**Consistency and inconsistency between the fundamental relationships on which different traffic assignment models are based**

25 Nov 2016, revised 9 June 2017, Accepted 24 Sept 2017

Malachy Carey\*, Paul Humphreys, Marie McHugh, Ronan McIvor

Ulster University Business School, Ulster University, Northern Ireland, UK.

 [ m.carey@ulster.ac.uk, pk.humphreys@ulster.ac.uk, ml.mchugh@ulster.ac.uk, r.mcivor@ulster.ac.uk ]

\*Corresponding author. Tel.: +44 (0)7526 683 889.

**Abstract:**

We compare the forms and properties of different macroscopic behavioural relationships on which different static and dynamic traffic assignment models (STA and DTA) have been based, namely travel time-flow functions, travel time-occupancy functions and flow-occupancy functions or flow-density functions. Given any one of these three relationships we can derive the other two, so that they appear mathematically equivalent. These three behavioural relationships focus on different variables, but they are meant to be observing the same traffic and are not meant be inconsistent or incompatible with each other. However, we see that the forms and properties usually assumed for any one of these three relationships are often inconsistent or incompatible with the forms or properties that are usually assumed for the other two, or that are considered acceptable for the other two.

It might be thought that any inconsistencies between these relationships are due to modelling different phenomena, however that does not explain some of the surprising differences or inconsistencies, for example, those we find implied by linear or piecewise linear functions or by the best-known travel time-flow function, the BPR function. Such differences or inconsistencies are seldom mentioned in the literature though they are important because these models are so central in network traffic assignment. They are important when considering or comparing the solutions or predictions from the various traffic assignment models. But even if not comparing solutions, it seems important to know whether the model being used is inconsistent with another model that is widely accepted and widely used for similar scenarios. The results do not suggest discontinuing the use of any of the three forms of relationships or the corresponding assignment models referred to above. Rather, they suggest that more attention should be given to noting such potential inconsistencies, incompatibilities or limitations when considering, reporting or comparing results or predictions from traffic assignment models. In some cases the results indicate the direction in which the model biases the predictions compared to those from an alternative model which could have been used. In some cases, alternative properties or functional forms could or should be developed or used.

**Keywords**: travel time functions; flow-density; flow-occupancy; speed-density; dynamic traffic assignment; static traffic assignment; model consistency; congestion

**1. Introduction**

We consider and compare the macroscopic relationships on which the classic static traffic assignment (STA) model and various classes of dynamic traffic assignment (DTA) models have been based. More specifically, we consider and compare the following.

1. The travel time-flow functions which, together with conservation equations, are the basis of the traditional static traffic assignment (STA) models, which have been in use for sixty years since Beckman *et al*. (1956). These have been embedded into many commercial software packages which have been used in thousands of applications in many countries worldwide.
2. The travel time-occupancy functions which, together with conservation equations, are the basis of a widely discussed class of dynamic traffic assignment (DTA) models (e.g., see list of references in Carey *et al*. (2014)).
3. The flow-occupancy functions or flow-density functions which, together with conservation equations, are the basis of increasingly widely used classes of dynamic traffic assignment models based on the cell-transmission model (CTM) (Daganzo (2004, 2005a, 2005b)) and later transmission models, e.g. the Link Transmission Model (LTM), Yperman (2007), the General Link Transmission Model (GLTM), Gentile (2010), the Two-regime Transmission Model (TTM), Balijepalli *et al*. (2014), Ngoduy *et al*. (2016). These flow-occupancy functions are often referred to as the fundamental diagram of traffic flow (Haight (1963), Gartner *et al*. (2007), Transportation Research Board (2011)).

The functions referred to in (a)-(c) above can be written as 𝜏 = *f*(*q*), 𝜏 = *h*(*x*) and *q* = *g*(*x*) respectively, where where *q* is flow rate (in vehicles per unit time), is the time taken to traverse a road segment (or cell or link) and *x* is the occupancy of the segment. There are just three variables in the above three relationships (a)-(c) and these three variables are linked by a well-known identity (1.2) below. Each of the three relationships ((a)-(c)) contains exactly two of these three variables hence, given any one of the three relationships, we can derive the other two relationships by substituting from the identity (1.2), hence the three relationships are mathematically equivalent. There are various ways to show this, for example as follows. A well-known identity, which is often referred to as the fundamental identity or fundamental equation of traffic flow, relates three traffic flow characteristics *k*, *q* and *v*, thus

 *q* = *kv* (1.1)

where *q* is flow rate (vehicles per unit time), *k* is traffic density (vehicles per unit distance), and *v* is the space mean speed. Equation (1.1) should not be confused with the flow-density or flow-occupancy function or curve *q* = *g*(*x*) in (c) above, which is an empirical relationship often referred to as the fundamental relationship or fundamental diagram of traffic flow. For a road segment of length *L*, density *k* can be written as *k* = *x*/*L* where *x* is the occupancy of the segment, and space mean speed *v* can be written as *v = L*/ where is the time taken to traverse the segment. Substituting these in (1.1) gives *q = x*/ or

 *x* = *q*. (1.2)

Equation (1.2) also has a very simple natural interpretation when the three variables in it are constant over time, since in that case it states that the number of vehicles in a road segment (or cell or link) equals the number of vehicles entering it per unit time multiplied by the time that they take to traverse it. However, when we use the identity (1.2) to substitute in any one of the three relationships from (a), (b) and (c) above, to obtain another of these three relationships, that does not mean that we are assuming that the variables *x* or *q* or are held constant over time. We are using equation (1.2) in a comparative statics sense. That is, given a point on one of the curves (a), (b) or (c) we use (1.2) to find the corresponding point on another of the curves. That is how these curves are normally used. They show a relationship between points on the curves but do not include the dynamics of moving from one point to another on a curve.

Though the three bivariate relationships or functions ((a)-(c)) can be derived from each other, the form or properties that are usually assumed for any one of them in STA or DTA are not always consistent with the form or properties usually assumed for the other two. In comparisons, we find results that do not appear to be noted or discussed in the literature though they are significant and somewhat unexpected. For example, the most widely referenced and widely used time-flow function ((a) above) used in STA, namely the BPR function (Bureau of Public Roads (1964)), implies a travel time-occupancy function ((b) above) that eventually becomes strictly concave as density or occupancy increases. In contrast, in the literature this function is instead always assumed convex or linear and concavity does not appear to have been suggested previously. Concavity of the travel time-occupancy also implies that, as density or occupancy increases, the rate of increase of travel time decreases rather than increases, which again is the opposite of what is assumed in the literature. As another example, the flow-density or flow-occupancy function ((c) above) used in DTA (in the CTM, LTM, GLTM and TTM) has a downward bending part that is considered essential for describing congestion, queues and spillback. It is known that this downward-bending part implies that the time-flow function, on which STA models are based, has a backward-bending part, but in practice in STA this backward-bending part is omitted.

It may be thought that the differences between the three relationships ((a)-(c)) may be due to some fundamental differences in the phenomena being modelled. While that may be true in some cases, in many cases it does not explain the differences that are found (that is, it does not explain why, if we take one of the relationships and, from it, derive another of the relationships, the latter turns out to have a form that would be considered very unusual or unacceptable). In some cases, the form of the relationship seems to be chosen just to make the associated traffic assignment model easier to analyse or solve. For example, it seems that the travel time-occupancy functions ((b) above) are often assumed linear but only because this avoids potential first-in-first-out (FIFO) violations in the associated DTA model. On the other hand, when nonlinear travel time-occupancy functions are used they are assumed convex so that the associated DTA model will be well-behaved and solvable. As another example, the BPR travel time-flow function ((a) above) that is used in STA for networks implies a travel time-occupancy function ((b) above) that is initially convex and then becomes concave, though the latter form is never suggested or assumed in practice. As a further example, when travel time-flow functions ((a) above) are use in STA network models, the backward bending part of the curve is omitted to make the resulting STA models convex and hence analytically and computationally tractable.

More generally, when using any one of these relationships it is useful and important to know what its form or properties imply for the form or properties of the other two relationships. Knowledge of those implications could or should influence the choice of functional forms and properties or parameters. Even when the assumed properties of the different relationships are fully consistent with each other, it is still important to know that.

In traffic flow theory, the variables used are typically flow, speed and density at an instant in time and space. But, in traffic assignment models (STA and DTA) for networks, the variables are considered over a finite span of time or time step, and over a finite road length, a link or link segment or cell, and it is assumed that the above variables are constant over this time step and over the segment, link or cell. Thus, in STA, the link travel time-flow relationship is assumed constant over time. Similarly, in DTA, density is assumed constant over a finite distance, hence the variable used is occupancy (*x* = *kL*), where occupancy is density *k* times the length *L* of the segment. Thus, in this paper, as in the literature, we use a flow-occupancy function *q* = *g*(*x*) rather than a flow-density function though, when density is constant along a road segment, this is only a nominal change of variable. Similarly, a travel time-occupancy function ** = *h*(*x*) is used rather than its inverse, speed-density or speed-occupancy.

There is a substantial literature on each of the traffic assignment models that are based on the above three functional relationships ((a)-(c)) above, but in that literature these relationships are generally discussed separately. In particular, there is little discussion of how the form or properties of one of the relationships may limit, or be inconsistent with, the form or properties usually assumed for another of the relationships. Of the three relationships, the flow-occupancy function ((c) above) is now perhaps the most discussed and used in the traffic assignment literature. A recent review of flow-density functions is given in Carey and Bowers (2012). Some reviews of time-flow functions ((a) above) can be found in Branston (1976), Rose *et al*. (1989), Spiess (1990) and Akçelik (1991): time-flow functions have been in use for more than half a century so reviews of them tend to be quite old. Travel time-occupancy functions ((b) above) can be obtained by inverting the speed-occupancy functions (or speed-density functions) which are included in the Carey and Bowers (2012) review. They are also discussed separately in the many papers on DTA based on time-occupancy functions that are listed in in Carey *et al*. (2014), e.g. Xu *et al*. (1996, 1999) and Wu *et al*. (1995, 1998) used quadratic forms of these. Some papers do not directly discuss the above relationships but instead discuss an equivalent form or inverse of these, such as speed-flow (the inverse of time-flow) or speed-density (the inverse of time-density or time-occupancy).

Since in this paper we are considering three different functions or relationships concerned with some form of congestion, it is worth noting that the term “congestion” is used with different meanings in different branches of the literature. From its first mention in the economics literature (see Walters (1961), and discussion in Lindsey and Verhoef (2000), Small and Chu (2003)) the upward sloping part of the link travel time-flow ( = *f*(*q*)) curve has been referred to as congested. The same is true for the literature (in many hundreds of papers reports and applications) on static traffic and transport modelling and assignment (see for example Sheffi (1985)). The corresponding part of the flow-density function (*q* = *g*(*k*)) is also upward sloping, but in the traffic engineering literature it is referred to as *uncongested*. Also, in the economics literature any backward bending part of the time-flow function is usually referred to as hypercongested, while in static traffic modelling and assignment the backward bending part is generally not used at all, as it is considered transient or unstable and so of less interest for equilibrium modelling, and would cause analytical and computational problems in network modelling. In the traffic engineering literature only the downward sloping part of the flow-density function is referred to as congested, or sometimes as oversaturated. It corresponds to the backward bending (hypercongested) part of the time-flow function. To avoid possible confusion due to these inconsistent definitions of congestion we will avoid referring explicitly to congestion, except for a few times where it should not cause any ambiguity.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Derived | relationships. |  |
|  |  |  = *f*(*q*) | ** = *h*(*x*)  | *q* = g(*x*) |
| Given |  = *f*(*q*) | N.A. | Section 3. | Section 6.3 |
| relation- |  = *h*(*x*) | Section 6.1 | N.A. | Section 5.  |
| ships. | *q* = g(*x*) | Section 6.2 | Section 4. | N.A. |

**Table 1**. Relationships considered in this paper.

In Section 2 we set out some definitions as used in this paper. In Section 3 we consider the implications of travel time-flow functions = *f*(*q*) for travel time-occupancy functions = *h*(*x*). In Section 4 we consider the implications of flow-occupancy functions *q* = *g*(*x*) for travel time-occupancy functions = *h*(*x*). In Section 5 we consider the reverse of Section 4, namely the implications of time-occupancy functions = *h*(*x*) for flow-occupancy functions *q* = *g*(*x*). In Sections 3 to 5 we considered three pairs of relationship where we found significant differences or inconsistencies between the forms or properties usually assumed for one of these relationships and the corresponding forms properties usually assumed for another. However, as indicated Table 1, there are also three other pairs of relationships. In these three pairs, considered in Sections 6.1-6.3, the differences or similarities between the derived properties and the usually assumed properties are perhaps already better known, hence we set them out more briefly.

In summary: Section 3 considers implications of travel time-flow functions = *f*(*q*) for travel time-occupancy functions = *h*(*x*) and Section 6.1 considers the reverse of this. Section 4 considers the implications of flow-occupancy functions *q* = *g*(*x*) for travel time-occupancy functions = *h*(*x*) and Section 5 considers the reverse of this. Section 6.2 considers the implications of implications of flow-occupancy functions *q* = *g*(*x*) for time-flow functions = *f*(*q*) and Section 6.3 considers the reverse of this. Section 7 concludes.

Before proceeding, we note that in considering the functions 𝜏 = *f*(*q*), 𝜏 = *h*(*x*) and *q* = *g*(*x*), referred to above, we are considering these without any additional constraints that have been introduced in recent work in network modelling, such as capacity constraints on link outflows, to obtain vertical queues, or introducing node models that restrict link inflows or outflows, or enabling horizontal queues. Some of these features are discussed in Bliemer *et al*. (2014), which also goes further and, for the first time it seems, derives new static models or approximations directly from dynamic models, though the derived static models of course can not reflect the full range of behaviours of the dynamic models.

We also note that, throughout this paper, the travel time functions 𝜏 = *f*(*q*) and 𝜏 = *h*(*x*) and flow function *q* = *g*(*x*) are written without a time parameter *t*. When these functions have been used in dynamic traffic assignment models they are usually written as 𝜏(*t*) = *f*(*q*(*t*)), 𝜏(*t*) = *h*(*x*(*t*)) and *q*(*t*) = *g*(*x*(*t*)). That requires extending the definitions of the variables to indicate at what locations the quantities are observed or measured or defined. They are usually defined as follows.

𝜏(*t*) = *f*(*q*(*t*)) is usually defined as the segment travel time (the time taken to traverse a given road segment) for a vehicle/traffic that enters the segment at time *t* when the entry flow rate is *q*(*t*).

𝜏(*t*) = *h*(*x*(*t*)) is usually defined as the segment travel time for a vehicle/ traffic that *enters* the segment at time *t* if the amount/ volume of traffic in the segment (the segment occupancy) at time *t* is *x*(*t*).

*q*(*t*) = *g*(*x*(*t*)) is usually defined as the exit flow rate from a road segment at time *t* if the amount/ volume of traffic in the segment at time *t* is *x*(*t*).

All three of the above relationships are approximations when the variables in the relationships are varying over time. In reality, we would expect the left-hand side variables to also depend on the location or distribution of traffic along the road segment, but that is not included in these equations. For example, if the traffic on a road segment at time *t* is located mainly near to the segment exit then, until that traffic has exited, the exit flow rate at time *t* may in reality continue to be *higher* than indicated by *q*(*t*) = *g*(*x*(*t*)): the latter implicitly assumes that the traffic *x*(*t*) is uniformly distributed along the segment. Also, for the same scenario, for traffic entering the segment at time *t*, its travel time may be *lower* than indicated by 𝜏(*t*) = *h*(*x*(*t*)), because the higher density traffic near to the segment exit may have exited before the traffic entering at time *t* arrives there gets there, so that the traffic entering at time *t* may encounter only the remaining lighter traffic.

As already noted, in the rest of this paper we omit the “(*t*)” from all variables. The time variable or parameter *t* can be assumed to be implicitly present for all variables but no use is made of it in the paper, e.g., derivatives or difference are never taken with respect to *t*. The paper compares various traffic behavior functions or relationships and one can think of these comparisons as being for the static versions. That allows us to focus on the inherent differences, consistencies and inconsistencies that arise even in the static or comparative statics context. The approximations referred to above for the dynamic context are additional considerations and are already discussed in the literature (e.g. Carey and Ge (2004, 2005) and elsewhere).

**2. Preliminary definitions**

To avoid possible misunderstandings, we here define four terms as used in this paper, since they may be used with somewhat different meanings elsewhere.

*Occupancy*. Let the occupancy of a road segment denote the number of vehicles in the segment, that is, its length times the traffic density along the segment, where density is the number of vehicles per unit distance. It is often defined in this way, e.g. in the cell-transmission model (Daganzo (1994,1995a, 1995b)). Note that in the traffic engineering literature the term “occupancy” has also been used in other ways. It has sometimes been used to refer to the vehicle length plus the distance to the next vehicle in front and sometimes refers to the ratio of the sum of vehicle lengths to the roadway length. For loop detectors it is often defined as the percentage of time that a point or segment of a roadway is occupied by a vehicle.

*Convex about the origin*. This is referred to in Propositions 1(d), 7 and 8(e) and in Sections 6.1 and 6.2. A function is said to be strictly “convex about the origin” (e.g. see Bazaraa *et al*. (2006)) if, for all points on the curve, a straight line (a ray) from the origin to that point is everywhere below the curve up to that point and is everywhere above the curve after that point, which implies that (a) the ray cuts the curve only once and (b) if the curve is denoted by say *y = f*(*x*) then the slope of the ray is *f*(*x*)/*x* and is greater than the slope of the curve at the point where the ray cuts the curve, i.e., .

This form of convexity is a useful alternative to the usual convexity or concavity property for a curve, since it allows a function to have both convex and concave segments. Note that a curve can be convex but not convex about the origin and, conversely, can be convex about the origin but not convex. A function is said to be *weakly* convex about the origin if the above definition is weakened to allow a ray from the origin to coincide with a final (straight line) segment of the curve, in which case, for that part of the curve, .

*Quasiconvex*, *quasiconcave* and *quasilinear*. Quasiconvexity is used in Propositions 1(d), 3(iv) and quasiconvexity, quasiconcavity and quasilinearity in Proposition 4b(ii)(d). Consider any two points on a curve or function. The curve is said to be strictly quasiconvex if all points on the curve lie below the higher of the two end points, for every pair of end points. More formally, following Bazaraa *et al*. (2006) a strictly quasiconvex function can be defined as follows. Let *S* be a nonempty convex set. A function is said to be strictly quasiconvex if for each  with , we have < maximum for all , .

If the < is replaced by then is said to be quasiconvex rather than strictly quasiconvex.

The following properties are useful in this paper.

1. The negative of a quasiconvex function is said to be quasiconcave.
2. All convex functions are also quasiconvex but not all quasiconvex functions are convex.

Similarly, all concave functions are also quasiconcave but not all quasiconcave functions are concave.

1. If a curve is strictly increasing or strictly decreasing then it is (strictly) quasiconvex and (strictly) quasiconcave and is said to be (strictly) quasilinear.

Fig. 1(b) and the top figure in Fig. 2 are quasiconvex, quasiconcave and hence quasilinear, and are strictly so after any initial flat segment. Similarly, the top halves of several figures in this paper (e.g. Figs 1 and 3 to 8) are convex and hence also quasiconvex and are strictly quasiconvex after any initial flat segments. They are also strictly increasing and hence strictly quasiconcave and quasilinear, after any initial flat segment.

*Linear equations, functions or curves*. To avoid repeatedly distinguishing between affine functions and linear functions (or curves or equations) we will refer to both as linear, following the common practice in say linear regression, linear curve fitting, linear algebra, linear demand curves, etc. When we really mean a linear function, *y* = *bx*, rather than an affine *y* = *a* + *bx*, we will say so.

**3. Implications of commonly used time-flow functions = *f*(*q*) for time-occupancy functions = *h*(*x*)**

We make the same assumptions for the time-flow function  as are usually found in the literature on static traffic assignment. That is, it starts from and is everywhere positive, nondecreasing and convex, may converge to a vertical asymptote ( as or may be unbounded ( as .

The travel time-occupancy function has been widely used in dynamic traffic assignment and, in that literature, it has always been assumed convex. However, below we present two examples where the derived from the time-flow function is concave, or switches from convex to concave as *x* increases. The first of these examples is any piecewise linear time-flow function and the second example is the BPR travel time-flow function (Bureau of Public Roads (1964)) which is the longest standing, best known and most widely used and referenced travel time function. As illustrated in Fig. 1(b), the derived travel time-occupancy function has a property that does not appear to have been suggested or proposed for flow-occupancy functions for DTA models, namely it eventually becomes concave as link density or occupancy increases.

**Example 1: Time-occupancy function implied by a linear or piecewise linear time-flow function .**

Let the time-flow function be linear and increasing, thus  = *a* + *bq* with *a* and *b* positive. Substituting the identity *q = x*/, from (1.2), for *q* gives , and . The latter derivative is positive and, to show that the former is also positive, rewrite as and note for all hence . Hence occupancy *x* is an increasing strictly convex function of . Hence its inverse (the time-occupancy function ) is an increasing strictly *concave* function of *x* that starts from (*x*,) = (0,a) and as , and .

If is piecewise linear, rather than a single straight line as above, then the time-occupancy function becomes a series of consecutive increasing strictly *concave* functions of .

**Example 2: Time-occupancy function implied by a BPR time-flow function .**

The BPR travel time-flow function (Bureau of Public Roads (1964)) can be written as

 (2)

where *a*, *c* and are positive constants and *b* >1 so that the function is everywhere increasing and convex. The parameter is the free-flow travel time, to which (2) reduces as flow *q* goes to zero. The parameter *c* is often referred to as the flow capacity, but when *q* = *c* then (2) reduces to , so *c* can be defined as the flow that causes the travel time to increase by a fraction *a* above its free-flow value. The originally proposed, and still often used, values for *a* and *b* are 0.15 and 4 respectively. But a range of other larger and smaller values have also been estimated and used for *a*, *b* and *c*, depending on the type of roadway. Here we assume only that *a*, *c* and > 0 and *b* > 1.

Substituting the identity *q = x*/, from (1.2), into the BPR function (2) and rearranging gives

 (3)

which is the inverse of the corresponding time-occupancy function. This can not be rearranged or inverted to give as an explicit function of *x* but it can of course be solved numerically for for any given *x*. The properties of the time-occupancy function, the implicit inverse of (3), are set out in the following proposition and illustrated in Fig. 1.



**1(a).** BPR travel time-flow function .



**1(b)** Time-occupancy function implied by BPR time-flow function .

**Fig. 1.** Relationship between BPR time-flow function and implied time-occupancy function.

**Proposition 1.** The BPR travel time-flow function (2), as illustrated in Fig. 1(a), is upward sloping and strictly convex for all and is unbounded, i.e.,  as . It implies a travel time-occupancy function , the inverse of (3), as illustrated in Fig. 1(b), that has the following properties.

(a). when *x* = 0 and as . Also, is strictly convex for all 0 < *x* < and strictly concave for all *x* ≥ where is an inflection point.

(b1). when *x* > .

(b2). from above as from above and when *x* = .

(b3). from above as .

(c). At the inflection point in (a) above, and .

(d). It is quasiconvex and is also strictly ‘convex about the origin’ (as defined in Section 2 above).

**Proof.** From (2), and hence the BPR function (2) is upward sloping and strictly convex for all , since *a*, *c* and  are positive and *b* > 1. Also, in (2), as .

(a). The first sentence in (a) follows immediately from (3). To check for convexity or concavity of the time-flow curve we first consider the derivatives of its inverse (3), thus

 (4)

and

 (5)

Recall that *a*, *c* and are positive constants and *b* >1. Since , from (2), the expression before the square brackets in (5) is always positive hence the sign of is the sign of the expression in square brackets. Setting the latter to > 0 or < 0 and simplifying gives if and only if and if and only if , where . This implies is strictly concave when and strictly convex when if . It follows immediately that the reverse is true for the inverse function, that is, is strictly convex when and strictly concave when .

(b1). The BPR function (2) implies when hence and hence, from (4), when . This means that is a one-to-one mapping hence can be inverted to give (b1), i.e., when

(b2). From (3), as , from above. But from the first line of (4) we also have implies (to see this, recall that *b* > 1 hence the exponent (1/*b* – 1) in (4) is negative). It follows that implies . But from (b1), is a one-to-one mapping, hence we can invert to give . It follows that as , as in (b2). Also, when then which reduces (4) to 0, hence (b2) holds.

(b3). From (3), as , But from the first line of (4) we also have implies . But from (b1), is a one-to-one mapping, hence we can invert to give . Hence we have, as , as in (b3).

(c). This follows immediately on substituting , at the inflection point, from (a) into (3) and (1.2).

(d). From (b1), is everywhere increasing hence, from the definition of quasiconvexity in Section 2, it is quasiconvex.

To show that is strictly convex about the origin, it is sufficient to show that its gradient at any point is always less than the gradient of a straight line from the origin to that point. From (3) we have and substituting this into (4) we obtain

. From (2), for all *q* > 0 and hence for all *x* > 0. Hence, for all *x* > 0, the preceding expression in square brackets is >1 hence , which implies . That is, the slope of is everywhere less than the slope of a straight line from the origin, so is strictly convex about the origin. ◼

The properties of stated in the above proposition, in particular property (a), are illustrated in Fig. 1. The switch from convex to concave occurs at the inflection point at . To ensure that the BPR function (2) is convex rather than concave we must have *b* > 1, hence if we let *b* take any value from 1 to then implies . If, for example, *b* = 4 then the inflection point is at , that is, concavity starts at 60% above the free-flow travel time . More generally, if *b* = 1.5, 2, 3, 4 or 5 then the inflection point is at = 1.2, 1.33, 1.5, 1.6 or 1.66 respectively, that is, concavity starts at 20%, , 50%, 60% or  respectively above the free-flow travel time . Such increases above the free-flow travel time often occur in heavy or congested traffic hence the flows and travel times would often be on the concave portion of the travel time-occupancy curve.

Above we considered the properties of time-occupancy functions that are derived from two forms (linear and BPR) of time-flow functions. We now consider the properties of time-occupancy functions that are derived from more general forms of time-flow functions that have been proposed or used in the literature.

**Proposition 2.** Let be a positive convex function, starting from when . Let it be increasing for all , except for a possible initial constant segment .

Then for the corresponding , derived via (1.2):

(i). starts from when , with gradient 0 and thereafter is strictly increasing when is increasing and is constant if is constant, i.e. for any initial constant segment of .

(iia). can be convex or concave or switch between convex and concave, i.e. can be positive or negative or switch between positive and negative.

(iib). If any segment of is an upward sloping straight line then the corresponding segment of is strictly *concave* rather than convex.

**A note on notation**. In this paper, except in the proof below, all derivatives are with respect to *x*, hence we can abbreviate to  and to . However, in the proof of the present proposition we take derivatives of = *f*(*q*) with respect to *q* and derivatives of  = *h*(*x*) with respect to *x*, writing the first derivatives of these as and  respectively.

**Proof.** **(i).** To find the properties of , we first derive the first and second derivatives of with respect to *x*. Since we do not know the form or properties of , we proceed indirectly, using whose properties we already know (assumed in the proposition), to obtain the derivatives of with respect to *x*. Recall from (1.2) that . Substituting this in gives and differentiating with respect to *x* gives

 hence equivalently (6a)

 . (6b)

and bringing all to the left-hand side gives

 . (6c)

The right-hand side of (6c) is positive when is positive and is 0 when is 0, hence (i) follows.

**(iia).** To consider the convexity or concavity of consider the second derivative with respect to *x*, i.e. the first derivative of (6a) or (6b). Taking derivatives and simplifying eventually yields

and bringing all to the left-hand side gives

 (7)

The denominator in (7) is always positive hence the sign of is the sign of the numerator. To obtain the sign of the numerator, first consider the sign of . We do not know the sign of since it includes and, as we are seeking to infer the properties of , it follows that is as yet unknown. However, we can infer the sign of by a more circuitous route as follows.

From (6c) we know that is nonnegative since  is nonnegative and all the other terms are positive. Then in (6b), since and are nonnegative and all terms other than on the right-hand side of (6b) are nonnegative, it follows that must be nonnegative. More specifically, from (6c), when then and are positive and when is then .

Returning to the numerator in (7), we have seen that is positive (nonnegative). Also, is positive (nonnegative) since *f*(*q*) is convex, hence the first of the two terms in the numerator is positive (nonnegative). The second term (after the minus sign) in the numerator is also positive (nonnegative), since is positive (nonnegative). Hence the sign of the numerator depends on which of the two terms in the numerator is the larger. Hence the numerator in (7) can be positive or negative and, since the denominator in (7) is always positive, the quotient, and hence the left-hand side , can be positive or negative (as in Example 1 above) or switch from positive to negative as *x* increases (as in Example 2 above).

**(iib).** If a segment of is an upward sloping straight line and hence the numerator in (7) is negative and, since the denominator in (7) is always positive, it follows that the quotient, and hence will be negative, hence the corresponding segment of is strictly concave. ◼

**4. Implications of commonly used flow-occupancy functions *q* = *g*(x) for travel time-occupancy functions ** = *h*(*x*)**

In this section we consider implications of commonly used flow-occupancy functions *q* = *g*(x) for travel time-occupancy functions ** = *h*(*x*). In the next section, Section 5, we will consider the reverse of this. The *q* = *g*(x) flow-occupancy functions are usually assumed concave but, based on theory and empirical evidence, it is also often assumed that the right-hand tail is convex, with a negative gradient, so that option is included in Propositions 4 and 6 below. Travel time-occupancy functions ** = *h*(*x*) seem always to be assumed convex in theory and in numerical examples or applications. However, in Propositions 4 and 6 below we find that travel time-occupancy functions, derived from commonly used well-behaved concave flow-occupancy functions, can easily have both convex and concave segments, as illustrated in Fig. 2.

For convenience let *q* = *g*(x) be twice differentiable, though similar results can be obtained without this assumption. Assume as usual that *q* = *g*(x) is initially increasing and may eventually decrease. As noted in (1.2), when flow is constant over time the link occupancy is hence or since *q* = *g*(x). Then letting and respectively denote the first and second derivatives of *q* = *g*(x) with respect to *x*, and let and denote the first and second derivatives of with respect to *x* we have

 (8.1)

 . (8.2)

Hence ** = *h*(*x*) is convex if and only if the right-hand side of (8.2) is ≥ 0. Equation (8.1) is used in the proofs of Proposition 3(ii) and Proposition 4b (ii)(a), (iv) and (v) below and (8.2) is used in the proofs of Propositions 4b (ii)(b) and 6(ii)-(iv).

In the next proposition we show that the ** = *h*(*x*) corresponding to (derived from) *q = g*(*x*) is quasiconvex. Quasiconvex functions can have convex or concave segments but we do not consider that in the next proposition. Instead we consider it later in Propositions 4a and 4b, for ** = *h*(*x*) corresponding to piecewise linear *q = g*(*x*), and in Proposition 6 for ** = *h*(*x*) corresponding to more general nonlinear *q = g*(*x*).

**Proposition 3.**

[This proposition shows that the usual unimodal, continuously differentiable *q = g*(*x*) curve implies a strictly increasing and quasiconvex ** = *h*(*x*), as in for example Figs 7 and 9(iii).]

As is usual, assume flow-occupancy function *q* = *g*(*x*) starts from the origin, increases to a peak and declines thereafter. It is either concave throughout or may, at or beyond the peak, become convex. Also, assume that it is continuously differentiable. Then:

1. If *q = g*(*x*) has an initial linear segment (*q =* *bx*, *b* > 0) then the corresponding initial segment of ** = *h*(*x*) is , a constant, hence .
2. ** = *h*(*x*) is strictly increasing for all *x* > 0 (except for corresponding to a possible initial linear segment of *q = g*(*x*) as in (i)).
3. ** = *h*(*x*) is strictly quasiconvex.

**Proof**: The properties assumed for *q = g*(*x*) ensure that  for all *x* > 0, except for any initial linear segment of *q = g*(*x*). This can be seen by inspection of a flow-occupancy curve, noting that ** is the slope of a straight line from the origin to *q = g*(*x*) and ** is the slope of a tangent to *q = g*(*x*). For any initial linear segment of *q = g*(*x*) we have for.

[These conditions also ensure that *q* = *g*(*x*) is convex about the origin and is strictly convex after any initial linear segment.]

1. From (1.2), , hence when we have , hence
2. This follows from the properties of *q = g*(*x*), stated in the paragraph before (i) above. When then (8.1) implies , i.e. ** = *h*(*x*) is strictly increasing. For any initial linear segment , we have , from (i).
3. Any function that is increasing for all , except for a possible initial segment that is constant, satisfies the definition of quasiconvex for all . ◼

A strictly convex function is also strictly quasiconvex but the converse does not hold. That is, a quasiconvex function can be convex or strictly convex, or concave or strictly concave, or have both (strictly) convex and (strictly) concave segments. The above proposition does not indicate whether any or all of these outcomes are possible for ** = *h*(*x*) functions derived from standard *q* = *g*(x) functions. However, the following propositions (4a and 4b) show that all of these outcomes are in fact possible for such ** = *h*(*x*) functions when the corresponding *q* = *g*(x) functions are piecewise linear. Proposition 6 shows that they are also all possible when *q* = *g*(x) takes more general nonlinear forms.



**Fig. 2.** A convex piecewise linear *q* = *g*(*x*) implies a nondecreasing ** = *h*(*x*) with concave and convex segments.

The following Proposition 4a is needed in Proposition 4b. It shows that if ** = *h*(*x*) is derived from a piecewise linear *q* = *g*(*x*) and if, at a breakpoint of *q* = *g*(*x*), its left derivative is larger than its right derivative then the reverse is true at the corresponding breakpoint of ** = *h*(*x*).

Note that the left derivative of a function (e.g. *q* = *g*(*x*)) is larger than its right derivative if the function is concave (as in the lower half of Fig. 2) and is smaller if the function is convex.

Another way to summarise Proposition 4a is as follows. Proposition 4a(i) shows that, when ** = *h*(*x*) is derived from a *concave* piecewise linear *q* = *g*(*x*) then, at each breakpoint of ** = *h*(*x*), its left derivative will be *smaller* than its right derivative (as illustrated in Fig. 2). Proposition 4a(ii) shows that, if or when *q* = *g*(*x*) has a *convex* right-hand tail, then at each breakpoint of ** = *h*(*x*) corresponding to this tail of *q* = *g*(*x*), the left derivative of ** = *h*(*x*) will be *larger* than its right derivative.

**Proposition 4a.**

Let the flow-occupancy function *q* = *g*(*x*) be piecewise linear throughout, starting from the origin (0,0), rising to a flat peak and then declining again, as in the lower half of Fig. 2. Consider two possible cases, as follows.

Case (i): is piecewise concave throughout (as in Fig. 2 and in part (i) of this proposition),

Case (ii): somewhere at or beyond its peak, *q* = *g*(*x*) switches from piecewise concave to piecewise convex for its right-hand tail (as in part (ii) of this proposition).

Let the breakpoints of the piecewise linear be at , *i* = 1, … , *n*. Now consider the function implied by this . The breakpoints of will be at the same values of *x* as breakpoints of .

Then:

**(i).** In case (i), at each of the breakpoints of , the left derivative of will be *less* (less positive or more negative) than the right derivative.

**(ii).** In case (ii), at each of the breakpoints of , the left derivative of will be *larger* than the right derivative.

**Proof:** **(i).** The equations of the linear segments of to the left and right respectively of the *i*’th break-point can be written as and respectively where and are the slopes of these linear segments. Substituting the first of these two equations into from (1.2) gives the equation of the segment of to the left of *i*’th break-point , as . The first derivative of this is and at the breakpoint this reduces to . Similarly, the right derivative of at is . Hence < if < , which reduces to “if < ” hence “if > ”. But the assumption that is piecewise concave implies > , hence it follows immediately that < . That is, at each of the breakpoints of the left derivative of will be less than the right derivative.

**(ii).** The proof is the same as for part (i) above except, change each < to > and each > to < and also change the word concave to convex and convex to concave. (The inequality signs appear only in the second and third sentence from the end of the proof, and the words concave and convex appear only in the last two sentences of the proof.) ◼

**Proposition 4b.**

[This proposition considers the convexity/ concavity of a ** = *h*(*x*) implied by a piecewise linear *q* = *g*(x).]

Make the same assumptions about as in Proposition 4a above. Parts (i)-(v) of this proposition assume that *q* = *g*(*x*) is concave while part (vi) assumes that its right-hand tail may be convex.

Then, as illustrated in Fig. 2, ** = *h*(*x*) will consist of a series of segments with the following properties.

(i) The initial segment of ** = *h*(*x*) is constant (** = *b*) corresponding to the initial linear segment of *q* = *g*(*x*), namely *q* = (1/*b*)*x*.

After the initial linear segment of *q* = *g*(*x*) and ** = *h*(*x*), ** = *h*(*x*) has the following properties.

(ii). When *q* = *g*(*x*) is concave and *increasing* (, i.e. up to the peak of *q* = *g*(*x*), the corresponding ** = *h*(*x*) is (a) strictly increasing (), (b) is piecewise concave, i.e., it is strictly concave *between* its breakpoints, (c) is neither concave nor convex *at* its breakpoints, so it is neither concave nor convex overall, (d) is strictly quasiconcave, quasiconvex and hence quasilinear.

(iii). When *q* = *g*(*x*) is concave and *decreasing* (), i.e. beyond the peak of *q* = *g*(*x*), the results differ from those in (ii) only as follows. Part (a) continues to hold (but mention of the initial segment is no longer needed). In part (b), ** = *h*(*x*) is convex rather than concave between breakpoints. In part (c), ** = *h*(*x*) is strictly convex at breakpoints and hence is convex overall. Part (d) continues to hold (but mention of the initial segment is no longer needed).

(iv). When *q* = *g*(*x*) is constant (), i.e. at a flat peak of *q* = *g*(*x*), let *q* = 1/*b*. Then the corresponding ** = *h*(*x*) is a straight line with gradient *b*.

(v). If the final downward sloping linear segment of *q* = *g*(*x*) cuts the *x* axis at then and as , so that the final segment of converges to a vertical asymptote at .

(vi). If the downward sloping piecewise linear right-hand tail of *q* = *g*(*x*) becomes piecewise *convex* rather than piecewise concave (which is not illustrated in Fig. 2). Then, the results are the same as in (ii) (a), (c) and (d) and (iii)(b).

**Proof**: **(i):** Substituting *q* = (1/*b*)*x* into  from (1.2) gives ** = *b*.

**(ii).** (a). As was noted in the first paragraph of the proof of Proposition 3, for all *x* > 0, except for any initial linear segment linear segment *q* = *bx*. Substituting this in (8.1) gives

(b). Since *q* = *g*(*x*) is piecewise linear, except at the breakpoints hence, between breakpoints, (8.2) reduces to . Then, substituting (see proof of (ii)(a)) and (from assumptions), into this expression for gives between breakpoints, hence is strictly concave between its breakpoints.

(c). From (ii)(a), is an increasing function (). Also, from Proposition 4a(i), when is concave, then at each of the breakpoints of the left derivative of is smaller than the right derivative. But that is not compatible with concavity, hence is not concave at its breakpoints. It is also not convex at its breakpoints since it is concave and nonlinear between breakpoints.

(d). Since, from (a), is an increasing function (), after an initial flat segment, it is strictly quasiconcave, quasiconvex and hence quasilinear (see definitions in Section 2).

**(iii).** (a). The proof is the same as for (ii)(a) above, except that the mention of an initial linear segment linear segment *q* = *bx* is not needed here.

(b). Similar to the proof of (ii)(b) except that the assumption is replaced by assumption , which causes to change from to between breakpoints, hence become strictly convex rather than concave between its breakpoints.

(c). As in (ii)(a), at each of the breakpoints of the left derivative of is smaller (more negative) than the right derivative. That is consistent with convexity, hence is convex at its breakpoints and, from (iii)(b), it is also convex between breakpoints, hence is convex for all corresponding to the downward sloping concave part of .

**(iv).** When *q* = *g*(*x*) = 1/*b*, then hence (8.1) reduces to and (iv) follows immediately.

**(v)**. Let the final downward sloping linear segment of *q* = *g*(*x*) cut the *x* axis at . Then as , , and substituting these in , from (1.2), gives as . Also, for the final downward sloping linear segment of *q* = *g*(*x*), and substituting that and , , in (8.1), i.e. , yields as .

**(vi)**. See (ii) (a), (c) and (d) and (iii)(b). The main the difference between (ii) and (iii) on one hand and (vi) on the other hand is because in (ii) and (iii) the left derivative at each breakpoint of is less than the right derivative (from Proposition 4a(i)), while in (vi) the reverse holds, that is, the left derivative is larger than the right derivative (from Proposition 4a(ii). The latter is not compatible with convexity and, since is convex between breakpoints, it follows that is neither convex nor concave at breakpoints and hence neither convex nor concave overall (as in (ii)(c)). ◼

If the discretisation of *q* = *g*(*x*) into linear segments, in Proposition 4b, is refined to the continuous limit so that *q* = *g*(*x*) becomes a smooth nonlinear curve, then the results in Proposition 4b converge to those in Proposition 6 in which the curves are treated as continuously differentiable.

Trapezoidal flow-occupancy functions *q* = *g*(*x*) are a special case to the piecewise linear form illustrated in Fig. 2 and assumed in Propositions 4a and 4b. Since they have been widely used, in the cell-transmission model and elsewhere, we consider this special case here.

**Proposition 5.** [Convexity/ concavity of ** = *h*(*x*) when *q* = *g*(*x*) is trapezoidal.]

If the flow-occupancy function *q* = *g*(*x*) is trapezoidal, as in the bottom half of Fig. 3, then

(i) the travel time-occupancy function ** = *h*(*x*) is as in the top half of Fig. 3 and

(ii) it is convex for all *x* ≥ 0.

**Proof**: (i). This is a special case of Proposition 4a, in which only three segments from the bottom half of Fig. 2 remain, namely the first and last segments and the segment at the flat peak of *q* = *g*(x). Then ** = *h*(*x*) is reduced to the corresponding three segments in the top half of Fig. 2.

(ii). All three segments are separately convex, hence we need only show that the combined curve is convex at the two breakpoints. At the first breakpoint, a segment of ** = *h*(*x*) with slope 0 meets a segment with a positive slope hence the combined curve is convex.

For the second breakpoint of ** = *h*(*x*), the convexity follows as a special case of Proposition 4b(i). ◼



**Fig. 3.** Trapezoidal flow-occupancy *q* = *g*(*x*) implies a convex time-occupancy *τ* = *h*(*x*).

In Propositions 4a and 4b we considered the convexity or concavity of ** = *h*(*x*) when *q* = *g*(x) is piecewise linear. In the next proposition (Proposition 6) we consider the convexity or concavity of ** = *h*(*x*) corresponding to more general nonlinear forms of *q* = *g*(x).

**Proposition 6.** [Convexity / concavity of ** = *h*(*x*), assuming *q* = *g*(x) is continuously twice differentiable.]

Assume, as is usual, that the flow-occupancy function *q* = *g*(*x*) starts from the origin and is concave, as in Fig. 7 but, as in (iv) below, it may also have a convex right-hand tail. Also, for convenience, assume that it is continuously twice differentiable (so that it is convex if and concave if ).

1. If *q* = *g*(x) is initially a straight line through the origin (i.e. *q* = *bx*, *b* > 0). Then the corresponding initial segment of ** = *h*(*x*) is a constant namely ** = 1/*b*.
2. When *q* = *g*(x) is increasing and strictly concave (i.e. and ), then the corresponding = *h*(*x*) may be convex or concave. = *h*(*x*) has usually been assumed to be convex, but it will be concave ( if the magnitude of the second of the two terms on the right-hand side of (8.2) is less than that of the first term, otherwise it will be convex. Or, to state it differently, = *h*(*x*) will be concave if, in the second term on the right-hand side of (8.2), or does not have a “sufficiently” large negative value (i.e. if *q* = *g*(*x*) is not “sufficiently” concave) to make the magnitude of the second term larger than the magnitude of first term.
3. When *q* = *g*(*x*) is decreasing and concave (i.e. and ), or is non-increasing and strictly concave (i.e. and ), then ( so that ** = *h*(*x*) is strictly *convex*.
4. If *q* = *g*(x) becomes convex on its downward sloping right-hand tail (i.e. and ) then the corresponding = *h*(*x*) may be convex or concave, depending on the relative magnitudes of the two terms on the right-hand side of (8.2).

**Remark:** Part (i) above applies to any initial linear segment of *q* = *g*(x) and the remaining parts ((ii)-(iv)) refer to functions *q* = *g*(x) and ** = *h*(*x*) after any such initial linear segment. The proofs of (ii) to (iv) below make use of the first paragraph from the proof of Proposition 3. That paragraph shows that for all *x* > 0, except for any initial linear segment of *q = g*(*x*). In that case  in (8.2) is positive. That paragraph also notes that for any initial linear segment of *q = g*(*x*) we have . In that case in (8.2) is zero.

**Proof.**

1. Substituting from (1.2) into *q* = *bx* gives ** = 1/*b*.
2. From (8.2), . The second term (i.e. ) is positive because . The first term (including the minus sign) is negative, because is assumed and, as noted in the above Remark, . Thus in (8.2) is the sum of two terms, one negative and one positive, hence the sign of depends on the relative magnitudes of the two terms and (ii) follows.
3. These conditions are sufficient to ensure that each of the two terms on the right-hand side of (8.2) is nonnegative and that at least one of the two terms is positive, hence it follows from (8.2) that .
4. The proof is similar to that for part (ii). There we had and while here the signs ( > and < ) are reversed, thus and . In (ii), the first of the two terms on the r.h.s. of (8.2) was positive and the second was negative, while here the reverse is the case. As in (ii), the outcome (i.e. or ) again depends on the relative magnitudes of the two terms.  ◼

*The implied by the upward sloping part of q = g(x)*: *when is it convex and when concave?*

Based on empirical evidence it is generally assumed that the upward sloping part of *q* = *g*(*x*) is initially ‘mildly’ concave or is initially linear and then mildly concave. In the latter cases (mild concavity of *q* = *g*(*x*)) it might seem likely that *q* = *g*(*x*) would not be “sufficiently” concave to satisfy the condition in Proposition 6(ii), in which case the corresponding part of would be strictly concave. An example of this was provided in Proposition 4a and Fig. 2 by a typical concave piecewise linear *q* = *g*(*x*). We saw that, for each linear segment of *q*= *g*(*x*), the corresponding derived segment of the time-occupancy function was strictly concave, though in the literature it has always been treated as convex.

We now give an example to illustrate that this phenomenon ( concave or having concave segments) is not just due to *q*= *g*(*x*) having linear segments but can also occur if *q*= *g*(*x*) is nonlinear concave or has nonlinear concave segments, which are not “sufficiently” concave, as in Proposition 6(ii) above. A simple example is as follows. Let the flow-occupancy function *q*= *g*(*x*) be where and , which is everywhere upward sloping and strictly concave. The corresponding time-occupancy function , namely , is upward sloping and strictly *concave* rather than convex. This simple example is not entirely satisfactory since the time-occupancy function starts from the origin (0,0) though time-occupancy functions should start from (*x*,**) = (0,) where is the free-flow travel time for the road segment, link or cell. We can easily remedy this by replacing an initial segment of the flow-occupancy function with a linear segment , in which case the corresponding initial segment of will be .

In contrast to the above examples, we provide three examples (Examples 1, 2 and 3 in Section 5.3 below) in each of which **** is strictly *convex* and, in the corresponding *q* = *g*(*x*), the upward sloping part is concave. Stating this in reverse provides three examples where *q* = *g*(*x*) is “sufficiently” concave to satisfy the condition in Proposition 6(ii) and hence yield a convex time-occupancy function.

**5. Implications of commonly used travel time-occupancy functions for flow-occupancy functions .**

The scenario considered here is the reverse of that considered in Section 4. There we considered the implications of commonly used flow-occupancy functions *q* = *g*(x) for travel time-occupancy functions ** = *h*(*x*).

**5.1. Typical travel time-occupancy functions that yield flow-occupancy functions with no downward sloping (no ‘congested’ part).**

In this subsection we set out some typical or widely used travel time-occupancy functions and derive the corresponding flow-occupancy functions. In each case the flow-occupancy function has no downward sloping part and hence can not describe the range of behaviour described by a standard flow-occupancy function. This does not seem to have been previously discussed in the context of modelling DTA. Some further implications of the properties obtained in this subsection are set out in Appendix 1.

*Linear travel time-occupancy functions*



**Fig. 4.** Linear implies *q* = *g*(*x*) asymptotic to *q* = 1/*b*.

For reasons noted in Appendix 1, in DTA the most widely recommended form of the travel time-occupancy function is the linear form

  (9)

where *a* and *b* are positive. Using from (1.2) we obtain the corresponding flow-occupancy function

 (10)

as illustrated in Fig. 4. The first derivative of (10), i.e. is positive and the second derivative is negative since *a* and *b* are positive. Also, and as . Hence, as illustrated in Fig. 4, (10) is a positive, increasing, strictly concave function that starts from the origin and eventually becomes asymptotic from below to , so that it has no downward sloping part.

*Travel time-occupancy functions that converge to a straight line*

Instead of the purely straight-line travel time-occupancy function (9), to reflect initial free-flow conditions we may use one that initially curves upwards before becoming asymptotic to a straight line such as (9), as assumed in the following proposition.

**Proposition 7.** [Implications of for in Fig. 5.

For a travel time-occupancy function **,** illustrated in the top part of Fig. 5, assume that , and has a gradient that is initially zero or positive and thereafter positive, eventually becoming asymptotic from above to a straight line , , .

Then:

1. If is everywhere convex then is also convex about the origin.
2. If is convex about the origin, then the corresponding is upward sloping and eventually becomes asymptotic to *q* = 1/*b* from below, as in Fig. 5, where *b* is the gradient of the asymptote to .
3. If is convex (, then the corresponding is strictly concave.
4. Even if is not convex, but is convex about the origin, then may still be convex or strictly convex, depending on the relative values of , and .

**Proof.** (a). It is easily seen (by sketching a simple diagram), that a cord from the origin to  has a gradient and cuts the curve from below, and cuts it only once, so that which is the definition of convexity about the origin.

(b). Substituting from (1.2) into gives , hence . If then reduces to , which is positive. If then rewrite the expression for as hence the sign of is the sign of , since . But is assumed convex about the origin so , which again makes the sign of positive.

Also, recall from (1.2) that hence, as , and as . Hence as , that is, becomes asymptotic to from below.

(c). An expression for is given in (b) above and taking the derivative of that gives

 . Rearranging gives

 . (11)

Since is convex, , hence the second term in (\*), i.e. ), is negative. Also, from (a),  is convex about the origin, hence is positive. The other terms on the right-hand side of (11), i.e., , and are also all positive hence, from (\*), is negative, hence is strictly concave.

(d). Even if is not everywhere convex it can still be upward sloping and convex about the origin, so that . In that case, the right-hand side of (11) will still be negative if we exclude the final term ). Now consider the final term. Even if is at some points concave rather than convex, so that is negative rather than positive, which makes ) positive, the negativity of the rest of the right-hand side of (11) may be sufficient to outweigh this so that may still be still be negative. ◼



**Fig. 5.** asymptotic to implies asymptotic to , as in Proposition 7.

**Corollary.** Let  eventually become coincident with a straight line through the origin (i.e. , ) rather than just asymptotic to it. Then, as illustrated in Fig. 6, in the above Proposition 7 becomes coincident with , rather than just asymptotic to it and, subject to that change, the rest of Proposition 7 continues to hold. The converse also holds.

**Proof.** Make appropriate minor changes in Proposition 7 and its proof.  E.g., change the phrase “becomes asymptotic to”, which appears a few times, to “becomes coincident with”. ◼



**Fig. 6.** *q* = *g*(*x*) becomes horizontal if and only if  becomes a straight line  from the origin.

There are various ways to construct a function with the properties stated in the above Proposition 7.

E.g. 1. One such function is

 (12)

where *a*, *b*, *d* and *n* are positive, *n* need not be integer, and to ensure that the gradient is everywhere positive. It can be shown that this time-occupancy function starts from (0, ), is everywhere upward sloping and convex and becomes asymptotic to from above. The corresponding flow-occupancy function has the same properties as set out in Proposition 7.

E.g. 2. Another example of the above class of time-occupancy functions is

  (13)

where *a* and *b* are positive. This curve starts from (0, *a*) with initial gradient 0 and is everywhere increasing and strictly convex (i.e. for all , for all ) and as we have . Substituting (13) in from (1.2), the corresponding flow-occupancy function is

 (14)

It can be shown that (14) has the same properties as already set out in Proposition 7.

**5.2. More general implications of travel time-occupancy functions for flow-occupancy functions**

As noted earlier, the key relationships between the functions and can be obtained graphically or geometrically by drawing straight lines from the origin to as in Fig. 7.

Using , from (1.2) to substitute for *q* in gives , i.e. equals the (inverse of) the slope of a straight line from the origin to . This leads to simple relationships between and , stated in the following proposition and illustrated in Fig. 7.

**Proposition 8.** Let be a positive, increasing, continuously differentiable, convex function that starts from (x, = (0,, as in the top half of Fig. 7. Then starts from the origin (*x*, *q*) = (0,0) and:

(a). if and only if ,

(b). if and only if ,

(c). if and only if .

(d). Suppose that a tangent can be drawn from the origin to the curve . Since is also convex, only a single tangent can be drawn from the origin to the curve and it touches at a single point as in Fig. 7, or along a linear segment. Then as in the bottom half of Fig. 7, the derived is initially upward sloping, achieves a maximum and is thereafter downward sloping. The maximum of is where *b* is the gradient of the above tangent to .

(e). is ‘convex about the origin’ and is ‘strictly convex about the origin’ after a possible initial linear segment. (This still allows to be convex or concave, see Proposition 9 below.)



**Fig. 7.** Relationship between and , illustrating Propositions 8 and 9.

**Remarks.**

(i). is the slope of a straight line from the origin to the curve and is the slope of the curve at the same point, as illustrated in Fig. 7. Hence the three results (a)-(c) can be interpreted in simple geometric terms, which is useful in some proofs below. For example, (a) can be read as follows: if the slope of a straight line from the origin to the travel time-occupancy curve (i.e. ) is greater than the slope of the curve (i.e. ) at that point, then the corresponding flow-occupancy function is upward sloping at *x*, i.e. as in Fig. 7.

In (b), implies has reached its maximum and in (c), has become downward sloping.

(ii). Part (e) above refers to convexity about the origin, which was introduced in Section 2. In the present proposition, it is a usefully weaker property than convexity or concavity since it allows to be concave but to also have a convex right-hand tail, which has been observed to occur in practice.

**Proof.** (a)-(c). From (1.2), hence , hence hence the sign of is the sign of the right-hand side which is the sign of (). Then (a)-(c) follow immediately.

(d). This follows directly from (a)-(c) above.

[Draw a straight line from the origin to and consider the end point of the straight line, i.e., , moving along the curve starting from the vertical axis. The slope of the straight line is and the slope of the curve at the same point is . Up until the straight line becomes tangent to we have , when the straight line becomes a tangent to we have , and thereafter we have . Part (d) then follows immediately from parts (a)-(c).]

(e). We saw in (d) above, and in Remark (i) above, that the slope of at any point is always less than the slope of a straight line from the origin to that point on the curve, except for a possible initial linear segment of . This satisfies the definition of convex about the origin, given in Section 2 above. ◼

The above propositions show the implications of travel time-occupancy functions for the shape of the corresponding flow-occupancy function but, as noted in part (e) of the proposition, they do not consider whether this implies that the latter will be concave or convex. The next proposition addresses this.

**Proposition 9.** Let the travel time-occupancy function be a positive convex function that starts from and is everywhere increasing except for an initial segment which may be a horizontal straight line. Then, for the flow-occupancy function derived from the travel time-occupancy function :

1. is concave up to the maximum of (the peak flow).
2. Beyond the maximum point of , while is declining, its right-hand tail may switch from concave to convex.
3. The declining right-hand tail of is more likely to become convex the “more convex” is the right-hand tail of , i.e. the larger is in its right-hand tail.

**Proof**. (a). From (1.2), hence and we can show that Since is convex, is non-negative hence the third term in the square brackets is negative or zero. Hence ** is always negative if the sum of the first two terms is negative, i.e. if is negative or, equivalently, if . The later holds for all *x* up until *x* corresponding to the maximum of , i.e. up to , as was already noted in Proposition 8(a)-(c) and in Remark (i) just after that proposition. Hence is negative, and the flow-occupancy function is concave up to its maximum.

(b). If *x* is beyond the maximum of then, from Proposition 8(c) and Fig. 7, and the sum of the first two terms in the above expression for will be positive. But the third term is negative hence the overall sign of ** may be negative or positive. That is, beyond its peak, may remain concave or may become convex.

 (c). Also, since appears only in the final (negative) term in the above expression for , it follows that is more likely to become negative the larger is the component in that term. ◼

**5.3. Some travel time-occupancy functions that yield standard flow-occupancy functions , including a downward sloping part**

We now consider some nonlinear forms of the travel time function that yield a “standard” flow-occupancy function. That is, they correspond to a flow-occupancy function that starts from the origin, rises to a peak flow and then declines to or towards the horizontal axis. Some decline to a jam occupancy at zero flow (ex. 1 and 3 below and Table 2) while the others become asymptotic to the horizontal axis (Ex. 2 below) or to some flow level above the axis. In the latter cases, we can assume that in practice the function will be truncated at an appropriate jam occupancy.

Some of these functions are concave throughout (e.g. ex. 1 and 3) while the others are concave until somewhere beyond the peak flow rate and are then convex until jam occupancy (ex. 2 and 3) or until they become asymptotic to the horizontal axis or to some non-zero flow. In each of the examples we assume that parameters of the travel time function *h*(*x*) are chosen so that *h*(*x*) is positive and non-decreasing for .

|  |  |  |  |
| --- | --- | --- | --- |
|  | Ex. 1. Greenshields (1934) | Ex. 2. Polynomial. | Ex. 3. May & Keller (1968) and others. |
| Travel time-occupancy function, .  |  | ,  |   |
| Max (jam) occupancy. |  | None.  as  |  |
| Flow-occupancy function, . |  |  |  |
| Concavity of .  | Strictly concave. | Strictly concave until beyond the peak, then strictly convex. | If *n* ≤ 1, then concave.If *n* > 1, strictly concave until beyond the peak & then strictly convex.  |

**Table 2.** Travel time-occupancy functions that yield “standard” flow-occupancy functions (including a downward sloping part)

**Ex. 1**. Flow-occupancy function  derived from Greenshields (1934) functions.

The earliest published speed-density relationship is the linear from Greenshields (1934), which is still widely used particularly for illustrative purposes. Substituting the definitions and and rearranging gives which is defined over and is an increasing, strictly convex function that starts from and becomes asymptotic to the link capacity at

 Using (1.2), the corresponding flow-occupancy function is which is a strictly concave quadratic function starting from the origin, rising to a peak at and then declining to zero flow at jam occupancy at . ◼

**Ex. 2.** Flow-occupancy functions derived from quadratic or higher order polynomial time-occupancy function .

In numerical examples for DTA modelling, Xu *et al*. (1996, 1999) used a quadratic form and Wu *et al*. (1995, 1998) used a quadratic form These are special cases of a more general polynomial form , *n* ≥ 2, where we assume that all coefficients are positive. This is an everywhere increasing, strictly convex function and goes to as .

Substituting the above polynomial into from (1.2) gives the corresponding flow-occupancy function ]. It is easy to show that this function starts from the origin with slope , rises to a peak flow and then declines to become asymptotic to the horizontal axis. It is strictly concave until some point beyond the peak and then switches to strictly convex thereafter. ◼

**Ex. 3.** Flow-occupancy function derived from May and Keller (1968), Papageorgiou *et al*. (1989), May (1990).

May and Keller (1968) derived speed-density functions of the form from car-following models, where is the free-flow speed and is the jam density. In empirical work they found *n* > *m*, Papageorgiou *et al*. (1989) assumed *n >* *m* > 0 and May (1990), using an equivalent form, used *n* ≥ 1 and *m >* 0. This speed-density function is of particular interest since it reduces to the Greenshields (1934) function if *n* = 1 and *m* = 1, reduces to the Drew (1965) function if *n* = 1 and reduces to the Pipes (1967) function if *m* = 1.

 Using , and *k* = *x*/*L*, the corresponding time-occupancy function is . We can show that this function starts from with gradient 0 and is convex and increasing up to a vertical asymptote at the jam occupancy .

Using from (1.2) to substitute for in the above time-occupancy function we obtain the corresponding flow-occupancy function . This starts from the origin with gradient , rises to a peak and then declines to jam occupancy at . [If *n* < 1 it is concave throughout with a gradient of −∞ at jam occupancy, if *n* = 1 it has a gradient of at jam occupancy and if *n* > 1 it is strictly concave up to approaching jam occupancy and then becomes strictly convex with gradient 0 at jam occupancy.] ◼

**Ex. 4.** Flow-occupancy function derived from a piecewise linear travel time-occupancy function .



**Fig. 8.** Piecewise linear and corresponding .

An example of a piecewise linear and the corresponding is shown in Fig. 8. The derivation of the latter from the former is left to the reader. It is worth noting the following concerning concavity and convexity for Fig. 8.

(a) The fourth segment in is strictly convex rather than concave. If we replace the fourth linear segment in with two or more upwards sloping segments (with each segment steeper than the previous segment) then the corresponding segments in will be convex separately but the composite of the segments will be neither convex nor concave.

(b) The second segment in is strictly concave. If we replace the second linear segment in with two or more upwards sloping segments (with each segment steeper than the previous segment) then the corresponding segments in will be concave separately but the composite of the segments will be neither concave nor convex. ◼

**6. Other comparisons of the relationships used in traffic assignment models**

In Sections 3, 4 and 5 respectively we considered three of the pairs of relationships listed in Table 1 in the Introduction. In this section we consider the three remaining pairs of relationships listed in Table 1. These pairs are discussed more briefly than those in Sections 3, 4 and 5, because the differences between the results (the derived properties and the usually assumed properties) are perhaps better known. The pair of relationships considered in Section 6.1 are the reverse of the pair considered in Section 3. The pair of relationships considered in Section 6.2 are the reverse of the pair considered in Section 6.3.

**6.1. Implications of commonly used time-occupancy functions for time-flow functions**

We here assume the same properties for the time-occupancy function as are usually found in the literature. That is, we assume that it is everywhere positive, nondecreasing, likely to be convex and may converge to a vertical asymptote ( as from below) or may be unbounded ( as ).

Substituting , from (1.2), for *x* in the time occupancy function gives a time-flow relationship = *h*() in implicit form. Differentiating this with respect to and rearranging and interpreting we obtain the following three cases, which are illustrated in the two sub-figures in the top row in Figures 9(i), (ii) and (iii) respectively.



**Fig. 9(i).** Relationships between functions in Sections 6.1 to 6.3: case (i).



**Fig. 9(ii).** Relationships between functions in Sections 6.1 to 6.3: case (ii).



**Fig. 9(iii).** Relationships between functions in Sections 6.1 to 6.3: case (iii).

*Case* (i): = *h*(*x*) is convex and is strictly convex about the origin, e.g. as in Fig. 9(i). In that case the corresponding = *f*(*q*) is upward sloping and convex and converges to a vertical asymptote where is the slope of an asymptote to = *h*(*x*). Also, the derived = *f*(*q*) does not have any backward-bending part.

Case (i) holds, for example, for time-flow functions = *f*(*q*) derived from time-occupancy functions = *h*(*x*) that are linear or that converge to linear, as in Fig. 9(i).

*Case* (ii): = *h*(*x*) is convex and is weakly convex about the origin, e.g. as in Fig. 9(ii). In that case, the result is the same as in case (i) except that, instead of becoming asymptotic to , = *f*(*q*) increases up until and at that point = *f*(*q*) becomes a vertical line

It can be seen from Fig. 9(ii) that Case (ii) is a very special case, as it requires that the slope of the time-occupancy curve eventually coincides with a line through the origin.

*Case* (iii): = *h*(*x*) is convex and is not ‘convex about the origin’, e.g. as in Fig. 9(iii). In that case the corresponding = *f*(*q*) is upward sloping and convex up until and at that point = *f*(*q*) becomes backward-bending and concave.

Case (iii) holds, for example, for time-flow functions = *f*(*q*) derived from time-occupancy functions = *h*(*x*) that are quadratic, or higher polynomial functions, as in Fig. 9(iii).

In summary, if we start from time-occupancy functions = *h*(*x*) that have the range of shapes or properties that have been assumed for these functions, then the resulting or implied time-flow functions = *f*(*q*) inherit the full range of properties that have been assumed for these in the literature and in applications in STA. The implied time-flow functions do not have any additional unexpected or undesirable properties. Recall that was not true when we considered the reverse case in Section 3, that is, when we considered time-occupancy functions = *h*(*x*) derived from time-flow functions = *f*(*q*).

**6.2. Implications of commonly used flow-occupancy functions *q* = *g*(*x*) for time-flow functions = *f*(*q*)**

We assume the same properties for flow-occupancy function *q* = *g*(*x*) as are usually assumed in the literature. That is, it starts from the origin (0,0), rises to a peak and then declines. The curve is either (a) concave throughout, declining to a jam density, or (b) is concave to somewhere beyond the peak and may then become convex, but still downward sloping, until becoming asymptotic to the *x* axis or to a horizontal line above the axis, though in reality it must truncate at some finite jam density. Even if the right-hand tail of the curve becomes convex, the curve still satisfies the definition of ‘convex about the origin’ as given in Section 2.

It can be shown that the above flow-occupancy functions *q* = *g*(*x*) imply time-flow functions = *f*(*q*) as illustrated in Figs. 9(i)-(iii). The properties of the latter functions can be shown to be as follows.

1. Starts from (*q*, = (0,, where is the free-flow travel time and is the initial slope of the corresponding *q* = *g*(*x*) curve. It is an upward sloping convex curve until , where is the peak of *q* = *g*(*x*). It initially has a zero slope if the corresponding *q* = *g*(*x*) curve is initially a straight line through the origin. Up to the peak of *q* = *g*(*x*), as the slope of *q* = *g*(*x*) decreases the slope of = *f*(*q*) increases.
2. If *q* = *g*(*x*) achieves a peak at and then bends downward (as in Fig. 9(iii)) then = *f*(*q*) bends backwards at , and becomes concave. However, as and in the curve then, from (1.2), . That is, eventually becomes asymptotic to the vertical axis, which means that it becomes convex again.

It is well-known that in the basic STA model for networks, and in the many thousands of applications of it, the properties (i) above of are adhered but the backward bending part of in (ii) is not used. It is not used because it would yield a nonconvex bivariate = *f*(*q*), i.e. two values of for each value of *q*, which would cause serious analytical, computational and interpretational difficulties in STA models for networks.

Also in STA, instead of letting bend backwards at it is assumed that it becomes asymptotic to or to an asymptote larger than or becomes unbounded (i.e. and . This again was assumed for computational convenience in iterative solution algorithms, to ensure feasible solutions and to speed convergence.

**6.3. Implications of commonly used time-flow functions = *f*(*q*) for flow-occupancy functions *q* = *g*(*x*)**

To motivate the discussion, we first consider the flow-occupancy functions *q* = *g*(*x*) that are derived from the most widely used form of travel time-flow functions = *f*(*q*), namely the BPR time-flow functions. The properties of this *q* = *g*(*x*) are derived in Proposition 10 and illustrated in Fig. 10.

**Proposition 10.** (a). The flow-density function *q* = *g*(*k*) corresponding to the BPR travel time-flow function is the inverse of

 (15)

and is positive, increasing and strictly concave in *k*. Also,

(b). and as , so that the flow-density function *q* = *g*(*k*) has no upper bound even though its gradient goes to 0 as .

**Remark.** The above proposition can be restated in terms of occupancy *x* instead of density as , by simply replacing *k* with *x* and omitting the *L* throughout

**Proof.** Substituting ** = *x*/*q* = *kL*/*q* from (1.2) into the BPR function (2) gives (15). From (15),

 and ).

Since *a* > 0, *c* > 0 and and *b* > 1 the first and second derivatives are positive, hence (15) is increasing and strictly convex in *q*. This is the inverse of the flow-density function hence the latter is increasing and strictly concave in *k*.

(b). From above, as hence as . Also, from (15), as and (b) follows. ◼

It follows from the above proposition that the flow-density function has no congested part (no downward sloping part) hence no jam density and has no upper bound hence no flow capacity.



**Fig. 10**: Flow-occupancy function implied by the BPR time-flow function

Having considered the implications BPR time-flow function for flow-density or flow occupancy functions, we now consider the implications of other generally used time-flow functions = *f*(*q*) for flow-occupancy functions *q* = *g*(*x*).

The travel time-flow functions = *f*(*q*) used in STA models are always assumed to be convex and increasing or nondecreasing for all starting from where is the free-flow travel time. It is usually assumed that the travel time remains constant at its free-flow value for low values of *q* so that the slope of = *f*(*q*) is zero for low values of *q*, and is also assumed that the travel time may be unbounded ( as ) or go to a flow asymptote *q* = (i.e. as from below). For example, as seen in Proposition 1 for the widely used BPR function, as and there is no backward bending part. In reality, of course, flow can not go to infinity or exceed a certain value . We here consider time-flow functions = *f*(*q*) that have the above properties.

Substituting = *x*/*q*, from (1.2), for in the time-flow function = *f*(*q*) gives the occupancy-flow relationship *x* = *qf*(*q*), which is the inverse of the flow-occupancy function. Differentiating *x* = *qf*(*q*) gives

which is positive, since is assumed, hence the corresponding flow-occupancy function *q* = *g*(*x*) is either

(a) increasing for all *x* ≥ 0, if *q* is unbounded in the time-flow function = *f*(*q*), or

(b) is bounded above by , if = *f*(*q*) is bounded above by .

Also,

 (16)

But we have assumed that and (i.e., = *f*(*q*) nondecreasing and convex) hence, from (16), so is convex and its inverse *q* = *g*(*x*) is therefore concave.

In summary, the flow-occupancy curve *q* = *g*(*x*) implied by a convex increasing time-flow curve = *f*(*q*), starts from (0,0) and is concave. It is everywhere increasing and unbounded above if = *f*(*q*) is unbounded, it is everywhere increasing to an asymptote if = *f*(*q*) has an asymptote at (as in Fig. 9(i)), and is increasing to achieve a maximum at if = *f*(*q*) achieves a maximum at (as in Fig. 9(ii)). Thus, if = *f*(*q*) is bounded above, then both curves have the same bound *q* = . Also, if *q* = *g*(*x*) is derived from the above forms of = *f*(*q*), that is normally used in STA for networks, then it (*q* = *g*(*x*)) has no downward sloping part and hence has no jam occupancy.

**7. Summary and concluding remarks**

In this paper (Section 3) we derived the properties of travel time-occupancy functions that are implied or derived from travel time-flow functions. We showed (Proposition 1) that the most widely used travel time-flow functions (the BPR functions) imply time-occupancy functions that are initially convex but then become *concave* as density or occupancy increases, though concave time-occupancy functions do not appear to have been previously suggested in the literature. More generally (Proposition 2) we show that for standard well-behaved convex time-flow functions the derived time-occupancy functions can switch from convex to concave, and we derive conditions under which such switching can or cannot occur. For example, we see that even simple piecewise linear convex time-flow functions imply time-occupancy functions that consist of a series of consecutive *concave* segments.

In Section 4 we derived properties of travel time-occupancy functions that are implied by (derived from) standard flow-occupancy functions. We find that such travel time functions are strictly increasing, after a possible initial constant free flow travel time. They are also quasiconvex, which ensures that the minimum of the time-occupancy function is the free flow travel time and there are no other local minima. These quasiconvex functions can, and in many cases will, have both convex and concave segments, as shown in Propositions 4a and 4b and illustrated in Fig. 6. For example, a piecewise linear flow-occupancy function implies a travel time-occupancy function which is composed of a series of concave and convex segments, hence is neither concave nor convex. This is all somewhat surprising since, as far as we can ascertain, travel time functions with concave segments have not hitherto been suggested or proposed. However, since we have shown that such travel time functions are fully consistent with standard forms of flow-density functions, we suggest that they should be acceptable in practice. That is, we propose that a wider range of forms of travel time-occupancy curves be used in practice, to allow consistency and comparability with flow-density functions and hence with other forms of static and dynamic traffic assignment models.

In Section 5 we considered the reverse of Section 4, that is, we derived the properties of flow-density functions, or flow-occupancy functions, that are implied or derived from travel time-occupancy functions. Results include the following. In Section 5.1 we see that the most commonly used travel time-occupancy functions, namely linear, yield flow-density functions that do not have the downward sloping (‘congested’) part that is present in standard flow-density functions. Even if we relax the linearity assumption, so that the travel time-occupancy function is initially constant (free flow) or its slope is initially increasing, the same result holds. In Section 5.2 we see that, whether or not the derived flow-density function has a downward sloping part depends on whether, for the travel time-occupancy curve, it is possible to construct a tangent from the origin to the curve. In Section 5.3 we presented some classic travel time functions that imply standard flow-density functions, that is, flow-density functions that are initially upward sloping and concave and then downward sloping and concave and perhaps with a convex right-hand tail.

In Sections 3 to 5 we took the commonly used form and properties of one of the three relationships listed in Table 1 in the Introduction and considered what it implies for one of the other relationships listed in Table 1. In Section 6 we do the same for the remaining three pairs of relationships from Table 1. In Section 6.1 we consider the implications of commonly used time-occupancy functions = *h*(*x*) for time-flow functions = *f*(*q*). The range of forms or shapes that have been used for the former generate the full range of shapes that have been assumed for the latter. In Section 6.2 we consider the implications of commonly used flow-occupancy functions *q* = *g*(*x*) for time-flow functions = *f*(*q*). The former imply that the latter ( = *f*(*q*)) have a backward bending part, but when the latter ( = *f*(*q*)) functions are used in static traffic assignment models for networks, the backward bending part is omitted. In Section 6.3 we consider the reverse of the above, that is, implications of commonly used time-flow functions = *f*(*q*) for flow-occupancy functions *q* = *g*(*x*). The former implies that the latter will not have any downward sloping part so that, in the literature that uses the latter (flow-occupancy/density functions) these would not be considered as describing congested flows, and could not describe queues or spillback in DTA modelling.

**Appendix 1. Some implications of the form and properties of the time-occupancy relationship**

Consider travel time-occupancy functions that are linear (affine) or converge to linear as set out in Section 5.1. As noted there, these functions imply flow-density functions with no downward sloping (congested) part, which may make them unsuitable for many applications. This is significant since the most widely recommended travel time-occupancy functions are linear. The main reason for assuming linearity in travel time-occupancy functions in dynamic traffic assignment (DTA) models is that they are the only form of travel time function that has been shown to ensure a certain desirable first-in-first-out (FIFO) property (see Nie and Zhang (2005) and others). In DTA models that are based on using travel time-occupancy functions, satisfying FIFO depends on the shape of the travel time functions and on the profile of the link inflows (see next paragraph). To see this, note that if (a) there is a rapid fall in the inflow to a link at time *t* which causes a fall in the link occupancy *x*(*t*) and hence a fall in *h*(*x*(*t*)) and if (b) the gradient of *h*(*x*(*t*)) is sufficiently large then the fall in *x*(*t*) can cause such a large fall in the link travel time *h*(*x*()) that the next vehicle or cohort traverses the link faster than the preceding vehicle or cohort and hence exits before the latter, thus violating FIFO. Only for linear travel time functions can we be sure that such FIFO violations can not occur, regardless of the link inflow profile, as shown in Nie and Zhang (2005) and elsewhere. The problem is compounded by the fact that, when applying or solving a network model for DTA we in general can not know in advance what will be the profiles of inflows to each of the links in the network, since these inflows are determined endogeneously when solving the network assignment model. Hence, when using nonlinear travel time functions, we can not be sure in advance whether a FIFO violation will occur.

In view of the above, the decision to use, or not use, linear travel time-occupancy functions in a DTA model would appear to come down to a choice or trade-off between using a linear function, to avoid possible FIFO violations, and using a nonlinear function that would be less restrictive and more consistent with the usual properties of flow-density functions but could yield FIFO violations. Linear travel time functions also have the following further disadvantage. Nie and Zhang (2005) use the cell-transmission model to obtain a more accurate relationship between link occupancy and travel time and use that to show that linear travel time functions tend to overestimate link travel times due to a so-called double counting effect. Fortunately, there is a way to avoid the above trade off or dilemma concerning using linear or nonlinear travel time functions. Carey *et al*. (2014) has shown that there is another, quite simple, way to ensure FIFO in such DTA models, which will work even if the travel time-occupancy functions are nonlinear and ‘strongly’ convex. The method is, in outline, as follows. Vehicles entering a link at times *t* and *t*+1 respectively exit at times and respectively. If then, adding (*t*+1) to each side, gives , i.e., , hence FIFO is violated. To avoid or eliminate such a FIFO violation, simply increase the link traversal time of the follow-on vehicles to ensure that FIFO holds. That is, change their exit time from to . The exact amount by which to increase  is discussed in Carey *et al*. (2014). This is intended to better reflect how traffic behaves in practice: in practice: less dense follow-on traffic would like to travel faster than denser slower traffic ahead but is held back by the slower traffic ahead. In the usual formulations, is set equal to even if this means violating FIFO, which may not be physically possible for road traffic. The above method, from Carey *et al*. (2014), ensures that the link travel time takes account of slower traffic ahead so that . The procedure outlined in the above paragraph is applicable regardless of the linearity or nonlinearity, convexity or nonconvexity, of the link travel time functions. Hence we no longer have the trade-off dilemma outlined in the preceding paragraph. That is, we can use these nonlinear travel time-occupancy functions in DTA without running the risk of FIFO violations.

*Acknowledgements*. The authors wish to thank the three reviewers of this paper for their very thoughtful, helpful comments, corrections and suggestions. Only the authors (and the gremlins) are responsible for any errors that may remain.

**REFERENCES**

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

Balijepalli, NC, Ngoduy, D., and Watling, DP, (2014). The Two-regime Transmission Model for Network Loading in Dynamic Traffic Assignment. *Transportmetrica A* 10(7), 563-584.

Bazaraa, M.S., H.D. Sherali and C.M. Shetty (2006). Nonlinear Programming: Theory and Algorithms. Third Edition. Wiley-Blackwell.

Beckman,M.J., McGuire,C.B, and C.B.Winston (1956). Studies in the Economics of Transportation. Yale University Press, Connecticut.

Bliemer, M.C.J. and Raadsen, M.P.H. and Brederode, L.J.N. and Bell, M.G.H. and Wismans, L.J.J. (2014) *A unified framework for traffic assignment: deriving static and quasi dynamic models consistent with general first order dynamic traffic assignment models.* In: 5th International Symposium on Dynamic Traffic Assignment, 17-06-2014 - 19-06-2014, Salerno, Italy (pp. 1 - 38). <http://doc.utwente.nl/99053/> .

Bureau of Public Roads (1964). Traffic Assignment Manual. U.S. Dept. of Commerce, Urban Planning Division, Washington D.C.

Branston, D. (1976). Link capacity functions: a review. *Transportation Research* 10(4), 223-236.

Carey, M. and Y.E. Ge (2004), Efficient discretisation of link travel time models. *Networks and Spatial Economics* 4(3), 269-290.

Carey, M. and Y.E. Ge (2005), Convergence of a discretised travel-time model. *Transportation Science* 39(1), 25-38.

Carey, M. and M. Bowers (2012). A review of flow-density functions. *Transport Reviews* 32(1), 49-73.

Carey, M. P.K. Humphreys, M. McHugh and R. McIvor (2014). Extending travel-time based models for dynamic network loading and assignment, to achieve adherence to first-in-first-out and link capacities. *Transportation Research Part B* 65, 90-104.

Daganzo, C.F. (1994). The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory, *Transportation Research Part B*, 28(4), 269-287.

Daganzo, C.F. (1995a). The cell transmission model, Part II: network traffic, *Transportation Research Part B*, 29(2), 79-93.

Daganzo, C.F. (1995b). A finite difference approximation of the kinematic wave model of traffic flow. *Transportation Research Part B*, 29(4), 261-276.

Drew, D.R. (1965). Deterministic Aspects of Freeway Operations and Control. *Highway Research Record,*  99, 48-58.

Gartner, N.H., Messer, C.J. and A.K. Rathi (Eds.) (2007). Traffic Flow Theory: a state of the art report. http://www.tft.pdx.edu/docs.htm. AHB45 Committee on Traffic Flow Theory & Characteristics, Transportation Research Board, National Academy of Sciences.

Gentile G. (2010). The General Link Transmission Model for Dynamic Network Loading and a comparison with the DUE algorithm. In New developments in transport planning: advances in Dynamic Traffic Assignment (selected papers from the DTA 2008 Conference, Leuven), eds L.G.H. Immers, C.M.J. Tampere, F. Viti, Transport Economics, Management and Policy Series, Edward Elgar Publishing, MA, USA, pp. 153-178.

Greenshields, B.D. (1934). A study of traffic capacity, *Proc. Highway Research Board* 14, 448-474.

Haight, F. A. (1963). Mathematical Theories of Traffic Flow. Academic Press Inc., New York.

Lindsey, Robin, and Erik Verhoef (2000). Congestion modeling. In Handbooks in Transport, volume 1: *Handbook of Transport Modelling*, ed. by David A. Hensher and Kenneth J. Button, Elsevier, Amsterdam, ch. 21, 353-373.

May, A.D. and Keller, H. E. (1968). Evaluation of Single-and Two-Regime Traffic Flow Models, *Fourth International Symposium on the Theory of Traffic Flow Proceedings*, Karlsruhe, West Germany.

May, A.D. (1990). *Traffic Flow Fundamentals*.  Prentice Hall, 384 pages.

Ngoduy, D., Hoang, N.H., Vu, H.L., Watling, D. (2016). Optimal queue placement in dynamic system optimum solutions for single origin-destination traffic networks. *Transportation Research Part B*. In press.

Nie, X. and H.M. Zhang (2005). Delay-function-based link models: their properties and computational issues. *Transportation Research Part B* 39(8), 729–751.

Papageorgiou, M., J.-M. Blosseville and H. Hadj-Salem (1989). Macroscopic modelling of traffic flow on the Boulevard Peripherique in Paris. *Transportation Research Part B*, 23(1), 29-47.

Pipes, L.A. (1967). Car Following models and the fundamental diagram of Road Traffic, *Transportation Research* 1, 21-29.

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

Sheffi, Y. (1985). *Urban Transportation Networks*. Prentice-Hall, Inc., Englewood Cliffs, N.J., USA.

Small, K.A. and X. Chu (2003). Hypercongestion. *Journal of Transport Economics and Policy* 37(3), 319-352.

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

Transportation Research Board (2011). 75 Years of the Fundamental Diagram for Traffic Flow Theory. Greenshields Symposium, Woods Hole, Massachusetts, July 8–10, 2008. Transportation Research Circular E-C149, Traffic Flow Theory and Characteristics Committee, Transportation Research Board, Washington, DC.

Walters, A.A. (1961). The theory and measurement of private and social cost of highway congestion, *Econometrica* 29, 676-699.

Wu, J.H., Chen, Y. and M. Florian (1995). The continuous dynamic network loading problem: a mathematical formulation and solution method, presented at the 3rd Euro Working Group Meeting on Urban traffic and transportation, Barcelona.

Wu, J.H., Y. Chen, and M. Florian (1998). The continuous dynamic network loading problem: a mathematical formulation and solution method. *Transportation Research* 32B,173-187.

Xu, Y., Wu, J.H., Florian, M., Marcotte, P. and D.L. Zhu (1996). New advances in the continuous dynamic network loading problem. Publication CRT-96-26*.* University of Montreal.

Xu, Y.W., J. H. Wu, M. Florian, P. Marcotte, and D. L. Zhu (1999). Advances in the continuous dynamic network loading problem. *Transportation Science* 33, 341–353.

Yperman I. (2007). The Link Transmission Model for dynamic network loading, PhD Thesis, Katholieke Universiteit Leuven, Belgium.