Published article:

Carey, M, Balijepalli, C and Watling, D (2013) *Extending the Cell Transmission Model to Multiple Lanes and Lane-Changing.* Networks and Spatial Economics. 1 - 29. ISSN 1566-113X

http://dx.doi.org/10.1007/s11067-013-9193-7

EXTENDING THE CELL TRANSMISSION MODEL TO MULTIPLE LANES AND LANE-CHANGING

Malachy Carey, Chandra Balijepalli, David Watling

m.carey@its.leeds.ac.uk, n.c.balijepalli@leeds.ac.uk, d.p.watling@its.leeds.ac.uk

Institute for Transport Studies, 34-40 University Road, University of Leeds, Leeds LS2 9JT.

Key words: Cell transmission model; multiple lanes; lane changing; road traffic congestion; spillback; dynamic traffic assignment; spillback

Abstract

Macroscopic or flow-based dynamic traffic assignment (DTA) models normally treat traffic in each direction on a roadway as a single lane and, since they do not consider multiple lanes, they can not consider lane-changing behaviour. To investigate how the results may be affected by explicitly considering lanes and lane changing, we consider a road link that consists of two adjacent homogeneous lanes. We assume that traffic entering each lane already knows in which lane it wishes to exit at the end of the link, whether it wishes to exit in the same lane or in the other lane. We model the traffic flows in each lane using a cell transmission model but adapt it to allow for traffic moving from cells in one lane to cells in the other lane. The CTM is used because it handles the modelling of queues and their spillback in an intuitive and widely accepted manner, and our extensions of it allow congestion in one lane to spill back into adjacent lanes. In particular, we investigate how lane-changing and congestion are affected by varying the assumptions concerning two key behavioural parameters, namely the locations at which drivers wish to change lanes and the vehicle spacing needed for lane changing (gap acceptance) as compared to the spacing needed when staying in the same lane (car following). We conclude that there are many situations where modelling a link as a single lane will give a poor approximation to the underlying multi-lane behaviour, or be unable to capture issues of interest, and for those situations multi-lane modelling is appropriate.

1 Introduction

In macroscopic (flow based) modelling of time-varying traffic on road networks (dynamic traffic assignment), lanes and movement between lanes are typically ignored. Partly as a result, such models, despite their many advances, tend not to be as used yet in practical applications, in contrast with microsimulation based models, which do include lanes and lane changing. However, we concur with previous authors (e.g. in reviews by Boyce *et al*. (2001), Peeta and Ziliaskopoulos (2001), Carey and Watling (2003)) in believing that analytic macroscopic models provide several advantages, such as the possibility to explore theoretical questions of existence and uniqueness and the possibility to devise and explore the convergence properties of solution algorithms for efficiently calculating solutions. It has also often been noted that analytic macroscopic model have the advantage of need fewer parameters. In the present paper we take a first step towards developing such tractable, analytic network models that include lane-changing. In particular, we consider a single lane model, extend it to include more than one lane and movement of traffic between lanes, investigate how this affects the results and also consider how the results are affected by varying certain lane-change parameters. We here consider a single link rather than a network since there are more than sufficient issues and problems to explore for a single link.

Though there has been much valuable previous research related to lane-changing in a macroscopic context, it has not been concerned with producing models that are tractable for use in network models for dynamic traffic assignment (DTA). It has been concerned with investigating specific aspects or issues such as characteristics of lane-changing traffic, stability, density waves, phase transitions, deviations from first-in-first-out (FIFO), lane-changing intensity and weaving, multi-class traffic, stochastic parameters, and empirical evidence. Though valuable, this work has been too specific and complex for use in network modelling. There is too much of it to review here, hence we refer the reader to reviews and reference lists in Tang *et al*. (2009) and Jin (2010a, 2010b). It is perhaps worth mentioning early seminal work on lanes in macroscopic traffic models (Gazis *et al*. (1962), Munjal and Pipes (1971), Michalopoulos *et al*. (1984)). The work has perhaps not progressed as much as these early authors might have expected.

From the literature referred to above we see that at present there is not a plausible and tractable way of including lane-changing in DTA models, yet we might ask: does this really matter? On the contrary, it might be argued that lane-changing behaviour is too unpredictable or too detailed a notion to be considered for planning models, and may not lead to an improved representation of traffic reality at the level of detail required for planning applications. The pioneers of DTA modelling would have faced similar concerns (of adding too much detail for planning models) when proposing alternatives to the established static equilibrium methods, especially given the difficulty in estimating time-sliced origin-destination matrices. However, the argument in favour of dynamic over static assignment is not that we necessarily always need the *detail* of dynamic models, but that we now know that static models are systematically biased in their treatment of congestion phenomena, even at the gross level. Adopting a similar spirit in our work, therefore, we have been motivated to consider DTA with lane-changing in order to see whether, at the gross level, there is a systematic effect at the gross-level compared with DTA without lane-changing. If such an effect exists, then there are several implications, even if at the present our understanding of dynamic lane-changing phenomena is relatively immature. For example, even if we are unsure about the specific behavioural mechanisms that underpin lane-changing, then if they have a gross-level impact we should at least consider performing sensitivity tests of alternative assumptions, and in the future should devote more resources to a better empirical understanding of such phenomena.

Returning, therefore, to our aim of introducing lanes and lane-changing into macroscopic DTA modeling, a natural strategy is to start from an established credible macroscopic model for a single link that does not include lanes or lane-changing, and then introduce these into the model. As a starting point in the present paper we chose the cell-transmission model (CTM) for a number of reasons. First, the CTM closely approximates the well-known LWR model (Lighthill and Whitham (1955), Richards (1956)) which appears to be the most widely accepted model of traffic flows on a link. Second, the CTM divides the link into a series of cells, which is very convenient for modelling lane-changing along the link. Third, one of the most important reasons why drivers change lanes is to avoid or reduce delays due to queues and spillback of traffic, hence it is important that these be included in the model. The CTM, and the LWR model, are well suited to this as they handle the formation and dissipation of queues and spillback in a way that is consistent with traffic flow theory.

Laval (2003), Laval and Daganzo (2006) and Laval and Leclercq (2008) introduced multiple lanes into a single link model as follows. They took the LWR model which describes flows on a single lane, and extended this to single links consisting of two or more lanes in parallel. In the conservation equation at each point on the lane they include inflows from the neighbouring lanes and outflows to neighbouring lanes. In this way they obtain an extended LWR model that includes lane-changing. The model is stated as in continuous in time and space hence to obtain a computable model they discretise it over time and space. The resulting discrete model time-space model is, not surprisingly, similar to the CTM. It differs from the usual single lane CTM in that the outflow from any cell in any lane can flow into the next downstream cell in the same lane or in any adjacent lanes.

In this paper, unlike Laval and Daganzo, we start by extending the single lane CTM instead of extending the single lane LWR model. This seems less elegant, starting from the discretised model (the CTM) rather than the underlying continuous LWR model. However, we have various reasons for taking this approach, the main one being as follows. We wish to consider a range of behavioural rules and sub-models in a multi-lane context and it is much easier to do that by extending the CTM than by extending the LWR model. It is not immediately obvious how some of these rules or sub-models would be introduced in the LWR model. Some examples, which we consider in this paper, are as follows.

(a) Drivers may have to change lanes to get onto an appropriate path to their destination (mandatory lane changing) or may wish to changes lanes to get into a faster or more desirable lane (discretionary lane changing). For mandatory lane changing:

 (i) Drivers may wish to change lanes as soon as possible after entry to the link.
 (ii) The proportion of drivers who wish to change lanes may vary with distance along the link, increasing approaching the exit.
 (iii) When entering a link, each driver may plan to change lanes as soon as possible after a certain distance along the link, and this may be different for each driver
 (iv) Drivers may change lanes when look-ahead indicates lane changing may be more difficult further ahead.

 We model (i)-(iii) in Section 4 below and consider (iv).

(b) In a multi-lane context, a cell in a given lane may not have sufficient "receiving" capacity to take all of the traffic that wishes to move into it from the next upstream cell in its own lane and adjacent lanes. In that case we introduce various lane priority rules to reflect driver behavior, for example:
 (i) Traffic has priority and can not be held back by traffic trying to enter from an adjacent lane.
 (ii) The receiving capacity in the downstream cell is shared between those who wish to move into it from the same lane and from adjacent lanes and is shared in proportion to their numbers.
 (iii) Same as (ii) except that the sharing may be in some proportion not based on their respective numbers.

 We model each of these in Section 5 below.

The present paper is a continuation of work in Carey (2006) and Balijepalli, Carey and Watling (2010) and has similarities to work in Zong *et al*. (2012).

In this paper we focus on mandatory lane changing, since modelling that was sufficiently challenging and we wished to include detailed modelling, computation and experiments. If we included that for both mandatory and discretionary lane changing then the paper would exceed journal length. We choose mandatory lane changing rather than discretionary in this paper since it is of course essential, and in urban conditions mandatory is much more important than discretionary, while the reverse may be true for highways. Discretionary lane changing requires additional modelling and would also affect the results. We plan to consider discretionary lane changing in later work. Modelling discretionary lane changing will involve comparing speeds in adjacent lanes, and the speed in each lane can be derived/ computed from the density and the flow density function, both of which are available in the CTM which is used in this paper.

2 The (single-lane) cell-transmission model

The LWR model is usually stated as a conservation equation, expressed as a partial differential equation, together with a flow-density function. The latter can be written as $y(x,t) = f(k(x,t))$, where $y(x,t)$ and $k(x,t)$ respectively denote the flow rate (vehicles per second) and density (vehicles per unit distance) at time *t* at location *x*. Daganzo (1994, 1995a), when seeking to approximate the LWR model, assumed a piecewise linear (trapezoidal) flow-density function, as in Fig. 1, which can be written as

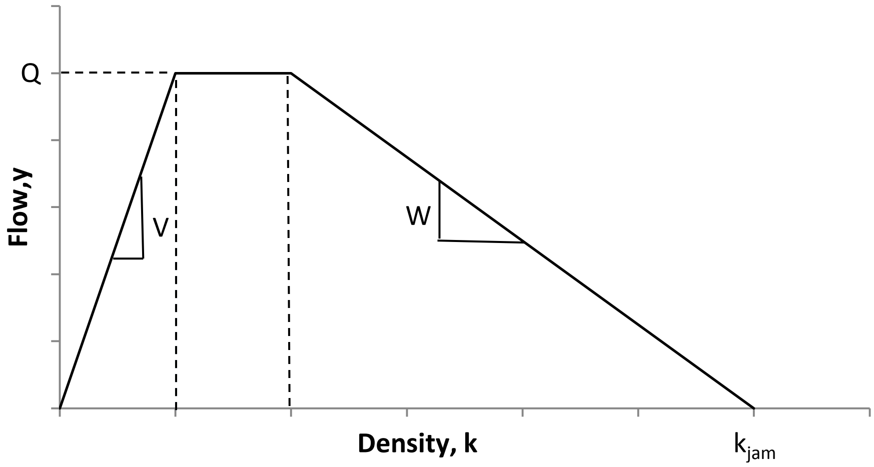

Figure 1. A trapezoidal flow-density curve.

$$y = Min\{Vk, y^{\max}, (k^{\max} - k)W\} \qquad 0 \le k \le k^{jam} \tag{1}$$

where $y^{\max}$ is the maximum flow rate, k^{jam} is the jam density, V is the free-flow speed and W is the wave speed. Using this, he showed that the LWR model can be closely approximated by dividing the link into homogeneous cells i = 1, …, I, and dividing time into time steps or time ticks t = 1, …, T, with the cell length chosen as the distance that can be traversed in one time step at free-flow speed. Then the LWR model is approximated by the flow equations

$$u_i^t = Min\{x_{i-1}^t,\ Q,\ (H - x_i^t)\delta\} \tag{2}$$

together with the conservation equations

$$x_i^{t+1} = x_i^t + u_i^t - u_{i+1}^t \tag{3}$$

where u_i^t is the flow from cell i-1 to cell i in time step t, x_i^t is the occupancy of cell i in time step t, Q is the maximum flow rate, H is the maximum holding capacity of a cell (at jam occupancy) and $\delta = W/V$. Empirical evidence indicates that $W < V$ hence $\delta < 1$. The three terms on the RHS of (2) indicate three traffic regimes namely uncongested flow, capacity flow and congested flow.

3 A two-lane cell-transmission model

Now consider a roadway consisting of two adjacent lanes L1 and L2, each carrying traffic in the same direction and again assume that each lane is homogeneous along its length, so that the parameters of the flow-density relationship are the same along its length. Thus, each lane is divided into I cells (i = 1, …, I) of equal length and with the cell end points being at the same locations in both lanes. To represent these lanes in the CTM, add subscripts L1 and L2 to the flow equation (2) and conservation equation (3), thus

$$u_{i,L1}^t = Min\{x_{i-1,L1}^t, Q_{L1}, (H_{L1} - x_{i,L1}^t)\delta_{L1}\} \text{ and } u_{i,L2}^t = Min\{x_{i-1,L2}^t, Q_{L2}, (H_{L2} - x_{i,L2}^t)\delta_{L2}\} \tag{4}$$

and

$$x_{i,L1}^{t+1} = x_{i,L1}^t + u_{i,L1}^t - u_{i+1,L1}^t + u_{i,L21(*)}^t - u_{i,L12(*)}^t \text{ and}$$

$$x_{i,L2}^{t+1} = x_{i,L2}^t + u_{i,L2}^t - u_{i+1,L2}^t + u_{i,L12(*)}^t - u_{i,L21(*)}^t \tag{5}$$

where

$x_{i,L1}^{t}$ and $x_{i,L2}^{t}$ are the occupancies of cell *i* in time step *t* in lanes L1 and L2 respectively,

$u_{i,L1}^{t}$ and $u_{i,L2}^{t}$ are the flows from cell *i*-1 to *i* in time step *t* in lanes L1 and L2 respectively, and

$u_{i,L21(*)}^{t}$ and $u_{i,L12(*)}^{t}$ are the flows from cell *i*-1 in lane L2 to cell *i* in lane L1, and from cell *i*-1 in lane L1 to cell *i* in lane L2 respectively, in time step *t*.

Let traffic move from one lane to the other as it progresses along the roadway. More specifically, traffic exiting from a cell can move directly into the next downstream cell in the same lane or in the adjacent lane. Let L11, L22, L21 and L12 denote traffic types, where the first numeral denotes the lane in which the traffic enters the link and the second numeral denotes the lane on which it exits, or wishes to exit, from the link. First consider traffic L11 and L22. Let

$x_{i,L11}^{t}$ and $x_{i,L22}^{t}$ denote the amounts of traffic of types L11 and L22 respectively in cell *i* in time step *t*.

$u_{i,L11}^{t}$ and $u_{i,L22}^{t}$ denote the flow of traffic types L11 and L22 from cell *i*-1 to *i* in lanes L1 and L2 respectively.

Now consider lane changing traffic type L21. Let L21(2) denote this traffic while it is still in lane L2 and L21(1) denote it after it has moved into lane L1. For simplicity in this paper we will consider only traffic type L21 (moving from lane L2 to L1) and not L12 (moving from lane L1 to L2). Thus,

$x_{i,L21(2)}^{t}$ is the traffic type L21 that is in still in lane L2 at time step *t* in cell *i*.

$x_{i,L21(1)}^{t}$ is the traffic type L21 that has already moved over into lane L1 by time step *t* in cell *i*.

$u_{i,L21(2)}^{t}$ is the flow from cell *i*-1 to *i* of traffic type L21 that is still in lane L2,

$u_{i,L21(1)}^{t}$ is the flow from cell *i*-1 to *i* of traffic type L21 that has already moved over into lane L1

$u_{i,L21(*)}^{t}$ is the flow of traffic type L21 from cell *i*-1 in lane L2 to cell *i* in lane L1 in time step *t*

The above notation implies that

$$u_{i,L1}^{t} = u_{i,L11}^{t} + u_{i,L21(1)}^{t} \tag{6.1}$$

$$u_{i,L2}^{t} = u_{i,L22}^{t} + u_{i,L21(2)}^{t} \tag{6.2}$$

We can also write a conservation equation for each of the separate traffic types L11, L22, L21(1) and L21(2) as follows.

$$x_{i,L11}^{t+1} = x_{i,L11}^{t} + u_{i,L11}^{t} - u_{i+1,L11}^{t} \tag{7.1}$$

$$x_{i,L22}^{t+1} = x_{i,L22}^{t} + u_{i,L22}^{t} - u_{i+1,L22}^{t} \tag{7.2}$$

$$x_{i,L21(1)}^{t+1} = x_{i,L21(1)}^{t} + u_{i,L21(1)}^{t} - u_{i+1,L21(1)}^{t} + u_{i,L21(*)}^{t} \tag{7.3}$$

$$x_{i,L21(2)}^{t+1} = x_{i,L21(2)}^{t} + u_{i,L21(2)}^{t} - u_{i+1,L21(2)}^{t} - u_{i,L21(*)}^{t} \tag{7.4}$$

$$x_{i,L1}^{t+1} = x_{l,L11}^{t+1} + x_{i,L21(1)}^{t+1} \;\;,\;\; x_{i,L2}^{t+1} = x_{l,L22}^{t+1} + x_{i,L21(2)}^{t+1} \tag{7.5}$$

We could add similar notation in brackets (i.e. (1), (2) and (*)) after the traffic type subscripts L11 and L22 for traffic types L11 and L22. However, since in this paper we are assuming that traffic types L11 and L22 do not change lanes, this notation would be redundant.

d_{L11}^t, d_{L22}^t and d_{L21}^t are the exogenous arrivals of these traffic types at the link entrance.

For reference, we summarise the scenario assumed in most of this paper as follows.

S1: Consider a link consisting of two adjacent lanes L1 and L2 with traffic in the same direction in both lanes. For traffic types arriving at the link the lane on which they enter the link and the lane on which the exit from the link are pre-specified. If the traffic arriving at, or waiting at, the entrance of a lane exceeds the inflow capacity of the first cell of the lane, the excess is held in an initial holding cell of arbitrarily large capacity until inflow capacity is available. Traffic within each lane moves forward governed by the cell-transmission model, as in Daganzo (1994, 1995b). Traffic movement between lanes is governed by various rules set out in this paper to describe driver behavior concerning lane changing.

The lane changing rules referred to in the preceding sentence, and set out below, extend the usual single lane CTM to multiple lanes. The LWR and the CTM (see (2) and (4)) are based on the idea that the flow at any location in space and time depends only on the density at that location and not the density or other factors at any later or earlier locations. However, in the real world, regardless of local densities or speeds, drivers may move into the lane they prefer to be in prior to exiting from the link, for example, move into the right lane if turning right or left lane if turning left. Or they may stay in their current lane for the same reason. Drivers may wait until near to the exit before changing into their correct lane for the exit, or they may change lanes much earlier in order to be sure that they do not arrive at the exit in the wrong lane. For example, if they see that congestion is worse further ahead in the adjacent lane they may move into it now as it may be more difficult to do so further ahead. Or drivers may have a preferred lane, e.g. the lane nearest to the kerb, even when it is more congested. All of the above factors are independent of local density or speed which is the basis of the CTM, hence lane changing in this paper is an extension of the CTM rather than an exact implementation of it.

4 Modelling the numbers who wish to change lanes at each cell and at each time step

To model this in the various forms of CTM presented or discussed below, label all traffic according to the lane on which it entered the link and the lane on which it plans to exit from the link and keep track of this labelled traffic as it progresses on its time-space paths through the link. This labeling and tracking of traffic is needed to enforce first-in-first-out (FIFO) for the traffic in each lane and in each cell within each lane. This extended notion of FIFO to include the consideration of lanes is a natural extension of the FIFO property generally held to be desirable for non-lane models (Carey, 2004). We will not set out this detailed labelling process in this paper, or in the algorithms below, just as it is not set out in the original Daganzo (1994, 1995a, 1995b) CTM articles or in later articles on the CTM. The process, without lane changing, is discussed in Carey, Watling and Balijepalli (2011). Enforcing FIFO involves labeling each traffic cohort in a cell according to the time at which they entered the cell and then letting them exit from the cell only in the same order as they entered it. A simplified version of FIFO enforcement that is sometimes used is as follows: if there are two or more traffic types in a cell and there is not sufficient exit capacity (receiving capacity) to let all this traffic move ahead into the next downstream cell in the current time step *t*, then let these two or more types exit in proportion to their numbers currently in the cell. However, this only ensures an approximation to FIFO since it does not fully take account of when the various traffic types entered the cell.

As a result of this labeling and tracking process, when the CTM process arrives at cell *i* at time *t* we know the (numerical) value of $x_{i,L21}^t$, the traffic currently (at time t) in cell *i* in lane L2 that intends to move into lane L1 now or at some stage before the exit from the link. Though drivers plan to change lanes, that does not mean that they plan to do so at the first opportunity. In view of that we introduce another variable, namely

$w^t_{i,L21}$ is the amount of traffic in cell *i*-1 of lane L2 at time *t* that wishes to change lanes (to the next downstream cell *i* in lane L1) as soon as possible.

By definition, $w^t_{i,L21} \le x^t_{i,L21}$.The following are some possible ways to estimate $w^t_{i,L21}$.

(i) *Drivers wish to change lanes as soon as possible after entry to the link.*

Perhaps the simplest way to estimate $w^t_{i,L21}$ is, set $w^t_{i,L21} = x^t_{i,L21}$, that is, assume that all traffic that plans to change lanes will try to do so at the first available opportunity.

(ii) *The proportion of drivers who wish to change lanes varies with distance along the link.*

Consider the traffic in cell *i* at time step *t* that wishes to eventually change lanes, i.e. traffic $x^t_{i,L21}$. In early cells this traffic may be in no hurry to change lanes but as it comes nearer and nearer to the end of the link it becomes more and more urgent to change lanes. This can be formalised by letting the numbers wishing to change lanes at cell *i* be $w^t_{i,L21} = (i/I)x^t_{i,L21}$. This assumes that the desire to change lanes increases linearly along the link, however, this assumption can be relaxed by writing $w^t_{i,L21} = f(i)x^t_{i,L21}$ where $0 \le f(i) \le 1$. The function $f(i)$ can be interpreted as the proportion or fraction of the traffic $x^t_{i,L21}$ that wishes to change lanes at cell *i*, or as the probability that an individual driver in the cell at time *t* will wish to change lanes at cell *i*. This allows the desire to change lanes to start from a higher level than $f(i) = 0$ or $f(i) = 1/I$. For example, $f(i)$ may be constant for all cells up until near the end of the link, at which point it may increase to 1. Letting $f(i)$ increase to 1 by the final cell(s) is appropriate since we assume that any traffic that has not yet changed lanes by that point will wish to do so. Note that since $0 \le f(i) \le 1$ we always have $0 \le w^t_{i,L21} \le x^t_{i,L21}$.

Two special cases for $f(i)$ are $f(i) = 1/I$, which reduces $w^t_{i,L21} = f(i)x^t_{i,L21}$ to $w^t_{i,L21} = (i/I)x^t_{i,L21}$, and $f(i) = 1$ which reduces $w^t_{i,L21} = f(i)x^t_{i,L21}$ to $w^t_{i,L21} = x^t_{i,L21}$ as in (i) above so that (i) is a special case of (ii).

(iii) *When entering a link, each driver plans to change lanes as soon as possible after a certain distance along the link.*

We can assume that, when they arrive at a link, some drivers may plan to change lanes soon after entering the link, some may plan to change a little further along the link and so on. Let $f_i(d^t_{L21})$ denote the fraction of traffic d^t_{L21} that plans to change lanes by cell *i*. Note that we assume that they plan to change lanes by a certain distance (cell) along the lane and not by a particular time or time step. For example, in the absence of better information, we can assume that the drivers lane-change plans are uniformly distributed along the link, in which case $f_i(d^t_{L21}) = (i/I)d^t_{L21}$.

Consider traffic in cell *i* at time *t*. The traffic that could potentially exit up to cell *i* by time *t* is $D^{*t}_{L21} = \sum_{t'=1}^{t-\tau(i,t,L2)} d^{t'}_{L21}$ where $\tau(i,t,L2)$ is the time taken to travel from the entrance up to cell *i* at time *t*. The numbers of these who wished to change lanes by cell *i* is $f_i(D^{*t}_{L21})$. The numbers that have actually changed lanes by time *t* at cell *i* is $U^t_{i,L21(*)} = \sum_{t'=1}^{t} u^{t'}_{i,L21(*)}$. Hence the numbers that wish to change lanes by cell i and have not yet done so is $w*^{t}_{i,L21} = f_i(D^{*t}_{L21}) - U^{t}_{i,L21(*)}$, which can be called the unsatisfied demand for lane changing. The numbers that wish to change lanes and are also available in cell at time t are

$$w^{t}_{i,L21} = \min\{ x^{t}_{i,L21},\ w*^{t}_{i,L21} \}. \tag{8}$$

This constraint is needed since the potential unsatisfied demand $w*^{t}_{i,L21}$ may not all have arrived at cell i by time t and may exceed the numbers $x^{t}_{i,L21}$ in cell i at time t.

The main complication in the above method for estimating $w^{t}_{i,L21}$ lies in computing the travel times $\tau(i,t,L2)$ at each cell i at time t. It is possible to compute these because the CTM computes the values of the variables for all cells i at each time step before proceeding to the next time step, so that at time step t the values of the variables for all preceding time steps are known.

(iv) *Change lanes when driver's look-ahead indicates lane changing may be more difficult further ahead*.

A widely observed phenomenon is as follows. There are two or more lanes, L1 and L2, and the traffic in L1 is congested and moving more slowly than in L2, or is jammed. Drivers in L1 do not move over into the faster lane L2, because they need to be in lane L1 at the next junction, exit. Also, drivers in lane L2 who need to move into L1 will typically move into it before the beginning of the congested slow moving section in L1. They do that because they know it will be more difficult to move into L1 further ahead where it is more congested. Also, speeding ahead in the faster lane and then moving into the congested lane at the last moment may be seen as 'queue jumping' and hence avoided. We do not report a formulation or results concerning the above phenomenon in this paper.

5 Modelling the flows within lanes and flows from lane to lane

In this paper we are assuming that, in the absence of lane-changing, traffic in each lane is described by the CTM. The CTM is based on a function or curve showing flow as a function of occupancy or density, for each cell in the lane. Density is the inverse of vehicle spacing, hence the fundamental equation can be stated as occupancy as a function of vehicle spacing or headway. The usual CTM describes flows on a single lane, without lane changing, hence implicitly assumes that vehicle spacing is based on car-following headways. However, when vehicles are changing lanes they are not adopting car-following headways. They are adopting lane-changing headway, or gap-acceptance headways, which are normally larger than car-following headways. Let

α = the ratio of lane-changing or gap acceptance headways to car-following headways.

In keeping with the above, the CTM for a single link or lane implicitly assumes that the receiving capacity of each cell i is based on vehicles having car-following headways. If the headways are larger than that, then the traffic will absorb a larger amount of the receiving capacities. More specifically, if lane-change traffic has headways that are α times larger, then it will absorb α times as much of the receiving capacity. In that case, lane-change traffic needs to be scaled up α times when checking that it does not exceed the receiving capacity. Thus, the usual receiving capacity constraint at time t, i.e.

(traffic sent to cell i from cell i-1) ≤ (receiving capacity of cell i)

can be re-written as

(traffic sent to cell i in lane L from cell i-1 in same lane)
+ (α) (traffic sent to cell i in lane L from cell i-1 in adjacent lane)

≤ (receiving capacity of cell i in lane L)

Or, more formally, $u^t_{i-1,L11} \le r(i,L1,t)$ is replaced by if $[u^t_{i-1,L11} + \alpha\, u^t_{i-1,L21}] \le r(i,L1,t)$,

The scenarios (i)-(iii) in the previous section describe drivers' plans or intentions to change lanes, but drivers may or may not be able to fully execute those plans since they are in competition for lane space with drivers already in the adjacent lane. More formally, there are two types of traffic in cell *i*-1 that are competing to move into cell *i* in lane L1 at time step *t*, which we will refer to as cell (i,L1,t). The two types are $x^t_{i-1,L1}$ in cell *i*-1 in lane L1 and $w^t_{i,L21}$ in lane L2. If there sufficient capacity (receiving capacity) in cell (i,L1,t) for both to enter it (i.e. if $x^t_{i-1,L1}$ + $w^t_{i,L21} \le r(i,L1,t)$) then let them do so. However, if the capacity $r(i,L1,t)$ is not sufficient to allow all of both types to enter (i.e. if $x^t_{i-1,L1}$ + $w^t_{i,L21} > r(i,L1,t)$) then share out the limited capacity $r(i,L1,t)$ between the two traffic types in proportions $F^t_{i,L1}$ and $F^t_{i,L21}$, i.e. let an amount $F^t_{i,L1}\, r(i,L1,t)$ of the first traffic type and an amount $F^t_{i,L21}\, r(i,L1,t)$ of the second traffic type enter cell (i,L1,t), where $F^t_{i,L1} + F^t_{i,L21} = 1$.

The above fractions can be referred to as priority factors, priority fractions, sharing fractions or splitting fractions. They can be estimated or defined in various ways and three different ways can be stated briefly as follows.

Definition (i). Let the fractions be $F^t_{i,L1}$ = 1 and $F^t_{i,L21}$ = 0.

Definition (ii). Let the fractions $F^t_{i,L1}$ and $F^t_{i,L21}$ be proportional to the numbers of these two traffic types wishing to move into cell (i,L1,t), so that the numbers moving into this cell are then proportional to the numbers wishing to move into it.

Definition (iii). More generally, let the fractions $F^t_{i,L1}$ and $F^t_{i,L21}$ take any nonnegative values that sum to 1.

(i) and (ii) are special cases of (iii). These three definitions of the priority fractions are intended to reflect a range of possible driver behaviors regarding lane changing. These definitions are used to govern lane-changing in algorithms for a two-lane CTM set out in subsections (i)-(iii) below. The sections (i)-(iii) below correspond to the definitions (i)-(iii) above.

(i) *Absolute priority for traffic in lane L1*

The above assignment, $F^t_{i,L1}$ = 1 and $F^t_{i,L21}$ = 0, can be interpreted as ensuring that traffic in lane L1 has absolute priority in moving forward in lane L1. Traffic from lane L2 can move over into lane L1 if and only if, while it is moving over, it does not restrict the ability of traffic already in lane L1 to move forward. Traffic in lane L1 will not have to reduce speed to let in traffic from lane L2. Of course, *after* this traffic from lane L2 has moved over into lane L1 it will affect the flow and speed of traffic in lane L1 as it move into downstream cells i+1, i+2, etc.

> Priority rule R1: If the sum of the two types of traffic wishing to move into cell (i,L1,t) exceeds its receiving capacity $r(i,L1,t)$ then let all of traffic already in lane L1 move first before allowing any of the lane-change traffic moves.

This priority rule can be implemented as follows.

> Apply the single-lane CTM to lane L1, ignoring any demand for lane-changing into lane L1 cell (i,L1,t). Then, if there is still unused receiving capacity in cell (i,L1,t), assign it to traffic that wishes to move into this cell from lane L2. Then apply the single-lane CTM to the remaining traffic in lane L2.

This two-lane CTM is set out more fully for scenario S1 in the following algorithm.

Algorithm A1: A two-lane CTM for scenario S1, with priorities given by rule R1.

Step 1: Apply the single-lane CTM to lane L1 to compute the flow $u_{i,L1}^{t}$ from cell *i*–1 to *i*. Subtracting this from the receiving $r(i, L1, t)$ capacity in cell (i,L1,t) leaves the remaining receiving capacity $r(i, L1, t) - u_{i,L1}^{t}$ = Min$\{Q_{L1}, (H_{L1} - x_{i,L1}^{t})\delta_{L1}\} - u_{i,L1}^{t}$.

Step 2: Compute the flow $u_{i,L21(*)}^{t}$ from lane L2 to lane L1 as follows. This flow should not exceed the demand for lane-changing , i.e. $u_{i,L21(*)}^{t} \le w_{i,L21}^{t}$. Also, $\alpha\, u_{i,L21(*)}^{t}$ should not exceed any remaining receiving capacity of cell (i,L1,t) as computed in Step 1. Combining these two requirements gives

$$\alpha\, u_{i,L21(*)}^{t} = \text{Min}\{\alpha\, w_{i,L21}^{t},\ Q_{L1} - u_{i,L1}^{t}, (H_{L1} - x_{i,L1}^{t})\delta_{L1} - u_{i,L1}^{t}\}.$$

Step 3 to 11: Steps 3 to 11 are the same as in Algorithm 2 below, hence we do not set them out again here.

□

In the solution provided by the above algorithm A1, and the other algorithms below, it is possible that not all of the traffic that planned to change lanes, from lane L2 into lane L1, i.e. d_{L21}^{t} ,will have done so before reaching the exit of the link. The unsatisfied demand for lane changing may still be positive in the final cell $i = I$. No model or algorithm could entirely eliminate that possibility, since more traffic may plan to move between lanes than the capacity of the receiving lane will allow. However, as noted just after algorithm A2 below, if some traffic wishes to change lanes from some cells or series of cells, and is unable to do so, that automatically triggers a self adjusting mechanism to make it more likely that the traffic can change lanes in the next cell and so on for successive cells. Also, in later work we plan to introduce additional features to reflect the fact that, as they approach the exit end of a link, drivers who wish to change lanes may adjust their behaviour to make lane changing more likely. They may become more aggressive about pushing into the traffic in their target lane.

(ii) *Traffic priority based on the relative quantities that are competing for lane space*

Priority rule R2: If the sum of the two types of traffic wishing to move into cell (i,L1,t) exceeds its receiving capacity $r(i, L1, t)$ then let this capacity be shared out between the two types of traffic in proportion to their demands (numbers).

To implement this, introduce priority factors or fractions

$$F_{i,L1}^{t} = x_{i-1,L1}^{t} / [\, x_{i-1,L1}^{t} + \alpha\, w_{i,L21}^{t}\,] \text{ and } F_{i,L21}^{t} = (1 - F_{i,L1}^{t}) = \alpha\, w_{i,L21}^{t} / [\, x_{i-1,L1}^{t} + \alpha\, w_{i,L21}^{t}\,] \quad (9)$$

to share out the receiving capacity $r(i, L1, t)$ if it is less than the demand for entry to cell (i,L1,t).

Algorithm A2: A two-lane CTM for scenario S1, with priority fractions in the receiving lane given by rule R2 and (9).

Step 0: Initialisation. Set i = 1 and t = 1. Initialise cell occupancies, i.e. set $x_{1,L1}^{t} = \bar{x}_{1,L1}^{t}$ and $x_{1,L2}^{t} = \bar{x}_{1,L2}^{t}$ for time steps t = 1, … , T, and set $x_{i,L1}^{1} = \bar{x}_{i,L1}^{1}$ and $x_{i,L2}^{1} = \bar{x}_{i,L2}^{1}$ for cells i = 1, … , I. Also, initialize all other variables.

Step 1: Compute the priority factors $F_{i,L1}^{t}$ and $F_{i,L21}^{t}$ from (9), to use to share out the (receiving) capacity $r(i, L1, t)$ in cell (i,L1,t).

Step 2: Compute $u_{i,L1}^{t}$ and $u_{i,L21(*)}^{t}$ by sharing out the receiving capacity $r(i, L1, t)$ as follows.

Compute the receiving capacity of cell i in lane L1, i.e. $r(i, L1, t)$ = min $\{Q_{L1}, (H_{L1} - x^t_{i,L1})\delta_{L1}\}$. Then,

if [$x^t_{i-1,L1} + \alpha\, w^t_{i,L21}$] $\leq r(i, L1, t)$, set flows ($u^t_{i,L1} = x^t_{i-1,L1}$ and $\alpha\, u^t_{i,L21(*)} = \alpha\, w^t_{i,L21}$), and

if [$x^t_{i-1,L1} + \alpha\, w^t_{i,L21}$] $> r(i, L1, t)$, set flows ($u^t_{i,L1} = F^t_{i,L1}\, r(i, L1, t)$ and $\alpha\, u^t_{i,L21(*)} = F^t_{i,L21}\, r(i, L1, t)$).

Step 3: Take the outflow $u^t_{i,L1}$ computed in Step 2 and decompose it into its two components $u^t_{i,L11}$ and $u^t_{i,L21(1)}$ (see equation (6.1)) by applying FIFO rules (see first paragraph of Section 4).

Step 4: Subtract the flow $u^t_{i,L21(*)}$, computed in Step 2, from the occupancy $x^t_{i-1,L2}$ of cell i–1 in lane L2 and add it to the occupancy $x^t_{i,L1}$ of cell i in lane L1.

Step 5: Apply the single-lane CTM to lane L2 to compute the flow $u^t_{i,L2}$ from cell i–1 to i within lane L2.

Step 6: Take the outflow $u^t_{i,L2}$ computed in Step 5 and decompose it into its two components $u^t_{i,L22}$ and $u^t_{i,L21(2)}$ (see equation (6.2)) by applying FIFO rules (see first paragraph of Section 4).

Step 7: Apply the conservation equations (5) and (7.1)-(7.4) to compute/ update the amount of each traffic type L1, L11 and L21(1) in lane L1 and traffic types L2, L22 and L21(2) in lane L2 for the next time step t+1. The aggregate number in each cell i in lanes L1 and L2 is then given by (7.5).]

Step 8: Compute / update $w^t_{i,L21}$ to time step t+1: this is needed in Step 2 of the next iteration (iteration t = t+1) and is needed to update the priority factors $F^t_{i,L1}$ and $F^t_{i,L21}$ in Steps 1 and 2. Various ways to define $w^t_{i,L21}$ are proposed and discussed in (i) to (iii) in Section 4. Each of these methods requires $x^t_{i,L21(2)}$, which was computed in Step 7 above.

Step 9: If i < I, set i = i+1 and return to step 1.

Step 10: Set i = 1. If t < T, set t = t+1 and return to step 1.

Step 11: Stop. The above one-pass process provides solution values for the cell inflows, outflows and occupancies for each traffic type in each cell in each time step.

□

It can be shown that the assignment in Step 2 is continuous, that is, the computed flows $u^t_{i,L1}$ and $u^t_{i,L21}$ vary continuously as [$x^t_{i-1,L1} + w^t_{i,L21}$] increases from less than $r(i, L1, t)$ to greater than $r(i, L1, t)$.

Note that the definition of the factors $F^t_{i,L21}$ and $F^t_{i,L21}$ in (ii) implies the following self-adjusting property. If not as much traffic has changed lanes up to cell i as would wish to, because of insufficient receiving capacity in lane L1, that increases the amount of traffic waiting to change lanes in the next cell, i.e. $w^t_{i,L21}$ increases. That in turn tends to increase the value of the priority factor $F^t_{i,L21}$ = $\alpha\, w^t_{i,L21}$/[$x^t_{i-1,L1} + \alpha\, w^t_{i,L21}$] and that in turn increases amount of traffic that will change lanes in the next cell, given by $u^t_{i,L21}$ where $\alpha\, u^t_{i,L21(*)} = F^t_{i,L21}\, r(i, L1, t)$. Thus, if traffic is unable to change lanes in some cell or series of cells, that automatically triggers a self adjusting mechanism to make it more likely to change lanes in the next cell and so on for successive cells.

(iii) *Other priority rules for traffic changing lanes*

In the above algorithm A2 it is assumed that the priority factors $F^t_{i,L1}$ and $F^t_{i,L21}$ (with $F^t_{i,L1} + F^t_{i,L21}$ = 1) are computed using the proportionality rule R2. Now consider an alternative priority rule.

> Priority rule R3: If the sum of the two types of traffic wishing to move into cell (i,L1,t) exceeds its receiving capacity $r(i, L1, t)$ then let this capacity be shared out between the two types of traffic in proportions $F^t_{i,L1}$ and $F^t_{i,L21}$, which are restricted only by nonnegativity and $F^t_{i,L1}$ + $F^t_{i,L21}$ = 1.

If rule R3 is used instead of R2 in above algorithm A2 then Step 2 has to be revised, since otherwise it may send more to the receiving lane than is available to send. The revised algorithm, which includes algorithm A2 as a special case, is as follows.

Algorithm A3: Same as Algorithm A2 but using priority rule R3 instead of R2.

The only change needed to convert Algorithm A2 to A3 is, in Step 2, replace

"If [$x^t_{i-1,L1} + \alpha\, w^t_{i,L21}$] > $r(i, L1, t)$, set [$u^t_{i,L1}$ = $F^t_{i,L1}\, r(i, L1, t)$ and $\alpha\, u^t_{i,L21(*)}$ = $F^t_{i,L21}\, r(i, L1, t)$]"

with the following.

If the demand for entry to cell (i,L1,t) exceeds its receiving capacity $r(i, L1, t)$, i.e.

if [$x^t_{i-1,L1} + \alpha\, w^t_{i,L21}$] > $r(i, L1, t)$

then either

(a) the demands $x^t_{i-1,L1}$ and $\alpha\, w^t_{i,L21}$ both exceed their allowed flows $F^t_{i,L1}\, r(i, L1, t)$ and $F^t_{i,L21}\, r(i, L1, t)$ respectively (the allowed flows are their allowed shares of the receiving capacity $r(i, L1, t)$), or

(b) the demand $x^t_{i-1,L1}$ is less than its allowed flow $F^t_{i,L1}\, r(i, L1, t)$ and the demand $\alpha\, w^t_{i,L21}$ exceeds its allowed flow $F^t_{i,L21}\, r(i, L1, t)$, or

(c) the demand $\alpha\, w^t_{i,L21}$ is less than its allowed flow $F^t_{i,L21}\, r(i, L1, t)$ and the demand $x^t_{i-1,L1}$ exceeds its allowed flow $F^t_{i,L1}\, r(i, L1, t)$.

In case (a) set [$u^t_{i,L1}$ = $F^t_{i,L1}\, r(i, L1, t)$ and $\alpha\, u^t_{i,L21(*)}$ = $F^t_{i,L21}\, r(i, L1, t)$],

in case (b) set [$u^t_{i,L1}$ = $x^t_{i-1,L1}$ and $\alpha\, u^t_{i,L21(*)}$ = $r(i, L1, t) - u^t_{i,L1}$], and

in case (c) set [$\alpha\, u^t_{i,L21(*)}$ = $\alpha\, w^t_{i,L21}$ and $u^t_{i,L1}$ = $r(i, L1, t) - \alpha\, u^t_{i,L21(*)}$]. □

As noted above, priority rule R2 is a special case of R3. Priority rule R1 is a special case of R3 with $F^t_{i,L1}$ = 1 and $F^t_{i,L21}$ = 0. Since R1 and R2 are special cases of R3 they can be handled by the Algorithm A3.

6 Numerical experiments

In this section we provide selected results from a more extensive body of numerical experiments that have been conducted. The purpose of the experiments reported in Section 6.1 is primarily to illustrate the lane-changing approach proposed, and to make some contrast with approaches that do not consider lanes or lane-changing. In Section 6.2, we then explore the impact of varying two key behavioural assumptions/parameters of the lane-changing model. While clearly we would expect such assumptions to have 'local' effects, e.g. in terms of the particular cells where lane-changing may take place, it is not clear whether they have any overall systematic effect at the level of detail that is typically of interest for DTA applications. That is to say, in Section 6.2 we try to answer the question: is the effect of varying the lane-change assumptions purely local, or are there substantial impacts on the out-flow profile from the link and/or on the link travel time profile? Finally, in Section 6.3 we aim to make a comparison with what we might expect to obtain were we to apply a single-lane CTM to approximate the multi-lane situation. Thus we try to answer the question: in broad terms does it really matter that we model lanes and lane-changing, or are we simply adding unnecessary detail and complication to the problem.

In our experiments we consider a single, homogeneous, two-lane road. A trapezoidal flow-density relationship is assumed to hold along the length of each lane. The problem is discretized in time and space by dividing each lane of the link into forty cells, such that in free-flow conditions vehicles traverse exactly one cell in one time step. The parameters of the trapezoidal flow-occupancy function in equation (4) for each of these 40 cells are assumed to be

$$Q_{L1} = Q_{L2} = 100,\ H_{L1} = H_{L2} = 600,\ \delta_{L1} = \delta_{L2} = 0.25.$$

The exogenous arrivals of the different traffic types at the link entrance are (in the notation of Section 3) assumed to be:

$$d_{L11}^{t} = \begin{cases} 80 & \text{for } t = 1,2,...,40 \\ 0 & \text{otherwise} \end{cases} \quad d_{L22}^{t} = \begin{cases} 16 & \text{for } t = 1,2,...,40 \\ 0 & \text{otherwise} \end{cases} \quad d_{L21}^{t} = \begin{cases} 64 & \text{for } t = 1,2,...,40 \\ 0 & \text{otherwise} \end{cases}.$$

That is to say, the *total* inflow for each lane L1 and L2 is a uniform 80 per time step, and since this is less than the capacity of 100 then if the traffic all stayed in the lane in which it entered the link, it would pass through the entire link at free-flow speed (i.e. in a travel time of 40 time increments). However, as specified above, we instead assume that 80% (64 out of 80) of the vehicles that enter lane L2 wish to change lane at some point along the link in order to exit in lane L1, and since 80 + 64 = 144 is greater than the 100 capacity of lane L1, it must be the case that a queue will eventually form in lane L1 and spill back along lane L1. The queue will eventually dissipate at some time after the inflows drop to zero (i.e. some time after time step 40), and so we must consider the model over some longer time-span, sufficiently long that all flows entering by time step 40 have exited the link. In all the results presented we have assumed traffic priority rule R2, and hence have adopted Algorithm A2 from Section 5.

6.1 Illustration of lane-changing logic

In order to demonstrate the features of the model described, we shall select a particular example that is convenient to illustrate since most of the features of interest occur in a smaller number of cells at the beginning of the link. This example uses model (i) in Section 4 to represent how drivers wish to change lanes, i.e. they all wish to change lanes as soon as possible. In terms of the ratio of lane-changing headways to car-following headways (see Section 5), we have assumed $\alpha = 1$.

In Figure 2 we show a kind of "bird's eye view" of the cells across the whole link (on the horizontal axis), with snapshots taken every ten time increments. For each snapshot, the spatial profiles show the receiving capacity of each cell, the inflows to the cell from upstream cells in L1 and L2, and the potential inflows from lane L2. The latter 'potential inflow' is the number in the cell who wish to change lanes as soon as possible. Two main phases can be identified. In the first phase (which stops shortly after time t = 40), there are cells for which the total amount of traffic that wishes to move into the cell from the preceding cell in lanes L1 and L2 exceeds the cell's receiving capacity. The receiving capacity is shared out among the two types of inflow (as in priority rule R2) so that some traffic that wishes to change lanes has to postpone it until later cells, thus spreading lane changing further along the link. On the other hand, with such a strong desire to change lanes as soon as they enter the link, it can be seen that nearly all of the lane-changing is completed less than one quarter way along the link. If we had assumed the demand for lane-changing is governed by (ii) instead of (i) from Section 4 then lane-changing would spread further along the link. In the second phase (starting between t = 40 and 50), all lane changing has been completed, but the early part of lane L1 is still congested from the earlier activity. After this period when congestion has cleared, any remaining traffic on the link is able to exit at the free flow speed of one cell per time step, and so for time steps t = 60, 70, 80, 90 we simply see the wave shifting along the link at ten cells per ten time steps.

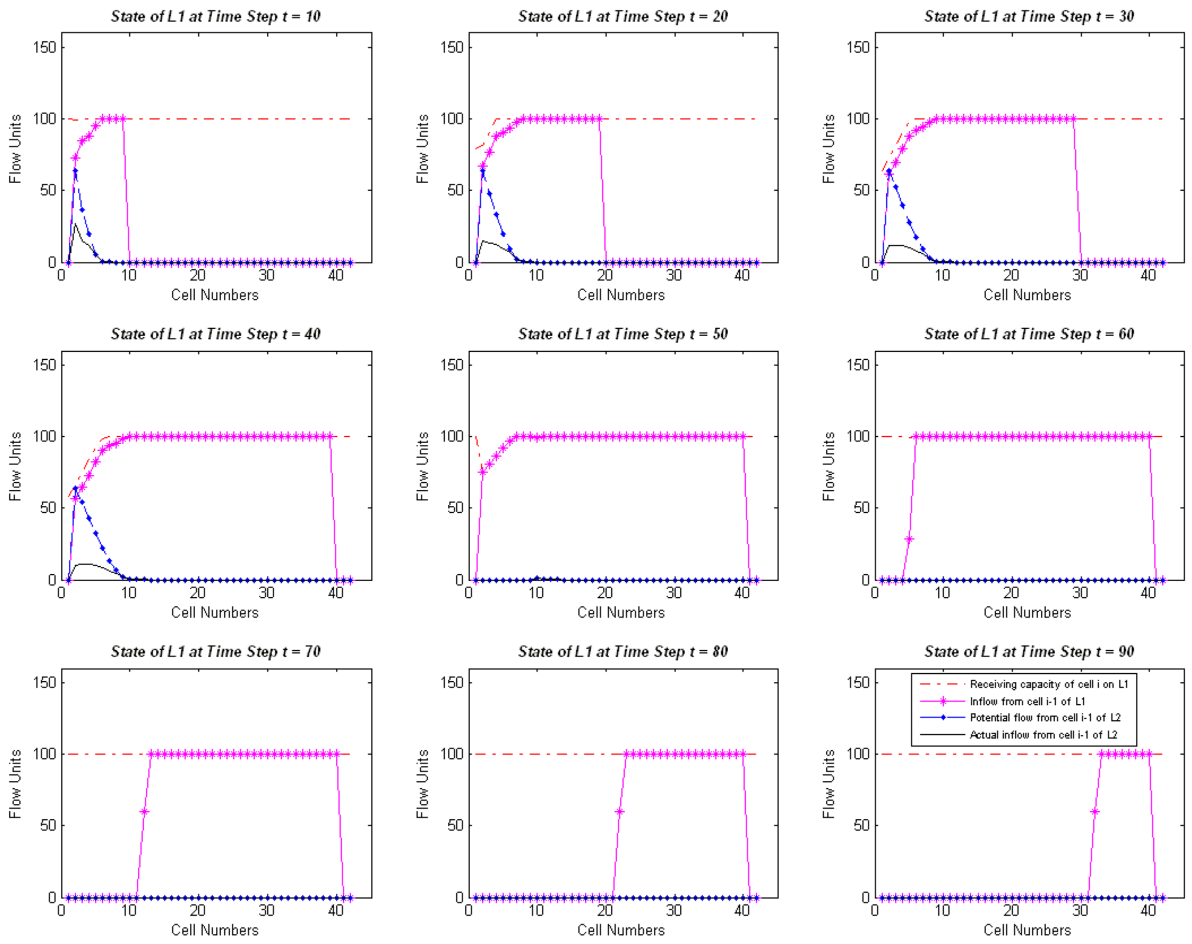

Figure 2. Temporal snapshots of the spatial profile of lane-changing activity along the link ($\alpha = 1$; $y_i = 1$, i.e. drivers wish to change lanes as soon as possible).

In Figure 3 we display the same example, but now in terms of temporal profiles (rather than spatial profiles), focusing on the early cells in which the main lane-changing activity takes place. In this Figure we can see how the receiving capacity for cells 2 to 5 falls and rises again over time in response to congestion, and how this total capacity is shared at any one time between the flows wishing to transfer from L1 and L2. The “potential” lane L2 to L1 transfer (i.e. the traffic currently in cell i in lane L2 that wishes to transfer to lane L1 as soon as possible) increases over time for any given cell, since the amount that can transfer in each time step is less than the cell inflow. It drops to zero when all relevant traffic has transferred. In contrast, for any given time step, the potential transfer decreases from cell to cell as transfers deplete the stock of traffic awaiting transfer. As noted above in discussing Figure 2, the completion of lane-changing activity signifies a qualitatively different phase of activity for the link. Thus lane-changing gives rise to temporal phenomena along the link, even in a case such as this where the lane-changers arrive uniformly (and in a uniform proportion to other traffic) at the link entry.

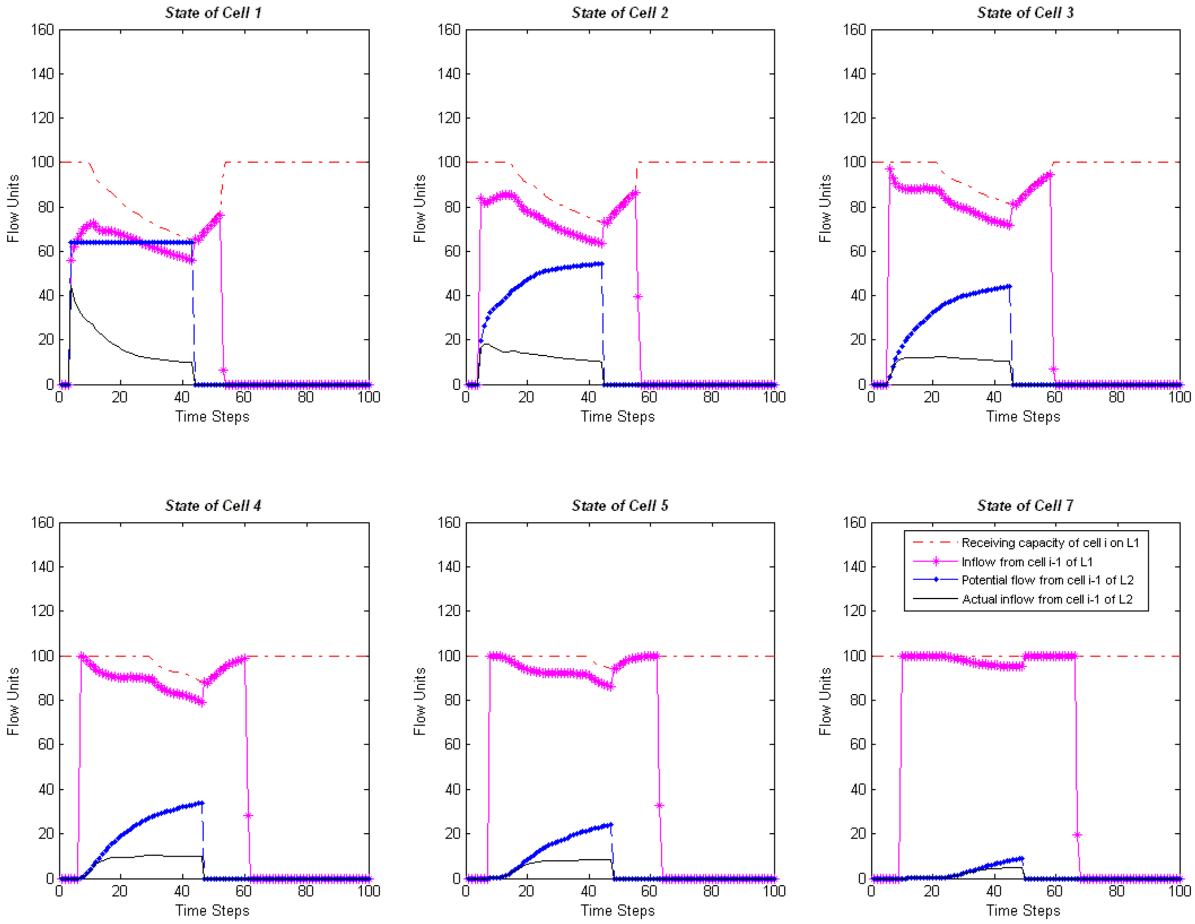

Figure 3. Cell-by-cell profiles of lane-changing activity in early cells of the link
(α = 1; y_i =1, i.e. drivers wish to change lanes as soon as possible).

Finally, in Figure 4 we contrast the states of the lanes L1 and L2, by illustrating the spatial profiles of occupancies at a number of time snapshots. In interpreting these graphs it should be noted that in our assumed flow-density relationship the downward sloping (congested) part starts at occupancy of 200 hence cells with occupancies above 200 are in a congested state. Even though there is no inflow after t = 40, the early cells in link L1 are so congested by t = 40 that it takes until t = 60 for the early cells in the link to revert to their uncongested state; after t = 60, traffic exits from lane L1 at the assumed maximum capacity rate (100 flow units per time step). These plots further emphasise that although the cells are identical across lanes in terms of their physical, flow-density characteristics, we see a very different relative usage of the cell capacity across the lanes, over time and space. This behaviour could not be captured by a single-lane CTM based on some kind of aggregate flow-density relationship (aggregated across lanes); this is an issue that we shall return to in Section 6.3.

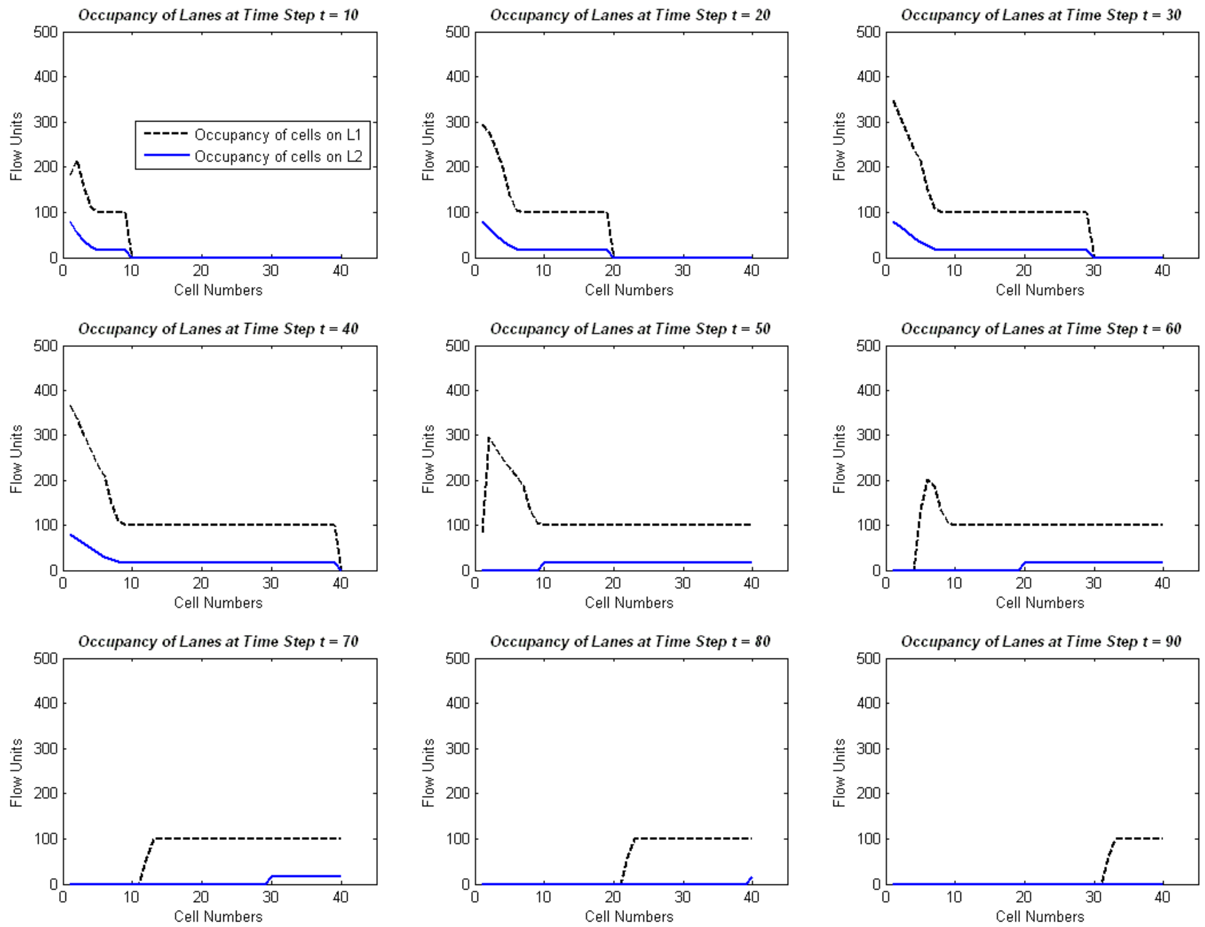

Figure 4. Temporal snapshots of spatial profiles of cell occupancies in lanes L1 and L2 ($\alpha = 1$, $y_i = 1$).

6.2 Sensitivity of outflow and travel time profiles to varying α or y_i, and implications for DTA

In this section we examine the sensitivity of the model outputs to two key behavioral parameters of lane-changing. Due to space limitations, it is not possible to show graphs equivalent to those in Section 6.1 for all values of these parameters. We will instead focus only on those results and graphs that are most relevant for implementing DTA for a network. Since the inflow profiles are taken as given, the results and graphs most relevant for DTA are (i) the link outflow profiles and (ii) the lane travel time profiles.

We first consider the effects of varying the parameter α, the ratio of vehicle spacing needed for lane-changing relative to that needed for car-following. Figure 5 shows the link out-flow profile and link traversal time profile for lane L1, for $\alpha = 1$, 2 and 3. Increasing α from 1 to 2 to 3 means doubling and then trebling the amount of receiving capacity that these vehicles require in the receiving cell in lane L1. That increases congestion and hence travel times in lane L1, as we see in the right hand panel in Fig 5: the travel time profile shifts upwards as α is increased. The increased travel times cause the outflow profiles in the left panel to shift downward and to the right.

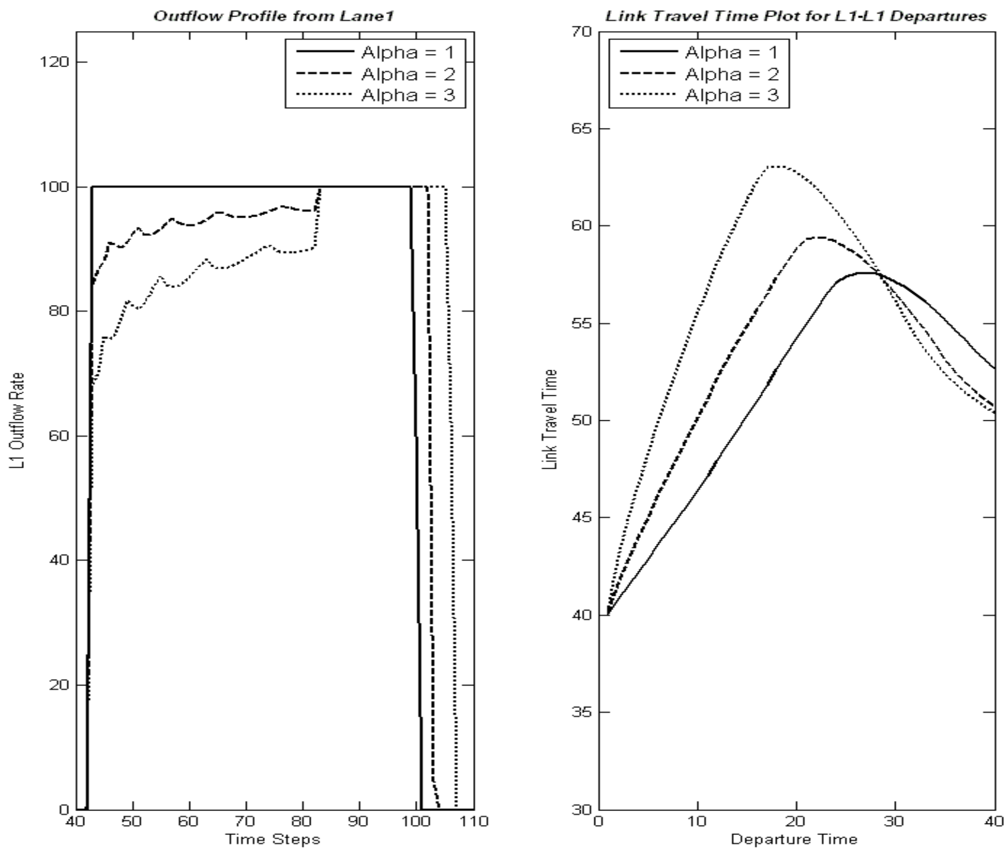

Figure 5. Lane L1 outflows and L11 travel time profiles for various values of α. (In all cases y_i = 1.)

As well as the above graphs illustrating the effects of varying α, we can generate similar graphs based on changing the assumptions concerning the parameter y_i , which defines how soon drivers wish to change lanes. For example in Section 4 we considered assumption (i), in which drivers wish to change lane as soon as possible after entering the link, and assumption (ii) which spreads the drivers' wish to change lanes further along the link. The choice of assumption tends to affect the profiles of outflows and travel times but are not graphed here.

Varying the parameters α and y_i affects not only the traffic that wishes to change lanes (traffic type L21) but also the traffic that is not changing lanes (traffic types L11 and L22). In Table 1 we show the effect on traffic L11 of varying α and y_i .
Effects on traffic L11 of increasing α *while holding y_i fixed*: In Table 1 column 2, α is three times its value in column 1 while y_i is held fixed. This triples the space needed in lane L1 for each vehicle that moves from lane L2 into L1, hence triples the amount of CTM receiving capacity that they use up in lane L1. That increases congestion spillback in lane L1 and, as we see in Table 1 columns 1 and 2, significantly increases the travel time span (from 101 to 107) and increases the total travel time of traffic type L11 by 7.75%.
Effects on traffic L11 of changing y_i while holding α *fixed*: In Table 1 columns 1 and 3, we changed y_i from y_i = 1 (i.e. change lanes as soon as possible) to y_i = *i*/I (i.e. the desire to change lanes increases linearly from 0 to 1 along the link). The latter causes lane changing to spread out more along the link instead of all wishing to rush into lane L1 asap. That in turn reduces congestion in lane L1 and reduces the travel time span (from 107 to 102) and reduces the total travel time of traffic type L11 by 6.1%.

	α = 1, y_i=1 (change lanes asap)	α = 3, y_i=1 (change lanes asap)	α = 3, y_i=i/I (spread lane changing along the link)
Time spread (time step at which last vehicle leaves lane L1)	101	107	102
Total travel time for demand type L11	1.6497×10^5	1.7775×10^5	1.6689×10^5

Table 1. Effects of varying lane-change parameters on travel times of non-lane-changing traffic L11

In the above figures and discussion we saw that varying the behavioral parameters α and /or y_i can significantly change the link outflow profiles and/ or the link or lane travel time profiles. However, we can also show that even when changes in α or y_i do not significantly change these profiles, they may still substantially change the location and spread of congestion *within* the link. To see this, compare Figs 6 and 7, which represent scenarios in which the travel demand profiles are the same and the outflow profiles are very similar. In Fig 6, y_i = 1, i.e. drivers wish to change lanes asap, and as a result in Fig 6 we see that the congestion is located early on the link. In Fig 7, y_i =i/I, i.e. drivers' desire to change lanes increases more gradually along the link and, as a result, in Fig 7 congestion occurs further along the link. Thus, though the travel time profiles are much the same in both cases, the location of congestion within the link is quite different in the two cases. This may be of more interest for traffic engineering and control purposes than for DTA purposes.

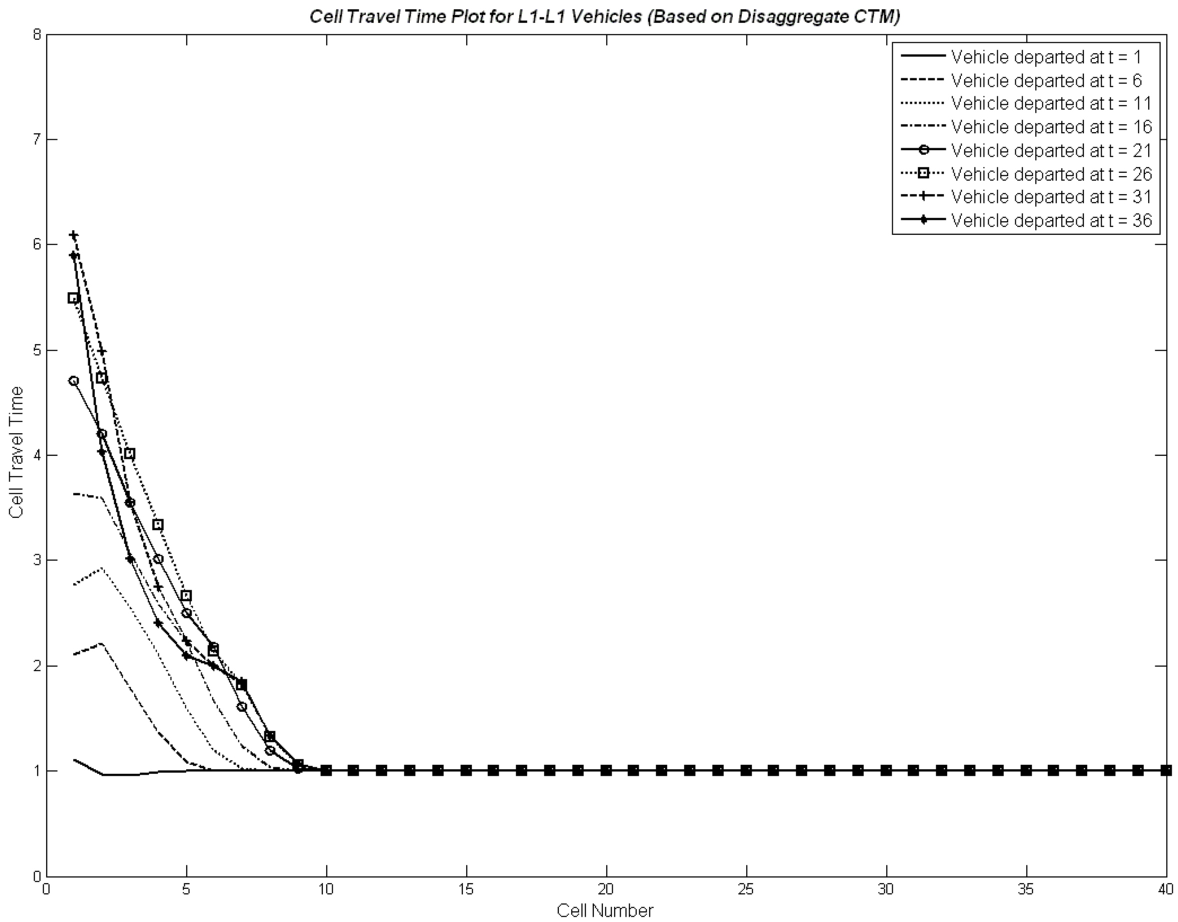

Figure 6. Cell travel times for traffic in lane L1, for various sample departure times
(α = 1; y_i = 1, so that L21 drivers wish to change lanes asap).

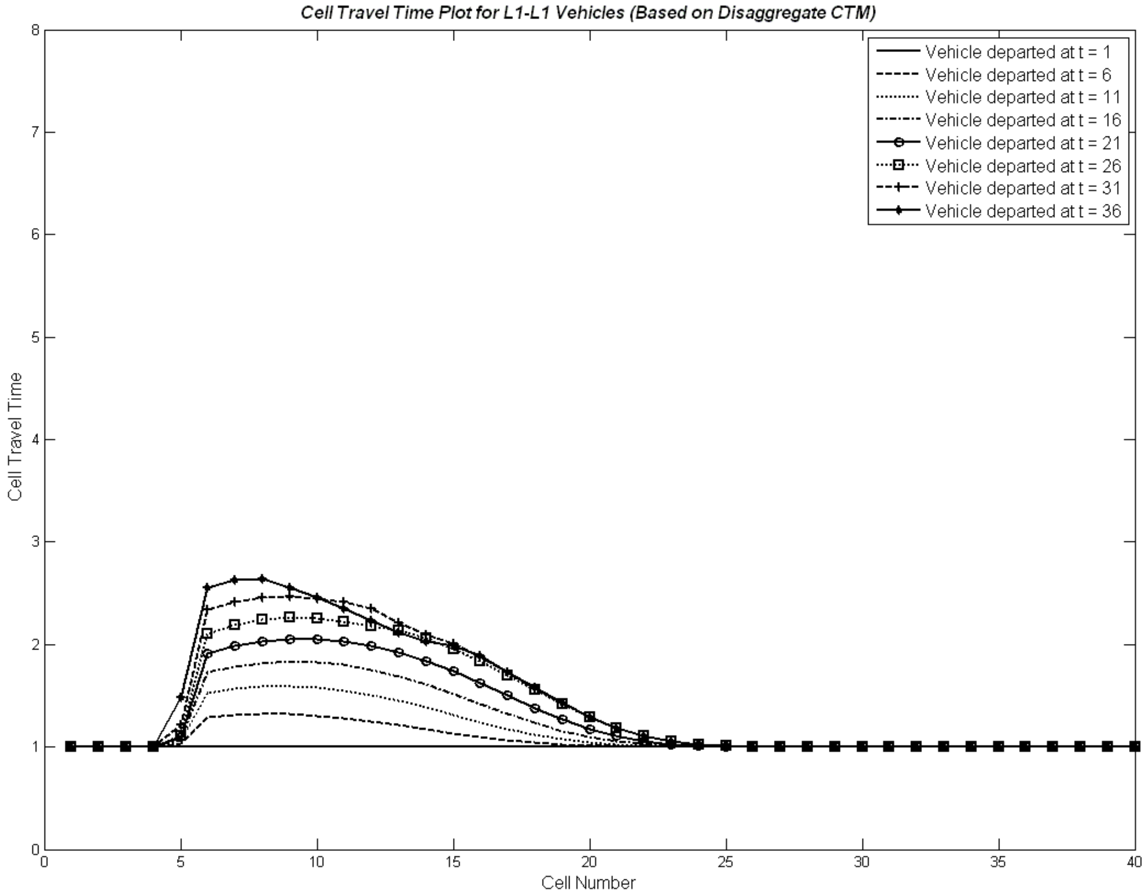

Figure 7. Cell travel times for traffic in lane L1, for various sample departure times ($\alpha = 1$; $y_i = i/\mathrm{I}$, so that L21 drivers wish to change lanes more uniformly along the link).

In summary, the lane changing behavior of drivers significantly affects the spatial and temporal dynamics of vehicle propagation on a road link. The larger are the gaps required by lane-changing drivers, the larger the travel times. Similarly, the spread of locations at which drivers first wish to change lanes affects congestion and the profiles of outflows and travel times. These profiles are key inputs into DTA modeling hence DTA results are affected by lane-changing behavior and by the precise assumptions that are made concerning it.

6.3 Comparing the two-lane CTM with a single-lane surrogate

The basic CTM represents traffic as if moving in a single lane, but is widely used in DTA models, and in other contexts, to represent links which in the real world links consist of multiple lanes. This raises the question therefore, can we capture at least approximately some of the effects we have observed in Sections 6.1/6.2 with a single-lane CTM, and if so is this approximation sufficiently good that modelling lanes and lane-changing is an unnecessary complication? In order to make such a comparison, it would be useful to make reference to some documented process of calibration for the standard CTM, especially in view of how it might be used to represent multi-lane roads, yet we are not aware of such a documented process that is available in the public domain. Therefore we must make some proposals ourselves as to how such a 'calibration' might take place.

Method 1.

In the numerical example considered in the subsections above we have two adjacent homogeneous lanes with each lane having a capacity of 100 flow units per time step. If we wish to model this as a single composite lane then a natural assumption would be to let the capacity of the single composite lane be the sum of the capacities of the two lanes that compose it, thus a capacity of 100 + 100 = 200. Also, since the two lanes are homogeneous the composite lane is homogeneous and has a capacity of 200 all along the lane. Applying the same simple logic to the travel demands yields a total demand of 160 flow units per time step. Then in a CTM context, since the capacity of each cell (200) exceeds the constant travel demand (160) in every cell, the traffic will flow from cell to cell along the full length of the link will be at the free flow speed of one cell per time step, and will therefore take a travel time of 40 time steps to traverse the link. In that case there is no congestion, which is clearly very different from all of the scenarios and results in the previous Sections 6.1 and 6.2. This way of forming the composite, single-lane model is by no means nonsensical, and would appear to fit with the intended logic of the CTM, but from this simple example we can see that it can entirely fail to capture important congestion phenomena, even at an aggregate level. The inclusion of lanes and lane-changing is therefore not just about adding detail, but it also qualitatively affects the aggregate properties of the link even when we are not ultimately interested in lanes and lane-changing.

While not wishing to diminish the impact of this simple comparison, we also considered how we might adjust the single-lane CTM to capture at least some of the congestion phenomena observed in the two-lane models in the preceding subsections. However, any such approach runs into the following fundamental problem. In the two-lane CTM we obtained a range of different results depending on what values we assumed for two behavioural parameters α and y_i . But a single-lane model does not contain such parameters hence any single-lane approximation cannot replicate or approximate the range of different results obtained from the two-lane model. If the single-lane approximation gives a good approximation to the two-lane model for some value of α and y_i then it must give a poor approximation for other values of α and y_i .

Method 2.

Despite the above problem, we developed what seems a very reasonable single-lane approximation to the two-lane model. This is based on constructing a flow-occupancy function for the composite single lane as follows. In the two-lane scenarios above it is assumed that the traffic enters the link in two lanes (L1 and L2) but it all exits in one lane (L1). Hence to reflect this in a single composite link:

(a) Let the flow-occupancy function for the first cell of the composite link be the sum of the flow-occupancy functions for the two separate lanes.
(b) Let the flow-occupancy function for the final cell of the composite link be the flow-occupancy function for lane L1.
(c) Then for all other cells of the aggregate link let the flow-occupancy function be obtained by linear interpolation between the flow-occupancy functions for the first and last cells obtained in (a) and (b).

To sum the flow-occupancy functions in (a)
let the cell flow capacity be the sum of those from the two component lanes,
let the cell occupancy at capacity (maximum) flow be the sum of those from the two component lanes, and
let the holding capacity (or jam occupancy) be the sum of those from the two component lanes.

Since it was assumed that the flow-occupancy functions for the two separate lanes are the same, it follows from (a)-(c) above that the flow-occupancy functions for the cells of the composite link taper down from cell i = 1 to the final cell i = 40. For example, the flow capacity tapers down from 200 to 100 and the cell holding capacity tapers down from 1,200 to 600. This is illustrated in Fig 8.

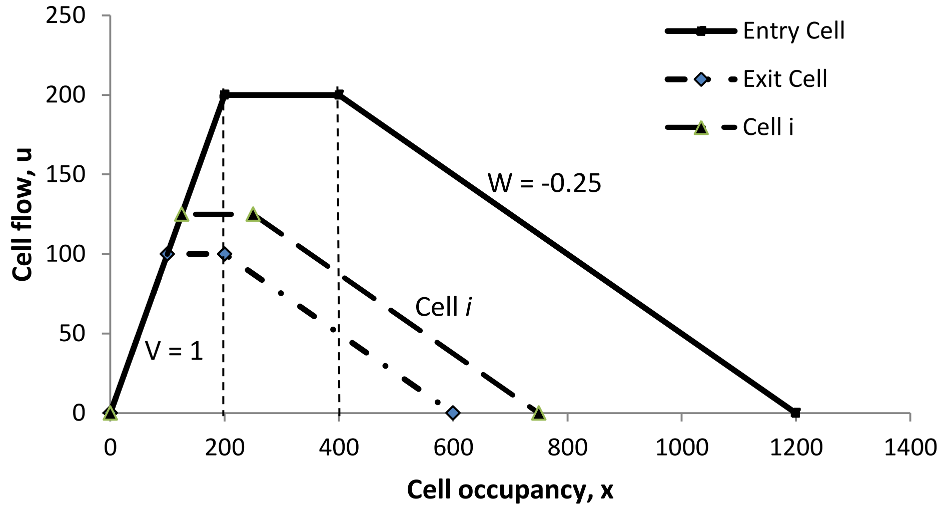

Figure 8. Flow-occupancy relationship for the single-lane model.

As the flow capacity for the single-lane model falls linearly from 200 for cell 1 to 100 at cell 40 hence by cell 23 it has fallen to 142.5 which is just below the travel demand level of 144 and from then on the capacity is below the demand level. As a result, flow from cell 23 onwards is congested (on the downward sloping part of the flow-occupancy curve) and this congestion spills back over time to cells before cell 23. This outcome is very different from most of the results for the two-lane model in previous sections. In Fig 9 we see that the congestion occurs towards the end of the link while for the two-lane model the congestion started and ended much earlier on the link as shown in Figs 6 and 7.

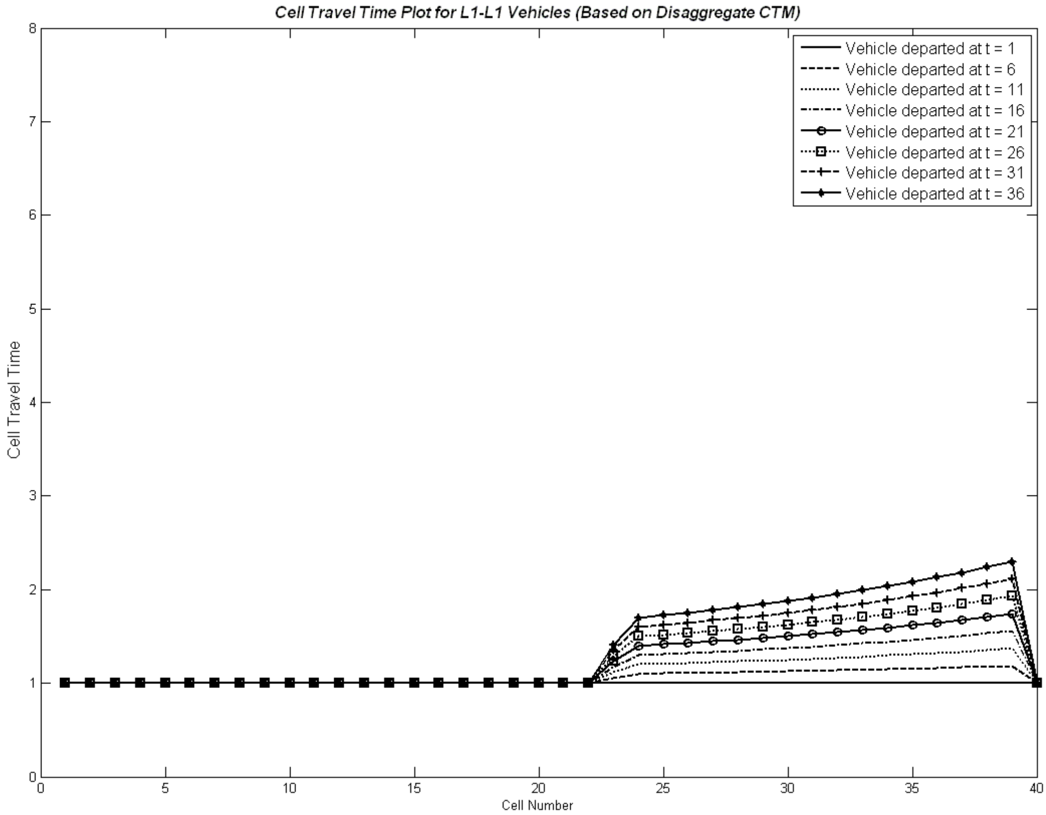

Figure 9. Cell travel times in a single-lane CTM based on Method 2 above, for various sample departure times

Concluding remarks

In summary, then, what may we conclude about the relationships between our multi-lane CTM with lane-changing and a CTM based on a single-lane approximation or surrogate? First let us note that there may be many multi-lane situations in which a simple aggregation of the lanes into a single surrogate lane, as in method 1 above, will give a reasonable approximation. This seems true if the lanes are largely independent of each other or if there is little systematic movement of traffic between lanes. Even if there is some weaving between lanes by vehicles overtaking each other, it may be possible to allow for that in the parameters of a flow-occupancy function, without introducing an explicit multi-lane model.

However, for multi-lane scenarios in which there are substantial systematic movements of traffic from lane-to lane, the problems and difficulties noted above arise if we try to model this with a single-lane surrogate. The above 'tapering' method for constructing a single-lane surrogate is applicable only for specific contexts in which the link exit capacity is less than the inflow capacity. Even in that case, it can not replicate all of the results from the two-lane model in the previous sections. Its parameters could perhaps be manipulated so that it gives a closer approximation for specific scenarios, but it seems that the only way to validate such results may be to compare them with results from a multi-lane model. But that means constructing a multi-lane model, in which case it seems there is no need for the single-lane surrogate.

In scenarios where the above tapering method is not applicable, different ad hoc approximation methods would be needed to construct single-lane surrogates. But again these would need such detailed knowledge of the particular lane changing context, and would need validated against a multi-lane model, that it may be preferable to just use a multi-lane model.

In conclusion, if there are systematic movements of traffic across lanes and we use a single-lane surrogate to represent the multi-lane context, as in Methods 1 or 2 above, then we:

- would have no way to represent or obtain the significant differences in results that we have observed in the multi-lane case that depend on the behavioural parameters of lane-changing;
- would be unable to see how the drivers' travel time and congestion experience, when traversing a link, depend on their choice of entry and exit lanes; and
- would not be guaranteed to capture even coarsely the 'average' congestion phenomena.

Since multiple lanes are such a common feature of urban roads, we conclude that trying to approximating these using only single lane models can yield poor approximations and an inability capture or describe some important behavior and phenomena.

7 Concluding remarks

In this paper we extended the usual single-lane cell-transmission model to apply it to two adjacent lanes with substantial movements of traffic moving between the lanes. We considered two lanes for simplicity but the approach can be extended to more than two lanes. We also assumed a trapesoidal flow-density function for each lane, since that is usually assumed for the single-lane CTM. We further assume that the demand or inflow profile for each lane is known and that, when vehicles enter the link, they already know in which lane they wish to exit at the end of the link.

The effect of lane-changing on congestion and on travel times depends not only on the flow-occupancy functions, the numbers who wish to enter and exit from each lane over time and the numbers who wish to change lanes but also depends on at least two other behavioural parameters the effects of which we investigate. The first of these parameters represents how the drivers' desire to change lanes varies along the link. We consider two possibilities namely (a) drivers wish to change lanes as soon as possible after entering the link and (b) drivers' wish to change lanes increases from 0 on entering the link to 1 (i.e. must change lanes) just before exiting from the link. Of course, even when drivers wish to change lanes they may not all be able to do so at their preferred location, due to congestion in the target lane – if the "receiving" capacity of the target cell is less than the numbers who wish to enter it, then some drivers have to postpone changing lanes until next downstream cell in which sufficient capacity is available. The second behavioural parameter represents the ratio of (a) the receiving capacity that is needed to receive a vehicle changing from an adjacent lane to (b) the receiving capacity that is needed to receive a vehicle moving forward in the same lane. This ratio can be interpreted as the ratio of the vehicle spacing needed for gap acceptance to the vehicle spacing needed for car following.

The results of various experiments with the model are shown in Section 6. Section 6.1 illustrates the behavior of the model for a basic case and Section 6.2 illustrates the effects of varying the two behavioural parameters referred to above. Section 6.3 considers what is perhaps the most interesting issue, namely, can a single-lane model give a sufficiently close approximation to a multi-lane model or must we instead resort to using multilane models such as developed in this paper? More specifically, if we wish to model traffic behavior (flows and travel times) in a link consisting of two or more adjacent lanes, can this be approximated sufficiently closely by a suitable a single-lane surrogate model? The advantage of a single-lane model is of course that is simpler to implement and, presumably because of that, all or almost all macroscopic (flow-based) DTA models implicitly assume that each link is a single lane. The conclusions that we come to concerning the two questions above is that in many important cases a single-lane approximation will give at best a very poor approximation and in those cases multi-lane modeling would be needed. Also, a multilane model may need to be used even more widely since we may not know in advance for which links a single-lane approximation would be sufficiently accurate. Further discussion of this issue is given in Section 6.3.

In summary, then, we believe we have shown that lane-changing assumptions *may* have an enormous impact on the overall out-flow and travel times of traffic, even at the gross level of the link. The impact is sufficiently great, we believe, that even if our current level of understanding lane-changing is poor then as a minimum we need to perform sensitivity tests to explore how sensitive policy decisions are to lane-changing assumptions. Also by developing a model that is able to accommodate a range of alternative lane-changing assumptions, we hope to activate more empirical research into the dynamics of lane-changing; often such research is 'demand-driven', and our approach establishes the need for such research and an outlet/use for it when it is completed.

The present paper focussed on extending the CTM to two-lane roadways, since that already provided sufficient challenges in investigating various issues concerned with lane changing and congestion. In future work we expect to extend this to more than two lanes since that raises additional issues. In Sections 3 to 5 we extended the CTM to multiple lanes and in Section 6 we compared that with adapting the parameters of a single lane CTM to try to capture some elements of a multi-lane context. We hope that our comparisons may stimulate interest – even those who use a standard single lane CTM must take a view as to how it is attempting to represent the kind of real-life, multi-lane phenomena, some of which we considered. There may be other possible ways, intermediate between the standard single-lane CTM and a full multi-lane CTM, to approximate such behaviour. For example, one possibility (suggested by a reviewer) is to divide a single link into three sub-links, one upstream and two in parallel in downstream, each corresponding to a lane. The flow ratio between the upstream link and the two downstream links can be determined similarly based on lane-changing decisions, with all lane-changing occurring where the upstream link diverge. One could change the length of the downstream links according to where the lane-changing is likely to occur.

In the present paper we focus on modelling flows on a single link. In later work we plan to extend this to a network. A key issue in extending to a network is whether or not within link lane-changing is assumed to be an integral part of path choice. If it is assumed to be an integral part of path choice that means that each traveler has to decide, when departing from the origin, at which cells (which locations within each link) s/he wishes to change lane. That implies choosing from a potentially intractable number of possible paths through the network. For a single link with multiple lanes, with each lane having multiple cells, there is a very large number of possible paths through the link since, in principle, traffic could weave about between lanes all along the link. If this is extended to a series of links, as in a network, the number of possible paths grows rapidly as it is the product of the numbers of possible paths through each of the individual links. An alternative, which is much more tractable, and probably better reflects traveller behavior, is to assume that the location of lane changes is not an integral part of the traveller's initial path choice. Assume instead that travelers decide which links to include in their path, and may even decide in which lane they plan to enter each link and in which lane they plan to exit from each link, but postpone the decision as to where to change lanes within links until they are actually on the link. Based on that, we can separate the within link modelling from the network modelling and can iterate between these until convergence is achieved. Such methods will be further developed in future work.

In the two preceding paragraphs we discussed extending the work in this paper, first extending it to multiple lanes or other possible approaches to capturing multi-lane behaviour and second extending it to networks. Some other directions for future research on this topic include the following, which are not in any particular order of importance. First, as discussed in the last paragraph in the Introduction, the present paper focussed on mandatory lane changing and should be extended to include discretionary lane changing. Second, traffic network microsimulation models typically include some lane-changing features and it would be desirable to explore how these could be simplified or approximated for use in macroscopic DTA models. That would enable transfer of ideas and promote consistency between microscopic and macroscopic modelling of time-varying traffic on networks. It is worth noting that microscopic models do not provide all the answers, since they face the same problems, namely a relatively poor empirical understanding of lane-changing behaviour, as well as further unknown relationships or parameters. In CTM based DTA there are of course topics other than lane changing that could also benefit from further investigation. One is the approximation of discrete signal controls and timings with the continuous flows and parameters used in DTA.

Acknowledgement

We wish to thank the two referees for their thoughtful comments, and thank the UK Engineering and Physical Science Research Council (EPSRC) for funding this research via grant EP/G051879. An earlier version of this paper was presented at the Fourth International Symposium on Dynamic Traffic Assignment (DTA2012), held at Martha's Vineyard, Massachusetts, USA, 4-6 June 2012.

References

Balijepalli, C., M. Carey and D. Watling (2010). Introducing lanes and lane-changing in a cell-transmission model. Presented at the *Third International Symposium on Dynamic Traffic Assignment* (DTA2010), held in Takayama, Japan, 29-31 July 2010.

Boyce D., Lee D.-H. & Ran B. (2001). Analytical models of the dynamic traffic assignment problem. *Networks and Spatial Economics*, 1(3-4), 377-390.

Carey, M. (2004). Link Travel Times I: Desirable Properties. *Networks and Spatial Economics*, 4, 257-268.

Carey, M. (2006). Introducing lane-changing into the CTM and DTA. Presentation at *UCL Traffic Flow Workshop – Sept 2006*, Centre for Transport Studies, UCL, 11-12 Sept. 2006. http://www2.cege.ucl.ac.uk/cts/TrafficFlow/pdfs/Carey.pdf

Carey, M., C. Balijepalli and D. Watling (2012). The cell transmission model with multiple lanes and lane-changing. Presented at the *Fourth International Symposium on Dynamic Traffic Assignment* (DTA2012), held on Martha's Vineyard, Massachusetts, USA, 4-6 June 2012.

Carey, M. and D. Watling (2003), Introduction to a special issue on dynamic traffic assignment. *Networks and Spatial Economics* 3(3), 403-406. (Due to a printing error, it appears in Issue 3(4) of the journal.)

Carey, M., D. Watling and C. Balijepalli (2011). Ensuring first-in-first-out in cell transmission models. Submitted to Transportation Research Part B.

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.

Gazis, D. C., R. Herman, and G. H. Weiss (1962) Density oscillations between lanes of a multilane highway. *Operations Research* 10, 658-667.

Jin, W.-L. (2010a). Macroscopic characteristics of lane-changing traffic. *Transportation Research Record: Journal of the Transportation Research Board*.

Jin, W.-J. (2010b). A kinematic wave theory of lane-changing traffic flow. *Transportation Research Part B* 44, 1001-1021.

Laval, J.A. (2003). Some properties of a multi-lane extension of the kinematic wave model. *TRB 2003 Annual Meeting* CD-ROM. Also, ITS Technical Report UCB-ITS-WP-2003-1, University of California Berkeley.

Laval, J.A. and C.F. Daganzo (2006). Lane-changing in traffic streams. *Transportation Research Part B 40(3)*, 251–264.

Laval, J.A. and L. Leclercq (2008). Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. *Transportation Research Part B* 42, 511–522.

Michalopoulos, P. G., D. E. Beskos, and Y. Yamauchi (1984) Multilane traffic flow dynamics: some macroscopic considerations. *Transportation Research Part B* 18, 377-395.

Munjal, P. K. and L. A. Pipes (1971) Propagation of on-ramp density perturbations on unidirectional two- and three-lane freeways. *Transportation Research Part B* 5, 241-255.

Peeta, S. and A.K. Ziliaskopoulos (2001). Foundations of dynamic traffic assignment: the past, the present and the future. *Networks and Spatial Economics* 1(3-4), 233-265.

Tang T.Q., Wong S.C., Huang H.J. and Zhang P. (2006) Macroscopic modeling of lane-changing for two-lane traffic flow. *Journal of Advanced Transportation*.

Zhong, R., A. Sumalee, T. Panl and W.H.K. Lam (2012). A cell transmission model with lane changing and incorporation of stochastic demand and supply uncertainties for freeway traffic state estimation. Submitted to *Networks and Spatial Economics* and presented at the Fourth International Symposium on Dynamic Traffic Assignment, held Masachusetts, USA, 4-6 June 2012.