed by Weng and Meng (2011).

Table 2
Examples of work zone capacity records from previous studies (Weng and Meng, 2011).

	(x_1)	(x_2)	(x_3)	(x_4)	(x_5)	(x_6)	(x_7)	(x_8)	(x_9)	(x_{10})	(x_{11})	(x_{12})	(x_{13})	(x_{14})	(x_{15})	(x_{16})
1	11.0%	0	Medium	Urban	2	2	Short-term	Night	3.6	Left	Long	Clear	0	Yes	22	Maryland
2	3.2%	–	–	Urban	1	1	Short-term	–	–	Left	–	–	–	–	–	Texas
3	13.2%	–	–	Urban	1	1	Short-term	–		Right	–	–	–	–	–	Texas
4	8.5%	–	–	Urban	2	2	Short-term	–	–	Left	–	–	–	–	–	Texas
5	3.7%	–	–	Urban	3	1	Short-term	–	–	Left	–	–	–	–	–	Texas
6	–	–	Medium	Urban	1	2	Short-term	–	–	Left	–	–	–	–	–	Texas
7	–	–	Light	Urban	2	1	Short-term	–	–	Left	–	–	–	–	–	Texas
8	–	–	Medium	Urban	3	1	Short-term	–	–	Left	–	–	–	–	–	Texas
9	–	–	Heavy	–	3	1	Short-term	–	–	–	–	–	–	–	–	California
10	25.0%	–	Medium	Rural	1	1	Short-term	Day	–	Right	–	–	–	–	–	Indiana
11	17.4%	–	Light	Rural	1	1	Short-term	Night	3.6	–	Short	Clear	–	Yes	–	South Carolina
12	–	–	Heavy	Urban	2	1	Short-term	Day	–	Left	–	Clear	–	–	–	San Antonio
13	4.5%	–	Medium	Rural	1	1	Long-term	Night	3.3	Right	–	–	–	–	–	North Carolina
14	–	–	–	Urban	2	1	Long-term	–	–	–	–	Rainy	–	–	–	Toronto
15	5.0%	–	Heavy	Urban	2	1	Long-term	Day	3.6	Left	–	Clear	1	–	–	Florida
16	5.0%	–	Heavy	Urban	2	1	Long-term	Day	3.6	Right	–	Clear	1	–	–	Florida
17	–	–	Heavy	Rural	2	2	Long-term	Day	–	Left	–	Clear	–	–	–	Wisconsin
18	–	–	Heavy	Rural	1	2	Long-term	Day	–	Left	–	Clear	–	–	–	Pearland

The symbol "–" represents the "missing" value assigned for the unavailable variable in a work zone record.

centage) into categorical levels in tree building by using a discretization method based on the *F*-test. Such discretization method may reduce the accuracy because of the shortcoming of the *F*-distribution assumption that is inappropriate in some domains of these factors.

2.2.3. Simulation approaches

In general, parametric and non-parametric approaches are the two efficient methods to estimate work zone capacity when actual work zone capacity data can be collected from the field sites. However, they are data-driven based approaches so that they highly rely on the quantity and quality of collected work zone capacity data. The poor data quality may reduce the reliability and accuracy of the estimation results. Table 2 shows some collected work zone capacity records used in the study of Weng and Meng (2011). From the table, it can be seen that many input values of variables are missing in each record. This will definitely affect the reliability of the estimation results. In addition, it is difficult to examine the effect of a specific factor on work zone capacity by collecting field data under different work zone circumstances. For example, it is unrealistic to set different distance of warning signs, different lane distributions, different speeds upstream of the work zone entrance and different merge strategies (late merge or early merge) in the field sites. To evaluate the effects of these factors, simulation approaches are considered to be a good choice. In the past, the simulation approach has already been applied to estimate freeway work zone capacity under various network configurations (Ping and Zhu, 2006) and under various lane closure configurations (Arguea, 2006; Chatterjee et al., 2009). Heaslip et al. (2011) used the simulation approach to estimate the capacity of work zones located on arterial streets. From the simulation results, it was found that the installation of a work zone may increase capacity when the intersection (prior to the installation of work zone) is congested.

In order to accurately use the microscopic simulation models or tools for traffic analysis in work zones, it is necessary to calibrate them to match the field conditions (such as lane configuration and queue lengths) by adjusting the parameters of simulation models or tools. However, it is not an easy task to choose appropriate simulation model parameters which could replicate the actual field conditions in work zones. In other words, it is difficult to calibrate the simulation models or tools. In addition, the simulation models and tools require high levels of computational resources and time (Edara and Cottrell, 2007). Therefore, in order to improve the computational efficiency, some researchers (e.g., Heaslip et al., 2009) proposed a hybrid approach to estimating work zone capacity by integrating the simulation and mathematical methods.

3. Work zone traffic delay

In general, work zone traffic delay can be defined as the difference between travel time on a roadway segment without work zones and the actual longer travel time in work zones. The existing approaches applicable for estimating traffic delay in work zones can be categorized into three types: macroscopic analytical approaches, macroscopic simulation approaches, and microscopic simulation approaches, as shown in Fig. 4 and Table 3. This section includes a brief discussion of these approaches which have been used for work zone traffic delay estimation in previous studies.

3.1. Macroscopic analytical approaches

Deterministic queuing approaches are the common macroscopic analytical approaches that are most widely used for estimating traffic delay. They are often illustrated using the diagram shown in Fig. 5. The critical inputs are the demand volume, the freeway capacity, the work zone capacity, and the work zone duration.

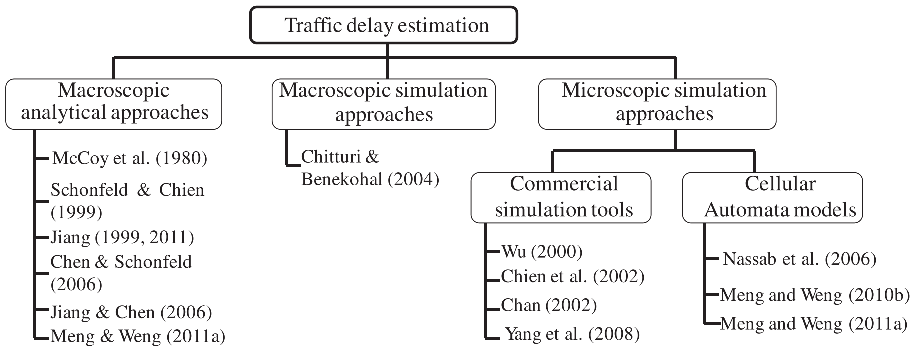

Fig. 4. Existing approaches to estimating traffic delay in work zones.

Table 3
Analysis of literature on approaches to estimating work zone traffic delay.

References	Approaches	Remarks
Chan (2002)	Microscopic simulation approach	PARAMICS is applied to estimate work zone traffic delay
Chen and Schonfeld (2006)	Macroscopic analytical approach	Deterministic queuing approach is applied to estimate work one traffic delay
Chien et al. (2002)	Microscopic simulation approach	The CORSIM software is applied to estimate work zone traffic delay
Chitturi and Benekohal (2004)	Macroscopic simulation approach	FRESIM, QuickZone and QUEWZ are applied to estimate work zone traffic delay
Jiang (1999)	Macroscopic analytical approach	An enhanced deterministic queuing approach taking into account deceleration and acceleration delay is applied to estimate traffic delay
Jiang (2001)	Macroscopic analytical approach	A queue-discharge rate table is used to improve the queuing delay estimation accuracy
Martinelli and Xu (1996)	Macroscopic analytical approach	Deterministic queuing approach is applied to estimate traffic delay
McCoy et al. (1980)	Macroscopic analytical approach	Deterministic queuing approach is applied to estimate traffic delay
Meng and Weng (2013a)	Macroscopic analytical approach	An improved deterministic queuing model is developed to estimate traffic delay
Meng and Weng (2010)	Microscopic simulation approach	CA model is applied to estimate work zone traffic delay
Schonfeld and Chien (1999)	Macroscopic analytical approach	Deterministic queuing approach is applied to estimate traffic delay
Tang and Chien (2008)	Macroscopic analytical approach	Deterministic queuing approach is applied to estimate traffic delay
Wu (2000)	Microscopic simulation approach	The INTEGRATION software is applied to estimate traffic delay
Yang et al. (2008)	Microscopic simulation approach	A hybrid method integrating a macroscopic analytical model with a microscopic simulation tool is applied to calculate work zone delays

For this approach, many researchers (e.g., McCoy et al., 1980; Schonfeld and Chien, 1999; Jiang and Adeli, 2003) assumed that there is no queue formed when the approaching traffic flow is less than work zone capacity. Therefore, traffic delay equals to the moving delay which is caused by work zone speed limits in this situation. For example, McCoy et al. (1980) considered that traffic delay is equal to the time lost when vehicles travel through the work zone. The time lost is taken to be a function of the difference between the average overall speed of the two-lane two-way no-passing operation and that of the normal four-lane divided operation. When traffic flow exceeds work zone capacity, queues will form at upstream of the work zone. In this situation, traffic delay is regarded as the sum of queuing delay and moving delay in work zones (McCoy et al., 1980). Taking into account the variations of traffic speed, Meng and Weng (2013a) developed an improved deterministic queuing model to estimate work zone traffic delay. They also remedied two major flaws on the queuing delay and moving delay estimation formulae in some previous studies (Chen and Schonfeld, 2006; Tang and Chien, 2008). The first flaw is that some previous studies ignored a fact that a queue may completely disappear before the end of a time interval, when the arriving traffic flow rate at that time interval is lower than work zone capacity. This could cause their queuing delay estimation at that time interval being overestimated. The second flaw is that these studies assumed that the vehicle departure rate is equal to the arriving traffic flow, when the arriving traffic flow rate is lower than work zone capacity. However, the vehicle departure rate may be greater than the arriving rate when there is a queue.

It should be noted that these studies implicitly provided different definitions of traffic delay. More specifically, moving delay was regarded as the traffic delay in some studies (e.g., McCoy et al., 1980; Schonfeld and Chien, 1999; Jiang and Adeli,

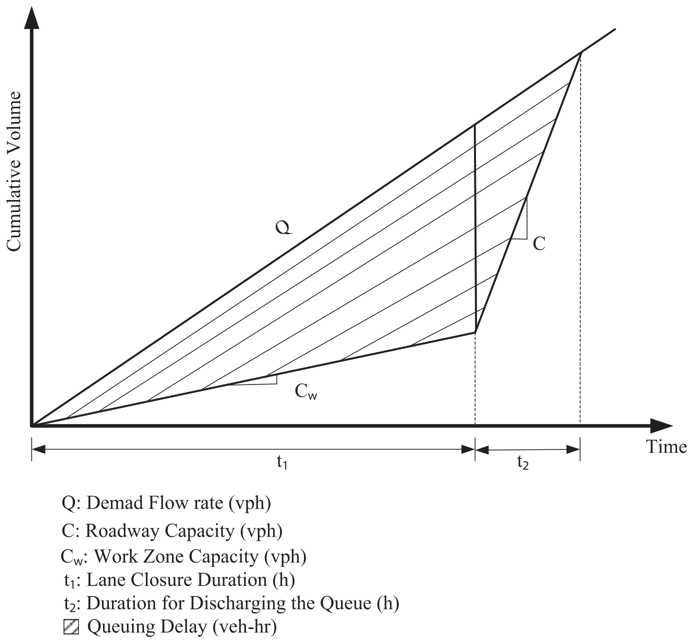

Fig. 5. Deterministic queuing approach.

2003) while other studies (e.g. Chen and Schonfeld, 2006; Tang and Chien, 2008) took the sum of queuing delay and moving delay as traffic delay in work zones.

Although the deterministic queuing approach is widely used because of its simplicity, it has the following two major limitations. First, vehicles will execute deceleration maneuvers before entering work zone, and then accelerate their speeds after passing through the work zone in reality. However, the deterministic queuing approach does not take into account this part of delays caused by the deceleration and acceleration maneuvers. To complement for such a weak point, Jiang (1999) proposed an enhanced analytical delay estimation approach by taking into account deceleration and acceleration delay in addition to the delays which were calculated using the deterministic queuing approach. Second, vehicle queues still form when traffic flow is below work zone capacity because of the randomness of traffic flow. In other words, the deterministic queuing approach could underestimate the queuing delay. Therefore, a queue-discharge rate table instead of the deterministic queuing approach is used by Jiang (2001) in order to improve the accuracy in estimating queuing delay.

3.2. Macroscopic simulation approaches

In addition to macroscopic analytical approaches, there are several macroscopic simulation tools available for estimating traffic delay in work zones. QUEWZ, FRESIM and QuickZone are the most widely used macroscopic simulation tools. Chitturi and Benekohal (2004) compared the performance of these three tools in estimating traffic delay in work zones using the field data collected from eleven freeway work zones in Illinois. It was found that both QUEWZ and FRESIM could overestimate the vehicle speeds under queuing conditions. However, QuickZone was found to underpredict queue lengths and delays in comparison to the field data.

3.3. Microscopic simulation approaches

A number of simulation software is able to simulate work zone traffic and they have already been applied to estimate traffic delays in work zones in previous studies. ARENA is a microscopic simulation model with an advanced animation module which has been applied to estimate work zone traffic delay by Maze and Kamyab (1999). However, it is found that this software could underestimate work zone traffic delay because it simply uses the deterministic queuing approach to estimate the queuing delay at the work zone upstream. In addition to ARENA, some other microscopic traffic simulation tools including the INTEGRATION (Wu, 2000), CORSIM, PARAMICS (Chan, 2002) and VISSIM have been used by researchers to estimate traffic delay in work zones. It should be pointed out that the CORSIM is the most widely used while commercial software PARAMICS and VISSIM are the two tools which could provide higher estimation accuracy than the other microscopic simulation software. Compared with the macroscopic analytical and simulation approaches, the commercial microscopic simulation tools could provide more accurate traffic delay estimates because they are able to model complex traffic dynamics at the individual vehicle level. However, they require high computational resources and long execution times to model traffic flow operations.

To improve computational efficiency, one unique method is to integrate the concept of deterministic queuing theory and the microscopic simulation software to estimate work zone traffic delay (Chien et al., 2002). Yang et al. (2008) proposed such a hybrid approach that integrates a macroscopic analytical model with a microscopic simulation tool to calculate work zone delays. In unsaturated traffic conditions, the CORSIM is used to estimate traffic delay while the deterministic queuing model is employed for the saturated and oversaturated conditions. Nevertheless, microscopic traffic simulation tools (e.g., CORSIM) may not accurately represent driver behavior when drivers are approaching to a work zone. This is because some software simulate a work zone as a prolonged incident blockage. When modeling a lane blockage in CORSIM, the tool assumes that drivers have no knowledge of the approaching blockage and that there is no transition taper. Although other software (e.g., INTEGRATION) do a better job of capturing appropriate lane changing behavior in work zones, they do not allow users to modify work zone configurations. In other words, the estimated traffic delays from these software do not vary with different work zone configurations.

Another efficient method to improve computational efficiency is to develop a cellular auotomata (CA) model to simulate work zone traffic and estimate the resulting traffic delays. Nassab et al. (2006) is the first to use CA model to simulate vehicles' lateral behaviors near a partial lane closure. However, their CA model was too coarse to be applied for the real case because the length of transition area was assumed to be two cells in their model, which is inconsistent with the real-world work zone configurations. In fact, driver behavior in work zones varies with traffic flow and work zone configurations (Al-Kaisy and Hall, 2001). To account for this fact, the randomization probability parameter which is used to represent driver acceleration and deceleration behavior can be modeled as a function of work zone configurations and traffic flow. Based on such an improvement, the CA model could provide slightly higher estimation accuracy of traffic delay than the PARAMICS in work zones (Meng and Weng, 2010). In microscopic traffic simulation software, a lane change maneuver is assumed to be completed in one second. Such an assumption is inconsistent with the field observation because it takes at least two seconds for a vehicle to complete a lateral movement (Gundaliya et al., 2008). Therefore, in order to avoid the unrealistic lane change duration, the CA model still needs to be improved, e.g., by adding a rule to control lateral traveling speed (Meng and Weng, 2011). By improving lateral traveling speed, Meng and Weng (2011) found that a CA model with an additional rule which is used to produce the realistic lateral speed could provide more accurate work zone traffic delay estimates, compared with a CA model not taking into account such a rule.

4. Conclusions and recommendations for future work

This paper has presented a thorough and comprehensive literature review focusing on the approaches to estimating the major work zone operational issues: capacity and traffic delay in work zones. In addition, several potential problems and gaps regarding each issue have been identified.

The existing approaches for work zone capacity estimation adopted in previous studies can be categorized into three types: parametric approaches, non-parametric approaches and simulation approaches. Table 4 presents the features of these work zone capacity estimation approaches. When a number of work zone capacity data can be collected from the field sites, both parametric and non-parametric approaches can be applied to estimate work zone capacity. In general, non-parametric approaches perform better than the parametric approach in terms of estimation accuracy. However, when there exist strong linear relationships between work zone capacity and its influencing factors, parametric approaches could outperform non-parametric approaches. It should be noted these two approaches require a large number of data. In addition, they are not able to explore the effects of those factors which are difficult to be taken into account in the field sites, such as the distance of warning signs, lane traffic distributions and merge strategies. Although simulation approaches are considered to be the best alternative in evaluating the impacts of these factors on work zone capacity, they require high computational resource and long execution time. Therefore, future studies might focus on a hybrid method which integrates non-parametric and simulation approaches for work zone capacity estimation, in order to increase computational efficiency and estimation accuracy.

It should be noted that the existing work zone capacity estimation approaches can only output one constant value for a given work zone. However, in reality, there are unpredictable fluctuations in work zone capacity because of variations in exogenous factors (i.e., driver familiarity, heavy vehicle percentage, weather change). In other words, the true value of work zone capacity cannot be known with a perfect confidence. Therefore, work zone capacity should be represented by means of a probability distribution rather than a constant value in future. More specifically, future studies should formulate the parameters (e.g., mean, standard deviation) of work zone distribution models as a set of functions of external factors such

Table 4
Comparisons of approaches to estimating work zone capacity.

Approaches	Estimation accuracy	Ease of use	Data requirements	Computational resources
Parametric approaches	Low	Simple	Large	Low
Non-parametric approaches	High	Complex	Large	Low
Simulation approaches	High	Complex	Small	High

as road type, number of closed lanes, number of opened lanes, lane closure location, work zone duration, work intensity and working time.

Regarding the traffic delay estimation in work zones, three types of approaches including macroscopic analytical approaches, macroscopic simulation approaches and microscopic simulation approaches have been widely used. Among these approaches, macroscopic analytical approaches are the simplest. If users place more focus on the estimation accuracy, a microscopic simulation approach will be a good choice. Among microscopic simulation approaches, the CA model is a promising and effective method in estimating traffic delay because it not only has short computation time but also replicates work zone traffic very well. Because of these reasons, the CA model could be the best alternative if high accuracy and short execution time are desired in estimating traffic delay (Meng and Weng, 2013b). However, it should be pointed out that the current CA models are too simple. For example, the CA model developed by Meng and Weng (2011) did not take road networks into account. Future studies should extend the model by incorporating a road network. In future, the CA model should also take into account the possible lane changing behavior that vehicles in the through lane may change into other through lanes in order to avoid merging conflicts from the closed lane.

Beside of the CA model, current commercial microscopic traffic simulation software (e.g., PARAMICS and VISSIM) are also frequently used for traffic delay estimation because of their ability to describe the dynamic vehicle and driver behavior at a highly detailed level. However, these software face a great challenge that it is time-consuming to calculate the state evolution of vehicles. The computation time consumed by microscopic traffic simulation increases very fast as the road network expands and the number of vehicles increases. In previous studies, the integration of macroscopic analytical models and microscopic simulation software are proven to be an effective method to reduce the computation time. The parallelization of microscopic traffic simulation might be another effective method to improve computation efficiency. It will be therefore important to employ a parallel microscopic traffic simulation method for estimating traffic delay in future. Recently, multi-agent based traffic simulation models have been developed which can provide fast dynamic and agent-based traffic simulation. Hopefully, more studies in the future will use agent-based traffic simulation approaches to estimating traffic delay in work zones.

Acknowledgements

The authors thank the overview paper editor and five anonymous reviewers for their thorough review of the earlier versions of the paper. This study is supported by the Chinese National Natural Science Foundation (71210001).

References

Adeli, H., Jiang, X., 2003. Neuro-fuzzy logic model for freeway work zone capacity estimation. Journal of Transportation Engineering 129 (5), 484–493.

Al-Kaisy, A.F., Hall, F.L., 2001. Examination of effect of driver population at freeway reconstruction zones. Transportation Research Record 1776, 35–42.

Al-Kaisy, A., Hall, F., 2003. Guidelines for estimating capacity at freeway reconstruction zones. Journal of Transportation Engineering 129 (5), 572–577.

Al-Kaisy, A.F., Zhou, M., Hall, F., 2000. New insights into freeway capacity at work zones: an empirical case study. Transportation Research Record 1710, 154–160.

Arguea, D.F., 2006. Simulation-based Approach to Estimate the Capacity of a Temporary Freeway Work Zone Lane Closure. Master Degree of Engineering Disertation, University of Florida, U.S.

Avrenli, K.A., Benekohal, R., Ramezani, H., 2011. Determining the speed-flow relationship and capacity for a freeway work zone with no lane closure. In: Proceeding: 90th Transportation Research Board Annual Meeting, Washington, DC.

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.

Chan, H.W., 2002. Traffic Delay Due to Lane Closure. Bachelor of Engineering Dissertation, Department of Civil Engineering, National University of Singapore.

Chatterjee, I., Edara, P., Menneni, S., Sun, C., 2009. Replication of work zone capacity values in a simulation model. Transportation Research Record 2130, 138–148.

Chen, C.H., Schonfeld, P., 2006. Work zone optimization for multiple lane highway resurfacing projects with time constraints and an alternative route. In: Presented at 85th Transportation Research Board Annual Meeting, CD-ROM #06-1812.

Chien, S.I., Goulias, D.G., Yahalom, S., Chowdhury, S.M., 2002. Simulation-based estimates of delays at freeway work zones. Journal of Advanced Transportation 36 (2), 131–156.

Chitturi, M., Benekohal, R., 2004. Comparison of QUEWZ, FRESIM and QuickZone with field data for work zones. In: Proceedings of the 83rd Transportation Research Board Annual Meeting, CD-ROM. Washington, DC.

Dixon, K.K., Hummer, J.E., Lorscheider, A.R., 1996. Capacity for North Carolina freeway work zones. Transportation Research Record 1529, 27–34.

Dudek, C.L., Richards, S.H., 1981. Traffic Capacity Through Work Zones on Urban Freeways. Report FHWA/TX-81/28+228-6. Texas Department of Transportation, Austin.

Edara, P. K., Cottrell, B.H., 2007. Estimation of traffic mobility impacts at work zones: state of the practice. In: Proceedings of the 86th Transportation Research Board Annual Meeting, CD-ROM, Washington, DC.

Grillo, L.F., Datta, T.K., Catherine, H., 2008. Dynamic late lane merge system at freeway construction work zones. Transportation Research Record 2055, 3–10.

Gundaliya, P.J., Mathew, T.V., Dhingra, S.L., 2008. Heterogeneous traffic flow modeling for an arterial using grid based approach. Journal of Advanced Transportation 42 (4), 468–491.

HCM, 2010. Highway Capacity Manual 2010. Transportation Research Board.

Heaslip, K., Kondyli, A., Arguea, D., Elefteriadou, L., Sullivan, F., 2009. Estimation of freeway work zone capacity using simulation and field data. Transportation Research Record 2130, 16–24.

Heaslip, K., Jain, M., Elefteriadou, L., 2011. Estimation of arterial work zone capacity using simulation. Transportation Letters: The International Journal of Transportation Research 3 (2), 123–134.

Jiang, Y., 1999. Traffic capacity, speed and queue-discharge rate for Indiana's Four-lane freeway work zones. Transportation Research Record 1657, 10–17.

Jiang, Y., 2001. Estimation of traffic delays and vehicle queues at freeway work zones. In: Proceedings of the 80th Transportation Research Board Annual Meeting, CD-ROM, Washington, DC.

Jiang, X., Adeli, H., 2003. Freeway work zone traffic delay and cost optimization model. Journal of Transportation Engineering 129 (3), 230–241.

Karim, A., Adeli, H., 2003. Radial basis function neural network for work zone capacity and queue estimation. Journal of Transportation Engineering 129 (5), 494–503.
Kayani, W.I., Bham, G.H., 2010. Work zone capacity estimation: an analytical review. In: TRB 89th Annual Meeting Compendium of Papers DVD, Washington, DC.
Kianfar, J., Edara, P., Sun, C., 2012. Analytical methods for deriving work zone capacities from field data. Journal of Transportation Engineering 138 (6), 809–818.
Kim, C., 2008. A Mechanistic Model of Work Zone Capacity. University of California, Davis.
Kim, S., Son, Y., 2003. Development of a new merge capacity model and the effects of ramp flow on the merge capacity. In: Proceeding of the 82nd Annual Meeting, CD-ROM, Washington, DC.
Kim, T., Lovell, D.J., Paracha, J., 2001. A new methodology to estimate capacity for freeway work zones. In: Proceeding of the 80th Transportation Research Board Annual Meeting, CD-ROM, Washington, DC.
Krammes, R.A., Lopez, G.O., 1994. Updated capacity values for short-term freeway work zone lane closure. Transportation Research Record 1442, 49–56.
Martinelli, D.R., Xu, D., 1996. Delay estimation and optimal length for four-lane divided freeway work zones. Journal of Transportation Engineering 122 (2), 114–122.
Maze, T., Kamyab, A., 1999. Work Zone Simulation Model – Companion Report for Traffic Management Strategies for Merge Areas in Rural Interstate Work Zones. ISU CTRE Management Project 97-12, Iowa Department of Transportation.
McCoy, P.T., Pang, M., Post, E.R., 1980. Optimum length of two-lane two-way no-passing traffic operation in construction and maintenance zones on rural four-lane divided highways. Transportation Research Record 773, 20–24.
Meng, Q., Weng, J., 2010. A Cellular Automata model for work zone traffic. Transportation Research Record 2188, 131–139.
Meng, Q., Weng, J., 2011. An improved Cellular Automata model for heterogeneous work zone traffic. Transportation Research Part C: Emerging Technologies 19 (6), 1263–1275.
Meng, Q., Weng, J., 2013a. Optimal subwork zone operational strategy for short-term work zone projects in four-lane two-way freeways. Journal of Advanced Transportation 47, 151–169.
Meng, Q., Weng, J., 2013b. Effects of work zone configuration, traffic flow and heavy vehicle percentage on traffic delay at work zones. Asia Transport Studies 2 (3), 239–252.
Nassab, K., Schreckenerg, M., Boulmakoul, A., Ouaskit, S., 2006. Impacts of different t types of ramps on the traffic flow. Physica A 369, 841–852.
Ping, W.V., Zhu, K., 2006. Evaluation of work zone capacity estimation models: a computer simulation study. In: The 6th Asia–Pacific Transportation Development Conference, 19th ICTPA Annual Meeting, Hong Kong.
Racha, S., Chowdhury, M., Sarasua, W., Ma, Y., 2008. Analysis of work zone traffic behavior for planning applications. Transportation Planning and Technology 31 (2), 183–199.
Sarasua, W., Davis, W.J., Chowdhury, M.A., Ogle, J.H., 2006. Estimating interstate highway capacity for short-term work zone lane closures: development of methodology. Transportation Research Record 1948, 45–57.
Schattler, K.L., Datta, T.K., 2003. Assessing the impacts of the dynamic early lane merge traffic control system. In: Institute of Transportation Engineers 2003 Annual Meeting and Exhibit (Held in Conjunction with ITE District 6 Annual Meeting), CD-ROM.
Schonfeld, P.M., Chien, I.J., 1999. Optimal work zone length for two-lane highways. Journal of Transportation Engineering 125 (1), 21–29.
Shaaban, K., Harb, R., Radwan, E., 2011. Evaluation of new simplified dynamic lane merging systems (SDLMS) for short-term work zone lane closure configuration. Jordan Journal of Civil Engineering 5 (3), 446–458.
Tang, Y., Chien, S., 2008. Scheduling work zones for highway maintenance projects considering a discrete time-cost relation. Transportation Research Record 2055, 21–30.
Tarko, A., Venugopal, S., 2001. Safety and Capacity Evaluation of the Indiana Lane Merge System. FHWA/IN/JTRP-2000/19. Joint Transportation Research Program, Purdue University.
U.S. Department of Transportation, 2003. Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD), Federal Highway Administration.
Weng, J., Meng, Q., 2011. Decision tree-based model for estimation of work zone capacity. Transportation Research Record 2257, 40–50.
Weng, J., Meng, Q., 2012. Ensemble tree approach to estimating work zone capacity. Transportation Research Record 2286, 56–67.
Wu, Y.Y., 2000. Traffic Delay at Expressways Due to Lane Blockage. Bachelor of Engineering Dissertation, Department of Civil Engineering, National University of Singapore.
Yang, N., Schonfeld, P., Kang, M.W., 2008. A hybrid methodology for freeway work-zone optimization with time constraints. Public Works Management & Policy, 1–12.
Zheng, N., Hegyi, A., Hoogendoorn, S., van Zuylen, H., Peters, D., 2011. A comparison of freeway work zone capacity prediction models. Procedia Social and Behavioral Sciences 16, 19–429.