

An exact modelling of the uniform control traffic delay in undersaturated signalized intersections

Wissam Fawaz^{1*} and John El Khoury²

¹Department of Electrical and Computer Engineering, Lebanese American University, PO Box 36, Byblos, Lebanon

²Department of Civil Engineering, Lebanese American University, PO Box 36, Byblos, Lebanon

SUMMARY

The average delay experienced by vehicles at a signalized intersection defines the level of service (LOS) at which the intersection operates. A major challenge in this regard is the ability to accurately estimate all the components underlying the overall control delay, including the uniform, incremental and initial queue delays. This paper tackles this challenging task by proposing a novel exact model of the uniform control delay component with a view to enhancing the accuracy of the existing approximate models, notably, the one reported in the Highway Capacity Manual 2010. Both graphical and analytical proofs are employed to derive exact closed-form expressions for the uniform control delay at undersaturated signalized intersections.

The high degree of accuracy of the proposed models is analysed through extensive simulations to demonstrate their abilities to exactly characterize the performance of real-life intersections in terms of the resulting vehicle delay. Unlike the existing widely adopted uniform delay models, which tend to overestimate the LOS of real-life intersections, the delay models introduced in this paper have the merit of exactly capturing such a LOS. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS: signalized intersections; control delay analysis; level of service; performance analysis

1. INTRODUCTION

The traffic delay observed at a signalized intersection is an important performance measure that is widely used in traffic engineering for the purpose of assessing the operational efficiency of an intersection. In particular, vehicle delay is at the heart of the determination of an intersection's level of service (LOS). In that particular regard, the methodologies described in the current Highway Capacity Manual (HCM2010) [1] are used to assign the LOS based on the average control delay incurred by vehicles at the intersection approaches.

According to Dion *et al.* [2], the delay at signalized intersections is the difference between the travel time that is experienced by a vehicle while traversing a signalized intersection and the travel time that the same vehicle would have experienced in the absence of traffic signal control. Similarly, delay in the HCM2010 is defined as 'the additional travel time experienced by a driver, passenger, bicyclist, or pedestrian beyond that is required to travel at the desired speed'. The signalized intersection chapter (i.e. Chapter 18) of the HCM2010 provides equations for calculating control delay, that is, the delay a vehicle experiences due to the presence of the traffic signal and conflicting traffic. The control delay equation comprises three elements, namely, uniform, incremental and initial queue delay values.

This paper is concerned with descriptive models of traffic flow rather than prescriptive ones, where the main focus is signal timing. The main reason for focusing on descriptive models is that these models are an essential element in the process of formulating optimal signal control strategies. Existing delay models, such as the one presented in [3], revolve mainly around two main tenets, a uniform one and an incremental one to help capture both the fluid and random characteristics of traffic flow.

*Correspondence to: Wissam Fawaz, Department of Electrical and Computer Engineering, Lebanese American University, PO Box 36, Byblos, Lebanon. E-mail: wissam.fawaz@lau.edu.lb

In this paper, we propose an exact formulation of the uniform component of traffic delay. This is particularly true because, to the best of our knowledge, the literature lacks such an exact model. It is widely known that the uniform delay component can be derived based on the following assumptions [4]: (i) the intersection is modelled as a D/D/1 queue (i.e. with deterministic inter-arrival time and deterministic service time) where the server goes on vacation during the red period and (ii) the queue is in statistical equilibrium and witnesses uniform arrivals at an arrival flow rate of λ as well as uniform departures at a departure flow rate of μ .

The rest of the paper is organized as follows. Section 2 positions the contribution delineated in this paper relative to the relevant open literature. Section 3 presents the currently widely used uniform delay formulation and highlights its limitations. A revised formulation is then delineated in Sections 4 and 5 following graphical and analytical approaches, respectively. Section 6 presents the numerical results. Finally, Section 7 concludes the paper.

2. RELATED STUDIES

A raft of previous studies looked into the control delay performance measure. A pioneering contribution in this respect was made by Beckmann *et al.* [5] and later Webster [3], who proposed a fundamental delay formula based on a simulation study of a one-lane signalized intersection. Following Webster's work, other delay models emerged, such as the ones due to Miller [6], Newell [7, 8], McNeil [9], Akcelik and Roupail [10], Heidemann [11] and Heidemann and Wegmann [12].

All of these models share a number of common aspects. Specifically, they all introduce approximate delay formulas that are composed of three main elements denoted by D_1 , D_2 and D_3 . The first element D_1 is the delay when the intersection is subject to uniform arrivals and departures, which is widely known as the uniform term. The second element D_2 captures the randomness of the arrival and departure processes, which is referred to as the incremental term. The third element D_3 is a calibration term that accounts for an initial queue formation as a result of pending demands from previous time periods.

Recently, Mukhopadhyaya *et al.* [13] considered the case of lane undisciplined traffic, in which vehicles and motorcycles form batches occupying the full width of a lane and exit the intersection one batch at a time. The mean delay in the presence of such batching was approximately analysed by using an extension of the Webster mean delay formula. Moreover, the concept of *maximum vehicle delay* was introduced by Smith [14] and defined as the maximum waiting time observed by any vehicle during a given cycle. Then, Lavrenz *et al.* [15] proposed to use the maximum vehicle delay to characterize the impact of timing adjustments on movements at the intersection. Autey *et al.* [16] compared the performance of four unconventional intersection schemes in terms of delay and capacity by simulations.

The present study differs from the aforementioned related studies in that it develops an exact characterization of the uniform component of the average delay experienced by vehicles at undersaturated signalized intersections with no width-wise clustering of vehicles (i.e. intersection is subject to lane disciplined traffic). The mathematical expressions derived in this paper are new to the best of our knowledge. Moreover, these expressions yield more accurate delay results relative to the most commonly used delay model, namely, the one adopted by the HCM2010 and many prominent traffic engineering books including [17]. For completeness, the uniform component D_1 of the existing delay model is given next:

$$D_1 = \frac{0.5C(1 - \frac{g}{C})^2}{1 - \frac{\lambda}{\mu}} = \frac{\lambda\mu r^2}{2(\mu - \lambda)} \times \frac{1}{\lambda C} = \frac{\mu r^2}{2C(\mu - \lambda)} \quad (1)$$

where r represents the effective red period, g the effective green period and C the total cycle length (i.e. $C = r + g$).

In what follows, the main limitations of the existing D_1 formulation (e.g. Equation (1)) are highlighted in Section 3, and then, revised expressions capturing the true values of the total uniform delays experienced by vehicles at undersaturated signalized intersections are proposed in Sections 4 and 5.

3. LIMITATION OF EXISTING FORMULATION OF UNIFORM DELAY: SAMPLE SCENARIO

The signalized intersection chapter (Chapter 18) of the HCM2010 provides equations for calculating control delay. The control delay equation is composed of three terms: uniform, incremental and initial queue delay terms that we denote for convenience by D_1 , D_2 and D_3 , respectively. More specifically, the delay formulation in the HCM2010 [1] is written as $D = D_1 + D_2 + D_3$. Equations defining D_2 and D_3 are beyond the scope of this paper and thus will not be discussed herein. Nonetheless, the average uniform delay D_1 (shown in Equation (1)) is of critical importance for the paper and therefore is given special consideration. D_1 can be obtained from the total uniform delay D_t by dividing D_t by the number of vehicles served within the analysis period C , the cycle time. This is especially true because D_t is the sum of the delays observed during the course of both the effective green and red phases. Traditionally, D_t is assumed in the HCM2010 and [17] to be given by the following expression:

$$D_t = \frac{\lambda\mu r^2}{2(\mu - \lambda)} \tag{2}$$

In this section, we show through a sample scenario, a knowingly unrealistic but simple-to-demonstrate idea, that Equation (2) underestimates the total delay value. So consider a scenario where vehicles arrive at a traffic signal along a one-lane approach. Suppose for demonstration purposes that the cycle length $C = 2$ seconds and that C consists of effective green and red periods of length $g = 1$ second and $r = 1$ second, respectively. In the considered scenario, vehicles are assumed to arrive uniformly at an arrival rate of $\lambda = 2$ vehicles/second and depart uniformly at a departure rate $\mu = 4$ vehicles/second. Given these input parameter values, the total uniform delay D_t evaluates through Equation (2) to $D_t = (2)(4)(1^2)/2(4 - 2) = 2$ seconds, which is the area of the triangle shown in Figure 1. Note that in Figure 1, the arrivals of vehicles to and departures from the system in the context of the scenario under study within an interval of length C are shown. The arrival time of the i th vehicle is designated by a_i , $i = 1, 2, 3, 4$, and the departure time of the i th arriving vehicle by l_i , $i = 1, 2, 3, 4$.

However, a closer look at the investigated scenario reveals that the correct value of the total uniform delay is 3.5 seconds as per the following justification. As a matter of fact, it can be easily shown that the four vehicles arriving during the time interval of length C experience total delays of 1.25, 1.0, 0.75 and 0.5 seconds. These delay values sum up to 3.5 seconds, which is 75% larger than the value calculated via the classical Equation (2). The reason for this non-negligible

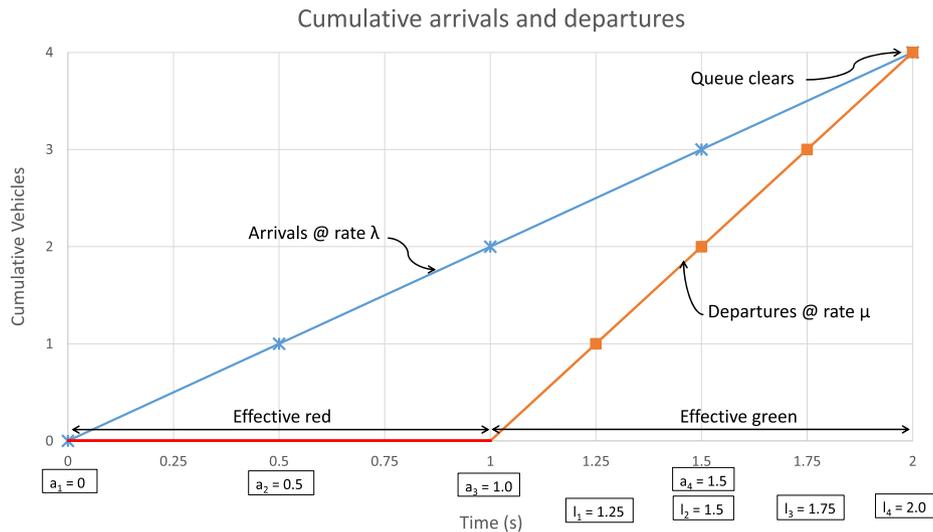


Figure 1. Arrivals and departures—continuous representation.

difference between the total delay value produced by Equation (2) and the real observed one is as follows. Vehicle arrivals and departures are discrete events and are basically approximated by a continuous graph like the classical one depicted in Figure 1. Yet a more adequate representation of arrivals and departures is one that makes use of step functions that better capture the real dynamics of the system, as illustrated in Figure 2.

We argue that such a discontinuous representation of the cumulative arrivals and departures is better suited than the classical continuous representation when it comes to total uniform delay computation. This observation is asserted by the following reasoning. In Figure 2, the $A(t)$ step function indicates the cumulative arrival process, and with each arrival, $A(t)$ increases by 1. Moreover, the $L(t)$ step function represents the number of vehicles that departed from the system up to time t . The total uniform delay under this condition can be obtained by summing up the areas of the squares lying between $A(t)$ and $L(t)$. In the studied scenario, there are 14 such squares, each of which having an area of 0.25. Consequently, the total uniform delay is found to be $14 \times 0.25 = 3.5$ seconds, which, unlike the value obtained from Equation (2), matches exactly the expected total uniform delay value.

In light of the previous discussion, it becomes clear that a revised—more precise—formulation of the total uniform delay is worth investigating. This issue is thoroughly addressed in the two subsequent sections, where exact closed-form expressions for the total uniform delay are derived through graphical and analytical analyses of the problem.

4. GRAPHICAL ANALYSIS

The sample scenario discussed in Section 3 is a special unrealistic case of undersaturation where $\lambda C = \mu g$. In this demonstration case, the number of vehicles served within the effective green period g matches the number of vehicles that arrived during the analysis cycle C . In other words, the time needed to clear the queue, referred to in the literature as t_c and defined as the time from the start of the effective green until queue clearance happens, is equal to the effective green period (Figure 3).

Without loss of generality, the derivation of the uniform delay component is carried out in this section under the assumption that the queue of vehicles forming at an intersection is cleared before the start of a new cycle. That is, the signalized intersection is assumed to be operating in undersaturated conditions, where $\lambda C = \mu g$. Hence, the graphical analysis of the delay will be performed incrementally, considering increasingly more general cases of undersaturation. In a first step, the specific case where $\lambda C = \mu g$ is considered. Then, the more general case where $\lambda C < \mu g$ is tackled in a second step.

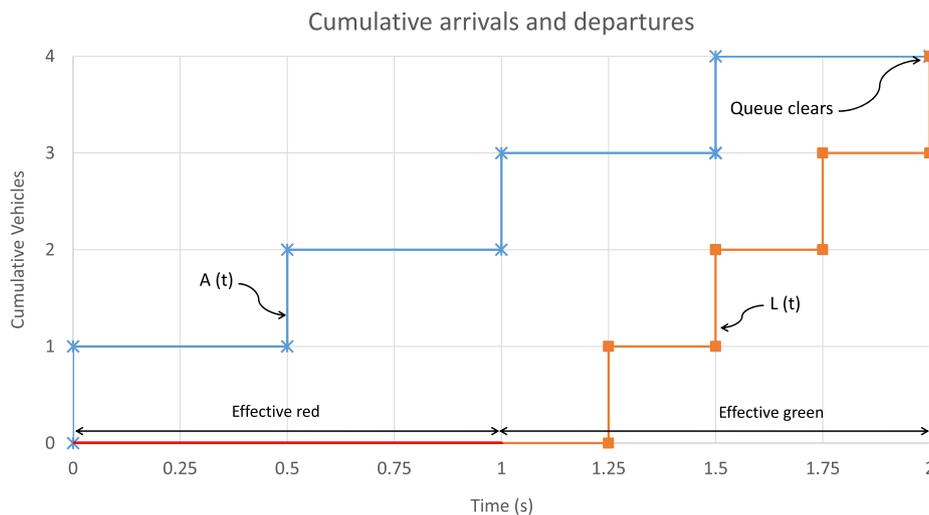


Figure 2. Arrivals and departures—discrete/discontinuous representation.

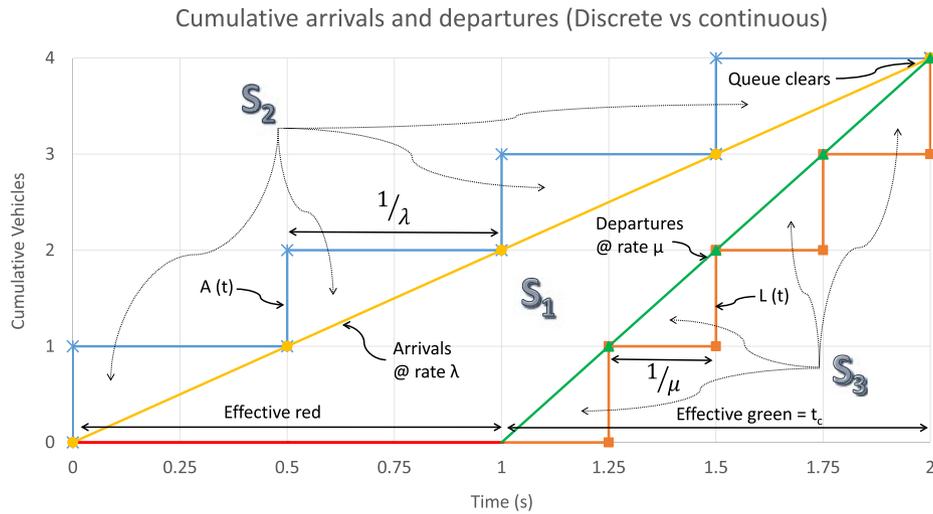


Figure 3. Discrete versus continuous arrival/departure times.

4.1. Case 1: $\lambda C = \mu g$

Figure 3 concurrently plots the cumulative continuous arrival/departure functions relating to the sample scenario studied in Section 3 and their discrete counterparts. Equation (2) corresponds to the area between the lines representing the cumulative continuous arrivals and departures. We denote this area by S_1 . It is clear that the classical formulation of the uniform delay component does not account for the areas of both the triangles formed between the cumulative discrete and continuous arrival functions and the triangles that exist between the cumulative discrete and continuous departure functions. We designate these unaccounted for areas by S_2 and S_3 , respectively. This explains why the uniform delay calculated using Equation (2) underestimated the total uniform delay by 75% in the context of the sample scenario considered in Section 3. As a distinguishing factor from the existing delay formulation given by Equation (2), our exact formulation of the uniform delay component takes into consideration all three areas, S_1 , S_2 and S_3 .

It is worthwhile noting first that t_c , the time to clear, can be obtained as follows:

$$\lambda(r + t_c) = \mu t_c \Leftrightarrow t_c = \frac{\lambda r}{\mu - \lambda} \tag{3}$$

Both S_2 and S_3 depend on the number of triangles N that sit between the continuous arrival/departure function and its discrete equivalent. N represents the total number of departures occurring during t_c and as a result can be written as follows: $N = \mu t_c$. Moreover, S_2 and S_3 depend on the area of the individual triangles. For S_2 , the triangles bound by the cumulative continuous and discrete arrival functions should be considered. The area of each one of these triangles is given as follows: $(\text{height})(\text{base})/2 = (1)(1/\lambda)/2$. Similarly, the area of the individual triangles underlying S_3 can be expressed as follows: $(\text{height})(\text{base})/2 = (1)(1/\mu)/2$. Putting these observations together, we obtain the following expressions for S_2 and S_3 :

$$S_2 = (\mu t_c) \frac{(1)(\frac{1}{\lambda})}{2} = \frac{\mu t_c}{2\lambda} \tag{4}$$

$$S_3 = (\mu t_c) \frac{(1) \left(\frac{1}{\mu}\right)}{2} = \frac{t_c}{2} \tag{5}$$

Summing up S_1 (given by Equation (2)), S_2 and S_3 results in the following exact expression for the total uniform delay component D_t :

$$D_t = \left(\frac{t_c}{2}\right) \left(\mu r + \frac{\mu}{\lambda} + 1\right) \tag{6}$$

Using the data of the sample scenario analysed in Section 3, the total delay can now be computed starting with t_c , $t_c = \lambda r / (\mu - \lambda) = (2)(1) / ((4) - (2)) = 1$ second, and continuing on with D_t , $D_t = (1/2)((4)(1) + (4/2) + 1) = 3.5$ seconds. The obtained value matches the result presented in Section 3 and thereby verifies the correctness of the newly obtained expression for D_t .

Finally, Equation (3) implies that D_t can be written as follows:

$$D_t = \frac{(r)(\lambda \mu r + \mu + \lambda)}{2(\mu - \lambda)} \tag{7}$$

4.2. Case 2: $\lambda C < \mu g$

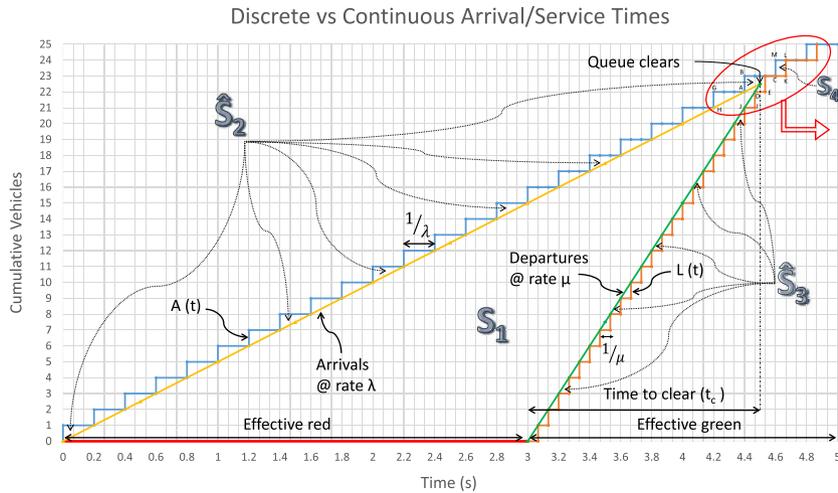
In this section, we turn to the more complex case in which $\lambda C < \mu g$. A sample scenario where this inequality holds is used to graphically illustrate the derivation process. For the same one-lane signalized approach considered earlier, suppose that the cycle length $C = 5$ seconds, of which $g = 2$ seconds and $r = 3$ seconds. Furthermore, vehicles are assumed to arrive uniformly at an arrival rate of $\lambda = 5$ vehicles/second and depart uniformly at a departure rate of $\mu = 15$ vehicles/second.

As shown in Figure 4a and b, the total delay is composed of five terms, which are defined and detailed as follows:

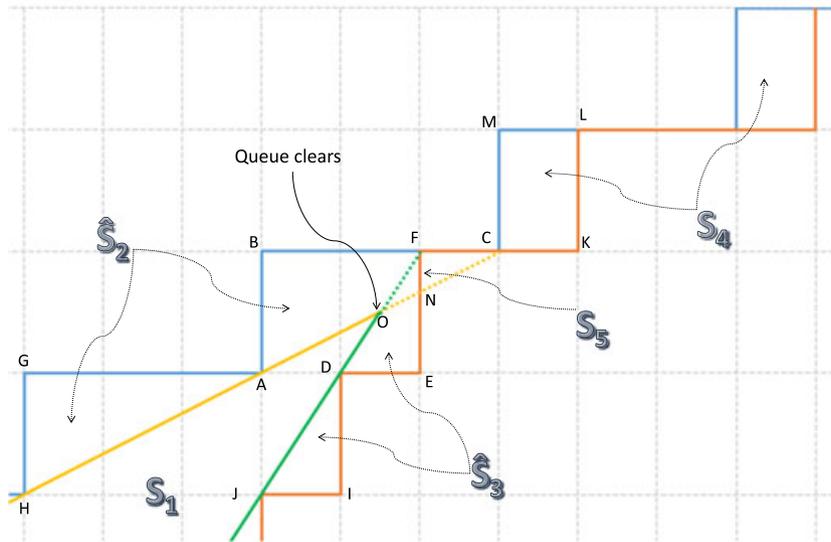
- 1 S_1 is the area of the big triangle bound by the continuous arrival and departure functions as shown in Figure 4a. S_1 corresponds to the classical expression of the total uniform delay D_t and is defined in Equation (2).
- 2 \hat{S}_2 is a variation of the S_2 area defined in Section 4.1. Similar to S_2 , \hat{S}_2 is concerned with the vehicles that experience queueing delay. Given that in this case, t_c is less than g , not all vehicles that arrive within the evaluation cycle C will experience queueing delay as is the case in the scenario discussed in Section 4.1. To calculate \hat{S}_2 , we multiply $N = \lceil \mu t_c \rceil$, the number of triangles bound by the $A(t)$ step function and the continuous arrival function (Figure 4a), by the area of each triangle. However, in this case, the number of triangles N may not be an integer, and therefore, the ceiling function is used to account for the total area of the last triangle [ABC] shown in Figure 4b. As such, \hat{S}_2 is given as follows:

$$\hat{S}_2 = \frac{\lceil \mu t_c \rceil}{2\lambda} \tag{8}$$

- 3 \hat{S}_3 is somewhat similar to the S_3 area defined in Section 4.1 and is also calculated as $N = \lceil \mu t_c \rceil$, the number of triangles sandwiched between the $L(t)$ step function and the continuous departure function (Figure 4a), times the area of each triangle. In a similar way, the ceiling function is used to account for the total area of the last triangle [DEF] depicted in Figure 4b. Thus, \hat{S}_3 is given by the following:



(a) Complete Cycle



(b) Zoomed at t_c

Figure 4. Graphical representation of delay at a signalized intersection approach.

$$\hat{S}_3 = \frac{[\mu t_c]}{2\mu} \tag{9}$$

- 4 S_4 has to do with the delay experienced by vehicles beyond queue dissipation and is equal to the number of vehicles that arrive after t_c multiplied by the average departure time (i.e. the sum of the areas of rectangles following and including the rectangle [CKLM] shown in Figure 4b). Hence, S_4 is expressed as follows:

$$S_4 = \frac{\lambda[(g - t_c)]}{\mu} \tag{10}$$

- 5 S_5 is the resulting excess delay caused by the use of the ceiling function in the evaluation of \hat{S}_2 and \hat{S}_3 to allow for the inclusion of the total areas of the triangles [ABC] and [DEF] shown in Figure 2.

To correct this error margin, the area of triangle [FNC] should be subtracted from that of triangle [ABC] (Figure 4b). In addition, the area of triangle [ONF] should be subtracted from that of triangle [DEF]. This excess delay S_3 (the sum of areas of triangles [ONF] and [FNC]) can be viewed graphically in Figure 4b and is calculated as the area of triangle [OFC] as follows:

$$S_5 = \frac{1}{2} \{ [\mu t_c] - (\mu t_c) \} \left\{ \frac{[\mu t_c]}{\lambda} - \frac{[\mu t_c]}{\mu}, -r \right\} \quad (11)$$

The total delay D_t is formulated as follows:

$$D_t = S_1 + \widehat{S}_2 + \widehat{S}_3 + S_4 - S_5 \quad (12)$$

Given the input parameter values presented here, the total delay resulting from the classical Equation 2 is 33.75 seconds, while the one obtained via Equation (12) is 36.93 seconds (about 10% difference). Finally, note that for the special condition where $t_c = g$, S_4 and S_5 become zero and the total delay equation presented in this section boils down to the same equation presented in Section 4.1 (e.g. Equation 7).

5. ANALYTICAL ANALYSIS

To provide further insight into the total uniform delay D_t , this section uses an alternative approach to derive an exact formulation of D_t . Instead of relying on graphs and figures as in Section 4, an expression describing D_t is obtained analytically.

Vehicles arriving during an evaluation cycle C can be classified into two broad categories, namely, the vehicles arriving during the red period of length r (seconds) and the ones reaching the signalized intersection during the green period of length g (seconds). Building on this observation, the total uniform delay D_t can be expressed as the sum of two components D_r and D_g , which represent the delay experienced by the vehicles from the red and green periods, respectively (i.e. the vehicles belonging to the two categories identified earlier).

5.1. Derivation of D_r

Given that vehicles arrive at the signalized intersection at a rate of λ (vehicles/second) and that the red period is r seconds long, the total number of vehicles found at the end of the red period is $N_r = \lambda r$. It is needless to say that λr is assumed to be an integer; otherwise, the largest integer value smaller than λr should be used, that is, the floor $[\lambda r]$. For the sake of simplicity of notation, the floor symbol is omitted in what follows.

Consider the i th arriving vehicle v_i , $i = 1, 2, \dots, \lambda r$. The total delay $D_{v_i}(\lambda, \mu, i)$ experienced by v_i depends on the residual red time $r - ((i - 1)/\lambda)$ as well as the amount of time required to serve all $i - 1$ vehicles that arrived prior to i . So $D_{v_i}(\lambda, \mu, i)$ can be written as follows:

$$D_{v_i}(\lambda, \mu, i) = r - \frac{i - 1}{\lambda} + \frac{i}{\mu} \quad (13)$$

Given that D_r is the sum of the delays experienced by all vehicles arriving during the red period, it follows that

$$D_r = \sum_{i=1}^{\lambda r} D_{v_i} \tag{14}$$

$$= \lambda r^2 + \frac{r}{2\mu} [\lambda + \mu + \lambda^2 r - \lambda \mu r]$$

5.2. Derivation of D_g

The vehicles that get to the signalized intersection during the green interval can be further subdivided into two subcategories, in particular, the vehicles that experience queueing delays and the ones that are subject to 0 queueing delay. The latter category corresponds to vehicles that arrive at the intersection after the dissipation of the queue that formed as a result of the accumulation of vehicles during the $r + t_c$ period. The number of such vehicles and as indicated in Section 4.2 is $\lambda \lfloor g - t_c \rfloor$. As such, the total delay experienced by these vehicles is $(\lambda \lfloor g - t_c \rfloor)(1/\mu)$.

On the other hand, the vehicles whose associated queueing delays are non-zero are mainly the vehicles that arrive during the first t_c seconds of the green period. Recall that t_c is the amount of time it takes for the queue size to hit 0 once the green period has begun. During these t_c seconds, a total of μt_c vehicles are discharged. But what is relevant to the derivation of D_g is the fact that within these t_c seconds, a total of λt_c new vehicles arrive at the intersection. It is these vehicles whose delays are of interest when computing D_g . The $\mu t_c = \lambda(r + t_c)$ equality from Equation 3 indicates that both the λr vehicles that accumulated in the red period and the newly arriving λt_c vehicles are cleared during t_c seconds.

Accordingly, some of the λt_c vehicles arriving during the first t_c seconds of the green period are affected by the service time of the λr vehicles inherited from the red period. Specifically, those vehicles that arrive at the intersection while at least one of the λr vehicles is still at the intersection incur an additional queueing delay matching the total service time of the already present vehicles. The number of such influenced vehicles can be easily shown to be $\lambda^2 r / \mu$ given that it takes $\lambda r / \mu$ seconds to clear all λr vehicles spilling over from the red period. The remaining $\lambda t_c - (\lambda^2 r / \mu)$ vehicles are only influenced by the service time of the vehicles from the green period queueing in front of them.

To further illustrate these ideas, consider a real-life scenario where $\lambda = 0.5$ vehicles/second (1800 vehicles/hour), $\mu = 0.8$ vehicles/second (2880 vehicles/hour), $g = 20$ seconds and $r = 10$ seconds. The sequence of events occurring during the course of an evaluation cycle C is given in Figure 5.

In this context, the number of vehicles that accumulate during the red period is $\lambda r = 5$ vehicles. These vehicles start receiving their service at the beginning of the green period, and $\lambda r / \mu = 6.25$ seconds later, the last of these vehicles successfully traverses the intersection. During that time, a total of $\lceil \lambda^2 r / \mu \rceil = 4$

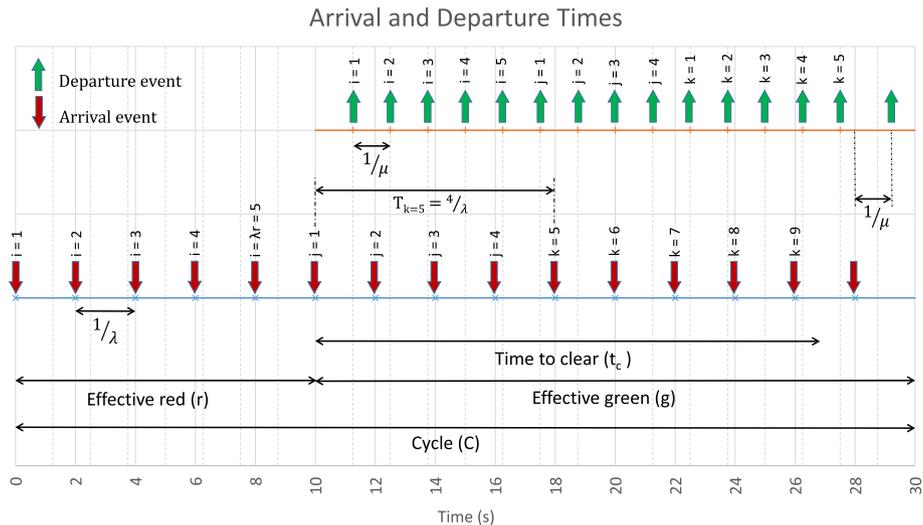


Figure 5. Illustration of events pertaining to D_{gs} .

newly arriving vehicles are influenced by the five vehicles spilling over from the red period. However, the remaining $\lceil \lambda t_c \rceil - \lfloor \lambda^2 r / \mu \rfloor = 5$ vehicles are only influenced by the service time of the vehicles arriving during the green period.

In light of the earlier discussion, D_g can be written as follows:

$$D_g = \sum_{j=1}^{\frac{\lambda^2 r}{\mu}} D_{v_j} + \sum_{k=\frac{\lambda^2 r}{\mu}+1}^{\lceil \lambda t_c \rceil} D_{v_k} + (\lambda \lfloor (g - t_c) \rfloor) \frac{1}{\mu} \tag{15}$$

where D_{v_j} represents the delay of the j th vehicle influenced by the vehicles that accumulated in the red period. In this case, this j th influenced vehicle waits until all vehicles from the red period along with the $j - 1$ vehicles preceding it are serviced before it can itself gain access to the intersection. For this reason, D_{v_j} is given by

$$D_{v_j} = \frac{\lambda r}{\mu} - \frac{j - 1}{\lambda} + \frac{j}{\mu} \tag{16}$$

D_{v_k} in Equation (15) is the delay experienced by a vehicle v_k (one of the five vehicles mentioned earlier —see variable k in Figure 5) that arrives at the intersection following the departure of all λr vehicles that were inherited from the red period. D_{v_k} depends on the number of vehicles that exist in the queue at the arrival instant of v_k . Let $t_k = (k - 1) / \lambda$ be the amount of time that elapses from when the green period begins until v_k shows up. For example, for $k = 5$ corresponding to the first non-influenced vehicle, $t_5 = 4 / \lambda$, as highlighted in Figure 5. It is clear then that D_{v_k} can be expressed as follows:

$$\begin{aligned} D_{v_k} &= \left[k - \left(t_k - \frac{\lambda r}{\mu} \right) \mu \right] \frac{1}{\mu} \\ &= \frac{1}{\lambda \mu} [k(\lambda - \mu) + \mu + \lambda^2 r] \end{aligned} \tag{17}$$

In summary, one can easily prove based on Equations (16) and (17) that D_g is given by the following expression:

$$\begin{aligned} D_g &= \frac{\lambda^2 r}{2\mu} \left[\frac{\lambda r}{\mu} \left(\frac{\lambda}{\mu} + 1 \right) + \frac{1}{\mu} \right] \\ &+ \sum_{k=\frac{\lambda^2 r}{\mu}+1}^{\lceil \lambda t_c \rceil} \frac{k}{\mu} - \frac{k - 1}{\lambda} + \frac{\lambda r}{\mu} + (\lambda \lfloor (g - t_c) \rfloor) \frac{1}{\mu} \end{aligned} \tag{18}$$

6. NUMERICAL RESULTS

In this section, we present a number of real-life scenarios to illustrate the high degree of accuracy achieved by the proposed uniform delay formulation. In the context of the considered scenarios, a comparison is made between the uniform delay values obtained from the classical equation given in Section 3 and those obtained using the delay models delineated in Sections 4 and 5.

At a first step, five examples of real-life service channels are presented in Table I. The first scenario represents a one-lane through movement at a signalized intersection. It is important to highlight in this respect that through movement refers to a situation under which vehicles go straight through the intersection with no vehicles making left or right turns. This through lane assumes a morning peak (AM) approach flow (arrival rate λ) of 900 vehicles/hour. The second scenario is identical to the first one

Table I. Relevant scenarios of real-life service channels.

Scenarios	λ (vehicles/second)	μ (vehicles/second)	g (seconds)	r (seconds)
AM-thru (1)	1/4	19/36	25	15
PM-thru (2)	1/3	19/36	30	15
Grade crossing (3)	1/12	5/12	480	120
Pedestrians (4)	1/8	5/12	240	60
Bridge (5)	5/72	7/18	3300	300

but is based on an afternoon peak flow (PM) of 1200 vehicles/hour. The cycle length C is assumed to be 40 and 45 seconds for the first and second scenarios, respectively, with effective green and red periods as per the values listed in Table I.

The third scenario assumes an at-grade crossing between a light rail transit system, similar to the one present in San Diego and San Francisco and a city street. The gates are closed for traffic during the train passage (effective red). In some cases, the transit stop occurs at the grade crossing, resulting in the closure of the street while the transit is loading/unloading passengers. A typical scenario is considered under this condition. Particularly, the average closure time is assumed to be 2 minutes, and the transit service headway between two consecutive trains is supposed to be 10 minutes (e.g. street traffic gets an effective green of 8 minutes). For this scenario, an approach volume of 300 vehicles/hour per lane is considered, and the capacity of the lane (service rate, μ) is assumed to be 1500 vehicles/hour, ignoring thus the effect of nearby signalized control on the lane capacity. The fourth scenario in Table I deals with a one-lane per direction street near an airport terminal or school zone. The street traffic is stopped for pedestrians to cross from the airport terminal to the parking on the opposite side of the street or for students to cross the street to the bus stop. In this case, the signal is assumed to have a 5-minute cycle ($g=4$ minutes and $r=1$ minute). For this scenario, the approach volume per lane is considered to be 450 vehicles/hour, and the lane service rate is assumed to be 1500 vehicles/hour. In the fifth scenario, we consider a moveable bridge over a water body that opens to allow the passage of vessels. Moveable bridges are of many types, including bascule, swing, table, thrust and vertical lift bridges. Such bridges still operate in cities like New Jersey, Chicago and London. While the bridge is open, traffic is stopped on the street. In that particular regard, we consider a one-lane per direction street over the bridge with a lane service rate of 1400 vehicles/hour. An average bridge opening and closing time of 5 minutes is assumed per passage, which occurs once in an analysis hour. In this case, the signal is assumed to have a 60-minute cycle ($g=55$ minutes and $r=5$ minutes). For this scenario, the approach volume per lane is supposed to be 250 vehicles/hour.

The results of the average uniform delay per vehicle (seconds) for all five scenarios as achieved by the classical uniform delay model, the proposed graphical/analytical models and simulations are shown in Table II. Note that an in-house Java-based simulator was developed to conduct extensive simulations in order to validate the delay values obtained from the different delay models. The simulator is based on the one we developed to generate the results reported in [18]. It is obvious from the reported results that the classical delay equation underestimates the uniform queueing delay value. For instance, the classical delay equation was found to underestimate the uniform delay by 50% in the context of scenario 1. The results also prove that the proposed graphical/analytical delay models along with the simulator converge to the same exact value of the average uniform delay.

Table II. Comparison of average uniform delay per vehicle (seconds).

Scenarios	Classical	Graphical	Analytical	Simulation	% difference
1	5.3	8.0	8.0	8.0	50
2	6.8	9.2	9.2	9.2	36
3	15.0	18.6	18.6	18.6	24
4	8.6	11.8	11.8	11.8	37
5	15.2	18.4	18.4	18.4	21

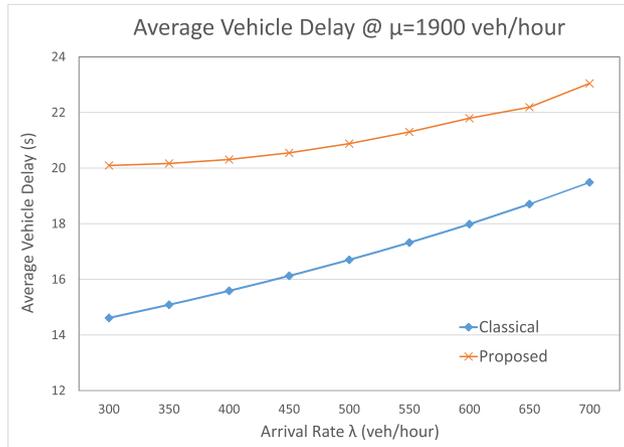


Figure 6. Average delay per vehicle for a one-lane approach to a signalized intersection— $C = 65$ seconds; $g = 25$ seconds; $r = 40$ seconds.

Given that the graphical and analytical models produced identical average uniform delay values, only one of the two models is used in what follows. Figure 6 concurrently plots the average uniform vehicle delay of an approach to an intersection as achieved by the classical delay model and the proposed delay formulation versus the arrival rate. Arrival rates ranging from 300 to 700 vehicles/hour are considered at the approach coupled with a standard saturation flow rate (service rate) of 1900 vehicles/hour. The data presented in the figure clearly show that the classical uniform delay equation considerably underestimates the average uniform delay as compared with the proposed delay formulation. In addition to varying the arrival rate, sensitivity tests were conducted to account for various cycle lengths and red times. Three common cycle lengths (60, 90 and 120 seconds) were considered to conduct the sensitivity analysis. For each cycle length scenario, the arrival rate was assumed to vary from one-fourth to two-thirds of the capacity (i.e. conditions of undersaturation). In each individual scenario of cycle length and arrival rate, the effective red time interval was increased in increments of 5 seconds. A total of 84 scenarios (28 cases for each cycle length) were analysed to assess the average uniform delay per vehicle using both the classical formulation and the proposed one. A sample of the sensitivity analysis results is shown in Figure 7 for a 90-second cycle length. In this figure, the per cent increase in the average uniform delay between the classical formulation and the proposed one is plotted as a function of two varying parameters, namely, the arrival rate and effective red interval.

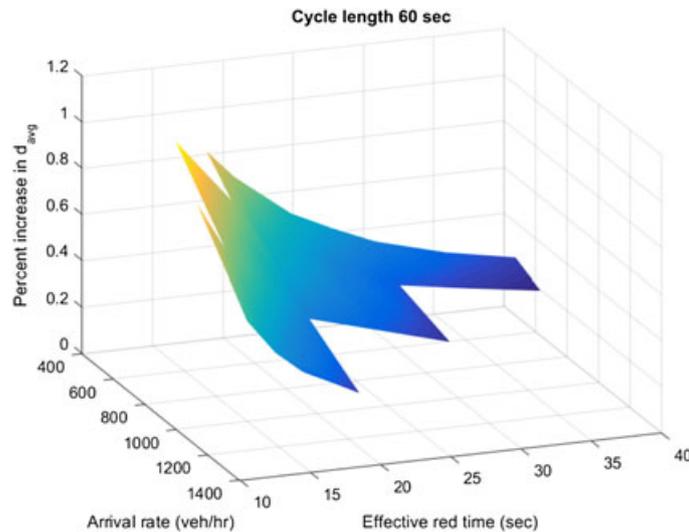


Figure 7. Analysis scenario with cycle length = 60 seconds.

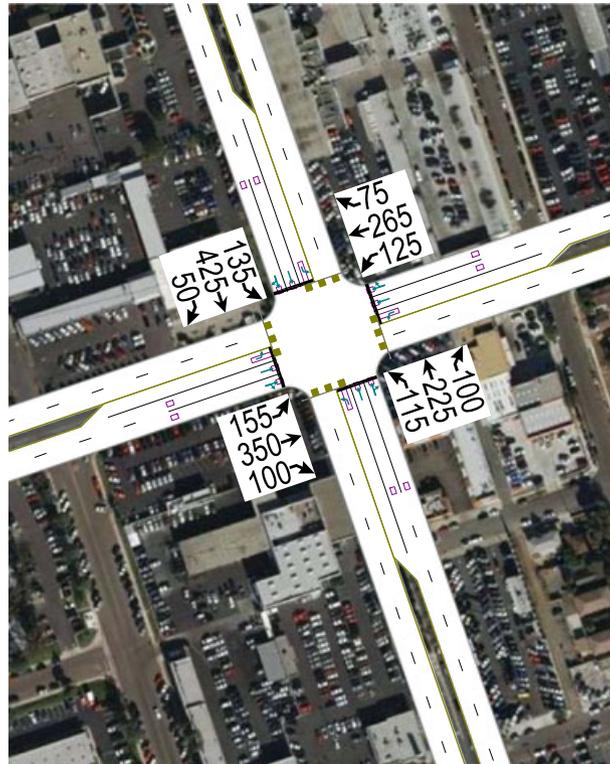


Figure 8. Real-life intersection layout with traffic volumes.

On average, the proposed uniform delay formulation produced delays that are 50% higher than the ones obtained using the classical delay formulation for the presented cycle length scenario. The results inherently show that the classical delay formulation underestimates the delay relative to the proposed formulation for undersaturated conditions.

Last but not least, we present a real-life intersection layout from San Diego, CA, with traffic volumes collected for the PM peak period in October 2010. As shown in Figure 8, the intersection layout has two through/shared-right lanes and one left-turn pocket lane on all approaches. Furthermore, the intersection has a standard four-approach layout (eastbound, westbound, northbound and southbound). The city of San Diego operates this signal using leading left turns, that is, left-turning vehicles get the green indication before the through/shared-right vehicles per direction (east–west or north–south). The

Table III. Traffic volumes and various average delay values pertaining to the real-life intersection.

	EB		WB		NB		SB	
	L	T/R	L	T/R	L	T/R	L	T/R
λ	155	406	125	297	115	252	135	460
μ	1805	3490	1805	3491	1805	3443	1805	3553
g	6	17	6	17	5	16	5	16
r	54	43	54	43	55	44	55	44
D_2	38.4	1.3	19.9	0.8	30.5	0.7	50.8	1.8
D_3	0	0	0	0	0	0	0	0
D_1	26.6	17.4	26.1	16.8	26.9	17.4	27.2	18.5
D	65.0	18.7	46.0	17.6	57.4	18.1	78.0	20.3
D'_1	40.4	21.8	42.0	22.3	44.2	24.0	43.1	22.6
D'	78.8	23.1	61.9	23.1	74.7	24.7	93.9	24.4

EB, eastbound; NB, northbound; SB, southbound; WB, westbound.

Table IV. Average delay and LOS achieved at the real-life intersection by the classical and proposed delay models.

	Classical	Proposed
d_1	30.7	38.5
LOS	C	D

LOS, level of service.

traffic volumes were collected for all movements at the intersection and are reported in Figure 8 and Table III. Note that the through movement flows and the right-turn flows shown in Figure 8 do not add up to the flows shown for the through/shared-right movements in Table III. The reason for this difference is as follows. A portion of the right-turn flows are not considered in the signal timing analysis as they make a right turn while the signal indication is red. Those right-turn flows are subtracted from the lane group flows. Because left turns can only turn left when the signal indication is green, the volumes in the figure and table match up exactly.

Using the HCM2010 methodology to derive the average delay per vehicle for the intersection, the three delay components (D_1 , D_2 and D_3) are calculated using the classical formulation and presented in Table III. Then, the uniform delay component is computed using the proposed formulation and labelled D'_1 in that same table. This process is repeated for all lane groups. A final average delay per vehicle for the intersection d_1 is then calculated and shown in Table IV along with the corresponding LOS. Using the new delay formulation, we prove that the intersection is in reality operating at a lower LOS than that estimated using the current HCM2010 methodologies. The results clearly show that the classical delay equation underestimates the uniform delay component and thus provides inaccurate LOS assessment at signalized intersections. Conversely, the proposed delay models accurately capture the dynamics of the real-life intersection, yielding an exact estimation of the LOS offered at the intersection.

7. CONCLUSION

Throughout this paper, we presented a revised formulation for the uniform delay experienced at undersaturated signalized intersections. The proposed formulation differs from the classical uniform delay equation that is widely used in the open literature in that it develops an exact characterization of the uniform component of the delay experienced by vehicles at any controlled service channel. The mathematical expressions obtained to this end are new to the best of our knowledge and yield more accurate delay results as compared with the most commonly used delay model, particularly, the one adopted by the HCM2010 and many prominent traffic engineering references. We also showed that the proposed formulation precisely captures the uniform delay compared with the classical, which proved to significantly underestimate the delay component, leading to inaccurate assessment of the LOS at undersaturated signalized intersections.

The flow of traffic is analogous to the flow of a compressible fluid. As such, shock wave theory can also be employed to analyse the flow of vehicles at a signalized intersection. The main difference between a shock wave analysis and a queueing analysis such as the one performed in this paper lies in the way vehicles are assumed to queue at the intersection stop line. While queueing analysis supposes vertical queueing of vehicles, shock wave analysis assumes that vehicles are queued horizontally one behind each other, that is, each vehicle is assumed to occupy some physical space. Even though a shock wave analysis of the uniform component of traffic delay is left as future work, it is worthwhile providing some insights into such analysis of the uniform delay. In addition to queue formation and dissipation, under shock wave analysis, different types of shock waves should be considered. More specifically, a first wave defining the boundary between incoming traffic and queued vehicles is observed. Then, two other waves are generated when the signal turns green. The first wave corresponds to the front of the surge of vehicles that leave the intersection at saturation flow rate at the beginning of the green interval. Moreover, there is a second wave that divides the vehicles in the queue from those that have started to accelerate forward. The last wave defines the end of the platoon of vehicles leaving

the intersection at saturation flow rate. By following an approach similar to the graphical one delineated in this paper and by accounting for the aforementioned waves, a future study should be able to achieve a shock wave-based formulation of the uniform traffic delay observed at undersaturated signalized intersections.

Indeed, the formulation performed in this paper has a generic fundamental significance that goes beyond the specific context of signalized intersections. This is especially true because it can be applied to any general system exhibiting the same behaviour as a signalized intersection. Owing to this generality, any further results that will be derived from the models presented in this paper have a potential significance for other related fields and areas of study.

REFERENCES

1. Transportation Research Board. *Highway Capacity Manual-2010*: TRB. Washington D.C., USA, 2010.
2. Dion F, Rakha H Kang Y. Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections. *Transportation Research Part B: Methodological* 2004; **38**:99–122.
3. Webster F. *Traffic Signal Settings—Paper No. 39*. Her Majesty Stationary Office: London, 1958.
4. N Roupail, A Tarko, L Jing. Traffic flow at signalized intersections. *Traffic Flow Theory Monograph, Ch. 9, Transportation Research Board*, 2001.
5. Beckmann M, McGuire C Winsten C. *Studies in the Economics of Transportation*. RAND Corporation: New Haven, Connecticut, 1956.
6. Miller A. Settings for fixed-cycle traffic signals. *Operational Research Quarterly* 1963; **14**:373–386.
7. Newell G. Queues for a fixed-cycle traffic light. *Annals of Mathematical Statistics* 1960; **31**:589–597.
8. Newell G. Approximation methods for queues with application to the fixed-cycle traffic light. *SIAM Review* 1965; **7**:223–240.
9. McNeil D. A solution to the fixed-cycle traffic light problem for compound Poisson arrivals. *Journal of Applied Probability* 1968; **5**:624–635.
10. Akcelik R, Roupail N. Estimation of delays at traffic signals for variable demand conditions. *Transportation Research Part B: Methodological* 1993; **27**:109–131.
11. Heidemann D. Queue length and delay distributions at traffic signals. *Transportation Research Part B: Methodological* 1994; **28**:377–389.
12. Heidemann D, Wegmann H. Queueing at unsignalized intersections. *Transportation Research Part B: Methodological* 1997; **31**:239–263.
13. S Mukhopadhyay, M Pramod, P Kumar. An approach for analysis of mean delay at a signalized intersection with indisciplined traffic. *Workshop on Intelligent Transportation Systems, 7th International Conference on Communication Systems and Networks*, January 2015.
14. W Smith. Signalized corridor assessment. Masters thesis, Purdue University, 2014.
15. S Lavrenz, C Day, A Hainen *et al*. Characterizing signalized intersection performance using maximum vehicle delay. *Transportation Research Board*, November 2014.
16. Autey J, Sayed T El Essawey M. Operational performance comparison of four unconventional intersection designs using micro-simulation. *Journal of Advanced Transportation* 2013; **31**:536–552.
17. Mannering F, Washburn S. *Principles of Highway Engineering and Traffic Analysis*. (5th edn) John Wiley and Sons: NJ, USA, 2013.
18. Khabbaz M, Fawaz W Assi C. A simple free-flow traffic model for vehicular intermittently connected networks. *IEEE Transactions on Intelligent Transportation Systems* 2012; **13**:1312–1326.