State Dependent Travel Time Models, Properties, & Kerner's Three-phase Traffic Theory

J. MacGregor Smith*
jmsmith@ecs.umass.edu

F. R. B. Cruz†
fcruz@est.ufmg.br

November 21, 2012

***Abstract* — One of the most important problems in today's modelling of transportation networks is an accurate estimate of travel time on arterial links, highway, and freeways. There are a number of deterministic formulas that have been developed over the years to achieve a simple and direct way to estimate travel times for this complex task. Realistically, however, travel time is a random variable. These deterministic formula are briefly reviewed and also a new way to compute travel time over arterial links, highway, and freeways, is presented based on an analytical state dependent queueing model. One of the features of the queueing model is that it is analyzed within the context of the theoretical three-phase traffic flow model. We show that the model provides a quantitative foundation alternative to qualitative three-phase traffic flow theory. An important property shown with the model is that the travel time function is not convex, but a sigmoid S-shaped (*i.e.* logistics curve). Extensive analytical and simulation experiments are shown to verify the S-shaped nature of the travel time function and the use of the model's method of estimation of travel time over vehicular traffic links as compared with traditional approaches. Finally, it is shown that the point-of-inflection of the S-shaped curve represents the threshold point where the traffic flow volume switches from Free Flow to Congested Flow.**

Keywords — Transportation; State dependent queue; travel time; approximations.

1 Introduction

"Therefore, queueing theories cannot be relied upon for a correct description of congestion on unsignalized freeways, and the term 'queue' will henceforth not be used to describe traffic on freeways", Boris S. Kerner, **The Physics of Traffic**, p. 85, Springer-Verlag, Berlin, 2004.

Most design and analysis studies of transportation networks need to estimate in some way the travel time over the traffic network segments. This is a fundamental but difficult problem because of congestion, natural variability of rates of travel, time of day, road and weather conditions, driver characteristics, and vehicle types. The traffic flow process is essentially a stochastic (random) process. Since this is a random process, one needs to estimate the probability distribution of the number of vehicles along a roadway in order to compute the desired performance measures of the roadway such as throughput, queueing delays, average number of vehicles along the segment and the utilization of the roadway segment. It is important to have an accurate formula since the use of this formula is critical to transportation planning and traffic assignment activities.

1.1 Motivation and Purpose

In this paper, we illustrate the special properties of a state dependent queueing approach to traffic modeling. Part of the reason for this is to show that through the queueing approach, it helps explain the often frustrating experience that drivers face when there is severe congestion on roadways. On another account, with the queueing approach, it begins to explain quantitatively some of the fundamental assertions of Kerner's three-phase traffic theory (Kerner, 2004).

Figure 1 illustrates the sudden disruptive effects and decreasing flow volume (*i.e.* throughput) which can occur while driving along a roadway at the lone occupant speed (maximum speed), $V_1 = 50$ mph respectively for a freeway stretch of one mile and one lane wide where the jam density is 220 vehicles/mile. Thus, as illustrated in Figure 1, there is a halt to the monotonic increase in flow volume (throughput), then a sudden decrease, and finally

*Department of Mechanical and Industrial Engineering, University of Massachusetts Amherst, Massachusetts 01003

†Departamento de Estatística, Universidade Federal de Minas Gerais, 31270-901 - Belo Horizonte - MG, Brazil. Partially funded by Brazilian Agencies CNPq, FAPEMIG, and PRPq-UFMG.

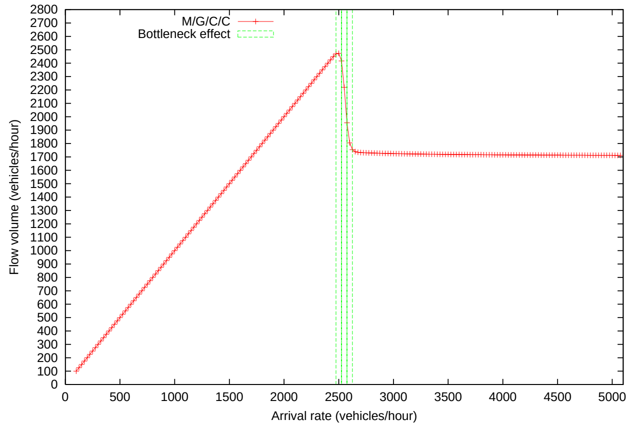


Figure 1: Flow volume (throughput) for a freeway stretch of one mile long and one lane wide for a maximum speed of $V_1 = 50$ mph and a jam density of 220 vehicles/mile

a leveling off. There is a myriad of flow rates (see the bottleneck effect region in Figure 1) between the crucial input arrival rate and the saturated flow volume. From a queueing perspective, this is a remarkable phenomenon. Most plots of throughput in finite buffer queueing studies are smooth monotonic concave functions that reach a plateau and do not descend, whereas this throughput curve is quasi-concave at best. These phantom effects of sudden slow-downs are also related to the empirical approach and understanding of the three-phase flow of traffic theory developed by Kerner (2004).

Another feature of this paper is to develop an approach to understanding what type of travel time calculations will result from the queueing approach. This is motivated by the following situations and examples.

For example, in estimating the amount of fire insurance for a community's fire protection, insurance companies often utilize a simple formula to estimate the response time for an engine company to respond to a fire. While the fundamental response time is a stochastic process, the insurance companies will often use a deterministic formula to estimate this response time. In fact, one formula often used is based upon linear-regression research by Hausner et al. (1974) and is the following:

$$T = 0.66 + 1.77D,$$

where:

$T :=$ time in minutes to the nearest 1/10 of a minute;

$0.66 :=$ a vehicle acceleration constant for the first 1/2 mile traveled;

$1.77 :=$ a vehicle speed constant validated for response distances ranging from 1/2 miles to 8 miles;

$D :=$ Euclidean distance based upon two-dimensional Cartesian coordinates from the engine company location to the fire location.

This formula was derived in a RAND Corporation study in New York city and is still used as a predictive tool. The formula is somewhat dated and is derived from empirical data in Yonkers, New York, but appears to be a very practical one (Insurance Services Office, 2011).

Another situation is in the traffic planning area, which is considered a crucial component since stable transportation systems are one of the main factors that contribute to a higher quality of life. Indeed, Buriol et al. (2010) propose an algorithm to solve a user equilibrium model (which describes the behavior of users on a given traffic network), while maintaining the system optimum model solution (which describes a traffic network operating at its best operation). This algorithm is based on a convex *approximation* to represent the cost of traveling along each arc, as a function of the flow on the arc (more on traffic assignment models on the classical book of Sheffi, 1985). This is a relevant area of research since any improvement is significant because of the many millions of dollars that are spent every day on traffic issues, as reported by Arnott and Small (1994).

There have been many formula developed over the years for estimating roadway traffic travel times. Why is there any need for another formula? First of all, it is important that the theoretical formula reflect as close as possible what happens in reality. If the understanding of the travel time phenomenon is not empirically and theoretically sound, then all the models based on the formula are deficient.

Secondly, if one can easily compute a more accurate estimate of travel time in congested environments, then this will be an important boost to its application in traffic modeling and traffic network design. As might be expected, one does not get something from nothing. In order to achieve a more accurate estimate, the tradeoff is that more computation work has to be done, yet the real benefit here will be a better approximation to the quantitative measure for congested highway traffic. With the advent and proliferation of powerful computer processors (even in cell phones and their web inter-connectivity), the computation times should not become extraordinary.

1.2 *Outline of Paper*

Section 2 presents a survey of relevant literature about travel time models and discusses two-phase and three-phase traffic theory of Kerner (2004) in contrast to the theory which derives from the state dependent approach. Section 3 presents the state dependent model and its derivation and computation of travel times along highway segments. Section 4 presents a comparison of the computation of travel times in particular the Bureau of Public Roads (1964) formula and a formula by Akçelik (1991) along with the state dependent model. Section 5 examines some incidents and bottlenecks and compares the state dependent approach with the other travel time models. Section 6 summarizes and concludes the paper.

2 Problem and Literature Review

In traffic assignment and evacuation planning models, a travel time delay function is necessary in order to express the relationship between the traffic flow volume and the expected delay on the traffic link. Numerous formulas have been constructed over the past 40 years for this purpose, among some of them are the contributions by the following authors: BPR (Bureau of Public Roads, 1964); Davidson (1966); Rose et al. (1989); Spiess (1990); Akçelik (1991); and Dowling et al. (1998). There are even other formulas not mentioned, but the above sample is considered representative.

2.1 *Criteria of Formula*

In one of the previous papers, Spiess (1990) has recommended a series of seven guidelines for a well-behaved delay function. If one posits a function $f(x)$ where x represent the ratio of volume to capacity v/k then Spiess recommends the following properties the function should maintain (Horowitz, 1991).

1. $f(x)$ should be strictly monotone increasing.
2. $f(0) = 1$ and $f(1) = 2$. In other words, the function should yield the free flow travel time at zero volume and twice the free flow travel time at capacity.
3. $f'(x)$ should exist and be strictly monotone increasing, *i.e.* the derivative is representative of a convex function. A convex function is desirable for optimization purposes.
4. $f'(1) = \alpha$, the exponent in the BPR function, *i.e.* the function has only a few well-defined parameters.
5. $f'(x) < M\alpha$, where M is a finite positive constant, *i.e.* the function should be finite for all volumes.
6. $f'(0) > 0$, the derivative is positive at zero volume.
7. The evaluation of $f(x)$ should not take more computation time than does the evaluation of the corresponding BPR function.

These guidelines are interesting because they can be used to compare alternative formulas. There is some argument that criterion #2 should be relaxed (Horowitz, 1991), since there is no intuitive rationale for such a criterion besides to ensure compatibility with the well known BPR type functions (Spiess, 1990), and that the third and seventh standards are inhibiting and should be dropped. The seventh criterion is important since if the calculations are excessive, then it will create convergence problems in the traffic assignment calculations.

Before we present the state dependent queueing model used to compute travel time, let us review the basic way of incorporating congestion in traffic flow models through a two-phase traffic flow model, then a three-phase traffic flow model.

2.2 *Two-Phase Traffic Models*

In two-phase traffic flow models, the argument is that there are only two phases of vehicular traffic flow: i) *Free Flow* and ii) *Congested Flow.* There are empirical data which show a strong correlation between the flow rate Q and vehicle density ρ, so that there is an upper boundary at the maximum point of free flow that corresponds to also a critical density value (see Kerner, 2004, pages 22–26). This two-phase traffic flow is illustrated in Figure 2. Some of the formula based upon two-phase traffic flow theory are reviewed in Section 4.

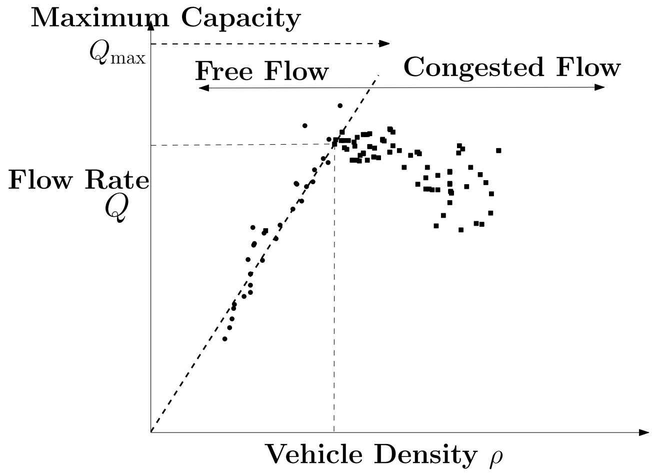


Figure 2: Correlation between flow rate Q and vehicle density ρ and the two-phase traffic flow model (free flow and congested flow)

2.3 *Three-Phase Traffic Models*

As we shall argue, three-phase traffic flow models originate with Kerner's research (Kerner, 2004) and the proposition that three-phases occur in traffic modeling, so that in *Congested Flow* there are two additional phases, *Synchronized Flow* and *Wide-Moving Jam Flow* as follows.

i) Free Flow:
   1) Free Flow Traffic;

ii) Congested Flow:
   2) Synchronized Flow;
   3) Wide-Moving Jam Flow.

One of the features of Kerner's work (Kerner, 2004) is that he only seems to be interested in the empirical foundation for the three-phase traffic flow models. This notion of phase transitions as we shall argue is similar to what happens for the state dependent model. Notice that we do not mean to imply that the three-phase models are widely accepted and that consistency with them lends credence to our state dependent model. Notable studies along these lines include Schónhof and Helbing (2009) and Treiber et al. (2010), which calls the theory into question on theoretical and empirical grounds. These are studies regarding travel time estimation in the context of considerably recent and on-going research on travel time monitoring.

One of Kerner's main hypotheses is that there is an infinite number of highway capacities of free flow at a bottleneck. These infinite capacities are bounded between a minimum and maximum capacity (see Kerner, 2009, chapter4). We have no problem with this argument, however, we want to show that the phase transitions can occur without a bottleneck.

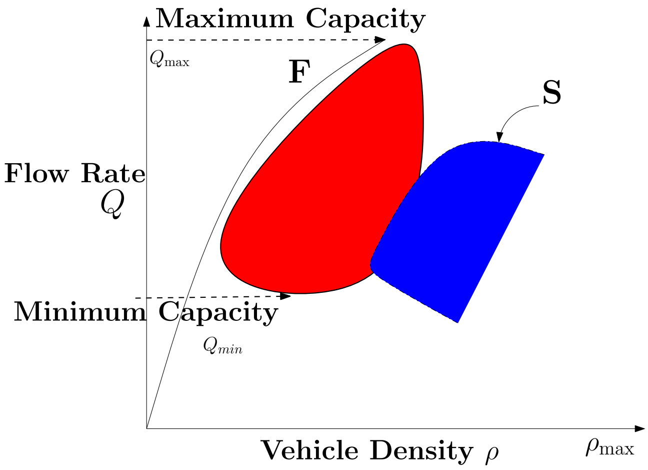


Figure 3: Relationship between the flow rate Q, the vehicle density ρ, and Kerner's three-phase flow theory (free-flow, F, and synchronized-flow, S)

Figure 3 illustrates the relationship between the flow rate of vehicular traffic and the density of vehicles in the three-phase traffic flow theory. This diagram is modeled after Kerner's book (see Kerner, 2009, p. 76). $F \rightarrow S$ illustrates the free flow (F) until there is some bottleneck at which the synchronized flow (S) occurs. Within the S phase, self compression of the vehicles comes into play and the density increases and this self compression is due to what Kerner calls the "pinch effect". There are minimum and maximum capacities (*i.e.* throughputs) which then occur. In fact, there are infinite capacities within a range of critical capacities.

In contrast to Figure 3 as was shown earlier, Figure 1 illustrates that at the "pinch point", there is also a sudden drop in capacity and eventually it levels off to a constant capacity (throughput) because the traffic has reached the jam density of the roadway segment. What Figure 1 clearly shows is that based upon a single road segment, not even a bottleneck, the state dependent model indicates the *threshold effects* of moving from free-flow (F) to a wide moving jam (J) with an intermediary threshold stage (*i.e.* $F \rightarrow S \rightarrow J$). State dependent traffic flow models offer a "quantitative" explanation to the qualitative observations of Kerner's theory since the queueing theory indicates a crucial relationship between the blocking effects of the finite queue, the arrival flow rates, and the ultimate capacity (*i.e.* throughput) of the finite queue. A bottleneck can induce a wide moving jam, but the wide moving jam can occur without the bottleneck! The roadway itself is a bottleneck.

Let us explain in some detail in the following section the state-dependent $M/G/c/c$ model and its properties and why we feel that a traffic flow theory based upon $M/G/c/c$ queueing models is worthy of examination.

3 $M/G/c/c$ State Dependent Model

We assume that travel time T is a *random variable* and that it is a function of the volume of traffic along the road segment. This road segment has finite length and width and these together with the vehicle sizes create a finite queueing situation, as seen in Figure 4. Thus, it is important here in this model to account for the length of the roadway segment not only its width (# of lanes) as this will directly affect the physical capacity of the road segment. Figure 4 illustrates four traffic lanes (road segments) which underly the basic relationship of the $M/G/c/c$ approach in which M stands for a queueing system with stationary Poisson arrivals at some rate λ, G for a general service time distribution (which will be considered *state dependent* in a sense that will be made clear soon), c servers in parallel, and room for at most c customers *including* those in service (*i.e.* no queue).

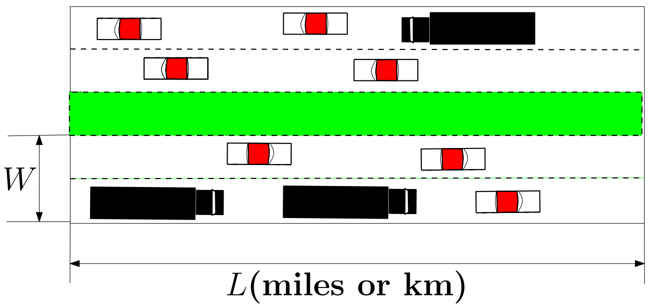


Figure 4: Finite queueing situation created by a road segment with finite length and width together with the vehicle sizes

3.1 Notation

At this stage of the paper, it is useful to have a list of important notation used in the discussion. Since the audience is both traffic researchers and operations researchers, we try to provide the notation that is common parlance.

M := Markovian arrivals (*i.e.* Poisson arrivals);

G := State dependent general service time probability distribution;

c := Number of servers and physical capacity of an $M/G/c/c$ queue;

L := Length of the road segment (miles or km);

W := Width of the road segment (# lanes);

ρ := jam density parameter, in vehicle/mile-lane (normally ranging from 185-265 vehicle/mile-lane);

Λ := External Poisson arrival rate to the network;

λ_j := Arrival rate to internal network node j;

p_c := Blocking probability in an $M/G/c/c$ queue (*i.e.* probability that arrivals are rejected because the queueing system is full);

Q := Capacity or vehicular flow volume (*i.e.* throughput of a highway segment);

θ := Mean throughput rate which is equivalent to Q above;

T := Travel time along a road segment or a path network;

WIP := Average work-in-process or number of vehicles in the queueing system.

3.2 Basic Relationships

In the basic road segment model, the maximum number of vehicles allowed on a road segment is captured by the (jam density) c is the following expression:

$$c = \rho \times L \times W. \tag{1}$$

That all highway segments have finite capacity (which we will define as c) based upon the jam density, ρ, the geometry of the roadway (*i.e.* its length, L, and number of lanes, W), and natural geography and topography where the roadway resides is perhaps an obvious observation. However, the implications of the finiteness are very critical for all that is to follow. Since there is a bounded length and width and depending upon the jam density, the highway can only handle a finite number of vehicles. One could argue that the bounding of the length is arbitrary, yet there are natural features of a highway such as changes in slope, plateaus and valleys, and the presence or absence of on- and off-ramps which naturally suggests a finite bound to L.

The state dependent model is a stochastic model which requires one to compute the probability distribution of the number of vehicles along a roadway as a function of the density of vehicular traffic traveling down the road segment. The speed-density curves which normally describe the vehicle speed along a segment will be used by the state dependent model to calculate the probability of the number of vehicles along the road segment. Once this probability distribution is found, the mean time to traverse the segment can be computed.

It is important to point out that the state dependent model is both a combined macro-level as well as a micro-level traffic model. It incorporates the macro behavior of individual cars through the state dependent curves, yet models each individual car traveling the highway segment. Each individual vehicle along the highway segment has its speed *i.e.* service rate adjusted "dynamically" to the number of vehicles traveling along the segment. As one vehicle enters or leaves the road segment, all the other vehicles still on the segment have their speeds adjusted. Whatever speed density curve is appropriate to the highway segment can be utilized in the state dependent model. Another important notion for the $M/G/c/c$ model is that there is no queue or queue discipline. The vehicle is either on the highway segment or it is not.

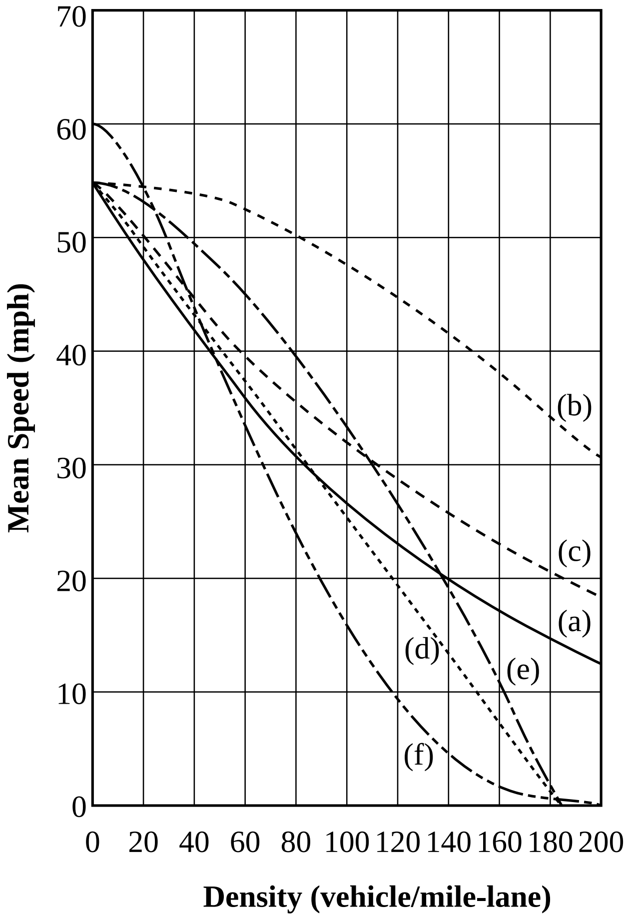
Figure 5 on the left presents an exponential approximation of empirical $M/G/c/c$ state-dependent traffic flow model, (a) (Jain and MacGregor Smith, 1997), and empirical distributions for vehicular traffic flows, (b) (Edie, 1961), (c) (Underwood, 1961), (d) (Greenshields, 1935), (e) (Drake et al., 1967), and (f) (Transportation Research Board, 2000). Figure 5 on the right illustrates experimental curves, (a) through (f), that relate the walking speed of pedestrians to crowd density, based on various empirical studies that illustrate that at a mean density of 3 pedestrians/m^2 walking is reduced to a shuffle, and at 5 pedestrians/m^2, forward movement essentially comes to a halt, as reported by Tregenza (1976). The striking similarity to the speed-density curves indicates that the results developed in this paper for vehicle modeling applies also to pedestrian traffic flow modeling (Yuhaski and Macgregor Smith, 1989). It is felt that many other particulate flows have also this speed decaying behavior as a function of density.

Classical macroscopic traffic studies by Greenshields and Greenberg have documented these linear and exponential decay functions for vehicular traffic (see May, 1990), and the state-dependent models can capture this macroscopic behavior. An important observation here is that any empirical curve can be utilized in the segments of the traffic studies through the queueing representation. One is not restricted to one single fixed speed-density curve. This yields a great deal of flexibility for the approach, which is not readily affordable in other approaches, except perhaps simulation (Cruz et al., 2005, 2010a). The problem with simulation is that it is very expensive from a computational running time and storage viewpoint to dynamically update the service rate of each vehicle in the traffic segment as a function of density of traffic in the segment. This is not a problem for the analytical models.

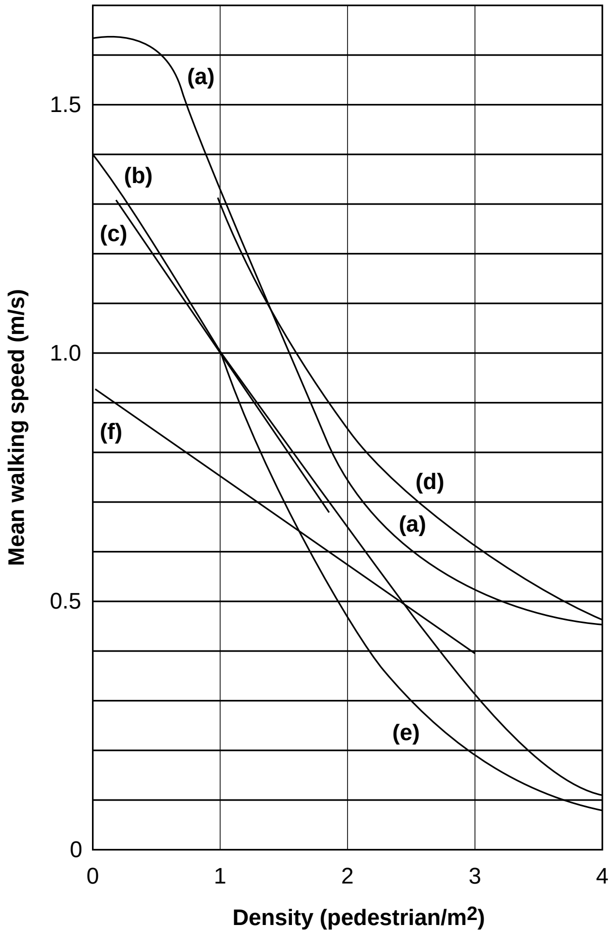
Over many years, a generalized model of the $M/G/c/c$ Erlang loss queueing model for this service rate decay which can model any service rate distribution (linear, exponential, ...) has evolved (Cheah and MacGregor Smith, 1994). It is a special case of an Erlang Loss model.

3.3 Derivation of Probabilities

The basic probability distribution for a linear congestion model and an exponential congestion model will be shown. The probability distribution is critical to the performance measures and especially the travel time func-



a) Exponential approximation of empirical $M/G/c/c$ state-dependent traffic flow model, (a) (Jain and MacGregor Smith, 1997), and empirical distributions for vehicular traffic flows, (b) (Edie, 1961), (c) (Underwood, 1961), (d) (Greenshields, 1935), (e) (Drake et al., 1967), and (f) (Transportation Research Board, 2000)



b) Experimental curves, (a) through (f), that relate the walking speed of pedestrians to crowd density, based on various empirical studies reported by Tregenza (1976)

Figure 5: Approximation of empirical vehicular/pedestrian speed-density curves

tion. Some of the details of this development are deferred to previous articles (Jain and MacGregor Smith, 1997; Yuhaski and Macgregor Smith, 1989).

3.4 Linear Congestion Model

The linear congestion model is based on the idea that the service rate of the servers in the $M/G/c/c$ queueing model is a linear function of the number of occupants in the system. This is basically the approach of Greenshields (see May, 1990). It is worthwhile mentioning though that linear/exponential models which do not contain the loss feature are also possible, for instance $G/G/1$ models (Vandaele et al., 2000).

A linear congestion model can be developed where the vehicle speed of a single vehicle is V_n and as it approaches the jam density c of the road segment $V_n \to 0$. For this reason, since a vehicle population of $n = c + 1$ is impossible, $V_n = 0$ is set for all $n \geq c + 1$.

Thus, if V_1 is the average travel speed of a lone vehicle and $V_{c+1} = 0$, then:

$$V_n = \frac{V_1}{c}\,(c + 1 - n)\,. \tag{2}$$

Equation (2) gives the vehicle-speed of n vehicles in a single lane, which is also the number of busy servers. Note that the service rate, r_n, of each of n vehicles in the lane, is the average of the inverse of the time it takes these individuals to traverse the length of the lane; therefore,

$$r_n = V_n/L. \tag{3}$$

Using Equation (2), this gives us,

$$r_n = \frac{V_1}{cL}(c + 1 - n). \tag{4}$$

The service rate of the queueing system (overall) is equivalent to the number of servers in operation (*i.e.*, occupied) multiplied by the rate of each server. Since all n servers, in a state-dependent $M/G/c/c$ queueing model, operate at the same rate, r_n, we have that μ_n, the overall service rate of the system when there are n vehicles in the lane, may be defined as:

$$\mu_n = nr_n = \frac{V_1}{cL}(c+1-n)n. \tag{5}$$

Expressions for the state probabilities are derived by substituting the expression for μ_n, Equation (5), into the Chapman-Kolmogorov equations for solving the probabilities of a single queue

$$p_n = \frac{\lambda^n}{\Pi_{i=1}^n \mu_i} p_0, \tag{6}$$

for $n = 1, \cdots, c$, and

$$\frac{1}{p_0} = 1 + \sum_{n=1}^{c} \left\{ \frac{\lambda^n}{[\Pi_{i=1}^n \mu_i]} \right\}, \tag{7}$$

to obtain Equations (8) and (9):

$$p_n = \frac{\lambda^n}{\left(\frac{V_1}{cL}\right)^n \Pi_{i=1}^n (c-i+1)i}\, p_0, \tag{8}$$

$$\frac{1}{p_0} = 1 + \sum_{n=1}^{c} \left\{ \frac{\ell^n}{\Pi_{i=1}^n (C-i+1)i} \right\}, \tag{9}$$

where V_1 is the free-flow speed, and $\ell = \lambda cL/V_1$. Note that L is expressed in meter (or miles) and λ is expressed in hour^{-1}. Thus, Equations (15) and (16) gives us the desired probability distribution for the linear congestion model.

3.5 Exponential Congestion Model

In developing the exponential congestion model, assume that, r_n, the service rate of each of the n occupied servers, is related to the number of vehicles by an exponential function. The form of the exponential function is based on the equation for the vehicle-speed, as depicted by the following relation.

$$V_n = V_1 \exp\left[-\left(\frac{n-1}{\beta}\right)^{\gamma} \right]. \tag{10}$$

Parameters β and γ are found by fitting points to the curve in Figure 5. Parameters β and γ will be referred to as the scale and shape parameters respectively. Fitting the points $(1, V_1), (a, V_a)$, and (b, V_b) gives one the algebraic relationships shown below:

$$\gamma = \ln\left[\frac{\ln(V_a/V_1)}{\ln(V_b/V_1)} \right] / \ln\left(\frac{a-1}{b-1}\right), \tag{11}$$

$$\beta = \frac{a-1}{\left[\ln\left(\frac{V_1}{V_a}\right)\right]^{1/\gamma}} = \frac{b-1}{\left[\ln\left(\frac{V_1}{V_b}\right)\right]^{1/\gamma}}. \tag{12}$$

By carefully approximating the positions of three representative points among the curves in Figure 5, the following sample coordinates are utilized:

$$\begin{aligned}
&V_n = 62.5 \text{ mph} \Rightarrow \text{at density} \\
&\qquad \delta = 1/LW \text{ vehicles/mile-lane } \Rightarrow n = \delta LW = 1, \\
&V_n = 48.0 \text{ mph} \Rightarrow \text{at density} \\
&\qquad \delta = 20 \text{ vehicles/mile-lane } \Rightarrow n = a = 20LW, \\
&V_n = 20.0 \text{ mph} \Rightarrow \text{at density} \\
&\qquad \delta = 140 \text{ vehicles/mile-lane } \Rightarrow n = b = 140LW.
\end{aligned}$$

See Jain and MacGregor Smith (1997) for further details. One should select different points to fit the curve when the free-flow speed V_1 changes.

Combining Equations (3) and (10) gives,

$$r_n = \frac{V_1}{L} \exp\left[-\left(\frac{n-1}{\beta}\right)^{\gamma} \right], \tag{13}$$

where V_1 is the free-flow speed. Therefore, one can express the overall service rate of the $M/G/c/c$ queueing model as

$$\mu_n = nr_n = n\frac{V_1}{L} \exp\left[-\left(\frac{n-1}{\beta}\right)^{\gamma} \right]. \tag{14}$$

Equations for the state probabilities are obtained by substituting the expression for μ_n, Equation (14), into Equations (6) and (7) to obtain Equations (15) and (16):

$$p_n = \frac{\lambda^n}{\Pi_{i=1}^n i\left(\frac{V_1}{L}\right) \exp\left[-\left(\frac{i-1}{\beta}\right)^{\gamma} \right]} p_0, \tag{15}$$

where

$$\frac{1}{p_0} = 1 + \sum_{n=1}^{c} \left\{ \frac{\lambda^n}{\Pi_{i=1}^n i\left(\frac{V_1}{L}\right) \exp\left[-\left(\frac{i-1}{\beta}\right)^{\gamma} \right]} \right\}. \tag{16}$$

Note that V_1 can be expressed in miles per hour (mph) or kilometers per hour (km/h), L is expressed in miles or meters, and λ is expressed in hour^{-1}.

3.6 Final Remarks

For both the linear and exponential congestion models the expected delay (*i.e.*, the expected travel time), $E(T)$, may be found by computing the average number in the queue, $E(N)$, then dividing it by the arrival rate, λ. Indeed, from Little's Law, one knows that under steady state conditions the average number of items in a queueing system equals the average rate at which items arrive multiplied by the average time that a item spends in the system, $E(N) = \lambda E(T)$ (about Little's Law, the reader may want to check Little and Graves, 2008, for a thorough discussion).

Since there exists a finite capacity queue, the *effective* arrival rate (*i.e.* excluding the items that are blocked) must be considered instead, which is defined in terms of the *external* arrival rate λ and the blocking probability

p_c, $\tilde{\lambda} = \lambda(1 - p_c)$, which gives that $E(N) = \tilde{\lambda}E(T)$. Finally, since $t \equiv E(T)$ then one achieves:

$$t = \frac{E(N)}{\tilde{\lambda}} = \tilde{\lambda}^{-1} \sum_{n=1}^{c} n p_n. \tag{17}$$

It would be most fortunate if one could achieve a closed form expression of the travel time delay formula, but this does not appear to be possible because t is a complex function of c in the probability calculations. One could fix c and derive such a function, but it would be tedious to deal with all these functions since there could be thousands of such functions.

4 Expected N and T Calculations

In this section of the paper, the travel time calculations of the three different models over a freeway segment are illustrated for varying lengths $L = 1, 2, 5, 10$ miles and variations in the traffic volumes. Let us introduce two alternative ways of computing travel time which we will contrast with the $M/G/c/c$ approach.

4.1 Alternative Ways of Computing Travel Time

4.1.1 BPR Formula

The BPR formula (Bureau of Public Roads, 1964) is probably the most well-known formula used in estimating travel delay. It was developed in 1964 using data from the Highway Capacity Manual (Transportation Research Board, 2000). The formula is:

$$t = t_f[1 + \alpha(x)^{\beta}],$$

where:

t := predicted travel time over the road segment;

t_f := travel time at the free-flow speed;

v := traffic volume (synonymous with traffic flow rate);

k := practical capacity (usually defined as 80% of actual capacity, in vehicles/h);

x := ratio of volume to capacity, *i.e.* $x = v/k$;

α := ratio of free flow speed to the speed at capacity (typically $\alpha = 0.15$);

β := indicates how abruptly the curve drops from the free-flow speed (typically $\beta = 4.0$).

4.1.2 Criticism of the BPR formula

First of all, we must recognize the importance of the BPR formula (Bureau of Public Roads, 1964), since it is probably the most widely used one. This classical formula has been considered by a legion of researchers (*e.g.* see Prashker and Bekhor, 2000; Bell and Cassir, 2002; Braess et al., 2005; Nagurney and Qiang, 2009; Buriol et al., 2010; Zheng and Liu, 2010, to cite a few)) and it is not but recently that comparisons questioning the supremacy of the BRP formula started to appear in the literature (see García-Ródenas et al., 2006; Ghatee and Hashemi, 2009; Cruz et al., 2010b).

Since the BPR formula is an empirically based model, there have been a number of criticisms of the values of the parameters set in the model. Dowling et al. (1998) recommend an updated version of the parameters. Thus, $\alpha = 0.05$, for signalized facilities, and 0.20, for all other facilities, and $\beta = 10.0$.

The resulting speed-flow curve is flatter than the original BPR curve for v/k ratios < 0.70 and the new curve drops more quickly in the vicinity of capacity $v/k = 1.00$. We will defer to Dowling et al. (1998) for their experiments and rationale for these adjustments.

4.1.3 Akçelick's Formula

Akçelik (1991) developed his travel delay formula based on the steady-state delay equation for a single channel queueing system. A time-dependent form of the delay equation was then arrived at using a coordinate transformation method:

$$t = t_f + \left\{ 0.25T \left[(x-1) + \sqrt{(x-1)^2 + 8\frac{J_A}{QT}x} \right] \right\},$$

where:

t := average travel time per unit distance (hours/km);

t_f := travel time at the free-flow speed (hours/km)

x := ratio of volume to capacity (*i.e.* v/k, also called the degree of saturation);

T := the flow period (hours) (typically one hour);

Q := capacity or flow rate (vehicles/hour);

J_A := a delay parameter.

The delay parameter J_A ensures that the delay equation will predict the desired speed of traffic when demand is equal to capacity. In fact, Akçelik (1991) has the following equation to calculate this quantity:

$$J_A = \frac{2Q}{T}(t_c - t_f)^2,$$

where t_c is the rate of travel at capacity. The suggested values for the delay parameter are presented in Table 1.

4.1.4 Remarks

Notice that neither the BPR (Bureau of Public Roads, 1964) or Akçelik's (Akçelik, 1991) model takes into account the finite length and width of the road segments. We feel that this is a distinct disadvantage of these prior

Table 1: Suggested values for J_A (Akçelik, 1991)

Facility Type	t_f (km/h)	c (vehicles/h)	J_A
Freeway	120	2000	0.1
Expressway	100	1800	0.2
Arterial	80	1200	0.4
Collector	60	900	0.8
Local Street	40	600	1.6

models. While Akçelik's model is based on a single-channel $M/G/1$ queue, the model that follows is actually a queueing model for the traffic road segment, not an analogous model.

In the experiments that follow, a one-lane freeway road segment was assumed with $V_1 = 62.5$ mph $\equiv 100$ km/h. The capacity $k = 2400$ vehicles/h was set from the Highway Capacity Manual (Transportation Research Board, 2000). The BPR formula used had the modified parameter values $\alpha = 0.20$, $\beta = 10$, and for Akçelik's model the delay parameter and time value were set $J_A = 0.10$ and $T = 1.0$ respectively. For the $M/G/c/c$ experimental results, a jam density of 200 vehicles/mile-lane was assumed and all other parameters for the linear and exponential models were used as defined previously.

Additionally, a discrete-event digital simulation model was run to confirm the accuracy of the solutions generated. The simulation model was developed by Cruz et al. (2005) and implemented as C++ program. The simulation model considers a road segment as an $M/G/c/c$ state dependent queue, with Poisson arrival rate λ, general service rate G and limited capacity c. The simulations took place on a PC, CPU Pentium II 400 MHz, 256 MB RAM, under Windows NT 4.0 operating system. The simulation times were set to 20 hours, with a *burn-in* period of 10 hours. In order to compute 95% confidence intervals, 30 replications were performed. Longer and shorter simulation time settings were tried but the results (not shown) were not significantly different.

4.2 Linear State Dependent Model

As seen in Figure 6, probably the most significant result is that the travel time function of the $M/G/c/c$ model is S-shaped, it is not convex (see Guideline #3). This property is not necessarily bad since it means that the function is quasi-concave, but it is not convex. It is important to see from the curve that the travel time t resulting from the $M/G/c/c$ state-dependent model is not monotone increasing, but actually $\frac{dt}{d\lambda} \to 0$, *i.e.* the derivative of t with respect to the arrival rate λ goes to zero for some threshold value of the arrival rate (or traffic demand) λ. This threshold value of traffic demand λ is thus a limiting value for the volume to capacity ratio, $x = v/c$. It does appear from all the curves that in the linear state dependent model, there is an upper limit near 2100 vehicles/h approximately when the vehicles will start to slow down which is consistent with the estimated capacity $c = 2400$ vehicles/h. Therefore, the derivative of the travel time formula of the $M/G/c/c$ model goes to zero around this point.

This leveling off of the travel time function makes sense because as the traffic volume approaches the jam density, it will slow down and monotonically increase in value, but not stop altogether, since unless impeded by an incident, the traffic will keep moving. Notice that average speed is meant here since even without an incident traffic has the stop-and-go behavior in very heavy conditions.

Table 3 (in the appendix) shows analytical and simulated performance measures for the linear model, namely the blocking probability p_c, the throughput θ, the expected number of cars on the road link $E(N)$, the expected service time $E(T)$, and the cpu time to run the simulation. The confidence intervals are certainly too narrow to be noticed in the figures but are instructive because they are showing the low variability of the estimates and the close agreement between the $M/G/c/c$ model and the simulation.

Figure 7 illustrates a comparison of the linear analytical and simulation values for the 1, 2, 5, and 10 mile experiments respectively starting from the top left hand figure.

4.3 Exponential State Dependent Model

As in the linear state dependent model, the travel time curve for the exponential state dependent curve is also S-shaped. This is illustrated in Figure 8. For the exponential model, which is perhaps more realistic for traffic segments, it does not rise so steeply and abruptly as is the case for the linear model. The linear model more abruptly reduces the speed as the usage gets closer to the capacity C, unlike the exponential model, which is smoother.

One important aspect of the exponential $M/G/c/c$ model is that it always achieves Guideline #2 of Spiess's guidelines in that it predicts that the travel time is at least twice the free-flow time at capacity. Also, the $M/G/c/c$ model is more pessimistic than the other two models, but this makes sense since the travel speed on the link is dynamically adjusted by the speed-density curve as the traffic density increases.

As one can see, when the length of the freeway section is short, Akçelik's model (Akçelik, 1991) is similar to the state dependent model but when the length of the freeway is up to 10 miles, Akçelik's model does not capture the congestion delay. The BPR model (Bureau of Public Roads, 1964) seems to agree pretty well with the state dependent model in that the BPR curves follow closely the state-dependent model, except that they are more optimistic in the lower volumes and go off to infinity at higher volumes.

Table 4 in the appendix shows all performance measures and their respective confidence intervals, the blocking probability p_c, the throughput θ, the expected number of cars on the road link $E(N)$, the expected service

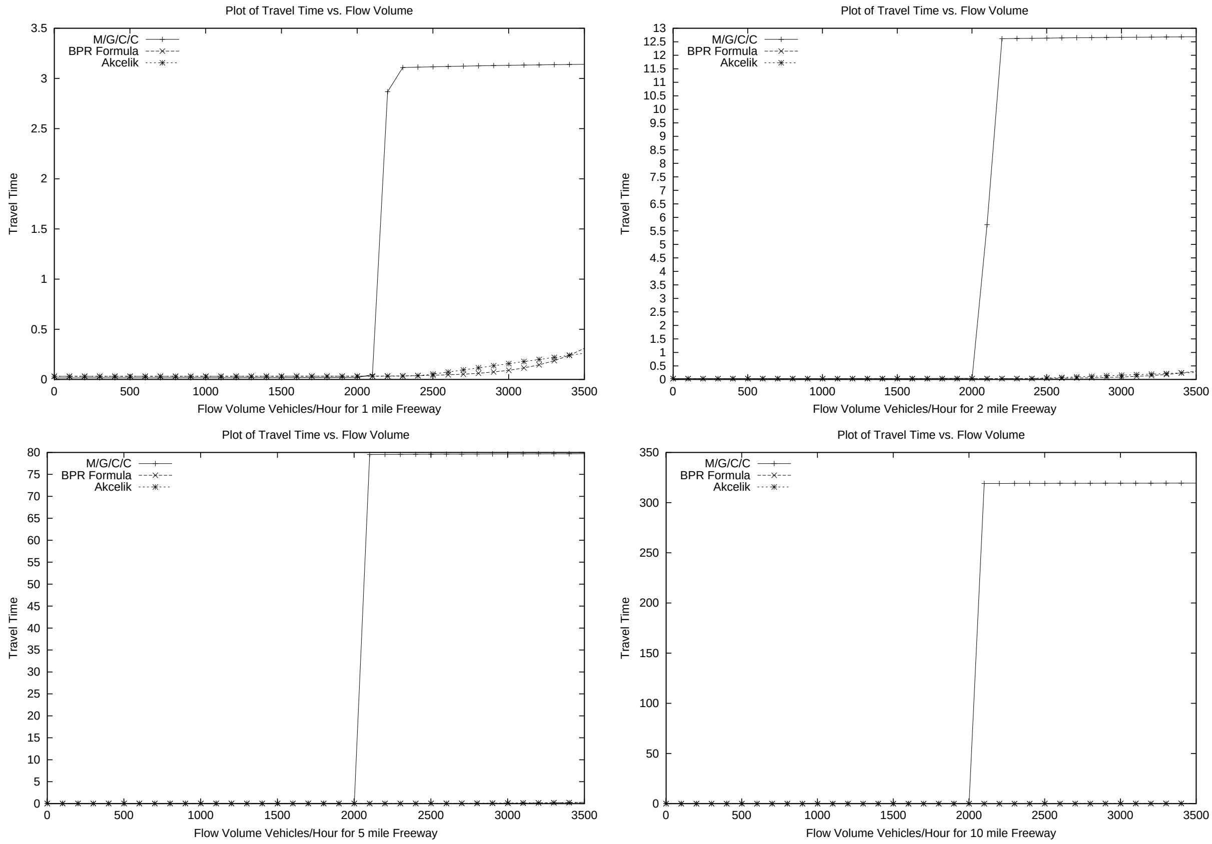


Figure 6: Travel time curves for the *linear* $M/G/c/c$ flow model, the BPR formula, and for the Akçelik's model, for single-lane freeways of 1, 2, 5, and 10 miles long, clockwise from the top left

time $E(T)$, and the cpu time to run the simulation.

Figure 9 illustrates a comparison of the exponential analytical and simulation values for the 1,2,5 and 10 mile experiments starting from the top left hand figure. As a result of this set of experiments, we have the first definition for calculating the travel time in an $M/G/c/c$ network model, which declares that:

Definition 1: *For a given highway segment with parameters L, W, and other parametric factors $(V_1, V_a, V_b, \ldots)$, the travel time function of the $M/G/c/c$ model is S-shaped*

$$T(\lambda) := \frac{T_{\max}}{1 + e^{-\kappa(\lambda-\lambda^*)}} \tag{18}$$

where:

$T_{\max}$:= maximum travel time;

λ := traffic flow volume;

λ^* := flow volume threshold at point-of-inflection;

κ := parameter integrating the $M/G/c/c$ model and the highway characteristics;

This expression for the travel time function is a solution to the differential equation $\frac{dT}{d\lambda}$. Notice that Equation (18) is just a logistic function (or logistic curve), which is a common sigmoid curve (Cavallini, 1993). The particular form of the S-shaped function is not unique. However, in queueing models, this form of the function is pretty useful. This definition follows from the nature of the $M/G/c/c$ model and the experimental results demonstrated in this paper. The expression argues that the travel time function $T(\lambda)$ is bounded and not infinite. No matter what state dependent curve is utilized, there would be maximum travel time threshold $T_{\max}$ for the given highway or roadway parameters. The above expression is suitable to deriving this expression from sample data. However, with the software for the $M/G/c/c$ model, the S-shaped curve can be directly generated.

Conjecture 1: *The travel time function between any two points on an un-signalized highway network for the $M/G/c/c$ model is an S-shaped function.*

While the proof of this is dependent upon isolating second derivative information of the $M/G/c/c$ function, this appears to be a very difficult problem since the $M/G/c/c$ travel time function is not in closed form. While the proof of this property of $M/G/c/c$ functions is to be determined, it is also expected that the general travel time function on a highway network is S-shaped, it does not go off to infinity as is expected in the BPR or related deterministic formulas. Some empirical evidence follows in the next section of the paper.

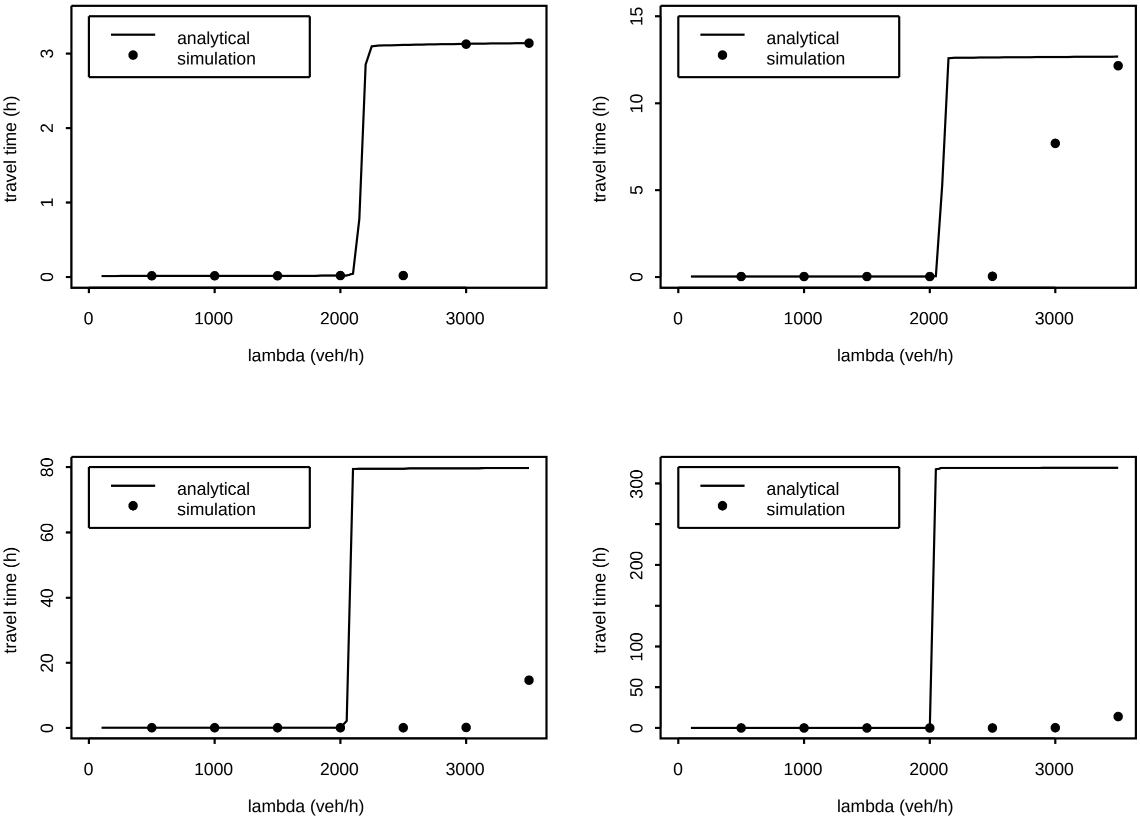


Figure 7: Travel time curves for the *linear* $M/G/c/c$ flow model and simulations for single-lane freeways of 1, 2, 5, and 10 miles long, clockwise from the top left

There have been many conjectures regarding the maximum flow rate $T_{\max}$ for a highway. See the Highway Capacity Manual (Transportation Research Board, 2000), exhibit 8-19 for a select set of maximum flow rate capacities for various metropolitan areas. What is provided in this paper is that now this maximum travel time $T_{\max}$ can be found for a particular arterial or highway segment with the understanding of the S-shaped travel time function and the $M/G/c/c$ state dependent model.

Finally, and most importantly, once one has the S-shaped curve, one can calculate the point of inflection for the curve and this will indicate the threshold traffic flow volume in which the transition stage from *Free Flow* to *Congested Flow* occurs. The point-of-inflection marks the switch from the convex to the concave part of the S-shaped function. This definition again follows from the basic properties of the S-shaped curve.

Definition 2: *The point of inflection of the S-shaped travel-time curve in the $M/G/c/c$ model is the value of the traffic volume at which the* Free Flow *traffic volume switches to* Congested Flow.

While it may be computationally difficult to find this point-of-inflection in the travel time function, the following conjecture indicates a key relationship between the point-of-inflection of the travel time curve and the throughput or flow rate function Q.

Conjecture 2: *The point-of-inflection of the travel time function $T(\lambda)$ corresponds to the maximum flow rate $\theta_{\max}$ of the throughput function.*

Although, no formal proof appears possible at the present time, the relationship holds experimentally as is shown in Figure 10. Figure 10 illustrates the correspondence between the throughput function and the travel time function for a freeway of 1 mile long and 1 car width wide. Roughly speaking at the flow volume of $\lambda \approx 2700$ vehicles/h, $\theta_{\max}$ equals the point-of-inflection of $T(\lambda)$. We have modeled this relationship between the throughput function and the travel time function for a series of V_1 values.

Thus, whether one ascribes to the two-phase or three-phase traffic flow theory, the point-of-inflection of the travel time function indicates the transition threshold for moving from Free Flow to Congested flow. From a practical point-of-view, if one can find the maximum throughput rate, then this corresponds to the point-of-inflection value in the travel time function. On the other hand, one can generate the S-shaped curve with the $M/G/c/c$ model, then identify the point-of-inflection from the curve itself.

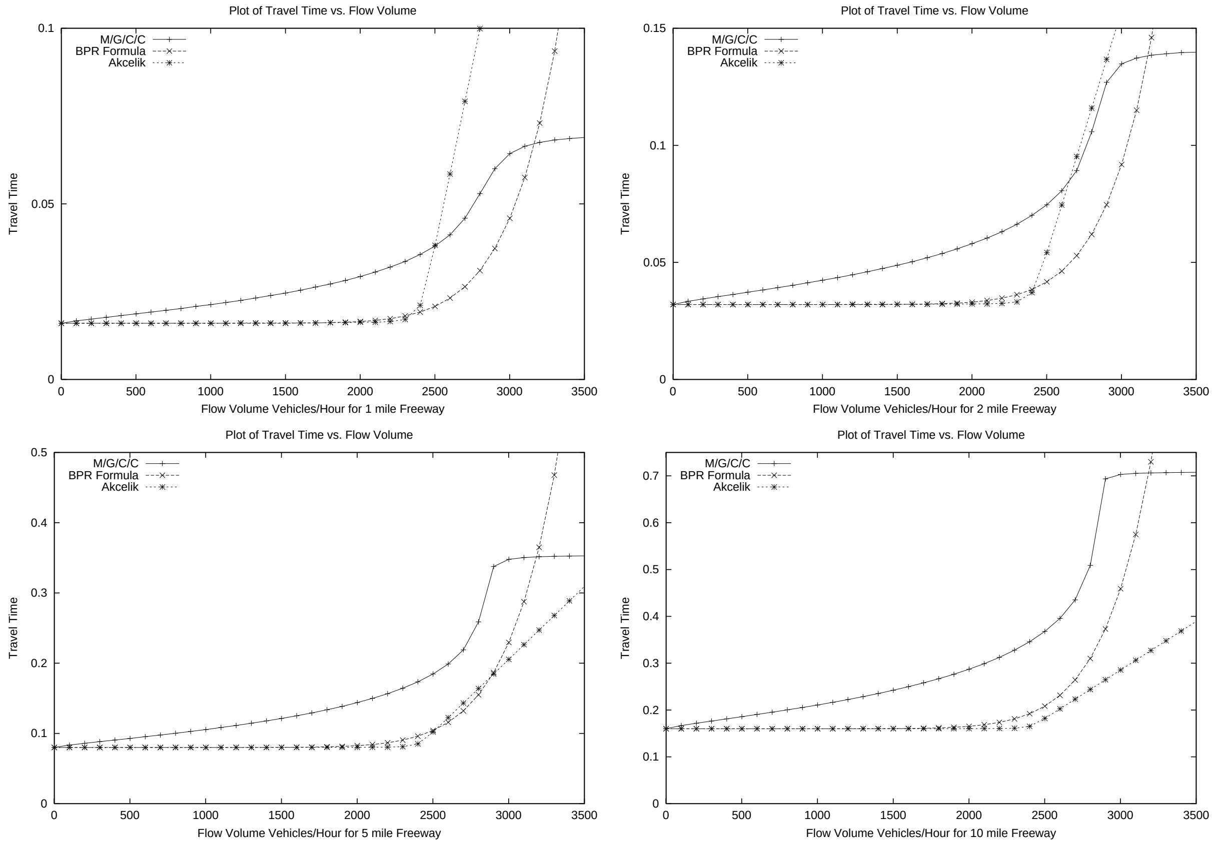


Figure 8: Travel time curves for the *exponential* $M/G/c/c$ flow model, the BPR formula, and for the Akçelik's model, for single-lane freeways of 1, 2, 5, and 10 miles long, clockwise from the top left

Also, from a practical perspective, knowing that the point-of-inflection of the travel time function corresponds to the average travel time T and to the maximum throughput $\theta_{\max}$ means that one can readily compute the travel time values along a path using Little's Law (Little and Graves, 2008) as:

$$T = \frac{L_s}{\theta_{\max}}, \tag{19}$$

$$T_{\max} = \frac{2L_s}{\theta_{\max}}, \tag{20}$$

where L_s is the average number of vehicles along the path. Equations (19) and (20) are very compact formulas that depend upon knowing L_s, but that is possible from the $M/G/c/c$ software tools or from fitting the logistics Equation (18) to sample traffic flow data.

Also, one might argue that the travel time function is not only for one segment of road traffic but a network of $M/G/c/c$ queues. However, the travel time function can be generated from an $M/G/c/c$ network model of traffic flows. This methodology for computing the $M/G/c/c$ network performance measures is discussed in several related papers (Mitchell and MacGregor Smith, 2001; Cruz et al., 2010b).

4.4 *Empirical Evidence of S-shaped Travel Time Curve*

As empirical evidence and verification of the S-shaped travel time model, Kerner (2004) did a study of vehicle travel times around an incident where a floating car vehicle was used to record travel times in and around the incident. The graph of the empirical travel times is indicated in Figure 11 which is modeled after the graph appearing in Kerner et al. (2005). Clearly, this represents an S-shaped curve where the three-phases are as earmarked. Figure 11 also includes the dissolution of the congestion and the return to normal free-flow travel time.

5 Examples

If we examine some incidents and bottlenecks (on-ramps) with the $M/G/c/c$ model, we can also compare the $M/G/c/c$ approach with the other travel time models when these problems occur. As we shall argue, this will further be corroborated with Kerner's empirical work.

More importantly, in utilizing the simulation model as opposed to the steady-state $M/G/c/c$ model, we can demonstrate the nonlinear oscillatory behavior of the bottleneck or incident on the traffic flow travel time dynamics.

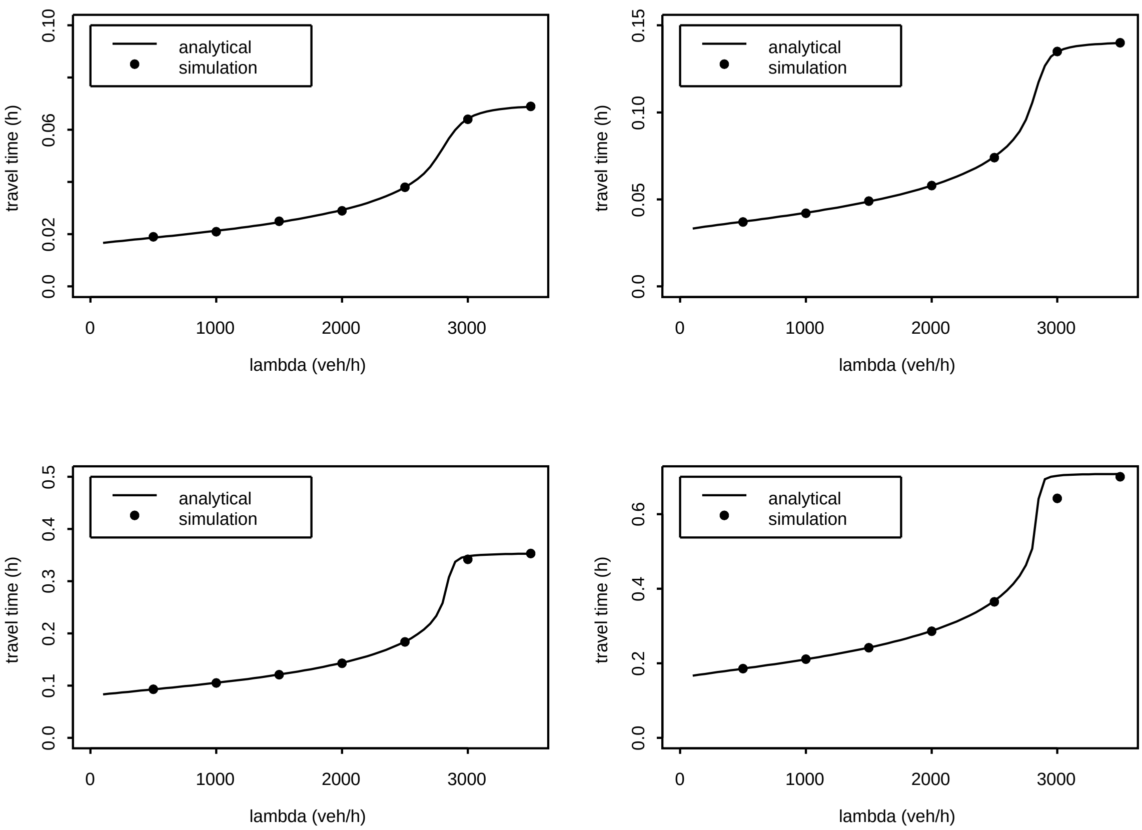


Figure 9: Travel time curves for the *exponential* $M/G/c/c$ flow model and simulations for single-lane freeways of 1, 2, 5, and 10 miles long, clockwise from the top left

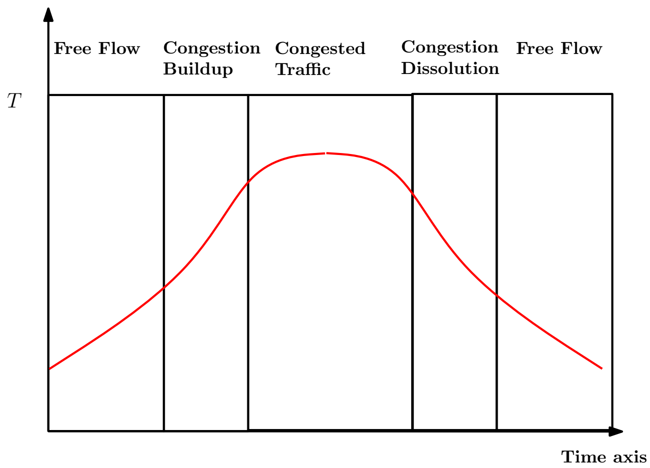


Figure 11: Kerner's empirical travel times around an incident in a road

5.1 Incidence Model

We are interested in utilizing $M/G/c/c$ models for modeling incidences vehicular traffic systems. A basic model for capturing the phenomenon of traffic incidence is given in Figure 12, where for a given traffic segment, the flow of vehicles is impeded by the reduction in capacity of the highway segment. The top half of Figure 12 represents a traffic segment which is unimpeded while the bottom half represents the reduction in capacity due to some incident. The incident will cause a merging of the two flows (part 1) at the point of incidence (part 2) where then blocking of the traffic flows will occur as a function of the volume of traffic arrival processes. Once the traffic emerges from the end of the incident (part 3), the traffic will proceed to even out into the two lanes. Five nodes are necessary to capture the traffic interruption due to the incidence, so the problem becomes quite complex. The situation sketched can be understood as an increase in local variability as the two flows need to merge and therefore will take more space per vehicle just before the incident; just after the incident the reverse phenomenon takes place.

We provide a sample of results of the incident. First, with the analytical models of the process (BPR and $M/G/c/c$), then another with a simulation model of this process. Let us assume that the traffic process is first a highway where traffic arrivals occur at λ for both lanes. The traffic segments are divided into three 1-mile-long segments, and when the middle segment loses capacity, the merge node is 1-lane wide. Otherwise, the nodes are 2-lane wide.

In Table 2, we present results for the basic network model for analytical and simulation approaches. It is

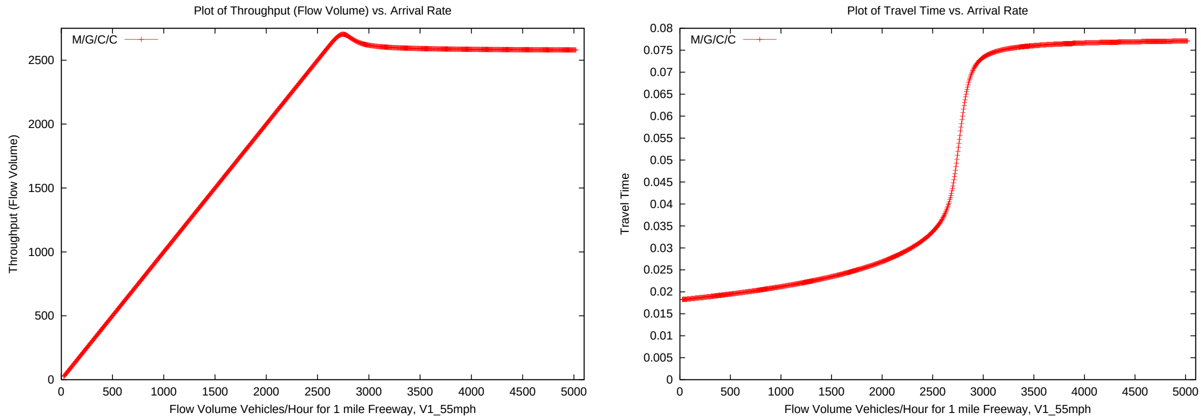


Figure 10: The maximum flow rate of the throughput function (left) and the point-of-inflection (right) of the travel time function for an $M/G/c/c$ model for a freeway stretch of one mile long and one lane wide for a maximum speed of $V_1 = 55$ mph

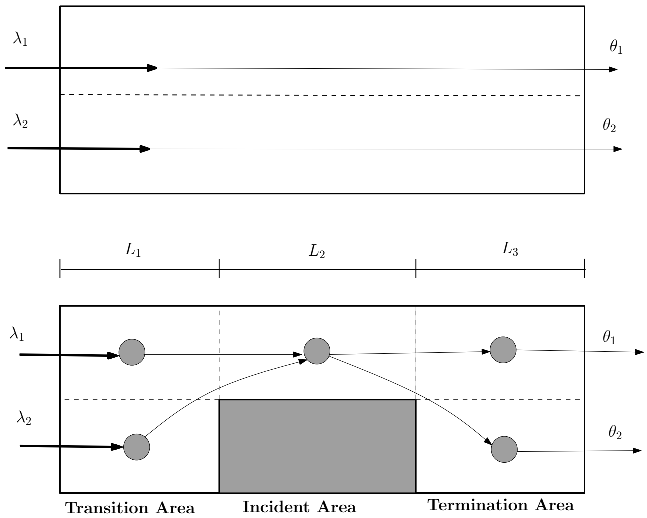


Figure 12: Basic incidence model for a given traffic segment where the flow is impeded by the reduction in capacity of the highway segment

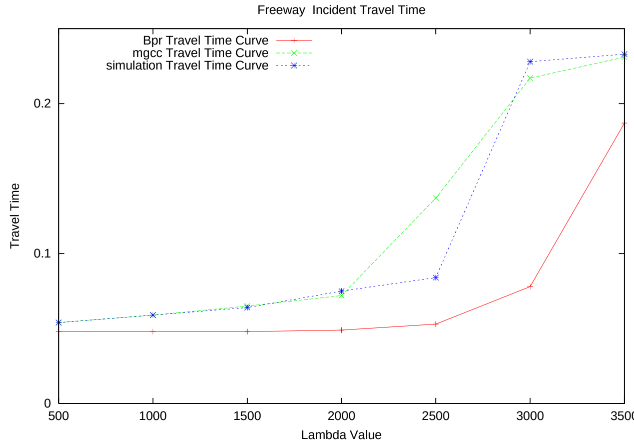


Figure 13: Travel times for the BPR function, the $M/G/c/c$ model, and the simulation model as functions of the arrival rate (λ) for the basic incidence model (Figure 12)

noticeable that under low arrival rates both analytical methods (BPR and $M/G/c/c$) seems to agree but the differences become large when the traffic is heavy. The $M/G/c/c$ model is also compared with simulations and the results are close and mostly within the standard error of means. Regarding the simulation model, we remark that the processing times are really very high as the traffic becomes heavy and, consequently, the number of vehicle entities in the simulation model explodes along with the CPU times.

Figure 13 illustrates the S-shaped curve of the incidence model for the $M/G/c/c$ travel time function along with the one from the BPR function and the simulation model. Notice that the BPR function way underestimates the congestion problems at the incident until the traffic is extremely heavy.

As a footnote to the above type of study, recently, many states in the U.S.A. have begun to require queueing analysis studies of highway work zones in order to estimate the travel delay and queueing length effects of these traffic flow interruptions. Not only for safety reasons, but for travel planning, construction, and overall management of the traffic incident, queueing studies are crucial.

5.2 Oscillatory Model

Finally, in a related measure of the $M/G/c/c$ model for modeling vehicular traffic flows, a small study of an on-ramp is proposed. What is intended to show here is that through the transient $M/G/c/c$ model, that the oscillatory behavior of traffic in the vicinity of an incident can be shown. This also corresponds to the oscillatory behavior of empirically based traffic flows that Kerner et al. (2005) have shown. Davis (2010) has also demonstrated the oscillatory behavior at a highway bottleneck using a different type of simulation model than the $M/G/c/c$ one described in this paper which also relies on the three-

Table 2: Analytical and simulated travel times for basic incidence model

	transition area t_1		incident area t_2		termination area t_3		t_{total}				
									simulation		
λ	BPR	$M/G/c/c$	BPR	$M/G/c/c$	BPR	$M/G/c/c$	BPR	$M/G/c/c$	mean	SE mean	cpu
500	0.016	0.017	0.016	0.019	0.016	0.017	0.048	0.054	0.054	0.0001	0h 0m 2s
1000	0.016	0.019	0.016	0.021	0.016	0.019	0.048	0.059	0.059	0.0001	0h 0m 7s
1500	0.016	0.020	0.016	0.025	0.016	0.020	0.048	0.065	0.064	0.0002	0h 0m 15s
2000	0.016	0.021	0.017	0.029	0.016	0.021	0.049	0.072	0.072	0.0003	0h 0m 27s
2500	0.016	0.076	0.021	0.038	0.016	0.023	0.053	0.137	0.084	0.001	0h 0m 49s
3000	0.016	0.129	0.046	0.064	0.016	0.024	0.078	0.217	0.228	0.001	1h 25m 48s
3500	0.016	0.138	0.155	0.069	0.016	0.024	0.187	0.231	0.233	0.001	1h 36m 35s

phase microscopic model of Kerner and Klenov (2002, 2003, 2009, 2010) made in the framework of Kerner's theory.

For the example, in Figure 14, an on ramp situation, actually, a three-node merge, with one lane wide and one mile long each is considered. The situation here is similar to the case presented in Figure 12, i.e., a locally increase in the variability as two flows need to merge.

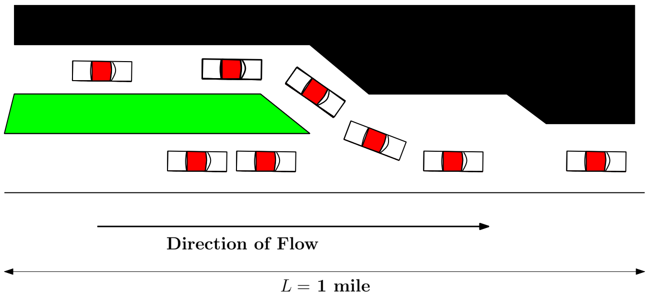


Figure 14: A three-node merge with one lane wide and one mile long representing an on ramp situation

In Figure 15, it is observed that when the arrival rate is moderate (500 and 1000 vehicles/hour) the number of vehicles exited is approximately stable and centered around the arrival rate. However, for heavy traffic, the number of exiting vehicles oscillates. Additionally, above 2000 vehicles/h, the number of vehicles exited does not increase any longer, indicating the system saturation (wide-moving jam). These theoretical results correspond to the empirical results of Kerner (Kerner, 2004, Figure 9.6, page 252), and the results of Davis (2010) in and around a bottleneck.

This type of oscillatory behavior for the exiting traffic occurs in other particle or state dependent models with $M/G/c/c$ queues and queueing networks, such as pedestrian flows in and around a bottleneck.

5.3 Open Questions and Future Research

There are a number of directions possible with this research. One is to recognize that the $M/G/c/c$ model is also directly applicable to modelling pedestrian networks (Yuhaski and Macgregor Smith, 1989), so many of the similar features of the travel delay function as shown in this paper apply to pedestrian dynamics. Some of the research already conducted for pedestrian networks has occurred in Mitchell and MacGregor Smith (2001) and MacGregor Smith (1991, 1994, 1996).

One can also use this model in traffic assignment applications and the authors are considering the design of an algorithm for such an application in future papers (Sheffi, 1985). The difficulty with this traffic assignment algorithm, however, is that the travel time function is non-convex, so care must be taken in designing the nonlinear programming methodology.

6 Summary and Conclusions

This paper has presented an overview of the problem of estimating travel time on road links. Three different models were examined for their ability to estimate the travel time over road links under various situations. It has been shown that the theoretical $M/G/c/c$ models support the empirically-based three-phase traffic flow model of Kerner (2004). The S-shaped curves of the $M/G/c/c$ state-dependent model are felt to be an important contribution to the understanding of travel-time behavior, since they provide a quantitative way to define the maximum flow rate of a highway segment.

References

Akçelik, R., 1991. Travel time function for transport planning purposes: Davidson's function, its time dependent form and an alternative travel time function. Australian Road Research 21 (3), 49–59.

Arnott, R., Small, K., 1994. The economics of traffic congestion. American Scientist 82, 446–455.

Bell, M. G. H., Cassir, C., 2002. Risk-averse user equilibrium traffic assignment: An application of game theory. Transportation Research Part B 36 (8), 671–681.

Braess, D., Nagurney, A., Wakolbinger, T., 2005. On a paradox of traffic planning [translation from the original in German, Braess (1968)]. Transportation Science 39, 446–450.

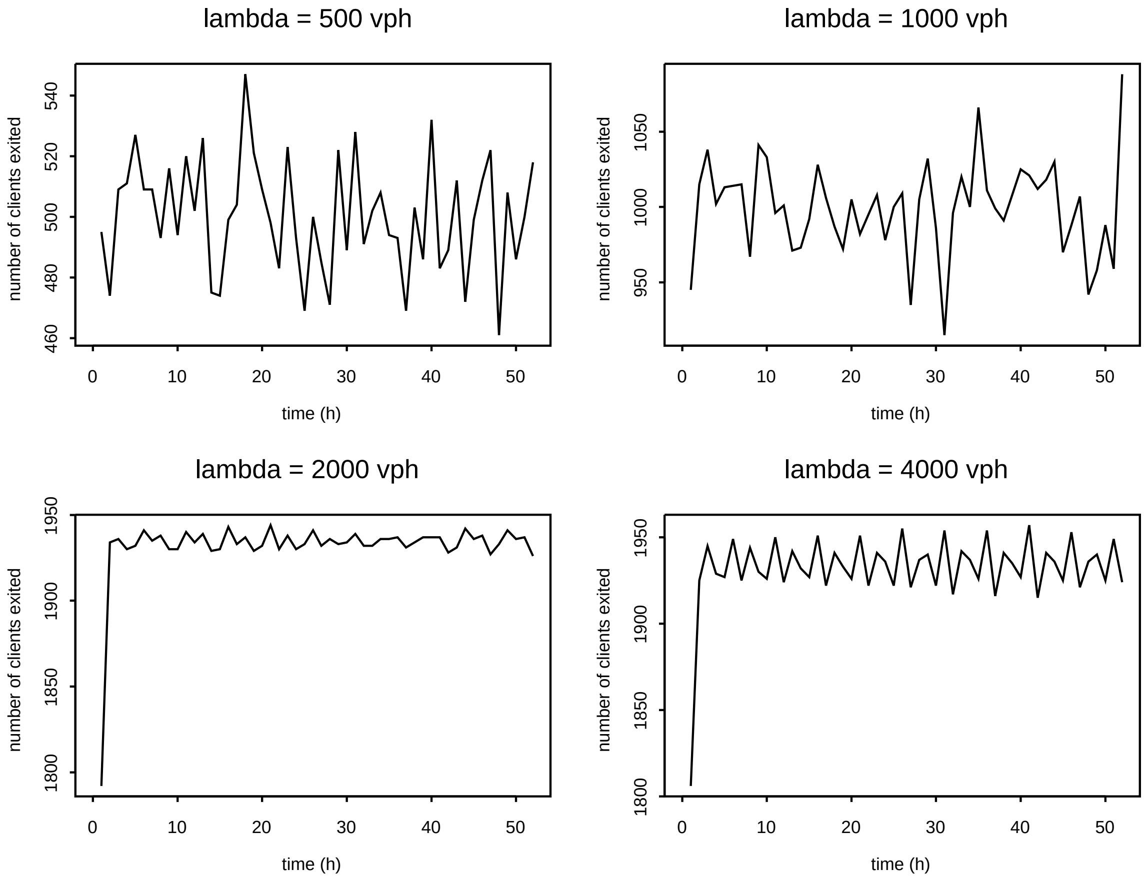


Figure 15: The oscillatory behavior of the number of vehicles exited (throughput) for several arrival rates ($\lambda = 500$, 1000, 2000, and 4000 vehicles/hour) for the on-ramp configuration (Figure 14)

Bureau of Public Roads, 1964. Traffic assignment manual. U.S. Department of Commerce.

Buriol, L. S., Hirsch, M. J., Pardalos, P. M., Querido, T., Resende, M. G. C., Ritt, M., 2010. A biased random-key genetic algorithm for road congestion minimization. Optimization Letters 4 (4), 619–633.

Cavallini, F., 1993. Fitting a logistic curve to data. The College Mathematics Journal 24 (3), 247–253.

Cheah, J.-Y., MacGregor Smith, J., 1994. Generalized $M/G/C/C$ state dependent queueing models and pedestrian traffic flows. Queueing Systems 15 (1), 365–386.

Cruz, F. R. B., MacGregor Smith, J., Medeiros, R. O., 2005. An $M/G/C/C$ state dependent network simulation model. Computers & Operations Research 32 (4), 919–941.

Cruz, F. R. B., Oliveira, P. C., Duczmal, L., 2010a. State-dependent stochastic mobility model in mobile communication networks. Simulation Modelling Practice and Theory 18 (3), 348–365.

Cruz, F. R. B., van Woensel, T., MacGregor Smith, J., Lieckens, K., 2010b. On the system optimum of traffic assignment in $M/G/c/c$ state-dependent queueing networks. European Journal of Operational Research 201 (1), 183–193.

Davidson, K. B., 1966. A flow travel-time relationship for use in transportation planning. In: Proceedings 3rd Australian Road Research Board Conference (Part 1). Melbourne, Australia, pp. 183–194.

Davis, L. C., 2010. Predicting travel time to limit congestion at a highway bottleneck. Physica A: Statistical Mechanics and its Applications 389 (17), 3588–3599.

Dowling, R. G., Singh, R., Cheng, W. W.-K., 1998. Accuracy and performance of improved speed-flow curves. Journal Transportation Research Record: Journal of the Transportation Research Board 1646, 9–17.

Drake, J. S., Schofer, J. L., May, A. D., 1967. A statistical analysis of speed density hypotheses. Highway Research Record 154, 53–87.

Edie, L. C., 1961. Car following and steady-state theory. Operations Research 9, 66–76.

García-Ródenas, R., López-García, M. L., Niño-Arbelaez, A., Verastegui-Rayo, D., 2006. A continuous whole-link travel time model with occupancy constraint. European Journal of Operational Research 175 (3), 1455–1471.

Ghatee, M., Hashemi, S. M., 2009. Traffic assignment model with fuzzy level of travel demand: An efficient algorithm based on quasi-Logit formulas. European Journal of Operational Research 194 (2), 432–451.

Greenshields, B. D., 1935. A study of traffic capacity. Highway Research Board Proceedings 14, 448–477.

Hausner, J., Walker, W., Swersey, A. J., 1974. An analysis of the deployment of fire-fighting resources in Yonkers, New York. Report Number R-1566/2-HUD/CY, RAND Corporation, Santa Monica, CA.

Horowitz, A. J., March 1991. Delay/volume relations for travel forecasting, based on the 1985 Highway Capacity Manual. Report FHWA-PD-92-015, FHWA, U.S. Department of Transportation.

Insurance Services Office, 2011. Response time considerations. URL `http://www.isomitigation.com`

Jain, R., MacGregor Smith, J., 1997. Modeling vehicular traffic flow using $M/G/C/C$ state dependent queueing models. Transportation Science 31 (4), 324–336.

Kerner, B. S., 2004. The Physics of Traffic. Springer, Berlin.

Kerner, B. S., 2009. Introduction to Modern Traffic Flow Theory and Control. Springer, Berlin.

Kerner, B. S., Demir, C., Herrtwich, R. G., Klenov, S. L., Rehborn, H., Aleksic, M., Haug, A., September 13-16 2005. Traffic state detection with floating car data in road networks. In: Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems. Vienna, Austria, pp. 700–705.

Kerner, B. S., Klenov, S. L., 2002. A microscopic model for phase transitions in traffic flow. Journal of Physics A: Mathematical and General 35 (3), L31–L43.

Kerner, B. S., Klenov, S. L., 2003. Microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks. Physical Review E 68, 036130.

Kerner, B. S., Klenov, S. L., 2009. Phase transitions in traffic flow on multilane roads. Physical Review E 80, 056101.

Kerner, B. S., Klenov, S. L., 2010. A theory of traffic congestion at moving bottlenecks. Journal of Physics A: Mathematical and Theoretical 43 (42), 425101.

Little, J. D. C., Graves, S. C., 2008. Building intuition: Insights from basic operations management models and principles. In: Chhajed, D., Lowe, T. J. (Eds.), Chapter 5 - Little's Law. Springer Science+Business Media, New York, NY, pp. 81–100.

MacGregor Smith, J., 1991. State dependent queueing models in emergency evacuation networks. Transportation Research Part B 25, 373–389.

MacGregor Smith, J., 1994. Application of state-dependent queues to pedestrian/vehicular network design. Operations Research 42, 414–427.

MacGregor Smith, J., 1996. Topological network design of state-dependent queueing networks. Networks 28, 55–68.

May, A. D., 1990. Traffic Flow Fundamentals. Prentice-Hall, Englewood Cliffs, NJ.

Mitchell, D. H., MacGregor Smith, J., 2001. Topological network design of pedestrian networks. Transportation Research Part B 35, 107–135.

Nagurney, A., Qiang, Q., 2009. A relative total cost index for the evaluation of transportation network robustness in the presence of degradable links and alternative travel behavior. International Transactions in Operational Research 16 (1), 49–67.

Prashker, J. N., Bekhor, S., 2000. Some observations on stochastic user equilibrium and system optimum of traffic assignment. Transportation Research Part B 34, 277–291.

Rose, G., Taylor, M. A. P., Tisato, P., 1989. Estimating travel time functions for urban roads: Options and issues. Transportation Planning and Technology 14 (1), 63–82.

Schónhof, M., Helbing, D., 2009. Criticism of three-phase traffic theory. Transportation Research Part B: Methodological 43 (7), 784–797.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall, Englewood Cliffs, NJ.

Spiess, H., 1990. Conical volume delay functions. Transportation Science 24 (2), 153–158.

Transportation Research Board, 2000. Highway capacity manual. Tech. rep., National Research Council, Washington, DC.

Tregenza, P. R., 1976. The Design of Interior Circulation. Van Nostrand Reinhold Company, New York, USA.

Treiber, M., Kesting, A., Helbing, D., 2010. Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts. Transportation Research Part B: Methodological 44 (8-9), 983–1000.

Underwood, R. T., 1961. Speed, volume, and density relationships: Quality and theory of traffic flow. Yale Bureau of Highway Traffic, 141–188.

Vandaele, N., van Woensel, T., Verbruggen, A., 2000. A queueing based traffic flow model. Transportation Research Part D: Transport and Environment 5 (2), 121–135.

Yuhaski, S. J., Macgregor Smith, J., 1989. Modeling circulation systems in buildings using state dependent queueing models. Queueing Systems 4 (4), 319–338.

Zheng, X., Liu, M., 2010. Forecasting model for pedestrian distribution under emergency evacuation. Reliability Engineering & System Safety 95 (11), 1186–1192.

A Tables

Table 3: Analytical and simulated performance measures *versus* arrival rate (linear)

		p_c			θ			$E(N)$			$E(T)$			
			simulation*			simulation*			simulation*			simulation*		
L	λ	anlt.	average	95% CI	anlt.	average	95% CI	anlt.	average	95% CI	anlt.	average	95% CI	*cpu (s)
1	500	0.000	0.000	[0.000;0.000]	500	498	[495 ; 500]	8.35	8.31	[8.27;8.35]	0.017	0.017	[0.017;0.017]	3.9
	1000	0.000	0.000	[0.000;0.000]	1000	1002	[998 ;1005]	17.5	17.6	[17.5;17.6]	0.018	0.018	[0.018;0.018]	12
	1500	0.000	0.000	[0.000;0.000]	1500	1496	[1492;1500]	27.9	27.8	[27.7;27.9]	0.019	0.019	[0.019;0.019]	24
	2000	0.000	0.000	[0.000;0.000]	2000	1996	[1991;2001]	40.1	40.0	[39.9;40.2]	0.020	0.020	[0.020;0.020]	42
	2500	0.974	0.000	[0.000;0.000]	64.2	2497	[2492;2503]	200	55.6	[55.4;55.8]	3.12	0.022	[0.022;0.022]	70
	3000	0.979	0.979	[0.978;0.979]	63.9	64.1	[63.5;64.8]	200	201	[199;203]	3.13	3.13	[3.13 ; 3.13]	50
	3500	0.982	0.982	[0.982;0.982]	63.7	62.5	[62.3;62.7]	200	196	[196;197]	3.14	3.14	[3.14 ; 3.14]	51
2	500	0.000	0.000	[0.000;0.000]	500	498	[495 ; 500]	16.7	16.6	[16.5;16.7]	0.033	0.033	[0.033;0.033]	5.6
	1000	0.000	0.000	[0.000;0.000]	1000	1002	[998 ;1005]	35.1	35.1	[35.0;35.3]	0.035	0.035	[0.035;0.035]	19
	1500	0.000	0.000	[0.000;0.000]	1500	1496	[1492;1500]	55.8	55.6	[55.4;55.8]	0.037	0.037	[0.037;0.037]	41
	2000	0.000	0.000	[0.000;0.000]	2000	1996	[1991;2001]	80.1	79.9	[79.6;80.2]	0.040	0.040	[0.040;0.040]	77
	2500	0.987	0.000	[0.000;0.000]	31.7	2497	[2492;2503]	400	111	[110;111]	12.6	0.044	[0.044;0.044]	142
	3000	0.989	0.884	[0.786;0.982]	31.6	347	[53.2; 640]	400	279	[239;319]	12.7	7.69	[6.03 ; 9.35]	158
	3500	0.991	0.991	[0.991;0.991]	31.5	32.3	[31.9;32.6]	400	392	[388;397]	12.7	12.2	[12.1 ; 12.2]	112
5	500	0.000	0.000	[0.000;0.000]	500	498	[495 ; 500]	41.7	41.6	[41.3;41.8]	0.083	0.083	[0.083;0.083]	10
	1000	0.000	0.000	[0.000;0.000]	1000	1002	[998 ;1005]	87.7	87.9	[87.5;88.2]	0.088	0.088	[0.088;0.088]	42
	1500	0.000	0.000	[0.000;0.000]	1500	1496	[1492;1501]	139	139	[139;140]	0.093	0.093	[0.093;0.093]	107
	2000	0.000	0.000	[0.000;0.000]	2000	1996	[1991;2001]	200	200	[199;200]	0.100	0.100	[0.100;0.100]	227
	2500	0.995	0.000	[0.000;0.000]	12.6	2497	[2492;2503]	1000	276	[275;277]	79.6	0.111	[0.111;0.111]	423
	3000	0.996	0.000	[0.000;0.000]	12.6	3000	[2995;3005]	1000	405	[402;407]	79.7	0.135	[0.134;0.135]	752
	3500	0.996	0.998	[0.997;0.998]	12.5	8.45	[8.11;8.78]	1000	124	[119;129]	79.7	14.7	[14.6 ; 14.8]	281
10	500	0.000	0.000	[0.000;0.000]	500	498	[495 ; 500]	83.5	83.1	[82.7;83.5]	0.167	0.167	[0.167;0.167]	20
	1000	0.000	0.000	[0.000;0.000]	1000	1002	[998 ;1005]	175	176	[175;176]	0.175	0.175	[0.175;0.175]	96
	1500	0.000	0.000	[0.000;0.000]	1500	1496	[1492;1500]	279	278	[277;279]	0.186	0.186	[0.186;0.186]	257
	2000	0.000	0.000	[0.000;0.000]	2000	1995	[1990;2001]	400	399	[398;400]	0.200	0.200	[0.200;0.200]	501
	2500	0.997	0.000	[0.000;0.000]	6.27	2498	[2493;2503]	2000	552	[551;554]	319	0.221	[0.221;0.221]	900
	3000	0.998	0.000	[0.000;0.000]	6.26	3000	[2995;3005]	2000	803	[799;807]	319	0.268	[0.267;0.269]	1519
	3500	0.998	0.999	[0.999;0.999]	6.26	3.78	[3.60;3.97]	2000	53.5	[51.0;56.1]	319	14.2	[14.0 ; 14.3]	605

Table 4: Analytical and simulated performance measures *versus* arrival rate (exponential)

		p_c			θ			$E(N)$			$E(T)$			
			simulation*			simulation*			simulation*			simulation*		
L	λ	anlt.	average	95% CI	anlt.	average	95% CI	anlt.	average	95% CI	anlt.	average	95% CI	*cpu (s)
1	500	0.000	0.000	[0.000;0.000]	500	498	[495 ; 500]	9.35	9.30	[9.24; 9.35]	0.019	0.019	[0.019;0.019]	6.9
	1000	0.000	0.000	[0.000;0.000]	1000	1002	[998 ;1005]	21.3	21.4	[21.3; 21.5]	0.021	0.021	[0.021;0.021]	19
	1500	0.000	0.000	[0.000;0.000]	1500	1496	[1492;1500]	36.9	36.7	[36.6; 36.9]	0.025	0.025	[0.025;0.025]	40
	2000	0.000	0.000	[0.000;0.000]	2000	1996	[1991;2001]	58.6	58.4	[58.1; 58.7]	0.029	0.029	[0.029;0.029]	73
	2500	0.000	0.000	[0.000;0.000]	2500	2497	[2492;2503]	95.0	94.7	[94.2; 95.3]	0.038	0.038	[0.038;0.038]	134
	3000	0.052	0.051	[0.050;0.053]	2843	2843	[2842;2844]	183	183	[182 ; 183]	0.064	0.064	[0.064;0.064]	313
	3500	0.188	0.187	[0.186;0.189]	2841	2841	[2841;2841]	196	196	[196 ; 196]	0.069	0.069	[0.069;0.069]	354
2	500	0.000	0.000	[0.000;0.000]	500	498	[495 ; 500]	18.6	18.5	[18.4; 18.6]	0.037	0.037	[0.037;0.037]	8.9
	1000	0.000	0.000	[0.000;0.000]	1000	1002	[998 ;1005]	42.4	42.4	[42.2; 42.6]	0.042	0.042	[0.042;0.042]	28
	1500	0.000	0.000	[0.000;0.000]	1500	1496	[1492;1500]	73.2	72.9	[72.6; 73.2]	0.049	0.049	[0.049;0.049]	64
	2000	0.000	0.000	[0.000;0.000]	2000	1996	[1991;2001]	116	115	[115 ; 116]	0.058	0.058	[0.058;0.058]	130
	2500	0.000	0.000	[0.000;0.000]	2500	2497	[2492;2503]	186	186	[185 ; 187]	0.075	0.074	[0.074;0.075]	280
	3000	0.055	0.054	[0.052;0.055]	2836	2836	[2836;2837]	382	382	[381 ; 383]	0.135	0.135	[0.134;0.135]	713
	3500	0.191	0.190	[0.189;0.192]	2830	2830	[2830;2830]	396	396	[396 ; 396]	0.140	0.140	[0.140;0.140]	754
5	500	0.000	0.000	[0.000;0.000]	500	498	[495 ; 500]	46.5	46.2	[46.0; 46.5]	0.093	0.093	[0.093;0.093]	15
	1000	0.000	0.000	[0.000;0.000]	1000	1002	[998 ;1006]	106	106	[105 ; 106]	0.106	0.106	[0.105;0.106]	59
	1500	0.000	0.000	[0.000;0.000]	1500	1496	[1492;1501]	182	181	[181 ; 182]	0.121	0.121	[0.121;0.121]	165
	2000	0.000	0.000	[0.000;0.000]	2000	1996	[1990;2001]	288	287	[285 ; 288]	0.144	0.144	[0.143;0.144]	374
	2500	0.000	0.000	[0.000;0.000]	2500	2498	[2492;2503]	461	460	[457 ; 463]	0.184	0.184	[0.184;0.185]	757
	3000	0.058	0.057	[0.055;0.059]	2826	2827	[2826;2827]	983	983	[982 ; 985]	0.348	0.348	[0.348;0.348]	1755
	3500	0.193	0.192	[0.191;0.194]	2823	2823	[2823;2824]	996	996	[996 ; 996]	0.353	0.353	[0.353;0.353]	1852
10	500	0.000	0.000	[0.000;0.000]	500	498	[495 ; 500]	92.8	92.4	[91.9; 92.9]	0.186	0.186	[0.185;0.186]	26
	1000	0.000	0.000	[0.000;0.000]	1000	1002	[998 ;1006]	211	211	[210 ; 212]	0.211	0.211	[0.211;0.211]	136
	1500	0.000	0.000	[0.000;0.000]	1500	1496	[1492;1500]	363	362	[361 ; 363]	0.242	0.242	[0.242;0.242]	360
	2000	0.000	0.000	[0.000;0.000]	2000	1996	[1991;2001]	574	572	[569 ; 574]	0.287	0.287	[0.286;0.287]	745
	2500	0.000	0.000	[0.000;0.000]	2500	2498	[2492;2503]	919	916	[911 ; 922]	0.368	0.367	[0.366;0.368]	1458
	3000	0.059	0.058	[0.057;0.060]	2822	2822	[2821;2823]	1984	1984	[1983;1985]	0.703	0.703	[0.703;0.704]	3691
	3500	0.194	0.193	[0.191;0.194]	2820	2821	[2821;2822]	1996	1996	[1996;1996]	0.708	0.708	[0.707;0.708]	3861