Transportation Planning and Technology, 2014
Vol. 37, No. 1, 3–19, http://dx.doi.org/10.1080/03081060.2013.844907

The effect of green time on stochastic queues at traffic signals

Nicholas B. Taylor* and Benjamin G. Heydecker

Centre for Transport Studies, University College London, Gower Street, London WC1E 6BT, UK

(*Received 10 March 2013; accepted 25 July 2013*)

Many analyses of traffic signal queues use Webster and Cobbe's formula, which combines the net effect of the red/green cycle with a term representing stochastic effects, idealised as an M/D/1 queue process having random arrivals and uniform service. Several authors have noted that this component should depend not only on demand intensity but also on throughput capacity in each green period, although an extra empirical term may partially allow for this. Extending the service interval in M/D/1 (M = Markovian, i.e. random, D = deterministic, i.e. uniform, 1 = one server) enables the effect to be reproduced, but no exact expressions for its moments are found. Approximate formulae for the extended mean exist but are accurate only near saturation. The paper derives novel approximations for the equilibrium mean and also variance and utilisation, using functions linking traffic intensity with green period capacity. With three moments, equilibrium probability distributions can be estimated for which a method based on a doubly nested geometric distribution is described.

Keywords: signal; queue; stochastic; M/D/1; variance; probability distribution

Introduction and background

Real signalised junctions are complicated by demand-responsive timings, conflicting movements and coordination. Microscopic simulation, which need only specify short-term individual behaviour, is used increasingly. Nevertheless, the formula of Webster and Cobbe (1966) for queue size or delay at an isolated signal is still widely regarded. In addition to red and green phase component, it contains a term representing stochastic effects, including transient overload. This is equivalent to an idealised queuing process where customers arrive randomly and are served at uniform intervals. Several authors, including Olszewski (1990), have pointed out that this overestimates the true stochastic queue component, which simulation shows falls, albeit slowly, with increasing throughput capacity in the green period. Webster's formula has an empirical term that may compensate for this effect and can be related to a more exact formula. However, it is difficult to integrate these with computationally efficient time-dependent approximate methods.

After describing the basic stochastic process, an extension is developed to take account of green period capacity. Simulation results based on it are used to derive approximations to the queue's equilibrium utilisation, mean and variance in forms compatible with time-dependent queuing methods, which are shown to be more consistent than earlier empirical approximations. A novel approach makes use of link functions between traffic intensity and green period capacity, allowing a broad physical

*Corresponding author. Email: nicholas.b.taylor@ucl.ac.uk

interpretation. Comparison with simulation shows that the results are accurate over a range of traffic intensities and green period capacities. A method of estimating a probability distribution from the three moments is described that may have wider application. Finally, the theoretical nature of the analysis is justified, and various conceptual and practical issues are discussed.

The mean queue size at a signal

Allsop and Hutchinson (1972) trace the origins of signal queue analysis back to A J H Clayton in 1940, as well as discuss the impact of different assumptions about arrival patterns. A well-established expression for the mean queue size L at a fixed-time traffic signal is Equation (1), ascribed to Webster (Webster and Cobbe 1966):

$$L = L_P + L_V + L_W = \frac{x\mu c(1-\Lambda)^2}{2(1-x\Lambda)} + \frac{x^2}{2(1-x)} - 0.65(x\mu c)^{\frac{1}{3}}\, x^{(2+5\Lambda)} \qquad (1)$$

The queue size in vehicles is composed of a phase term L_P visualised as growing linearly during the red phase and discharging linearly and fully during green, a stochastic and overload term L_V, and an empirical adjustment L_W. Other variables are defined below as follows:

- c signal cycle time;
- g green time within each cycle;
- r red time within cycle;
- Λ green cycle time ratio = g/c (avoiding possible confusion with demand rate λ);
- s saturation flow, the maximum flow rate across the stop-line;
- μ = gs/c, capacity of the movement taking account of the green/cycle ratio;
- x average degree of saturation or utilisation of service at the stop-line.

Webster's formula strictly applies to Poisson arrivals at an isolated signal, so excludes factors like minimum headway, vehicle actuation, platooning and coordination. However, the concern here is solely with the effect of green capacity on the stochastic term and its form in relation to existing time-dependent methods. Other variables used in the paper are the following:

$G = gs = \mu c$, the number of vehicles[1] that can pass in a single green period
$\rho = \lambda/\mu$, demand intensity, relative to capacity

Note that G is present effectively as μc in the first and last terms of Equation (1) but not the stochastic term. The degree of saturation x cannot exceed 1, so only the stochastic term can accommodate indefinite queue growth and then in principle only by being time-dependent. A more integrated approach is the derivation from first principles of Heidemann (1994), using a generating function and the results of Meissl (1963). Heidemann shows this gives results imperceptibly different from Webster and Cobbe's for values of G up to 45. If equilibrium conditions are imposed, the queue formula can be reduced to a form analogous to Equation (1) containing the identical phase term L_P:

$$L = L_P + L_{V[G]} + L_H \quad \text{where}\; L_H = \frac{-\Lambda(1-x)\left(2L_{V[G]} + \Lambda x\right) + \Lambda x}{2(1-\Lambda x)} \qquad (2)$$

The 'exact' stochastic term $L_{V[G]}$ here is a complicated expression that requires numerical evaluation of the roots of a function of degree G. However, when $G = 1$ it reduces to L_V. The third term, L_H, is free of empirical constants, and test calculations suggest it is generally small. However, Equation (2) is again inconvenient for time-dependent traffic modelling because of the intractable stochastic term, so a simpler expression is sought.

Dependence of the stochastic queue on green period capacity

Olszewski (1990) uses Markov simulation based on transition probabilities, allowing for a general arrival distribution and variable cycle time, to confirm observations by Newell (1965) and Miller (1969) that the mean size of the stochastic queue at a signal decreases systematically with increasing green period capacity[2] G. Although the decline is gradual, it can be substantial for long green times, and this trend appears to continue indefinitely. For example, at 90% saturation ($x = 0.9$), the mean queue with 40 second green is half that with a short (1–2 second) green. The US Highway Capacity Manual (HCM) (Rouphail, Tarko, and Li 1996) and Australian time-dependent formulae (Akçelik 1998) also contain empirical terms depending on green period capacity. Olszewski's Figure 3[3] shows that his EVOL simulation results compare well with Newell's formula, though less well with Akçelik's.

Some fundamental properties of queues and time-dependent approximation

All queues obey the time-dependent deterministic Equation (3) representing conservation of units (customers, vehicles etc.), where L_0 is the initial queue and x is the average utilisation or degree of saturation over the period of development $[0,t]$. If the demand intensity $\rho < 1$, the mean queue tends to an equilibrium value given by the Pollaczek–Khinchin (P–K) mean queue formula (4) (e.g. Kleinrock 1975):

$$L_d(x,t) = L_0 + (\rho - x)\mu t \tag{3}$$

$$L_e(x) = Ix + \frac{Cx^2}{1-x} \tag{4}$$

For equilibrium to exist, Equation (4) must be finite, so Equation (3) must also remain finite at equilibrium, implying that $\rho < 1$ and $x \to \rho$. Service occurs only when a queue is present, so at equilibrium, the *average* probability that the queue is zero is the complement of the utilisation:

$$\bar{p}_{0e} = 1 - \rho \tag{5}$$

Utilisation x represents the proportion of time that a queue is present at the stop-line or in service. The forms of Equations (3–4) show that it plays a crucial role in queue development, being the only quantity on the RHS that *must* vary with time and is capable in principle of producing any finite size of queue.

The coefficient I in Equation (4) reflects unavoidable service time, which conventionally applies only to priority junctions, while C depends on the coefficient of variation of service time.[4] For a signal $I = 0$, while theoretically $C = 0.5$, giving L_V as in Equation (1).

Empirically, C is found to be in the range of 0.5–0.6 (Burrow 1987). When Equations (3) and (4) are equated and solved for x or L, the result is the quasi-static 'sheared' approximation to time-dependent queuing, including stochastic effects, being defined for all ρ, since x never exceeds 1 (e.g. Kimber and Hollis 1979). This has some accuracy issues but is convenient for use in dynamic network assignment programs such as CONTRAM (Taylor 2003). Shearing the entire signal queue formula is problematic, as found by Han (1996). However, the methods described so far do not provide all the moments needed to determine queue size probability distributions, enabling the risk of long tailbacks, or spillback across upstream facilities, to be estimated. The rest of the paper, therefore, aims to obtain variance along with mean.

The M/D/1 process as an idealisation of the stochastic queue at a signal

The M/D/1 queue (M = Markovian, i.e. random, D = deterministic, i.e. uniform, 1 = one server) represents an idealised system where customers are served singly, but more than one random arrival can take place in each fixed average service time interval $1/\mu$. The effect of overflow from the red/green signal cycle is averaged out. In the M/M/1 process, by contrast, arrivals and service both occur randomly at given mean rates, which is more typical of a priority junction. Each process can be described by recurrence relations between queue state probabilities, which can be animated as Markov processes (Kendall 1951). Both yield closed-form equilibrium moments, including means in the P–K form Equation (4), making them convenient for use in time-dependent traffic modelling.

Arrivals at exponentially distributed random intervals result in the number of arrivals in each service period being Poisson distributed. For M/D/1, the probability of i customers being in the queue after $\mu t + 1$ service periods (i.e. that amount of throughput capacity) is the sum of the probabilities of $i + 1$ being in the queue at service point μt and no arrivals in the interval $[\mu t, \mu t + 1)$, i at μt and one arrival, $i - 1$ at μt with two arrivals, and so on. Hence, the probabilities of queue states $\{p_i\}$ where $i=\{0,1,2,\ldots\}$ evolve according to:

$$p_i(\mu t + 1) = \sum_{j=0}^{i+1} \frac{\rho^j e^{-\rho}}{j!} p_{i+1-j}(\mu t) \tag{6}$$

$$\Delta p_i(\mu t) = p_i(\mu t + 1) - p_i(\mu t) \tag{7}$$

At equilibrium, by definition, $\Delta p_i(\mu t) = 0$ for all i, so rearranging Equation (6) and allowing for the 'absorbing barrier' at $i = 0$, which corresponds to periods when the queue is zero and service is not utilised, the following equilibrium recurrence relations are obtained:

$$p_1 = (e^\rho - \rho - 1)p_0 \tag{8}$$

$$p_i = (e^\rho - \rho)p_{i-1} - \sum_{j=2}^{i} \frac{\rho^j}{j!} p_{i-j} \ (i > 1) \tag{9}$$

The first two terms in Equation (8) come from Equation (6) with $i = 0$, while the third can be obtained from Equation (6) by setting $i = -1$ (notional 'negative queue states' feature prominently later, and are discussed at the end of the paper). Moments of the equilibrium distribution can now be got from Equations (8–9) by evaluating the next highest time-dependent moment in each case:

$$\text{Probability of zero queue by evaluating} \sum_{i=0}^{\infty} i p_i : \quad p_{0e} = e^{\rho}(1-\rho) \tag{10}$$

$$\text{Mean queue by evaluating} \sum_{i=0}^{\infty} i^2 p_i : \quad L_e = \frac{\rho^2}{2(1-\rho)} \tag{11}$$

$$\text{Variance by evaluating} \sum_{i=0}^{\infty} i^3 p_i : \quad V_e = \frac{\rho^2(6-2\rho-\rho^2)}{12(1-\rho)} \tag{12}$$

The form of Equation (11) is that of the stochastic queue L_V in Equation (1) and is a particular case of Equation (4). However, Equation (10) is inconsistent with Equation (5). This is because it applies at the *end* of the green period, at which time, p_0 will be greatest since it excludes transient queues that have come and gone during the green period. That the end-of-green $p_{0\mathrm{e}}$ is always greater than the average-over-green $\bar{p}_{0e}$ (since $e^{\rho} \geq 1$) is consistent with this interpretation. This distinction does not arise where a queuing process can be formulated using infinitesimal service periods, as in the case of M/M/1.

Extending the M/D/1 queue process to general green period capacities

In this paper, we use a common subscript notation for queue moments, where *e* indicates equilibrium, [*G*] is added to indicate the dependence on green period capacity and if omitted *G* = 1 is assumed, and a final letter, e.g. M, distinguishes a particular origin, for example, the initial of an author, while an overscore indicates an average value and notional negative queue state indices are placed in brackets.

The basic M/D/1 process describes a situation where only one customer can be served in each green period (like a ramp-metre with a fixed cycle time). A more realistic model

Table 1. Conditions for specified final queue in green period with capacity *G*.

Initial queue state		Arrivals in green period
	Zero queue $i = 0$ at end of green period	
0		Up to G
1		Up to $G{-}1$
J		Up to $G{-}j$
$>G$		Not possible
	Non-zero queue $i > 0$ at end of green period	
0		Exactly $G + i$
1		Exactly $G{-}1 + i$
J		Exactly $G{-}j + i$
$>G{+}i$		Not possible

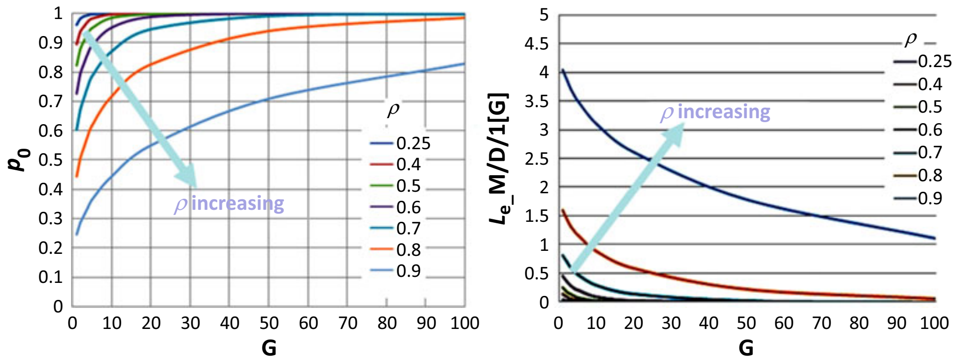

Figure 1. Dependence of M/D/1[G] equilibrium p_0 (left) and mean queue (right) on ρ, G from simulation (i_{max} = 10,000). Points measured from Olszewski (1990) are shown (EVOL).

should allow up to G customers to be serviced in each green period in which case the conditions for final queues of specified sizes are as given in Table 1.

Table 1 suggests that while the development of the queue size probability distribution $\{p_i\}$ for $i >> 0$ will be broadly similar to Equation (6), the expression for p_0 will be more complicated, involving several components. Cronjé (1983b) introduces negative net contributions in each cycle, from $-G$ to 0, representing actual arrivals minus green period capacity and sums these terms to give the value of p_0 at the end of the cycle. This amounts effectively to employing notional negative queue states down to $-G$. By a similar analysis to that for M/D/1, recurrence relations corresponding to Equations (6–9) are obtained:

$$p_i(\mu t + G) = \sum_{j=0}^{i+G} \frac{(G\rho)^j e^{-G\rho}}{j!} p_{i+G-j}(\mu t) \quad (i \geq -\mathrm{G}) \tag{13}$$

The presence of G on the LHS and the substitution of $G\rho$ for ρ on the RHS reflect the fact that all calculations relate to an idealised service period of G/μ in place of $1/\mu$. Equation (13) is valid for the notional states provided these include a notional zero state. The real (absorbing) zero state probability is got by summing all the notional probabilities, including (0), Equation (14) and states $I > G$ can be expressed in terms of *real* states alone:

$$p_0 = \sum_{i=-G}^{i=0} p_{(-i)} \tag{14}$$

$$p_i = e^{G\rho} p_{i-G}^{(e)} - \sum_{j=1}^{i} \frac{(G\rho)^j}{j!} p_{i-j} \quad (i>G) \tag{15}$$

However, there appears to be no equivalent formula for real states in the range $0<i<G$. Nevertheless, all states can be simulated incrementally using Equations (13–14). Figure 1 shows how Markov simulated[5] equilibrium p_0 (left) and mean queue size (right) vary with traffic intensity ρ and green period capacity G. These results are consistent with

Table 2. Representation of recurrence relations for G = 5 lower queue states.

G=5	**Initial state**											
Final state	0	1	2	3	4	5	6	7	8	9	10	11
(–5)	0											
(–4)	1	0										
(–3)	2	1	0									
(–2)	3	2	1	0								
(–1)	4	3	2	1	0							
(0)	5	4	3	2	1	0						
1	6	5	4	3	2	1	0					
2	7	6	5	4	3	2	1	0				
3	8	7	6	5	4	3	2	1	0			
4	9	8	7	6	5	4	3	2	1	0		
5	10	9	8	7	6	5	4	3	2	1	0	
6	11	10	9	8	7	6	5	4	3	2	1	0

Notes: Numbers in cells = numbers of arrivals in green period;
mean arrival rate = $G\rho = 5\rho$; departures in period = G exactly.

there being no positive lower limit on the mean queue size as G is increased (in practice, of course, delay on conflicting streams facing long red times might outweigh this!).

Table 2 visualises the raw recurrence relations for G = 5, where notional states appear in the final state (leftmost) column as bracketed indices, and the real initial states on which they depend are shaded. The figures in interior cells are numbers of arrivals, which translate into Poisson coefficients of the initial probabilities (columns) as in Equation (13).

The sum of terms in the column for an initial state k in Table 2 is given by:

$$K_k = \left[\sum_{i=0}^{\infty}(i+k-G)\frac{(G\rho)^i e^{-G\rho}}{i!}\right]p_k = [G\rho + k - G]\,p_k \tag{16}$$

The total of these columns must equal that of the final state (leftmost) column, so:

$$\sum_{1}^{\infty} ip_i - \sum_{0}^{G} ip_{(-i)} = L_{e[G]} - \sum_{0}^{G} ip_{(-i)} = \sum_{0}^{\infty} K_k = L_{e[G]} - G(1-\rho) \tag{17}$$

In Equation (17), $L_{e[\mathrm{G}]}$ is the mean steady-state queue. Therefore, the mean of the notional probability terms, which sum to the *real* p_0, satisfies Equation (18).

$$L_{(\,)} = \sum_{0}^{G} ip_{(-i)} = G(1-\rho) \text{ or } \quad \bar{p}_{0e} = \frac{1}{G}\sum_{0}^{G} ip_{(-i)} = (1-\rho) \tag{18}$$

Thus the *average* probability of zero queue during the service period is consistent with the deterministic queue Formula (3) and Equation (5). For G = 1, the simplicity of the formula for p_{0e}, Equation (10), is the happy result of there being two equations to solve for two unknowns. For $G > 1$, a simple formula for p_{0e} appears not to exist, and calculating the variance of notional states does not give a value for p_{0e} either, although the following hold:

$$p_{(-G)} = e^{-G\rho} p_{0e} \tag{19}$$

$$\mathrm{var}\{p_{(-i)}\}_{[G]} \rightarrow G\rho \ (p_{0e} \rightarrow 1) \tag{20}$$

Earlier empirical approximations to the mean stochastic queue

Given the computational cost of simulation, Miller (1969) proposed the following empirical approximation to the stochastic mean queue (with notation modified to be consistent with that used throughout this paper):

$$L_{e[G]M} = \frac{1}{2(1-\rho)} \exp\left(-\frac{4y}{3\rho}\right) \text{ where } y = (1-\rho)\sqrt{G} \tag{21}$$

Newell (1960) also devised an approximation, which Miller considered 'too complicated'. Cronjé (1983a) offers (without further explanation) a 'suggested modification to Newell':

$$L_{e[G]C} = \frac{I_a \rho \exp\left(-y - \frac{1}{2}y^2\right)}{2(1-\rho)} \text{ where } y \text{ is as defined in Equation.} \tag{22}$$

In Equation (22), I_a represents dispersal of arrivals, but the P–K Formula (4) does not normally allow for this, and it seems a somewhat arbitrary addition. We, therefore, drop the factor I_a in what follows. The Equations (21–22) appear to be optimised for 'heavy traffic', i.e. $\rho \approx 1$ but not exceeding 1, as is often the case in the literature, possibly because queue modelling is of most practical interest around saturation. In both cases, the exponential terms, which in principle ought to yield the M/D/1 mean when $G = 1$, do so approximately only for $\rho \approx 1$. However, the Cronjé–Newell method is accurate for higher values of ρ, making it attractive as the basis of an approximation adjusted to give improved results for smaller values of ρ. Rewriting Equation (22) as the basic mean queue multiplied by a factor f_C, as in Equation (23), an adjustment can be explored empirically by examining the behaviour of its error, Figure 2.

$$L_{e[G]C} = f_C(\rho, G) L_{e[G=1]} \quad \text{where } f_C(\rho, G) = \frac{\exp\left(-y - \frac{1}{2}y^2\right)}{\rho} \tag{23}$$

Since logarithm of the factor error escalates rapidly but linearly for values of $\rho <\sim 0.5$, the simplest correction is a bi-linear term embedded in an exponential function (24):

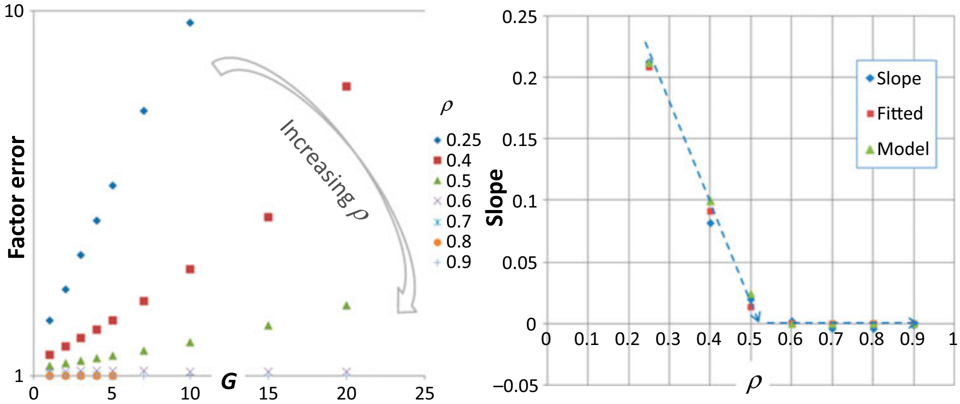

Figure 2. Errors in Cronjé–Newell factor (left) and relationship of slopes v. G to ρ (right).

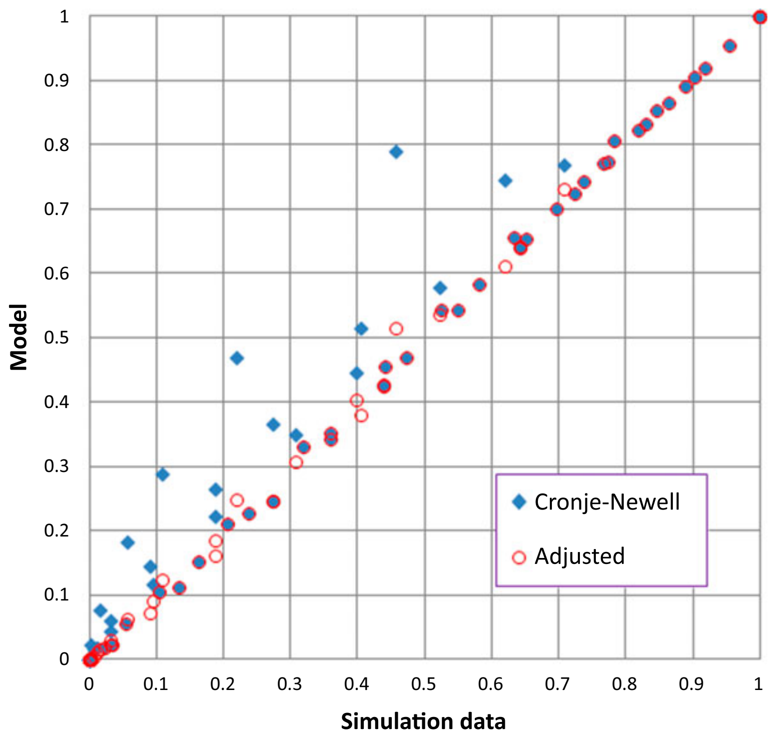

Figure 3. Performance of normalised original and adjusted Cronjé–Newell methods.

$$L_{e[G]est} = L_{e[G=1]} \min\left[\exp(\max(0.4 - 0.75\rho, 0)G)\frac{\exp\left(-y - \frac{1}{2}y^2\right)}{\rho}, 1\right] \tag{24}$$

Figure 3 shows that the adjusted approximation of Equation (24) outperforms the original Equation (23), but while queues produced by low traffic intensities $\rho <$ ~0.5 could be considered negligible for practical purposes, it is clear that some underlying structure has not been captured.

Link-function approach

The adjustment to Cronjé's method in Equation (24) is unsatisfying because apart from its one change of gradient, it conveys the message that no practical smooth function can represent the errors in a way that gives insight into an underlying structure. Experimentation reveals that the trends of $p_{0e[G]}$, $L_{e[G]}$ and $V_{e[G]}$ for various values of ρ and G can be made to overlap by transforming them using 'link functions' of the form of Equations (25–26).

$$z(h) = \frac{G+h}{\mu\tau_{re}(\rho)} \equiv \frac{G+h}{\tau_{re1}(\rho)} \tag{25}$$

$$\text{where } \tau_{re} = \mu^{-1}\left(1-\sqrt{\rho}\right)^{-2} \text{ is a relaxation time.} \tag{26}$$

For the sake of clarity, the dependence of z on ρ and G is, henceforth, 'understood'. The quantity τ_{re} is frequently cited as the stochastic relaxation time of a queue tending towards equilibrium. The variant $\tau_{re1} = \mu\tau_{re}$ is a dimensionless quantity which depends only on the demand intensity ρ. Hence the function z is also dimensionless and is independent of μ.

Figure 4 plots link-transformed $p_{0e[G]}$ against $z(2)$ for a range of values of ρ and G, showing how closely the points lie on a common parametric curve. For $G = 1$, p_{0e} and ρ are already related by Equation (10). Because those points are dispersed along the graph, extension to $G > 1$ is immediate by defining an 'effective ρ', η_0:

$$p_{0[G]est} = e^{\eta 0}(1-\eta_0) \quad \text{where} \tag{27}$$

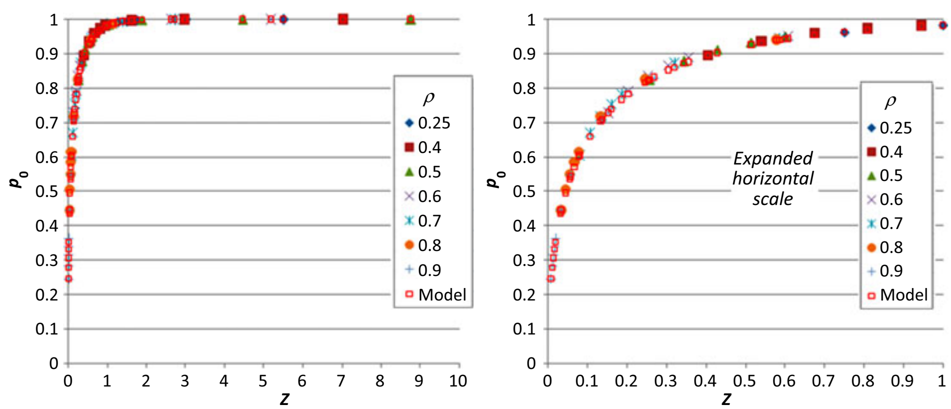

Figure 4. Plots of Markov-simulated p_{0e} against link function z which depends on ρ and G.

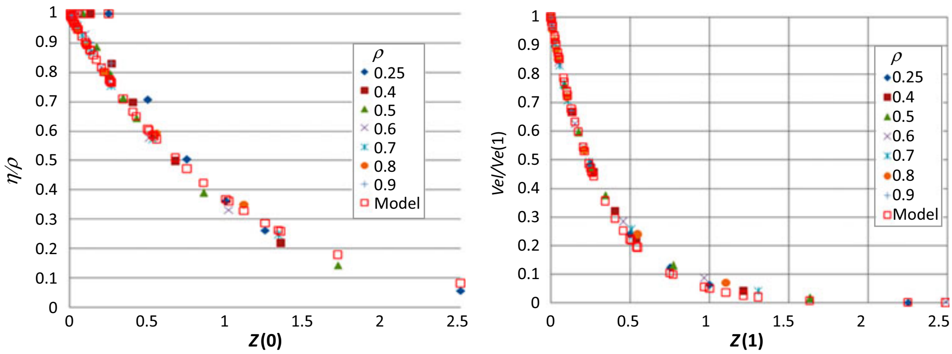

Figure 5. Plots of link-function mapping for simulated $L_{e[G]}$ (left) and $V_{e[G]}$ (right).

$$\eta_0 = \left(\max\left(1 - \sqrt{\frac{G+2}{3}}(1-\sqrt{\rho}),0\right)\right)^2 \tag{28}$$

To estimate the stochastic mean queue, an 'effective ρ' for each simulated queue value is estimated by inverting Equation (11). This quantity can be considered to represent the effective utilisation, distinct from the end-of-period p_0 given by Equation (27). For $G > 1$, as Figure 5 (left) shows, values of its ratio to ρ as a function of $z(0)$ fall roughly onto a common exponential curve, leading to Equations (29–30). The normalised equilibrium variance (right) also has an exponential appearance and can be approximated directly by Equation (31), an 'effective ρ' not being required. Upper limits ensure nominally correct results in the case $G = 1$. Note that both η_0 and $\eta_1 \to 1$ when $\rho \to 1$, so they behave like real traffic intensities.

$$L_{e[G]} = \frac{\eta_1^2}{2(1-\eta_1)} \quad \text{where} \tag{29}$$

$$\eta_1 = \rho \min\left[\exp\left(\frac{-G}{\tau_{re1}}\right),1\right] \quad \text{and also} \tag{30}$$

$$V_{e[G]} \approx V_{e[1]} \min\left[\exp\left(-3\left(\frac{G+1}{\tau_{re1}}\right)\right),1\right] \quad (\text{for } G > 1) \tag{31}$$

Figure 6, comparing Markov simulated and estimated moments, shows that that the family of approximations described earlier gives good results over the range of parameter values tested, namely, $\rho \in [0.25,0.9]$ and $G \in [1,100]$.

Estimating equilibrium queue size probability distributions

Based on his simulation results, Olszewski (1990) proposes Gamma or Negative Binomial distributions to represent the time-averaged queue size probability distribution over a peak. The Gamma distribution has been proposed elsewhere for describing queue

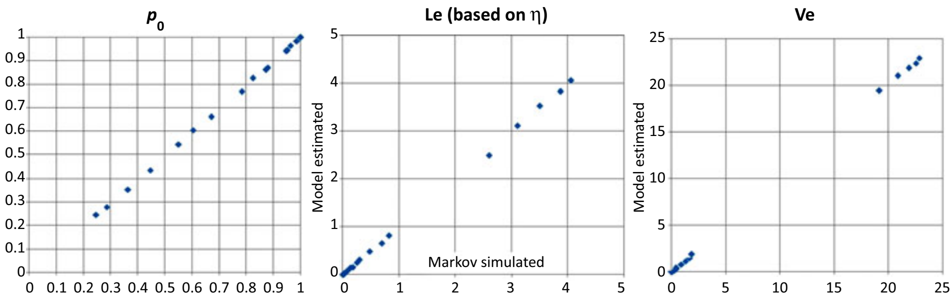

Figure 6. M/D/1[G] simulation and estimation compared for 25 pairs of ρ and G.

size distributions during oversaturated peaks (Halcrow Fox and Associates, under contract to TRL, unpublished report). It has several advantages including having two parameters for fitting and the exponential distribution (continuous equivalent of geometric) as a special case. The extended equilibrium probability distributions of M/D/1[G], including the notional terms, have a Gamma-like appearance, and manual fitting demonstrates that a close fit can be obtained for higher values of G, as shown by Figure 7. Poisson or LogNormal functions are unsatisfactory because they retain too much skewness, and their asymptotic behaviour is not exponential.

A simpler approach can be used for the real probabilities only by noting that, sufficiently far from the zero state, any equilibrium distribution is geometric, with a constant ratio between successive discrete state probabilities. This can be ascribed to a need for local invariance of the form of the distribution with change of viewpoint.

Certain equilibrium distributions in the literature have a nested form, involving two parameters instead of just ρ, where the geometric sequence applies only to queue sizes >0. However, three moments, p_0, L and V are needed to estimate queue size

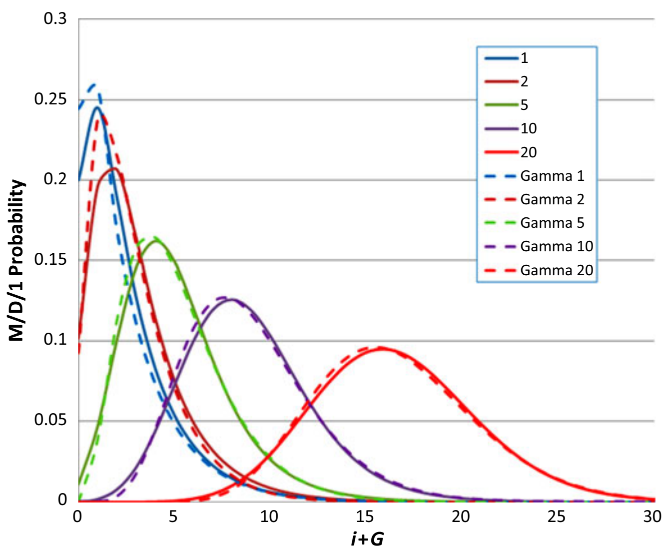

Figure 7. Gamma fits to M/D/1[G] extended distributions for ρ = 0.8, G up to 20.

probability distributions (Taylor 2005), and three parameters are required to fit three known moments. With three parameters as in Equation (32), the distribution becomes doubly nested geometric, whose first and second moments are given by Equation (33):

$$p_{0e} = 1 - \rho^* \quad p_1 = \rho^*(1 - \hat{\rho}) \qquad p_i = \rho^* \hat{\rho}(1 - \vec{\rho})\vec{\rho}^{i-2} \quad (i \geq 2) \tag{32}$$

$$L_e = \frac{\rho^*(1 + \hat{\rho} - \vec{\rho})}{1 - \vec{\rho}} \; V_e + L_e^2 = \frac{\rho^*(1 + 3\hat{\rho} - \vec{\rho}(2 + \hat{\rho} - \vec{\rho}))}{(1 - \vec{\rho})^2} \tag{33}$$

By inverting the moments in Equation (33), the ‘effective ρs’ are obtained as Equations (34):

$$\rho^* = 1 - p_{0e} \; \hat{\rho} = \frac{(V_e + L_e(L_e - 1))(1 - \vec{\rho})^2}{2\rho^*} \; \vec{\rho} = \frac{V_e + L_e(L_e - 3) + 2\rho^*}{V_e + L_e(L_e - 1)} \tag{34}$$

This will generally not work for time-dependent queue probability distributions, which can have Normal or bi-modal shape. An indicator of failure is that $\hat{\rho} > 1$. However, fitting a non-monotonic distribution where $p_1 > p_0$ is still possible, though not relevant to the present topic. Examples of fitted distributions for two values of G are given in Figure 8, showing that while p_1 is not fitted exactly, the shape of the distribution is fairly reproduced.

In principle, a fourth nesting level would allow p_1 to be fitted precisely. However, unlike p_0, which is related to the utilisation of service, p_1 cannot be calculated by analytical approximations such as shearing. Apart from this, a fourth moment, as required for a fully specified problem, is unlikely to be available from analytical approximations. So there would seem to be little benefit from making the method more complicated.

The doubly nested geometric distribution is limited to unimodal cases with mode ≤1. However, being based on general principles, it can be applied to any equilibrium distribution where all three ‘effective ρs’ are less than 1, such as those arising from Erlang-k bunched or staged arrivals or service. It cannot be assumed without justification, as is sometimes done, that equilibrium results can be applied to transient behaviour

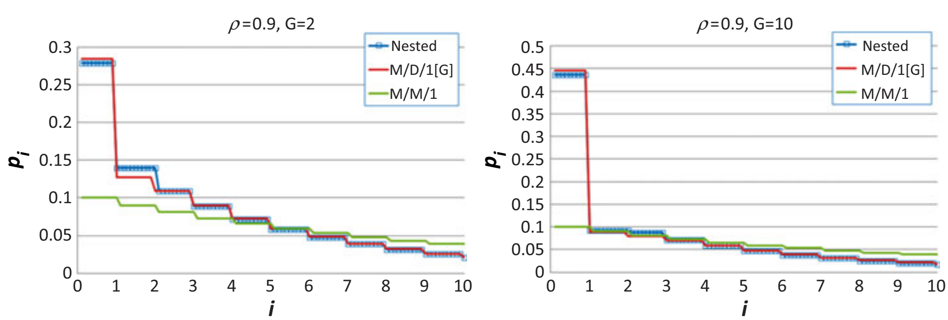

Figure 8. Fits between doubly nested geometric and Markov simulated M/D/1[G] distributions, with M/M/1 for comparison (lower graphs).

(Sharma 1990). One of us (Taylor) has also been investigating ways of fitting functions representing instantaneous probability distributions to dynamic queue moments.

Discussion

No new work has been found relevant to the topic of the effect of green period capacity on signal queues since the derivation by Heidemann (1994), possibly because the problem has been considered solved or of limited practical importance compared to the effects of optimisation, adaptive timing, platooning, coordination, etc. State-of-the-art reviews by Rouphail, Tarko, and Li (1996) and Akçelik (1998) quote Newell's approximation, and the empirically corrected HCM and Australian formulae developed in the 1980s, but do not propose any new methods, possibly because existing ones are considered sufficient.

While acknowledging this, the motivation for this paper lies in widening the range of queue types that can be handled by computationally efficient time-dependent methods like shearing, and in predicting variance and probability distributions along with means. That the green period effect emerges naturally from Heidemann's derivation indicates that it is more than just a convenient idealisation. Given that M/D/1 is taken to be a model of a stochastic signal queue, it would seem remiss to ignore the effect of something so essential.

Aside from computational efficiency, the need for simplifying approximations in queue calculation arises from the difficulty of describing even the simplest queue processes. Morse (1958) gives an exact formula for the time development of probabilities of the simplest M/M/1 random queue as a potentially infinite series of exponential or Bessel functions. Even this only describes development from a precise initial size. In realistic calculations, convolution with an initial probability distribution is required. Kleinrock (1975) calls this situation 'disheartening. While common methods exist for analysing queuing, such as the P–K transform approach (e.g. Kleinrock 1975), recent research tends to be microscopic and complex (e.g. Mirchandani and Zou 2007).

A feature of queuing, as of nature generally, is that simple relationships may emerge from a complex process through general principles of conservation and symmetry. Equilibrium results like that in Equation (4) are 'emergent' in the sense that exact formulae need not exist for the equilibrium mean of a general queuing process, since it is the limiting outcome of an infinite sequence of random events repeated infinitely many times. Yet all queue processes must obey deterministic conservation Equation (3), which, therefore, cannot say anything about any result that depends on the details of the process.

An equilibrium queue is defined only for traffic intensities $0 \le \rho < 1$ and is unbounded as $\rho \to 1$. Thus it should be possible to normalise equilibrium moments by factoring by finite expressions in ρ and G since this maps an infinite range into an infinite range. A physical interpretation is that random arrivals in green periods of different duration should on average have a similar impact on the stochastic queue at the end of a period, provided that the relationship between green time and the characteristic stochastic time scale is the same. The link-function approach exploits this symmetry. But constant factors are not necessarily appropriate for time-dependent queues. Expressing results in terms of generic formulae (3) and (4) with modified parameters – in the present case 'effective ρs' – allows the use of existing computationally efficient time-dependent methods.

On what basis can it be claimed that 'effective ρs' can be substituted directly into the time-dependent methods? Physically, the longer the green period, the more traffic can 'disappear without trace' before the queue is assessed at the end of the green period. This translates into a reduction in the effective demand intensity and degree of saturation, which are the critical variables in time-dependent queuing. The quasi-static principle relies on the rate at which information propagates through a queue being much greater than the rate of change of the queue itself so that the static relationship between queue size and degree of saturation can be assumed to apply approximately to dynamic cases. As long as these assumptions hold, it should be possible to apply methods like shearing with some confidence. Some observational validation of shearing was done by Kimber and Daly (1986). Validation of the extensions proposed here is part of a larger question concerning time-dependent methods, which involves considering dynamic probability distributions or at least their moments, currently being addressed by one of us (e.g. Taylor 2005).

The primary impact of the work described here is, therefore, greater consistency. The extent to which approximate time-dependent methods can accommodate realistic factors such as platooning, coordination and finite capacity queues has yet to be determined, though pursuance of research may depend on whether macroscopic modelling is considered likely to have a role as against microscopic simulation, an issue which is still disputed (Wood 2012). An ad hoc method of accounting for green waves was used in CONTRAM (Taylor 2003), but manipulation of parameters in an extended P–K mean formula representing arrival and service statistics would be preferable. On the other hand, a purely statistical model like the P–K formula may be inappropriate to cases where variability of demand is no longer simply random (Chow 2013). Whether efficient time-dependent methods can continue to absorb real-life complexities through the implementation of general principles may be a question for further research.

Conclusion

This paper responds to a perceived need to: (1) broaden the range of queue processes that can be modelled using computationally efficient time-dependent methods based on closed formulae; (2) incorporate variance and in particular (3) take account of the observation that the size of a stochastic signal queue depends on the green period capacity and not just the ratio of green to cycle time through the demand intensity.

An idealised stochastic signal queue process that takes account of green period capacity has been formulated by extending the basic M/D/1 model embodied in some existing signal queue formulae. Simulation using Markov methods has been used to generate test cases to compare with existing empirical approximations and to verify novel formulae for equilibrium values of the probability of zero queue, mean queue and variance in a form compatible with computationally efficient macroscopic time-dependent methods.

It is thought on structural grounds that the realism of transient behaviour using time-dependent methods should be similar to that for the basic process. Three moments enable probability distributions to be estimated and a simple doubly nested geometric approximation has been defined and demonstrated.

Acknowledgements

The work described in this paper forms part of the first author's research towards Ph.D. under the supervision of the second author. An early version of the paper was presented at the Universities Transport Studies Group (UTSG) Conference at Oxford, January 2013. Helpful comments by the editors and two anonymous reviewers and the general support of Dr Alan Stevens Transportation Chief Scientist at TRL are gratefully acknowledged.

Notes

1. Strictly, capacity in vehicles should vary with traffic composition.
2. *G* as used here corresponds to Olszewski's *B*.
3. We were unable to obtain permission to reproduce this, but some points taken from it are shown in Figure 1 later.
4. Some authors make *C* include the dispersion of arrivals also, but this does not emerge from the derivation.
5. The Markov simulation program used, which evaluates recurrence relations using small time steps, was developed by the first author with algorithm design and programming assistance by Neil H Spencer, then a sandwich student at TRL.

References

Akçelik, R. 1998. *Traffic Signals: Capacity and Timing Analysis*. Report ARR 123. Vermont South: Australian Road Research Board. [Most recent edition of review first published 1981].

Allsop, R. E., and T. P. Hutchinson. 1972. "Delay at Fixed-Time Traffic Signals [Parts I and II Respectively]." *Transportation Science* 6 (3): 260–305. doi:10.1287/trsc.6.3.260.

Burrow, I. J. 1987. *OSCADY: A Computer Program to Model Capacities, Queue and Delays at Isolated Traffic Signal Junctions*. TRL Report RR 105. Crowthorne House: Transport Research Laboratory.

Chow, J. Y. J. 2013. "On Observable Chaotic Maps for Queueing Analysis." In *Proceedings of 92nd TRB Annual Meeting, January 2013*, January 13–17. Washington, DC: Transportation Research Board.

Cronjé, W. B. 1983a. "Analysis of Existing Formulas for Delay, Overflow and Stops." *Transportation Research Record* 905: 89–93. Washington, DC: Transportation Research Board.

Cronjé, W. B. 1983b. "Optimization Model for Isolated Signalized Traffic Intersections." *Transportation Research Record* 905, 80–82. Washington, DC: Transportation Research Board.

Han, B. 1996. "A New Comprehensive Sheared Delay Formula for Traffic Signal Optimisation." *Transportation Research A* 30 (2): 155–171.

Heidemann, D. 1994. "Queue Length and Delay Distributions at Traffic Signals." *Transportation Research* 24B (5): 377–389.

Kendall, D. G. 1951. "Some Problems in the Theory of Queues." *Journal of Royal Statistical Society B (Methodological)* 13(2): 151–183.

Kimber, R. M., and P. Daly. 1986. "Time-Dependent Queuing at Road Junctions: Observation and Prediction." *Transportation Research* 20B (3): 187–203.

Kimber, R. M., and E. M. Hollis. 1979. *Traffic Queues and Delays at Road Junctions*. TRL Report LR 909. Crowthorne House: Transport Research Laboratory.

Kleinrock, L. 1975. *Queueing Systems*: Vol. 1 *Theory*. New York: Wiley Interscience.

Meissl, P. 1963. "Zufallsmodell einer Lichtsignalanlage mit mehrspurigem Stauraum [A Random Model of a Traffic Signal with Multilane Storage]." *Mathematik-Technik-Wirtschaft* 63 (1): 1–4 and 63 (2): 63–68.

Miller, A. J. 1969. *Some Operating Characteristics of Fixed-Time Signals with Random Arrivals*. Sydney: Institution of Highways and Traffic Research, University of New South Wales.

Mirchandani, P. B., and N. Zou. 2007. "Queuing Models for Analysis of Adaptive Signal Control." *IEEE Transactions on Intelligent Transportation Systems* 8 (1): 50–59. doi:10.1109/TITS.2006.888619.

Morse, P. M. 1958. *Queues Inventories and Maintenance*. New York: John Wiley.

Newell, G. F. 1960. "Queues for a Fixed-Cycle Traffic Light." *Annals of Mathematical Statistics* 31: 589–597. doi:10.1214/aoms/1177705787.

Newell, G. F. 1965. "Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light." *SIAM Review* 7: 223–240. doi:10.1137/1007038.

Olszewski, P. S. 1990. "Modelling of Queue Probability Distribution at Traffic Signals." In *Proceedings of the International Symposium of Traffic and Transportation Theory*, edited by M. Koshi, 569–588. New York: Elsevier.

Rouphail, N., A. Tarko, and J. Li. 1996. "Traffic Flow at Signalized Intersections." In *Chapter 9 of Traffic Flow Monograph*. Washington, DC: Transportation Research Board, pp. 9-1–9-32. http://www.fhwa.dot.gov/publications/research/operations/tft/chap9.pdf

Sharma, O. P. 1990. *Markovian Queues*. Chichester: Ellis Horwood.

Taylor, N. B. 2003. "The CONTRAM Dynamic Traffic Assignment Model." *Networks and Spatial Economics Journal – Special Issue on Dynamic Traffic Assignment* 3: 297–322. doi:10.1023/A:1025394201651.

Taylor, N. B. 2005. "Variance and Accuracy of the Sheared Queue Model." In *Proceedings of IMA Mathematics in Transport Conference*, edited by B. G. Heydecker, 259–278. Amsterdam: University College London, September 7–9.

Webster, F. V, and B. M. Cobbe. 1966. "Traffic Signals." *Road Research Technical Paper 56*. London: HMSO.

Wood, S. 2012. "Traffic Microsimulation – Dispelling the Myths." *Traffic Engineering and Control* 53 (9): 339–344.