Incorporating Practical Degree of Saturation in Capacity Estimation of On-Street, Mid-Block, Off-Line Bus Stops

Faheema Hisham, Jonathan M. Bunker, and Ashish Bhaskar

Abstract

The effectiveness of an on-street bus facility depends on the general traffic that shares the lane used by buses. The Transit Capacity and Quality of Service Manual (TCQSM) methodology was used to estimate facility bus capacity based on critical stop operation. Hisham et al. provided an improved understanding of performance of an on-street, mid-block, off-line bus stop by relating bus stop capacity to the adjacent lane traffic flow rate. Using the TCQSM methodology, bus stop capacity was determined by adopting an estimated operation margin that relates to a design bus stop failure rate. However, failure rate is theoretically ambiguous and difficult to quantify in practice, particularly under high volume traffic conditions. In contrast, degree of saturation is a direct measure of operating conditions experienced by buses using the stop, and by the adjacent lane general traffic, so it directly affects approach delay and queuing. The aim of this study was to better understand performance at on-street, mid-block, off-line bus stops by considering degrees of saturation of loading areas and the adjacent lane rather than design failure rate according to the TCQSM methodology, and to ensure that waiting times upstream of bus stops are kept to acceptable levels by determining bus stop maximum working capacity.

Arterial roads are a type of on-street bus (OSB) facility where buses and other forms of traffic share the same lanes. The performance of an OSB facility is highly dependent on the interaction between buses and other vehicles. When the adjacent lane at a bus stop carries a high general traffic flow rate, the interaction between buses and traffic will affect the capacity and quality of service of the bus facility. It is therefore essential to understand the operation of any potentially critical bus stop to understand and manage such a facility.

The Transit Capacity and Quality of Service Manual (TCQSM) includes a deterministic design capacity methodology for a critical bus stop. The critical stop is that which has the lowest bus capacity; therefore it governs the bus facility capacity. The TCQSM methodology was developed to properly estimate the achievable capacity with regard to operation of the critical bus stop. Capacity of the critical stop is equal to the product of the capacity of each of its loading areas, the number of effective loading areas, and a traffic blockage adjustment factor.

The TCQSM methodology includes an operating margin on the loading area dwell time to give the bus a reasonable time to accommodate any irregularities in the dwell time. According to the TCQSM theory, the sum of the mean dwell time and the operating margin is the maximum amount of time a bus can dwell on a loading area without creating a “bus stop failure.” A failure is defined by the TCQSM as a situation that arises when a bus arrives to use a loading area only to find another bus is still occupying it.

However, given the TCQSM theory, failure can be defined more accurately as a situation that arises when two buses, isolated from any other buses, arrive to use the loading area consecutively at a headway equal to the inverse of the specified capacity of the loading area, but the first bus dwells on the loading area for a duration longer than the average dwell time plus a specified operating margin, requiring the second bus to wait until the first bus has cleared the loading area. Importantly, from this more accurate definition, it may be deduced that the TCQSM theory presumes that any further failure is fully attenuated once the second bus reaches the loading area. This presumption, however, is incorrect unless a third bus arrives at a headway equal to or less than the inverse
of the specified capacity of the loading area minus the excessive dwell time of the first bus. Consequently, the TCQSM theory only considers an isolated case and does not allow for accumulation of delay from successive buses arriving unevenly.

The TCQSM theory considers failure rate (FR) by assuming that dwell times vary according to a normal distribution. Therefore, addition of the operating margin for the desired FR to the mean dwell time achieves the design dwell time, which is then used in determination of a loading area design capacity that reflects a desired level of operational reliability.

Following the logic of the clarification made about failure, the application of operating margin and associated FR can be feasible when bus arrivals are relatively evenly spaced. However, as bus arrival flow rate and adjacent lane general traffic flow rate increase, buses are more susceptible to unevenly spaced arrivals, which would increase interference between buses and would lead to increased failures (1). With respect to OSB operation, interference to general traffic will also yield higher sensitivity to loading area failures. Therefore, with respect to the loading areas of a bus stop, “failure” is an attribute that could occur not only with the dwell time, but also with both interference between buses and clearance times. To overcome this issue, this study considered “failure” with respect to all contributing factors.

Moreover, the TCQSM methodology accounts for traffic blockage effects at stops within the influence of signalized intersections. However, it states not to consider traffic blockage effects beyond their influence, which is the case for the on-street, mid-block, off-line (OS-MID-OFF) bus stops, which are the subject of this study. Hisham et al. identified that the TCQSM methodology does not address the relationship between stop bus capacity and flow rate of adjacent lane traffic at this type of stop (1). They developed the “Bus Stop Capacity with Adjacent Lane Requirements” (BCAL) model to overcome this limitation. Even though TCQSM allows for high traffic flow rate in the adjacent lane for re-entry capacity estimation, the theory does not relate actual conditions when the adjacent lane operates at saturation.

In response to the aforementioned issues, the aim of this study was to better understand performance at OS-MID-OFF bus stops by considering degrees of saturation (DOS) of loading areas and the adjacent lane rather than design FR according to the TCQSM methodology, and to ensure that waiting times upstream of bus stops are kept to acceptable levels by determining bus stop maximum working capacity.

To achieve the study aim, the current literature on bus facility capacity is reviewed in the next section. Existing theory to determine the theoretical critical bus stop capacity is provided in Section 3. Section 4 develops a theoretical model that considers adjacent lane constraints along with the practical DOS of the adjacent lane and the bus stop itself. Sections 5 and 6 implement the model developed and compares it with the TCQSM methodology. Section 7 offers conclusions and recommendations for further research.

**Literature Review**

Numerous studies have focused on the effects of high traffic volumes on bus stops. Some studied the location (4), design (5) and operation (6) of bus stops. However, capacity estimation of a bus stop is analytically most important because the critical stop governs the capacity of the whole transit facility (7).

Some studies related bus facility capacity to the location of bus stops. Lin et al. conducted research on improving capacity for exclusive bus lanes consisting of on-line bus stops (8). They found that moving the stop upstream of an intersection could increase capacities by 20% and that greater signal cycle times tend to reduce capacity. However, they found this to be inefficient owing to traffic delays that can be caused by a dwelling bus. Fitzpatrick et al. stated that off-line bus stops are separated from the traffic lanes to provide convenience to boarding and alighting passengers (9). At off-line stops, adjacent lane general traffic can pass without obstruction while a bus is dwelling. They are suitable at locations with high traffic volume, high speed roadways, or sections with a high number of boarding and alighting passengers. Zhao et al. investigated the combined effect of signalized intersection and off-line bus stops (10).

Levinson and St. Jacques (11) modified the *Highway Capacity Manual* (HCM) (12) equation for bus capacity estimation by use of field studies and simulations. They related bus capacity and FR to estimate maximum achievable capacity. They used a coefficient of variation of dwell times of 60% for the new estimate and concluded that maximum achievable capacity corresponded to a 25% FR. Several studies have determined facility bus capacity using a specific value of FR. However, for a bus stop with a single loading area Gu et al. defined FR differently from any other study (6). For uniform bus arrivals, they assumed that bus service time follows an Erlang-k distribution. The ratio of bus inflow (λ), to the loading area service time (μ) is set to be equal to $FR_{total}=\frac{C_{t}}{C_{s}}$; where $C_{t}$ is the coefficient of variation of the service time.

Some studies considered stochasticity and randomness in estimation of facility bus capacity. Ortiz and Bocarejo estimated capacity of the TransMilenio in Bogotá using a VISSIM microscopic simulation platform (13). They quantified the difference in capacities when randomness of bus system operations are included. Siddique and
Khan used NETSIM, a stochastic microscopic simulation platform to evaluate bus rapid transit facility capacity in Ottawa, Canada, under a 20-year horizon (14). With three scenarios presented, they compared estimated capacity with that of the TCQSM deterministic model to highlight the importance of incorporating stochasticity.

The TCQSM methodology includes a table of FR as a percentage, representing the probability that bus stop capacity is exceeded, and relates this to the desired level of service. This was developed through simulation under particular conditions. It also provides the theoretical means of estimating operating margin. Mathematically, bus stop capacity is greatest when FR is set to 50%, however, this would represent a case with a constant upstream queue. Therefore, based on simulation results they suggest 25% as the optimal FR to achieve a maximum design capacity. However, as discussed, the TCQSM definition of FR is theoretically ambiguous, and the means of prescribing FR is not sufficiently robust.

Therefore, some researchers have developed novel approaches instead of the FR approach in capacity estimation. Fernandez and Planzer identified that bus stop DOS is an important measure in bus stop capacity estimation (15). The DOS of the bus stop indicates how busy it is. This information can be used in designing bus stops to estimate a suitable combination of bus flow and passenger demand. However, the estimated DOS does not relate to its capacity estimation. Moreover, this estimation is particular to exclusive bus lanes. It would be useful to have a generic methodology at our disposal to estimate bus stop capacity, by which it is possible to control the DOS of the bus stop by ensuring an adequate service level in the bus stop area. Further, an OSB facility with an adjacent lane has not been considered in the literature and has been identified as a key gap in this research.

This study addresses two major gaps found in the literature review. First, the TCQSM theory to estimate stop bus capacity uses the approach of FR on dwell time. Even though several studies have used the TCQSM methodology by using the design FR, the accuracy of the definition of failure and its actual implications have not been sufficiently studied. DOS of a bus stop has been identified in the literature as a crucial parameter that ensures a desired level of service of a bus stop is reached, however it has not been incorporated into capacity estimation. Second, the TCQSM accounts for a reduction in bus stop capacity with an increase in adjacent lane general traffic flow rate only in relation to the bus re-entry gap acceptance process. However, the theory does not account for the time required by the adjacent lane to pass without exceeding practical saturation flow rate.

In seeking to achieve the central aim, this study also sought to fill the knowledge gap in understanding the operation of an OS-MID-OFF bus stop by building on the theoretical relationship between stop bus capacity and adjacent lane traffic volume developed by Hisham et al. (1) with consideration to practical DOS on the bus stop and adjacent general traffic lane.

**TCQSM Capacity of an On-Street, Mid-Block, Off-Line Stop**

TCQSM (2) presents a deterministic model to estimate theoretical capacity of a critical bus stop of a bus facility. A loading area is defined as a section of the stop that is designated for a single bus to stop and dwell to serve passengers. This study was concerned with an OS-MID-OFF bus stop whereby vehicles can only pass dwelling buses in the adjacent lane. Our testbed included a linear platform with two loading areas.

The TCQSM model implies that the service time per bus per loading area is equal to the sum of clearance time (s), dwell time component during green time (s), and the operating margin (s), which is extra time allocated to account for longer than average dwell time that could occur during dwell times longer than average. Therefore, bus stop capacity during a given period of time is expressed as follows:

\[
B = \frac{3600(g/C)}{t_c + t_d(g/C) + t_{om}} N_{el} f_{ib} \tag{1}
\]

where

- \(B\) = bus stop capacity (bus/h),
- \(g/C\) = green time ratio, equal to 1.0 for an OS-MID-OFF stop,
- \(t_c\) = clearance time (s),
- \(t_d\) = average dwell time (s),
- \(t_{om}\) = operating margin on dwell time,
- \(N_{el}\) = number of effective loading areas, and
- \(f_{ib}\) = traffic blockage adjustment factor, equal to 1.0 for an OS-MID-OFF stop.

Green time ratio is included in Equation 1 as a factor in the numerator of the loading area capacity quotient to reflect that the buses cannot access any loading area immediately upstream or downstream of a signalized intersection during red time periods, and as a factor of average dwell time to account only for the portion occurring during green time. If there is no immediate signalized intersection, which is the case at a mid-block bus stop, then the ratio \(g/C\) equals 1.0. Traffic blockage adjustment factor, \(f_{ib}\), represents the bus stop capacity reduction that could occur owing to general traffic requiring some capacity of the lane used by buses at a stop adjacent to a signalized intersection. However, at a mid-block bus stop away from the influence of a signalized intersection,
no traffic blockage adjustment applies so the factor equals 1.0.

Clearance time is the sum of the time taken by a bus to start up and travel its own length and the next bus to pull in, \( t_{su} \), and the re-entry delay to the bus while waiting for a gap in the adjacent lane, \( t_{re} \). Start-up time has a fixed value and corresponds to the mechanical and dimensional properties of the buses, whereas re-entry delay can vary depending on the stop attributes. When a bus stop is located away from an upstream traffic signal and outside the influence of a downstream traffic signal, traffic is assumed to arrive randomly past the bus stop.

In such a situation, the re-entry delay from gap acceptance is given by

\[
 t_{re} = \frac{3600}{c_{re}} + 900 \left[ \frac{N_{la}}{c_{re}} - 1 + \sqrt{\left( \frac{N_{la}}{c_{re}} - 1 \right)^2 + \frac{3600}{c_{re}} \left( \frac{N_{la}}{c_{re}} \right) \left( \frac{1}{450} \right)} \right] - 3.3
\]

(2)

where

- \( t_{re} \) is re-entry delay because of gap acceptance (s),
- \( c_{re} \) is capacity of the re-entry movement (vehicles per hour [vph]),
- \( N_{la} \) is number of actual loading areas,
- \( v \) is demand flow rate in the adjacent lane (vph),
- \( t_{ch} \) is critical headway of the re-entry movement (s) = 7.0s, and
- \( t_{f} \) is follow-up headway for the re-entry movement (s) = 3.3s.

Referring to Equation 1, the TCQSM model returns a design stop capacity on the basis of an operating margin on dwell time, calculated for a predefined FR. Although the model is deterministic, it assumes dwell time to vary normally. The FR is used in combination with dwell time variability and the average dwell time to provide an operating margin. This is calculated by selecting the standard normal variable for the predefined design FR and applying it along with coefficient of variation of dwell time \( (C_v) \) and the average dwell time to estimate the operating margin. Mathematically,

\[
 t_{om} = ZC_v t_d
\]

(4)

By predefining an FR, the TCQSM model assumes that no upstream queue will be formed for any proportion of the study period beyond an amount equal to the FR during dwell time for an isolated bus. The previous section discussed issues with this approach.

### Improved Model for Maximum Working Capacity of an On-Street, Mid-Block, Off-Line Stop

This section presents the development of a model that modifies the BCAL model developed by Hisham et al. by considering bus stop and adjacent general traffic lane DOS (1). This is called the “Bus Stop Maximum Working Capacity with Adjacent Lane Traffic” (BMWCA) model.

In the BMWCA model we consider loading area operation as being the fundamental building block of bus stop operation. The loading area net average total processing time per bus is given by

\[
 t_{la, net} = (t_{su} + t_{bhi} + t_{d} + t_{re})
\]

(5)

where \( t_{bhi} \) is a time increment resulting from interference between buses at the stop.

The loading area average total processing time per bus, \( t_{la} \), may be considered as the sum of fundamental components above, inclusive of processing margin, as follows:

\[
 t_{la} = (t_{su} + t_{bhi} + t_{d} + t_{re} + t_{pm})
\]

(6)

where \( t_{pm} \) is the processing margin on loading area net average total processing time per bus (as distinct from operating margin, which applies only to dwell time).

The BCAL model implies a maximum feasible DOS of the bus stop itself of 1.0, should a value of zero be assigned for the operating margin on dwell time. This would also be the case for the TCQSM theory. Similarly, should a value of zero be assigned for the processing margin in Equation 6, this would imply a maximum feasible DOS of the bus stop itself as 1.0. Processing margin can therefore be defined in relation to the loading area average total processing time per bus and a specified loading area maximum working DOS, \( X_{la,mw} \) as follows:

\[
 t_{pm} = t_{la}(1 - X_{la,mw})
\]

(7)

Equation 7 ensures that the loading area remains idle for a portion of loading area average total processing time per bus, which is equal to the additive inverse of the specified loading area maximum working DOS. From Equations 6 and 7 we define the processing margin as

\[
 t_{pm} = \frac{t_{la, net}(1 - X_{la,mw})}{X_{la,mw}}
\]

(8)

The time available for the adjacent lane to pass by the stop during the loading area average total processing time per bus is equal to the sum of its time components during which the bus does not obstruct the adjacent lane, and is given by...
\[ t_{al} = t_d + t_{re} + t_{pm} \] (9)

An important difference between the BMWCA model and the TCQSM model is that we acknowledge that general traffic in the lane adjacent to a given loading area has a theoretical capacity (vph), which Hisham et al. (1) defined as

\[ c_{al} = s_{al} \left( \frac{t_{al}}{X_{al}} \right) \] (10)

where \( s_{al} = \) adjacent lane saturation flow rate (vph).

The DOS of general traffic in the adjacent lane is given by

\[ X_{al} = \left( \frac{v_{al}}{c_{al}} \right) \] (11)

where \( v_{al} = \) adjacent lane general traffic arrival flow rate (vph).

Equation 6 requires bus re-entry delay to be quantified. The BMWCA model maintains the gap acceptance approach according to Equations 2 and 3. However, we acknowledge that the adjacent lane traffic is obstructed during start-up time and bus–bus interference time. Therefore, the merging bus driver will see a compressed traffic stream in the adjacent lane passing by the loading area during other times. For purposes of estimating re-entry delay owing to gap acceptance, Hisham et al. made the following adjustment to adjacent lane traffic flow rate for bus re-entry gap acceptance (1):

\[ v_{al}^* = v_{al} \left( \frac{t_{la}}{t_{al}} \right) \] (12)

Using Equations 10, 11 and 12, the adjusted adjacent lane traffic flow rate is given by

\[ v_{al}^* = X_{al} s_{al} \] (13)

Under this model, this adjusted value is applied to Equations 2 and 3 to calculate re-entry delay, \( t_{re} \).

The interference between buses at a bus stop may be reflected by a bus–bus interference factor as follows (1):

\[ f_{bbi} = \frac{N_{el}}{N_{la}} \] (14)

where \( N_{el} \) is the number of effective loading areas according to TCQSM values, and \( N_{la} \) is the number of actual loading areas.

Hisham et al. estimated the additional time component toward loading area average total processing time per bus, owing to bus–bus interference (s/bus), as a margin on the sum of the time components of loading area average processing time per bus, excluding processing margin (16), which we modified as follows:

\[ t_{bbi} = (t_{al} + t_{re} + t_d) \left( \frac{1}{f_{bbi}} - 1 \right) \] (15)

The system of Equations 5 through 15 allows us to determine the loading area average total processing time per bus, and the time available for the adjacent lane to pass by the stop during the average total loading area processing time per bus, provided that all inputs are known. The start-up time and average dwell time are typically inputs to an analysis. However, according to Equation 13, knowledge is required of the adjacent lane DOS to determine bus re-entry delay. The additional time component toward average total processing time per bus because of bus–bus interference (s/bus) can then be determined directly using Equation 15. Further, according to Equation 8, calculation of the processing margin requires knowledge of the loading area DOS. Therefore, these two unknown DOS must be resolved to solve the whole system of equations.

This must be undertaken in two steps. First, optimal adjacent lane flow rate must be determined, which is the highest adjacent lane flow rate at which the bus stop’s loading areas (assuming equal utilization) operate at their common practical DOS, \( X_{al,b} \), and the point at which the adjacent lane reaches its practical DOS, \( X_{al,p} \). We define practical DOS as the greatest value that yields an acceptable delay, and therefore uncongested operation.

It may be proven mathematically that this point is determined directly by

\[ v_{al,opt} = \left( X_{al,p} s_{al} \right) \left( 1 - X_{la,p} \left( 1 - \left( \frac{t_{re} + t_d}{t_{la,net}} \right) \right) \right) \] (16)

where re-entry delay, \( t_{re} \), is calculated using adjusted adjacent lane flow rate, \( v_{al}^* = X_{al,p} s_{al} \). Here \( X_{al,p} \) and \( X_{la,p} \) are both specified directly, as is discussed in the next section.

Second, the adjacent lane DOS and loading area maximum working DOS are determined. Here, \( v_{al} \) needs separating into two regimes, by comparing it with the optimal flow rate, \( v_{al,opt} \). Regime 1 occurs when \( v_{al} \leq v_{al,opt} \). Regime 2 occurs when \( v_{al} > v_{al,opt} \).

Adjacent lane working DOS is then calculated under each regime as follows:

\[ X_{al} = \begin{cases} \left( \frac{v_{al}}{v_{al,opt}} \right) \left( 1 - X_{la,p} \left( 1 - \left( \frac{t_{re} + t_d}{t_{la,net}} \right) \right) \right)^{-1}, & v_{al} \leq v_{al,opt} \\ X_{al,p}, & v_{al} > v_{al,opt} \end{cases} \] (17)

In Regime 1, \( X_{al} \) needs to be estimated as a function of re-entry delay and net average total processing time per bus. To calculate these two components of processing time the adjusted adjacent lane flow rate is used, which is recursively a function of adjacent lane DOS. Therefore,
it is necessary to apply the following objective function to determine the adjacent lane DOS for the given flow rate:

$$X_{al} = \arg\min_{X_{al, trial}}(\text{abs}(X_{al} - X_{al, trial}))$$

(18)

where a suitable initial trial value for the objective function is $$X_{al, trial} = \frac{1}{2}X_{al, opt}$$.

Loading area maximum working DOS is then dependent on adjacent lane flow rate regime as follows:

$$X_{la,mw} = \left\{ \begin{array}{ll}
\frac{X_{la,p}}{1 - \left( \frac{v_{al}}{v_{al, p}} \right)} & v_{al} \leq v_{al, opt} \\
\frac{1}{1 - \left( \frac{t_{la, net}}{t_{la, p, net}} \right)} & v_{al} > v_{al, opt}
\end{array} \right.$$

(19)

Once the loading area maximum working DOS is determined for the correct regime, the stop maximum working bus capacity may be determined as follows:

$$B_{x,mw} = \frac{3600X_{la}N_{la}}{t_{la, net}}$$

(20)

**Practical Degrees of Saturation at an On-street, Mid-Block, Off-line Bus Stop**

Equations 16 to 20 require that loading area practical DOS, $$X_{la,p}$$, and adjacent lane practical DOS, $$X_{al,p}$$, are specified to calculate the following: optimal adjacent lane flow rate, adjacent lane DOS when adjacent lane flow rate is less than optimal, and loading area DOS when adjacent lane flow rate is greater than optimal.

First adjacent lane practical DOS is considered. If we specify this value to be constant across all adjacent lane flow rates, then according to the aforementioned theory, when adjacent lane flow rate is less than optimal adjacent lane flow rate, adjacent lane working DOS will be less than practical DOS. When adjacent lane flow rate is equal to or greater than optimal adjacent lane flow rate, the working DOS will be equal to practical DOS.

We looked to an analog for arterial road bottleneck DOS. For the common constraint point of a signalized intersection, the typical default recommended practical DOS is equal to 0.9 (17). We consider that this value is suitable for the case of adjacent lane traffic passing a OS-MID-OFF bus stop on an arterial road.

Second we consider the loading area practical DOS. According to the aforementioned theory, in Regime 1 where adjacent lane flow rate is less than or equal to optimal adjacent lane flow rate, loading area working DOS is equal to practical DOS. In Regime 2, where adjacent lane flow rate is greater than optimal adjacent lane flow rate, loading area working DOS will decline from practical DOS to zero at the point where the adjacent lane reaches practical saturation, $$v_{al} = X_{al} X_{al,p}$$.

Along with its value, the assumption about whether loading area practical DOS should remain constant with adjacent lane flow rate in Regimes 1 and 2 requires careful consideration.

Bunker discussed that the processing of buses through a loading area of a bus stop has similar characteristics to the operation of an unsignalized intersection (18). However, the loading area as a server is subject to less fluctuation than the head of the queue on a minor movement at an unsignalized intersection. It was noted that the rate of increase in upstream average waiting time with DOS is less pronounced. However, the waiting time upstream of a loading area at a bus stop is more consequential, because of the effects of bus queuing on bus stop and adjacent lane operation. The following equation was developed to estimate upstream average waiting time:

$$t_{w,la} = 600T \left( X_{la} - 1 \right) + \left( X_{la} - 1 \right)^{2} + \frac{t_{la,net}X_{la}}{450T}$$

(21)

where system time $$T = 1.0 h$$.

Bunker discussed that Equation 21 is scalable (18). Where there are multiple loading areas, and assuming that all loading areas at the bus stop are equally utilized, the upstream average waiting time applies to all buses accessing the bus stop. It is important to note that in Equation 21, for a given adjacent lane flow rate, $$t_{la,net}$$ will be constant regardless of loading area DOS.

In Regime 1, adjacent lane flow rate is less than or equal to optimal adjacent lane flow rate. Equation 21 may be rearranged to determine an appropriate loading area practical DOS for a specified practical upstream average waiting time,

$$X_{la,p} = \frac{1 + \frac{t_{w,la,p}}{1200T}}{1 + \frac{2t_{la,net}}{3t_{w,la,p}}}$$

(22)

where $$t_{w,la}$$ is a specified practical upstream average waiting time, and $$t_{la,net}$$ corresponds to a given adjacent lane flow rate.

Loading area practical DOS should not cause excessive upstream average waiting time, particularly as adjacent lane flow rate approaches optimal, which corresponds to its own practical DOS. Bus drivers arriving at the stop to access a loading area should be able to do so within a time associated with the mechanical and geometric properties of the buses alone, and not components affected by demand fluctuation including dwell time, re-entry delay, and bus–bus interference time.
Therefore, we recommend that to determine optimal adjacent lane flow rate and associated loading area practical DOS, practical upstream average waiting time is limited to a value equal to the start-up time between buses.

Equation 17 enables loading area working DOS to be calculated in Regime 2. Equation 22 enables loading area upstream average waiting time to be calculated. Mathematically, these values will both be less than the respective values at optimal adjacent lane flow rate.

In Regime 1, where adjacent lane flow rate is less than or equal to the optimal value, Equations 17 to 19 and 22 may be solved recursively to determine both loading area working (equal to practical) DOS for a specified upstream average waiting time, and adjacent lane working DOS. Mathematically, loading area practical DOS will increase slightly as adjacent lane flow rate decreases.

Figure 1 shows a flowchart of the procedure for capacity estimation according to the methodology from Equations 1 through 22.

### Comparison between TCQSM Model and BMWCA Model

For direct comparison, we determine the theoretical bus capacity of an OS-MID-OFF bus stop using the TCQSM model based on Equation 1 and the BMWCA model of Equations 4 through 20 under conditions where adjacent lane general traffic has absolute priority over re-entering buses. We assign the following values: mean dwell time of 20s to reflect a typical bus stop operation; start-up component of clearance time of 10s for standard bus (19); re-entry delay using TCQSM default values of 7.0 s for critical headway and 3.3 s for follow-up headway; bus
stop designated to contain two actual loading areas and 1.85 effective loading areas according to TCQSM; TCQSM traffic blockage factor equal to 1.0. To estimate bus stop working capacity and DOS across a full range of adjacent lane flow rates using the BMWCA model, optimal adjacent lane flow rate must first be determined: the corresponding adjusted adjacent lane flow rate, $v_{opt}^a = X_{al,p}^a s_{al} = 1620$ vph. Using Equations 2 and 3 the corresponding re-entry delay, $t_{re} = 37.7$ s. Using Equation 15 the bus–bus interference time, $t_{bhi} = 5.5$ s. Using Equation 6 loading area net average total processing time per bus, $t_{al,net} = 73.2$ s. Using Equation 22 with a specified practical upstream average waiting time of 10 s, loading area practical DOS, $X_{al,p} = 0.17$. Finally, using Equation 16 the optimal adjacent lane flow rate, $v_{al,opt} = 1561$ vph.

We then use the routine illustrated in the flow chart of Figure 1 to determine bus stop working capacity and working DOS, as well as adjacent lane working DOS, all of which satisfy their practical limits, across a range of adjacent lane flow rate, $v_{al}$, between 0 vph and 1,620 vph. Figure 2 illustrates the relationship between bus stop working capacity and adjacent lane flow rate using the BMWCA model as well as the TCQSM model results for comparison.

The three marked curves in Figure 2 correspond to the BMWCA model; the green marked curve corresponds to a scenario with a recommended upstream average waiting time of 10 s (equal to the start-up time between buses). The two previously discussed regimes are apparent in each of the BMWCA curves, corresponding to each practical upstream average waiting time. The left side of the curve represents the regime where $v_{al} \leq v_{opt}$. When there is no adjacent lane flow, the upstream queue will comprise of buses only, in which the greatest value of maximum working capacity of the bus stop can be achieved. Depending on the assigned upstream average waiting time, the loading area DOS will be constrained. For instance, where there is no traffic flow in the adjacent lane, to limit the maximum waiting time to 10 s, the loading area will be able to operate at a maximum working DOS equal to 0.32 when the adjacent lane operates at a DOS equal to 0.

As adjacent lane flow rate increases, both bus re-entry time and bus–bus interference time increase, leading to an increase in the loading area net average total processing time per bus. According to Equation 22, when the assigned upstream average waiting time is held constant, as adjacent lane flow rate increases, the increase in the processing margin will result in a decrease in the loading area practical (and working) DOS that can be afforded. This in turn corresponds to an increase in required processing margin relative to loading area average total processing time per bus. As a consequence of these phenomena, loading area average total processing time per bus gradually increases with adjacent lane flow rate, resulting in a gradual reduction in bus stop maximum working capacity.

The right-hand side of the curve represents the regime where $v_{al} > v_{opt}$, such that the adjacent lane is operating at practical DOS. In Equation 22 the only variable parameter in this regime is adjacent lane flow rate. According to this equation, loading area working DOS reduces linearly to a value of zero when adjacent lane flow rate reaches practical saturation flow rate, $X_{al,p}^a$. The amount of time available to accommodate the processing of any buses on the loading area tends toward zero accordingly.

The yellow long-dashed and red short-dashed curves represent cases with 20-s and 30-s upstream average waiting times respectively. If the analyst considered an upstream average waiting time of 30 s to be tolerable for the bus stop configuration, almost twice the bus stop capacity could be achieved than that with a 10-s upstream average waiting time. Although higher upstream average waiting times can have high outputs in relation to bus stop maximum working capacity, because of increased travel time it can result in a worsening of quality of service both for passengers within the bus and for passengers waiting to board. High waiting times can also lead to excessive queue lengths for OSB operation, which can have an impact on the adjacent lane general traffic capacity and delay. For instance, at an adjacent lane flow rate of 600 vph, upstream average bus queue lengths will be 0.30, 0.93, and 1.73 buses for assigned upstream average waiting times of 10 s, 20 s, and 30 s, respectively.
Discussion

We highlight several key points from the results of the analysis of the testbed example illustrated in Figure 2.

- The BMWCA curves reduced more steeply than the TCQSM curves as adjacent lane flow rate increased, owing to the account made for the compressed adjacent lane stream during bus re-entry, and the adjacent lane capacity reduction from bus processing time requirements. TCQSM did not consider the condition where the adjacent lane reaches the point of saturation (1) whereas the stop capacity of the BMWCA model necessarily reached zero at its practical saturation flow rate (1,620 vph in this case) because there was insufficient time-space to accommodate any buses.
- Capacity at low adjacent lane flow rate was very similar between the TCQSM model and the BMWCA model for a 30 s upstream average waiting time. The BMWCA model showed that what may seem conservative when assuming a 5% FR using the TCQSM method still did not account for other stochastic influences, especially in relation to upstream queue formation because of upstream bus arrival headways.
- A feasible capacity estimation cannot be made when an operating margin is not used in the TCQSM model. By extrapolating the BMWCA family of curves, extremely excessive wait times would be expected.
- Bus stop maximum working capacity of the 10-s upstream average waiting time curve was approximately half that of the 30 s curve, across the entire Regime 1. Therefore, assignment of an appropriate upstream average waiting time is very important and must depend on site-specific conditions related to both acceptable passenger delay and the ability to store a bus queue upstream of the loading areas.
- Conditions in Regime 2 under the BMWCA model were highly volatile and should be avoided. Notwithstanding, we would not normally expect to see such high adjacent lane flow rates on an urban arterial road because they would correspond to fully saturated conditions at a mid-block location away from any influence of signalized intersections.

Conclusions

The study developed an empirical model to estimate stop bus capacity by incorporating the criterion of practical DOS, which is a novel approach toward bus stop capacity estimation and provides better insights from an operational perspective. In a manner similar to intersection analysis, we introduced practical DOS for the adjacent lane and the loading area to maintain a desired level of operational reliability in the bus stop area. To maintain
Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs
Faheema Hisham https://orcid.org/0000-0001-9703-0983
Jonathan M. Bunker https://orcid.org/0000-0002-4120-235X
Ashish Bhaskar https://orcid.org/0000-0001-9679-5706

References
1. Hisham, F., J. M. Bunker, and A. Bhaskar. Capacity Estimation of On-Street, Mid-Block, Off-Line Bus Stops Considering Yield-to-Bus Rule. Transportation Research Record: Journal of the Transportation Research Board, 2019. 2673: 269–278.
2. National Academies of Sciences and Engineering, and Medicine. Transit Capacity and Quality of Service Manual, 3rd ed. The National Academies Press, Washington, DC, 2013.
3. Jaiswal, S., J. M. Bunker, and L. Ferreira. Modelling Bus Lost Time: An Additional Parameter Influencing Bus Dwell Time and Station Platform Capacity at a BRT Station Platform. Proc., TRB 89th Annual Meeting Compendium of Papers, Washington DC, Transportation Research Board of the National Academies, January 10–14, 2010.
4. Ibeas, A., L. dell'Olio, B. Alonso, and O. Sainz. Optimizing Bus Stop Spacing in Urban Areas. Transportation Research Part E: Logistics and Transportation Review, Vol. 46, 2010, pp. 446–458.
5. Szeto, W. Y., and Y. Wu. A Simultaneous Bus Route Design and Frequency Setting Problem for Tin Shui Wai, Hong Kong. European Journal of Operational Research, Vol. 209, 2011, pp. 141–155.
6. Gu, W., Y. Li, M. J. Cassidy, and J. B. Griswold. On the Capacity of Isolated, Curbside Bus Stops. Transportation Research Part B: Methodological, Vol. 45, 2011, pp. 714–723.
7. St. Jacques, K. R. S., and H. S. Levinson. Operational Analysis of Bus Lanes on Arterials: Application and Refinement. Transportation Research Board, National Research Council, Washington, DC, 1997.
8. Lin, F.-B., P.-Y. Tseng, and C.-W. Chang. Capacities of Exclusive Bus Lanes with On-Line Linear Bus Stops on Urban Arterials. Proc., Eastern Asia Society for Transportation Studies, Washington, DC, Eastern Asia Society for Transportation Studies, 2011, p. 311.
9. Fitzpatrick, K., K. Hall, D. Perkinson, L. Nowlin, and R. Koppa. TCRP Report 19: Guidelines for the Location and Design of Bus Stops. Transportation Research Board of the National Academies, Washington, DC, 1996.

Author Contributions
The authors confirm contribution to the paper as follows: study conception and design: J. M. Bunker, F. Hisham; data collection: F. Hisham; analysis and interpretation of results: F. Hisham; draft manuscript preparation: F. Hisham, J. M. Bunker, A. Bhaskar. All authors reviewed the results and approved the final version of the manuscript.

this DOS, we introduced the “processing margin” as an additional component of loading area processing time per bus. This ensures that there is sufficient time between buses so that chaotic events such as bus–bus interference, bus bunching, and excessive queuing are taken into consideration.

Both the TCQSM and BMWCA models incorporate slack time components, however they are derived from two different concepts. On the one hand, the TCQSM operating margin is a fixed value that corresponds to a design FR on dwell time alone. However, this phenomenon of failure cannot be measured or observed directly. On the other hand, the BMWCA processing margin is a direct effect of the practical DOS of the loading areas of the bus stop, which accounts for the stochastic processing of buses, and acts as extra time allocated for bus stops at the loading area level, to maintain a desired performance.

The BMWCA model incorporates a recursive algorithm. The model first determines the highest adjacent flow rate at which the bus stop’s loading areas operate at their practical DOS. Specific to whether the adjacent lane flow rate is greater than, equal to, or less than the optimal adjacent flow rate, the DOS of the adjacent lane and the DOS of the loading area can be estimated. Practical DOS of the adjacent lane is an input for the BMWCA model. The BMWCA model proposes a methodology to estimate the practical DOS of the bus stop according to an assigned upstream average waiting time.

The results obtained by the BMWCA model showed that the upstream waiting time parameter is important in capacity estimation and quality of service because it directly affects the passenger either on the bus or waiting to board. Even though higher upstream waiting times result in bus stop maximum working capacities, the transit analyst must specify a suitable value such that the desired operational efficiency is achieved both in relation to passenger delay and upstream bus queue length. The case study demonstrated the strong need for setting policy about acceptable upstream average waiting time.

This study considered OS-MID-OFF bus stops. Future research will extend the BMWCA model for bus rapid transit stations, as well as off-line bus stops located at signalized intersections, particularly because these locations are common locations of candidate critical stops for OSB facilities.
10. Zhao, X. M., Z. Y. Gao, and B. Jia. 2007. The Capacity Drop Caused by the Combined Effect of the Intersection and the Bus Stop in a CA Model. *Physica A: Statistical Mechanics and its Applications*, Vol. 385, 2007, pp. 645–658.

11. Levinson, H., and K. St. Jacques. Bus Lane Capacity Revisited. *Transportation Research Record: Journal of the Transportation Research Board*, 1998. 1618: 189–199.

12. HCM 1985. *Highway Capacity Manual*, 1st ed. Transportation Research Board, Washington, DC, 1985.

13. Ortiz, M. A., and J. P. Bocarejo. Transmilenio Brt Capacity Determination using a Microsimulation Model in Vis-sim. *Proc., 93rd Annual Meeting of Transportation Research Board*, National Academies, Washington, DC, 2014.

14. Siddique, A. J., and A. M. Khan. Microscopic Simulation Approach to Capacity Analysis of Bus Rapid Transit Corridors. *Journal of Public Transportation*, Vol. 9, 2006, p. 10.

15. Fernandez, R., and R. Planzer. On the Capacity of Bus Transit Systems. *Transport Reviews*, Vol. 22, 2002, pp. 267–293.

16. Hisham, F., J. M. Bunker, and A. Bhaskar. Development of a Modified Bus Stop Capacity Model. *Proc., 97th Annual Meeting of the Transportation Research Board*, 7–11 January 2018, Washington DC, Transportation Research Board of the National Academies.

17. Akcelik, R. *Time-Dependent Expressions for Delay, Stop Rate and Queue Length at Traffic Signals*. Australian Road Research Board, Melbourne, 1980.

18. Bunker, J. M. High Volume Bus Stop Upstream Average Waiting Time for Working Capacity and Quality of Service. *Public Transport*, Vol. 10, 2018, pp. 311–333.

19. Levinson, H. S. *Operational Analysis of Bus Lanes on Arterials*. Transportation Research Board, Washington, DC, 1997.