Driving Behaviour: Models and challenges

Tomer Toledo

Faculty of Civil and Environmental Engineering,

Technion – Israel Institute of Technology, Haifa 32000, Israel

Tel: +972-4-8293080  Fax: +972-4-8295708  Email: toledo@technion.ac.il

Published in Transport Reviews 27(1), pp. 65-84, 2007
Bionote

Tomer Toledo is a Senior Lecturer in the Faculty of Civil and Environmental Engineering at the Technion – Israel Institute of Israel, Haifa 32000, Israel. Dr. Toledo’s research is in the areas of driver behaviour, large-scale traffic simulation, intelligent transportation systems and transportation network analysis.
Abstract

Driving behaviour models capture drivers’ tactical manoeuvring decisions in different traffic conditions. These models are essential to microscopic traffic simulation systems. This paper reviews the state-of-the-art in the main areas of driving behaviour research: acceleration, lane changing and gap acceptance. Overall, the main limitation of current models is that in many cases they do not adequately capture the sophistication of human drivers: they do not capture the inter-dependencies among the decisions made by the same drivers over time and across decision dimensions, represent instantaneous decision making, which fails to capture drivers’ planning and anticipation capabilities and only capture myopic considerations that do not account for extended driving goals and considerations. Furthermore, most models proposed in the literature were not estimated rigorously. In many cases, this is due to the limited availability of detailed trajectory data, which is required for estimation. Hence, data availability poses a significant obstacle to the advancement of driving behaviour modelling.
1 Introduction

Driving behaviour models capture drivers’ tactical manoeuvring decisions in different traffic conditions. These models are essential to microscopic traffic simulation systems and important to several other fields of transportation science and engineering such as safety studies and capacity analysis, in which aggregate traffic flow characteristics may be deduced from the behaviours of individual drivers. The literature on driving behaviour modelling focuses on a few key aspects: acceleration, lane changing and gap acceptance models that have been studied in the context of intersection crossing and within lane changing models. This paper reviews the state-of-the-art in these areas, and points out some of the limitations of current models and the challenges and needs for further research.

2 Acceleration models

Acceleration models can be broadly classified in two groups: car following models, which describe the acceleration drivers apply in reaction to the behaviour of their leaders (the vehicle in front), and general acceleration models, which also apply when drivers do not closely follow their leaders.

2.1 Car following models

The concept of car following was first proposed by Reuschel (1950) and Pipes (1953). Pipes assumed that the follower wishes to maintain safe time headway of 1.02 seconds from the leader. This value was derived from a recommendation in the California Vehicle Code. Using Laplace transformations, he developed theoretical expressions for the subject's acceleration given a mathematical function that describes the leader’s behaviour.
2.1.1 GM model

The stimulus-response framework that is the basis for many car following models was developed at the GM Research Laboratories (Chandler et al. 1958, Gazis et al. 1959 1961). According to this framework drivers react to stimuli from the environment. The response (acceleration) they apply is lagged to account for reaction time:

\[
response_n(t) = sensitivity_n(t) \times stimulus_n(t - \tau_n)
\]

(1)

t is the time of observation. \(\tau_n\) is the reaction time for driver \(n\).

The GM models assume that the car following stimulus is the leader relative speed (the speed of the leader less the speed of the subject vehicle). Several models, which differ in the specification of the sensitivity term, were developed. The non-linear GM model (Gazis et al. 1961) is the most general of these models:

\[
a_n(t) = \alpha \frac{V_n(t)^\beta}{\Delta X_n(t - \tau_n)} \Delta V_n(t - \tau_n)
\]

(2)

\(a_n(t)\) and \(V_n(t)\) are the acceleration and speed of the subject vehicle, respectively. \(\Delta X_n(t - \tau_n)\) and \(\Delta V_n(t - \tau_n)\) are the spacing between the subject vehicle and the leader and the leader relative speed, respectively. \(\alpha\), \(\beta\) and \(\gamma\) are parameters.

The GM models were initially estimated with trajectory data using correlation analysis (e.g. Chandler et al. 1958, Gazis et al. 1959). Other studies (e.g. May and Keller 1967) estimated the model parameters with macroscopic data and using the steady state speed-density equations derived from the model. However, reaction time is neglected with this
approach. More recently, Ozaki (1993) applied a sequential estimation procedure to estimate two sets of parameters, one for acceleration and the other for deceleration decisions. First, reaction times were estimated using regression analysis. Reaction time observations were approximated based on the time lag from when the subject becomes slower (faster for deceleration) than the leader to the beginning of the acceleration (deceleration) manoeuvre and the time lag from when the relative speed reached its maximum (minimum) value to the time the subject applies maximum acceleration (deceleration). He found that reaction times vary between acceleration and deceleration decisions, which may be explained by the activation of brake lights by a decelerating leader. The estimated reaction times were then used in correlation analysis to estimate the other parameters of the car following model. While the sequential estimation procedure is easy to implement is it statistically inefficient compared to simultaneous estimation using maximum likelihood methods. Furthermore, it is not clear that the time lags Ozaki identified correctly represent reaction times.

2.1.2 Extensions to the GM model

Over the years, several extensions and variations of the GM model were proposed.

**Acceleration and deceleration asymmetry.** Herman and Rothery (1965) noted that vehicles’ acceleration and deceleration capabilities are different. Therefore, they used different sets of parameters for the two decisions. The different sets may also account for differences in drivers’ alertness to an increase in the relative leader speed as opposed to a decrease in it. Estimation results (Treiterer and Myers 1974, Ozaki 1993, Ahmed 1999, Toledo 2002) support this distinction, which is also made in several traffic simulation implementations.
**Memory functions.** Lee (1966) proposed a model that hypothesizes that drivers react to the leader relative speed over a period of time rather than at an instant. The mathematical model is:

\[
a_n(t) = \int_0^t M(t-s) \Delta V_n(s) \, ds
\]

(3)

\( M(\ ) \) is a memory (or weighting) function, which represents the way drivers process information that they receive over time.

Lee proposed several forms of the memory function and analyzed the stability of the resulting response to periodic changes in the leader speed. Darroch and Rothery (1972) empirically estimated the shape of the memory function using spectral analysis. They found that a Dirac-delta function, which corresponds to the standard GM model, is a reasonable approximation. Implementation of these models in traffic simulation is considerably more complex because of the need to store an array of past conditions for each vehicle. It is therefore not surprising that this approach is not adopted in traffic simulators.

**Multiple car following.** Herman and Rothery (1965) and Bexelius (1968) hypothesized that drivers follow vehicles in front of their leader as well as the immediate leader. They proposed an additive linear model with different sensitivities to each of the leaders:

\[
a_n(t) = \sum_i \alpha_i \Delta V_{n,n-i}(t-t_n)
\]

(4)

\( \Delta V_{n,n-i}(\ ) \) is the relative speed with respect to the \( i^{th} \) nearest leader. \( \alpha_i \) are parameters.
Herman and Rothery (1965) report inconclusive results regarding the effect of the second-nearest leader. While this research direction received little attention in the literature, it could be useful in explaining the impact of the growing numbers of SUVs in the vehicle mix on traffic flow.

2.1.3 Spacing models

Spacing models hypothesize that the driver reacts to the leader spacing rather than to the relative speed. Newell (1961) assumed that the subject’s speed is a non-linear function of the spacing to the leader:

\[ V_n(t) = G_n\left[\Delta X_n\left(t - \tau_n\right)\right] \] (5)

The form of the \( G_n \) function specifies the car following behaviour. Newell (1961) studied the properties of the functional form:

\[ V_n(t) = V_{\text{max}}\left[1 - \exp\left(-\frac{\lambda}{V_{\text{max}}}\left(\Delta X_n\left(t - \tau_n\right) + D\right)\right)\right] \] (6)

\( V_{\text{max}}, \lambda \) and \( D \) are parameters. \( V_{\text{max}} \) and \( D \) can be interpreted as the maximum speed and the minimum space headway, respectively.

The acceleration the driver applies, which can be calculated by taking the derivatives of both sides of the equation above, corresponds to a non-linear GM model, with a sensitivity function that is an exponential function of the spacing. Newell studied the macroscopic properties of this model, but did not attempt to estimate the model parameters.
Kometani and Sasaki (1958, 1959) proposed a model based on the assumption that the subject speed depends on the leader space headway and on the leader speed. They studied the stability of the subject's motion in response to disturbances in the speed of the leader for two specifications: linear and quadratic in the leader speed. The linear formulation results in the following acceleration function, which simplifies to a linear GM model when $\beta = 0$:

$$a_n (t) = \alpha \Delta V_n (t - \tau_n) + \beta a_{n-1} (t - \tau_n)$$

(7)

$a_{n-1} (\cdot)$ is the leader’s acceleration. $\alpha$ and $\beta$ are parameters.

Hanken and Rockwell (1967) and Rockwell et al. (1968) developed a piecewise linear model. The data range was empirically divided in several regions defined by the space headway, leader speed and subject speed. The acceleration the driver applies deviates from the mean acceleration for the relevant region as a function of the deviation of the space headway, leader speed and subject speed from their respective means in the region:

$$a_n (t) = b^i_0 + b_1^i \Delta X_n (t - \tau_n) + b_2^i V_{n-1} (t - \tau_n) + b_3^i V_n (t - \tau_n)$$

(8)

$b^i_0$, $b_1^i$, $b_2^i$ and $b_3^i$ are parameters of the model within region $i$.

An analysis of the variance showed that the nonlinear effects of speed and spacing that the model captures were statistically insignificant, and so did not justify the piecewise specification.
2.1.4 Desired measures models

Several models were developed assuming that drivers try to attain some desired measure. Helly (1961) suggested that drivers respond to both the leader relative speed and the difference between the actual and desired space headway:

\[ a_n(t) = \alpha_1 \Delta V_n(t - \tau_n) + \alpha_2 \left[ \Delta X_n(t - \tau_n) - D_n(t - \tau_n) \right] \]

(9)

\( D_n \) is the desired space headway, which depends on the subject speed.

This model addresses a deficiency of the GM model that if two vehicles travel at the same speed, any value of the spacing between them is acceptable. Bekey et al. (1977) developed a similar model from optimal control considerations. Koshi et al. (1992) proposed a non-linear version of this model. Gabard et al. (1982) implemented it in the SITRA-B simulation model. In their model the desired space headway is given by:

\[ D_n(t - \tau_n) = L_{n-1} + V_n(t - \tau_n)T \]

(10)

\( L_{n-1} \) is the length of the leader vehicle. \( T \) is the desired time headway, which is assumed constant.

Aycin and Benekohal (1998) hypothesize that drivers try to attain a steady state relation with their leader within some time interval. The steady state is characterized by preferred time headways and speeds that are equal to the leader speed. To ensure a continuous acceleration profile, they compute the rate of change in the acceleration for the next simulation time step based on the current spacing, speeds and accelerations of the subject vehicle and the leader using equations of laws of motion. The model was calibrated as
follows: the preferred time headway was computed as the average of observations in
which the absolute relative speed was less than 5 ft/sec. Reaction time was assumed equal
to 80% of the preferred time headway. A driver was assumed to be car following if the
leader space headway was less than 250 ft. These values were selected based on values
found in the literature.

Bando et al. (1995) assumed that the acceleration drivers apply is proportional to the
deviation of their actual speed from a desired speed, which depends on the leader
spacing. Reaction times are ignored. The model is given by:

$$a_n(t) = \alpha \left[ DV_n(t) - V_n(t) \right]$$

(11)

$DV_n(t)$ is the desired speed. The following function was proposed, but no behavioural or
empirical justification has been presented:

$$DV_n(t) = \tanh(\Delta X_n(t) - 2) + \tanh(2)$$

(12)

Treiber et al. (2000) and Treiber and Helbing (2003) assumed that the acceleration is
affected both by the desired speed and the desired minimum space headway. The model
also incorporates the impact of vehicle capabilities, but ignores reaction time:

$$a_n(t) = a_{\text{max}} \left[ 1 - \left( \frac{V_n(t)}{DV_n(t)} \right)^\beta - \left( \frac{D_n(t)}{\Delta X_n(t)} \right)^2 \right]$$

(13)

$a_{\text{max}}$ is the maximum acceleration. $\beta$ is a parameter. The desired minimum space
headway is given by:

$$D_n(t) = \Delta X_{\text{jam}} + V_n(t) T_n(t) + \frac{V_n(t) \Delta V_n(t)}{2 \sqrt{a_{\text{max}} a_{\text{conf}}}}$$

(14)
\( \Delta X_{\text{jam}} \) and \( a_{\text{conf}} \) are the vehicle spacing at a standstill and the comfortable deceleration, respectively.

Addison and Low (1998) and Low and Addison (1998) proposed a model that combines a desired spacing term with the traditional GM term:

\[
a_n(t) = \alpha_1 \frac{V_n(t)^m \Delta V_n(t - \tau_n)}{\Delta X_n(t - \tau_n)} + \alpha_2 \left( \Delta X_n(t - \tau_n) - D_n(t - \tau_n) \right)^3
\]  

(15)

The desired space headway is given by:

\[
D_n(t - \tau) = \lambda V_n(t - \tau)
\]  

(16)

They studied the stability properties of the solution for a platoon of vehicles subjected to a periodic perturbation for a range of parameter values and showed that chaotic traffic behaviour may occur for some parameter values.

The main difficulty with desired measures models is that the latent nature of the desired measures makes their estimation more challenging. Most of the models described above were not empirically estimated with real-world traffic data.

2.1.5 Psycho-physical models

The psycho-physical model (Weidmann 1974, Leutzbach 1988) addresses two unrealistic implications of the GM models: the model assumes that drivers follow their leader even when the spacing between them is large, and that they perceive and react to small changes in the stimuli. Psycho-physical models introduced perceptual thresholds, which define the minimum value of the stimulus the driver will react to. The values of these
thresholds increase with the space headways, and so, capture both the increased alertness of drivers at small headways and the lack of following behaviour at large headways. Under these assumptions, a vehicle that travels faster than its leader will get closer to it until the deceleration perceptual threshold is crossed. The driver will then decelerate in an attempt to match the leader speed. However, the driver is unable to do this accurately and the headway will increase until the acceleration threshold is reached. The driver will again accelerate and so on. This model is able to explain the oscillating phenomenon observed in car following experiments. Several extensions, which utilize additional perceptual thresholds and zones to more accurately capture acceleration behaviours, such as free-flow, have been proposed (e.g. Ludmann et al. 1997, Schulze and Fliess 1997, Fancher and Bareket 1998, Kourjanski and Misener 1998, Misener et al. 2000). However, a framework for estimation of the model parameters has not been proposed, and so the model was only partially studied empirically. For example, Brackstone et al. (2002) estimate the perceptual threshold functions, but do not consider other parameters of the model. In most other models, perceptual thresholds are arbitrarily derived from the human factors literature.

2.1.6 Fuzzy logic models

Fuzzy logic models, which hypothesize that drivers can only approximately perceive the stimuli they react to, also attempt to capture the inherent imprecision in human behaviour. The model defines fuzzy states the driver may be in at any time. The values of the various stimuli are related to these fuzzy states through probabilistic membership functions. Rule-based logic determines the response (acceleration) the driver applies in each state. Kikuchi and Chakroborty (1992) use three stimuli to define the fuzzy state:
leader relative speed, space headway and leader acceleration. Brackstone et al. (1997) use the leader relative speed and the ratio of time headway to the desired time headway. The difficulty with these models is the definition of fuzzy sets, which are unobservable, and the calibration of the membership functions. Limited calibration efforts are reported in Brackstone et al. (1997) and Chakroborty and Kikuchi (2003).

2.2 General acceleration models

Gipps (1981) developed the first general acceleration model that applies to both car following and free-flow conditions. The maximum applicable acceleration is determined based on two constraints: the desired speed may not be exceeded, and a safe headway must be kept. The safe headway is the minimum that allows the driver to avoid a collision with the leader, if the leader applies emergency braking. Calculations are based on equations of laws of motion. Vehicle characteristics are captured through the use of maximum acceleration and deceleration values. Benekohal and Treiterer (1988) developed a similar model for the CARSIM model. Several acceleration values are calculated based on various considerations: the maximum and comfortable acceleration capabilities of the vehicle, attaining the desired speed, acceleration from a stopped position, car following and car following with an active collision avoidance constraint. The most constraining of these accelerations is used. Reaction times are randomly distributed in the population. Hidas (2002) also used a similar model structure. The driver calculates the acceleration for several conditions, which include car following, free-flow, lane drops and providing or receiving courtesy yielding.
Yang and Koutsopoulos (1996) developed another general acceleration model. The driver is assigned to one of three regimes based on time headway: emergency, car following and free-flow. In the emergency regime, the driver applies the necessary deceleration to avoid collision with the leader. The car following and free-flow regimes utilize a model developed by Ahmed (1999). The car following component is a generalization of the GM model that allows non-linearity in the stimulus term and different reaction times for the sensitivity and the stimulus:

\[ a_n(t) = \alpha \frac{V_n(t - \xi \tau_n)}{\Delta X_n(t - \xi \tau_n)} k_n(t)^{\delta} \Delta V_n(t - \tau_n)^{\rho} + \varepsilon_n(t) \]  

(17)

\( k_n(t) \) is the density of traffic ahead of the subject. \( \xi \in [0,1] \) is a sensitivity lag parameter. \( \varepsilon_n(t) \) is a normally distributed error term.

In the free-flow regime the driver tries to attain its desired speed:

\[ a_n(t) = \lambda [DV_n(t) - V_n(t - \tau_n)] + \nu_n(t) \]  

(18)

\( \lambda \) is a constant sensitivity term. \( \nu_n(t) \) is a normally distributed error term. \( DV_n(t) \) is the desired speed, which is a function of explanatory variables. Estimation results showed that the desired speed is affected by the vehicle type, macroscopic traffic characteristics (e.g. density) and the leader speed.

Both the time headway threshold and the reaction time are modelled as random variables. The parameters of all components of the model were jointly estimated using maximum likelihood estimation with data of individual vehicle trajectories. Zhang et al. (1998) also
implemented a model that defines multiple driving regimes: emergency, normal car following, uncomfortable car following and free-flow. Emergency braking is applied to avoid collision and bounded by the capabilities of the vehicle. Normal car following uses the non-linear GM model. A normal deceleration is used in uncomfortable car following, which applies when the normal car following model yields a positive acceleration and the headway is unsafe based on Pipes’ definition. Normal accelerations and decelerations are also applied in the free-flowing regime in order to attain the desired speed. In addition, drivers in mandatory lane changing situations may adapt their acceleration in order to be able to accept available gaps. The subject accelerates (decelerates) if either the total length of the adjacent gap is sufficient but the lag (lead) gap is too small or the total length of the adjacent gap is unacceptable but the gap between the lead (lag) vehicle and its leader (follower) is acceptable. However, the details of these models were not described.

Toledo (2002) developed a model that integrates acceleration and lane changing and allows drivers to accelerate in order to facilitate lane changing. The model assumes that drivers that wish to change lanes but reject the available gap select a target gap in traffic that they plan to merge into within a few seconds. Different acceleration models apply depending on the target gap choice. The specification of these models follows the GM stimulus-response framework. The stimulus drivers respond to, if unconstrained by their leader, is the distance between their current location and a desired position relative to the target gap. The acceleration sensitivity depends on the subject relative speeds with respect to the vehicles that define the target gap. These acceleration models were
estimated, jointly with the other components of the model, using trajectory data. Estimation results indicate that target gap accelerations differ significantly from those applied in other situations.

2.3 Cellular automata (CA) models

The models presented above have evolved to exhibit increased detail in order to capture various aspects of drivers' behaviour. However, this also leads to increased computational requirements. CA models, which use a minimal set of driving rules, and are therefore computationally very efficient, have been used for large-scale dynamic traffic modelling applications, such as transportation planning and real-time modelling. CA models are based on a discrete representation of both time and space. Road lanes are split into cells of equal size. Cell sizes (typically ~7.5 meters long) are chosen such that each vehicle occupies a single cell. The model describes the movement of vehicles from cell to cell in each time step (typically 1 second). Thus, speed can only assume a limited number of discrete values. While CA models may produce unrealistic behaviour at the microscopic level, the equivalency of their macroscopic behaviour to various macroscopic models has been demonstrated. A simple CA model (Biham et al. 1992) is formulated as follows:

$$V_n(t+1) = min\left(V_n(t) + 1, g_n(t), V_{max}\right)$$  \hspace{1cm} (19)

$V_n(t)$ is measured in integer number of cells. $g_n(t)$ is the number of open cells in front of the vehicle. $V_{max}$ is the maximum number of cells a vehicle can travel in a single time step.
Vehicles' positions are updated synchronously, i.e., all vehicles are moved simultaneously after their new speeds have been determined. The three terms in the speed function represent acceleration to the maximum speed, car-following and free-flow, respectively. For \( V_{\text{max}} = 1 \) the model corresponds to rule 184 in Wolfram's (1986) classification, and so the model is often referred to as CA-184.

Nagel and Schreckenberg (1992) presented a stochastic traffic CA (STCA) model, which adds randomness to the speed rules in order to capture fluctuations in keeping maximum speed, overreaction in braking and noisy accelerations (Nagel et al. 2003):

\[
\begin{align*}
\hat{V}_n(t+1) &= \min \left( V_n(t) + 1, g_n(t), V_{\text{max}} \right) \\
V_n(t+1) &= \begin{cases} 
\max \left( \hat{V}_n(t+1) - 1, 0 \right) & \text{w.p. } p \\
\hat{V}_n(t+1) & \text{otherwise}
\end{cases}
\end{align*}
\]

\( V_n(t+1) \) is an intermediate value. \( 0 \leq p \leq 1 \) is a random speed reduction parameter.

This model, with \( V_{\text{max}} = 5 \), is implemented in TRANSIMS (Nagel et al. 1998a). Several variations of the STCA model have also been proposed. Examples include a cruise control model that eliminates randomness in keeping maximum speed (Nagel and Paczheski 1995), slow-to-start rules in which vehicles are slower to accelerate from a standstill by requiring larger open space in front or by using different random speed reduction probability for stopped vehicles (Takayasu and Takayasu 1993, Barlovich et al. 1998), time-oriented models (TOCA) in which speed selection is based on time headway rather than open cells (Brilon and Wu 1999), anticipation-based rules that take into account the movement of the vehicle in front (Barrett et al. 1996) and using smaller cells
(e.g. 1.5 meters long) in order to obtain higher speed resolution and account for variability in vehicle lengths (Hafstein et al. 2004).

Another class of CA models are asymmetric stochastic exclusion processes (ASEP). These models differ from STCA in that vehicles are moved sequentially rather than simultaneously. The vehicle to be moved in each step is selected randomly. The simulation clock is advanced after $N$ vehicles are moved, where $N$ is the number of vehicles. Thus, on average each vehicle is moved once in each time step (Nagel 1998).

3 Lane changing models

Lane changing models often incorporate two steps: the lane selection process, i.e., the decision to consider a lane change and the lane choice, and the decision to execute the lane change. In most models, lane changes are classified as either mandatory (MLC) or discretionary (DLC). MLC are executed when the driver must leave the current lane. DLC are executed to improve driving conditions. Gap acceptance models are used to model the execution of lane changes.

3.1 Lane selection

Sparmann (1978) introduced the distinction between the wish to change lanes and the execution of the lane change. He also distinguished between changes to the nearside and to the offside. Changes to the offside are motivated by an obstruction on the current lane (e.g. slow vehicles) and by having better conditions on that lane. Changes to the nearside are motivated by not having obstructions on that lane. The execution of lane changes is determined by the available space in the target lane. The model implements psycho-
physical thresholds on the relative speed and spacing to define obstructions that drivers will respond to. Gipps (1986) presented a model that includes a hierarchy of considerations that determine the necessity and desirability of lane changes. Gipps defines three zones with respect to the distance to the intended turn that the driver needs to make to follow the path: when the turn is close, the driver selects the correct lane, unless it is blocked. When the turn is far away it has no effect on the lane selection. The driver then selects a lane by comparing the acceptable lanes with respect to, in order of importance, downstream lane blockages, lane use restrictions, locations of obstructions, presence of heavy vehicles and potential speed gain. The same rules also apply in the middle zone, when the intended turn is neither close nor remote, but only lane changes to the turning lanes or lanes that are adjacent to them are considered. The lane selection rules are evaluated sequentially, and so less important considerations are only evaluated if more important ones did not yield a lane choice. This deterministic rule priority system ignores trade-offs among the considerations (e.g. drivers would always avoid lanes with slow moving trucks, even if these lanes offer immediate speed advantage). Zone boundaries are also deterministic, ignoring variability among drivers and inconsistencies in the behaviour of a driver over time. No framework for estimation of the model parameters was proposed.

A similar model is implemented in SITRAS (Hidas and Behbahanizadeh 1999, Hidas 2002). Several reasons, which are evaluated in order of importance, may trigger lane changes: downstream turning movements, lane drops, lane blockages, lane use restrictions, speed advantage and queue advantage. Downstream turning movements, lane
drops and lane blockages may trigger MLC close to the point the lane change must be completed and DLC in the middle zone. Zone boundaries are the same as in Gipps’ model. MLC are also initiated by vehicles travel are in lanes they are not allowed to use (e.g. bus lanes). An attempt to obtain speed advantage or queue advantage, which are defined as the adjacent lane allowing faster travelling speed or having a shorter queue, will only result in DLC. A need for an MLC, for any reason, would override the priority system and terminate the lane selection process. However, the mechanism to resolve conflicting needs (e.g. when it is desirable to move in one direction for a turning movement, but in the other direction for speed advantage) is not described. Model parameters were not rigorously calibrated and no framework to perform this task was proposed.

CORSIM (Halati et al. 1997) also uses the classification of MLC or DLC. An MLC is executed by drivers that merge into a freeway, move to the correct lane to complete an intended turn or avoid a lane blockage or a lane drop. A DLC is executed when the driver perceives that driving conditions in the target lane are better, but a lane change is not required. The model considers three levels of decision-making: motivation, advantage and urgency. Drivers are motivated to change lanes when their speeds or leader headways drop below a tolerable threshold. The lane change advantage captures the benefits of moving to another lane and depends on the travel speeds and queue lengths in the two lanes. The lane change urgency depends on the number of lane changes required and the distance to the point where the lane change must be completed. The urgency factor affects gap acceptance decisions. A similar distinction between MLC and DLC is made in
MITSIM (Yang and Koutsopoulos 1996). Drivers execute MLC in order to connect to the next link on their path, bypass a downstream lane blockage, avoid entering a restricted-use lane and comply with lane use signs and variable message signs. The probability of initiating an MLC depends on the distance to the point the lane change must be completed, the number of lane changes required and traffic density. DLC are considered when the speed of the leader is below the subject’s desired speed. The driver then compares traffic conditions on the current and neighbouring lanes to select the desired lane. Unlike previous models, lane selection is based on a random utility model, which captures trade-offs between the various factors affecting this choice (e.g. speed advantage, presence of heavy vehicles and merging traffic). Ahmed et al. (1996) and Ahmed (1999) developed a general utility-based framework that captures both MLC and DLC situations. The lane changing process is modelled with three steps: a decision to consider a lane change, choice of a target lane and acceptance of gaps in the target lane. If an MLC situation does not apply or the driver chooses not to respond to it, a decision whether to consider a DLC is made using a two-step process: First, drivers examine their satisfaction with driving conditions in the current lane, which is affected by the difference between the subject speed and its desired speed. The model also captures differences in the behaviour of heavy vehicles and the effect of the presence of a tailgating vehicle. If the driving conditions in the current lane are not satisfactory, the driver evaluates conditions in neighbouring lanes and in the current lane in order to choose the target lane. The utilities of neighbouring lanes are affected by the speeds of the lead and lag vehicles in these lanes and the current and desired speed of the subject vehicle. A gap acceptance model is also included within the lane changing framework. In model estimation, the
choice to react to an MLC situation was not considered. Instead, parameters of the DLC and MLC component models were estimated separately. Gap acceptance parameters were estimated jointly with the other components for each case. The estimation was based on a maximum likelihood approach which used vehicle trajectory data and accounted for correlations among observation from the same driver caused by unobserved driver/vehicle characteristics. The DLC model was estimated with data collected from a freeway section. However, only trajectories of vehicles that changed to the offside lanes were used to guarantee that the lane changes are discretionary. The MLC model was estimated for vehicles merging from a ramp. Zhang et al. (1998) use similar definitions of MLC and DLC and the gap acceptance logic. The authors validated the model but did not suggest a framework for its calibration.

The above models assume a hierarchical structure of MLC and DLC with strict precedence of the first over the latter. Therefore, these models do not capture trade-offs between mandatory and discretionary considerations. Moreover, these models require knowledge of whether a vehicle is in an MLC situation or not. However, except for very special cases, the emergence of MLC situations is unobservable, and so appropriate models have not been estimated. To overcome these limitations, Toledo et al. (2003) presented a lane changing model that integrates mandatory and discretionary considerations in a single utility function for each lane. Lane choice and gap acceptance parameters were estimated jointly using observations of all vehicles on a freeway section. Estimation results indicate that explanatory variables related to the path plan, such as the distance to the intended exit and number of lane changes required to be in the correct lane
are important in lane selection and that significant trade-offs exist between mandatory and discretionary considerations.

Wei et al. (2000) developed a set of deterministic lane selection rules for drivers that turn into two-lane urban arterials and their subsequent lane changing behaviour based on observations from Kansas City, Missouri. Lane selection is determined by the location and direction of intended downstream turns. Drivers that intend to turn at the next intersection choose the correct lane. Drivers that intend to turn farther downstream choose the correct lane if it is the closest to the side they are entering the arterial from. If the correct lane is the farthest, lane choice is based on the aggressiveness of the driver. Drivers’ lane changing behaviour in the arterial is influenced by a similar set of rules. Analysis of the field observations showed that passing is an important behaviour that needs to be modelled. Vehicles already in the correct lane may undertake a passing manoeuvre (double lane change to the other lane and back) in order to gain speed. The model requires that both the adjacent gap in the other lane and the gap in the current lane between the subject’s leader and its leader be acceptable for passing to take place.

### 3.2 Gap acceptance

Gap acceptance models were initially developed to explain intersection crossing behaviour. They are also used in lane changing models, where drivers evaluate the gaps between the lead and lag vehicles in the target lane. Gap acceptance models are formulated as a binary choice problem. Drivers compare the available gap with an unobserved critical gap in order to either accept or reject it:
\[ Y_n(t) = \begin{cases} 1 & \text{if } G_n(t) \geq G_n^{*} (t) \\ 0 & \text{if } G_n(t) < G_n^{*} (t) \end{cases} \] (21)

\( Y_n(t) \) is the choice indicator variable with value 1 if the gap is accepted and 0 otherwise.

\( G_n(t) \) is the available gap and \( G_n^{*}(t) \) is the critical gap.

Critical gaps are modelled as random variables. Herman and Weiss (1961) assumed an exponential distribution, Drew et al. (1967) assumed a lognormal distribution and Miller (1972) assumed a normal distribution. Daganzo (1981) proposed a framework to capture critical gap variation in the population as well as in the behaviour of a single driver over time. He used a multinomial probit formulation appropriate for panel data to estimate parameters of the distribution of critical gaps. Mahmassani and Sheffi (1981) assumed that the mean critical gap is a function of explanatory variables, and so could capture the impact of various factors on gap acceptance behaviour. They estimated the model for a stop-controlled intersection and found that the number of rejected gaps (or waiting time at the stop line), which captures drivers’ impatience and frustration, has a significant impact on critical gaps. Madanat et al. (1993) used total queuing time to capture impatience. Cassidy et al. (1995) differentiated lags (the first gap) from subsequent gaps, and gaps in the near lane from gaps in the far lane. These variables significantly improved the fit of the model. Other parameters that may affect critical gaps include the type of manoeuvre, speeds of vehicles in the major road, geometric characteristics and sight distances, type of control in the intersection, presence of pedestrian, police activities and daylight conditions (e.g. Brilon 1988, 1991, Adebisi and Sama 1989, Saad et al. 1990,
Hamed et al. 1997). However, most of the discussion is qualitative and addresses macroscopic characteristics rather than microscopic drivers’ behaviour.

In the context of lane changing, Gipps (1986) assumed that drivers consider the lead gap and the lag gap separately, and that both gaps must be acceptable. Gaps are evaluated in terms of the deceleration required by the subject vehicle in order to follow the new leader and by the new lag to follow the subject vehicle. The required decelerations are acceptable if they are smaller than a threshold, which reflects vehicle capabilities and the urgency of the lane change. Kita (1993) estimated a logit gap acceptance model for the case of vehicles merging to a freeway from a ramp. He found that important factors are the length of the available gap, the relative speed of the subject with respect to mainline vehicles and the remaining distance to the end of the acceleration lane. Ahmed et al. (1996), within the framework of the lane changing model described above, assumed that both the lead and lag gaps must be accepted. The critical gap functional form guarantees that it is always non-negative:

\[
G_{n, g}^{cr, s}(t) = \exp\left(X_n^g(t)\beta^g + \alpha^g \nu_n + \varepsilon_n^g(t)\right) \quad g = \text{lead, lag} \tag{22}
\]

\(X_n^g(t)\) and \(\beta^g\) are vectors of explanatory variable and the corresponding parameters. \(\nu_n\) is a normally distributed individual specific random term that captures correlations among the decisions made by the same driver. \(\alpha^g\) is the parameter of \(\nu_n\). \(\varepsilon_n^g(t)\) is a normally distributed generic random term.

The model allows different gap acceptance parameters for DLC and MLC situations. Gap acceptance parameters were estimated jointly with other components of the model. Lead
and lag critical gaps under MLC situations were lower than under DLC situations. A similar formulation was also used in Toledo et al. (2003).

In heavily congested traffic conditions acceptable gaps may not be available. Ahmed (1999) developed a forced merging model for such situations, which assumes that drivers change lanes either through courtesy yielding of the lag vehicle or by forcing the lag vehicle to slow down. Important factors affecting this behaviour include the lead relative speed, the remaining distance to the point the lane change must be completed and existence of a total clear gap in excess of the subject vehicle length. Hidas and Behbahanizadeh (1999) and Hidas (2002) proposed an MLC model, which captures cooperation between the subject and lag vehicle in heavy congestion. The willingness of lag drivers to allow the subject vehicle to change lanes depends on their aggressiveness. Once the cooperation is established, the subject will start following the intended leader, and the lag will follow the subject. As a result, a gap will open in the target lane and the subject will be able to change lanes. In addition, Hidas (2005) distinguishes between cooperative lane changing as described above and forced lane changing, in which the subject forces the lag vehicle to decelerate.

3.3 Cellular automata (CA) models

CA models that incorporate lane changing behaviour have also been developed to model multilane traffic flow. The models incorporate conditions that capture the incentive and the safety of lane changing. A typical set of conditions (Rickert at al. 1996) is:
\[ g_n(t) < \min(V_n(t) + 1, V_{\text{max}}) \]
\[ g_{n,o}(t) > \min(V_n(t) + 1, V_{\text{max}}) \]
\[ g_{n,ob}(t) > V_{\text{max}} \]

\( g_n(t) \), \( g_{n,o}(t) \) and \( g_{n,ob}(t) \) are the number of open cells in front of the vehicle in the current lane and in the other lane, and behind vehicle in the other lane, respectively.

The first two conditions above verify that the driver's speed is constrained in the current lane and that the other lane provides better conditions. The third condition guarantees that space is available to lane change. If all conditions are met, the lane change will occur with some probability. Lane changing conditions, such as these, may be symmetric or asymmetric, i.e., different for the nearside and the offside. Nagel et al. (1998b) present a detailed summary of the various lane changing rules and their properties.

### 4 Challenges and research directions

The current emphasis in driving behaviour modelling is on improving the realism of models in order to increase the fidelity of microscopic traffic simulations not only at the macroscopic level but also at the microscopic level, which is increasingly important for applications such as safety and emissions. This goal may be facilitated in two major directions: increase the level of detail in the specification of models to better capture the complexity and sophistication of human decision-making processes, and improve the quality of data and the rigor of estimation of these models.

#### 4.1 Model specification

Some directions for improvement of the specification of state-of-the-art models are:
**Driving regimes.** Driving behaviour models incorporate an increasing number of driving regimes and situations to represent drivers' behaviour in different conditions. Acceleration models are evolving from simple car following to multi-regime models that also incorporate free-flow acceleration as well as various car following sub-regimes (e.g. acceleration and deceleration or reactive and non-reactive car following regimes). Lane changing models incorporate forced merging and courtesy yielding in addition to traditional MLC and DLC. This trend is likely to continue so that models would represent a more comprehensive set of behaviours real-world drivers may apply. The Introduction of multiple driving regimes requires definition of boundaries to determine which behaviour is active. For example, headway thresholds are used to determine whether a vehicle is in car following or free-flow, and the conditions that trigger MLC or forced merging are specified using various zones. However, often the values of these boundaries were either set arbitrarily or calibrated using ad-hoc procedures. These values are often used deterministically, ignoring the heterogeneity in the driver population. New models need to not only explain behaviours in multiple regimes, but also capture the boundaries and transitions between these regimes. Therefore, improved specifications and estimation approaches, which consider regime boundaries as random variables and calibrate their distributions jointly with the other parameters of the models are needed.

**Strategic pre-positioning.** Driving is a hierarchical process, which involves several levels of performance (Koppa 1997). Drivers make strategic trip planning and navigation decisions, such as selecting the trip schedule and path. These decisions affect their driving behaviour: drivers must prepare to be in the correct lanes to follow their path; the trip schedule affects desired speeds. It has been shown that the path plan is an important
factor affecting lane selection (e.g. Wei et al. 2000, Toledo et al. 2003). However, the effect of the trip schedule has not been incorporated in existing models. Furthermore, pre-positioning occurs in other cases as well. For example, drivers may avoid the nearside freeway lane to minimize interactions with weaving traffic, prefer specific lanes in urban arterials to avoid delays caused by turning traffic or avoid following a bus making stops.

**Extended field of view.** In most cases, existing models explain driving behaviour only using variables related to the immediate driving neighbourhood of the subject, such as the relative speeds and positions with respect to surrounding vehicles. However, drivers’ reaction to surrounding vehicles may also depend on their perception of the broader traffic conditions. For example, drivers may be slower to respond to a leader that accelerates to speeds that are above prevailing travel speeds, under the perception that any speed gain from closely following the leader cannot be sustained. Critical gap values may depend on traffic conditions in a similar way. Drivers may be willing to take higher risks and accept shorter gaps when traffic is denser realizing that larger (and safer) gaps are less likely to be available. The impact of these variables may be important. For example, Ahmed (1999) significantly improved the fit of the GM car following model by introducing traffic density as an explanatory variable.

**Inter-dependencies.** In order to model more sophisticated driving behaviour it is necessary to account for inter-dependencies among the various decisions drivers make, both over time and across decision dimensions. For example, acceleration behaviour may be affected by lane changing or intersection crossing decisions. Work in this direction has been done by Zhang et al. (1998) and Toledo (2002), who considered the effect of lane changing on acceleration behaviour and by Toledo et al. (2003), who captured trade-offs
between MLC and DLC considerations. However, most existing models still address various behaviours separately and independently.

**Planning and anticipation.** Drivers are able to anticipate the behaviours of other drivers and systems they interact with, and based on that, plan their actions ahead of time. However, most existing models assume that drivers make instantaneous decisions based on current or past conditions. Toledo (2002) captures drivers’ planning capabilities by assuming that drivers may conceive an action plan to execute a lane change. However, the details of the action plan are not made explicit since it is only defined by the gap in traffic the driver plans to merge into. In the context of intersection gap acceptance, Pollatschek et al. (2002) hypothesize that drivers anticipate the expected delay that they would incur if they reject the available gap, and use it in making their decision.

### 4.2 Model estimation and data requirements

Rigor in estimation of driving behaviour models requires that all parameters of the model, including reaction times and various thresholds be estimated jointly. Most models proposed in the literature, particularly lane changing models, were not estimated this way. Furthermore, some of the published estimation results are limited to simplified special cases. For example, Ahmed (1999) estimates parameters of an MLC model using data of vehicles merging from an on-ramp to a freeway, which may limit the applicability of the model. Furthermore, the estimation is separate from that of the DLC model. In many cases, deficiencies in model estimation result from limited availability of appropriate data. Estimation data needs to cover the important variables that may affect driving behaviours, which may be broadly classified into the following categories:
Neighbourhood variables, which describe the subject vehicle and its relations with surrounding vehicles, such as the subject speed, relative speeds and spacing with respect to the vehicles in front and behind it and lead and lag vehicles in adjacent lanes, the presence of heavy vehicles, and variables that capture prevailing traffic conditions, such as measures of densities and average speeds and their distributions by lane.

Trip plan variables, which capture the effect of the path plan and trip schedule. These may include distances to points where drivers must be in specific lanes to follow their path, the number of lane changes required to get to the correct lanes, indicators of whether the driver needs to take the next exit (or turn at the next intersection), whether the driver is ahead or behind schedule and so on.

Driving style and capability variables, which capture the individual characteristics of the driver, such as aggressiveness, reaction time and vision (e.g. sight distances), and of the vehicle, such as speed and acceleration capabilities. These characteristics normally not directly observed. However, their effects may still be captured with appropriate specification of the models, such as introduction of individual-specific effects that capture correlations between the various decisions drivers make.

Trajectory data, which consists of observations of the positions of vehicles at a high resolution of time (typically one second or shorter), provides useful information about some of these variables. Speeds, accelerations, lane changes and variables that capture the relations between the subject and other vehicles (e.g. relative speeds, time and space headways, lengths of gaps in traffic) can be extracted from the time series of positions. A wide range of technologies, such as aerial photography, video, GPS and cellular location
technologies have been utilized to collect trajectory data. Collection systems may be either fixed or moving. In a fixed system, a road segment is equipped with sensing systems, most often video based, which record the positions of all vehicles in the section. Moving systems include instrumented vehicles, which are equipped with sensing systems that record the position of the subject and its surroundings. Neither of these configurations is currently capable of collecting the complete set of data required for model estimation. A fixed collection system can provide information about the position of the subject vehicle and its relations with other vehicles. However, most available datasets only cover short road segments, up to 300-400 meters long. Thus, extended view variables, such as downstream densities and speeds cannot be accurately calculated. Geometry, weather, surface conditions and other similar factors are uniform within a short section, and so, their effects on driving behaviour cannot be captured. More significantly, only limited information about the effect of the path plan may be obtained since the path is not observed. In addition, no information about the driver and only limited information about the vehicle (e.g. length and width) are available. To overcome some of these limitations, datasets should be collected from longer sections with more versatile geometric characteristics. Furthermore, most of the currently available data was collected in freeways. Data in urban streets, in which other factors such as signals and signs, bus traffic and interactions with pedestrians may affect drivers’ behaviour, need to be collected.

Data collected by instrumented vehicles may alleviate some of the deficiencies of a fixed system. Observations of complete trips allow the effects of the path plan and trip
schedule to be captured. The effects of different road facilities and geometric designs may also be estimated. Driver and vehicle characteristics are also directly observed. However, these data currently only provide partial information about the driving neighbourhood. In many cases only the vehicle in front of the subject is observed. Therefore, while instrumented vehicles are a promising source of rich trajectory datasets, they are still unable to produce some of the fundamental variables required (e.g. lane changing behaviour requires observation of the vehicles in the adjacent lanes).

5 Summary

This paper presents a review of the state-of-the-art in the main areas of driving behaviour research: acceleration, lane changing and gap acceptance. Overall, the main limitation of current models is that in many cases they do not adequately capture the sophistication of human drivers: they do not capture the inter-dependencies among the decisions made by the same drivers over time and across decision dimensions, represent instantaneous decision making, which fails to capture drivers’ planning and anticipation capabilities and only capture myopic considerations that do not account for extended driving goals and considerations. More reliable traffic simulations require development of models that better capture these aspects.

Most of the driving behaviour models proposed in the literature were not estimated rigorously. In many cases, this was due to the limited availability of detailed trajectory data, which is required for model estimation. However, advances data collection technology and increased availability of detailed trajectory data, make the development and estimation of improved models feasible. In addition to the traditional video and film
methods, several technologies, such as instrumented vehicles and GPS systems have the potential to be useful for this purpose.

6 Acknowledgement

The author is a Horev Fellow supported by the Taub and Shalom Foundations. This work was partially supported by the EC under Marie Curie Grant No. 028526. The author would like to thank the anonymous referees for their comments that helped improve the quality of this paper.

7 References


Kikuchi, C., and Chakroborty, P., 1992, Car following model based on a fuzzy inference system. Transportation Research Record, 1365, 82-91.


Leutzbach, W., 1988, Introduction to the theory of traffic flow (Berlin: Springer-Verlag).


_Journal de Physique I_, 2, 2221-2229.


