Capacity drop through reaction times in heterogeneous traffic

Calvert, Simeon C.; van Wageningen-Kessels, Femke L.M.; Hoogendoorn, Serge P.

DOI
10.1016/j.jtte.2017.07.008

Publication date
2018

Document Version
Final published version

Published in
Journal of Traffic and Transportation Engineering (English Edition)

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jtte

Original Research Paper

Capacity drop through reaction times in heterogeneous traffic

*Simeon C. Calvert**, *Femke L.M. van Wageningen-Kessels*, *Serge P. Hoogendoorn*

Department of Transport & Planning, Delft University of Technology, Delft, The Netherlands

HIGHLIGHTS

- Various causes for the capacity drop have been identified.
- Reaction times in macroscopic models have been insufficiently considered.
- A discrete approach for reaction time induced capacity drop is proposed.
- An experimental case demonstrates the validity of the approach.
- Further research on hybrid causation of capacity drop is recommended.

ARTICLE INFO

Article history:
Received 2 January 2017
Received in revised form
6 July 2017
Accepted 7 July 2017
Available online 5 April 2018

Keywords:
Capacity drop
Traffic modelling
Heterogeneous traffic
Traffic flow

ABSTRACT

The capacity drop forms a major reason why the prevention of congestion is targeted by traffic management, as lower capacities are detrimental to traffic throughput. Various reasons describing the dynamics behind the capacity have been described, however one of these, reaction times, has had less explicit attention when modelling on a macroscopic flow level. In this contribution, a method to include the effect of reaction times for the capacity drop in heterogeneous traffic is proposed. The applied method further overcomes difficulties in including reaction times in a discrete time model through relaxation of the updating process in the discretization. This approach is novel for application in the considered first order approach, which is practise ready, contrary to many other models that propose similar approaches. The combination of the introduced method and the model form a solid development and method to apply the capacity drop based on this causation of the capacity drop. The results of the experiment case showed that the influence of traffic heterogeneity had a limited effect on the severity of the capacity drop, while it did influence the time of congestion onset. The influence of the reaction time on traffic showed greater capacity drop values for greater reaction time settings. The findings showed the method effective and valid, while the model application is also practise ready.

* *Corresponding author*. Tel.: +31 (0)624861512.
E-mail addresses: s.c.calvert@tudelft.nl (S.C. Calvert), F.L.M.vanwageningen-kessels-1@tudelft.nl (F.L.M. van Wageningen-Kessels), S.P.Hoogendoorn@tudelft.nl (S.P. Hoogendoorn).

Peer review under responsibility of Periodical Offices of Chang'an University.
https://doi.org/10.1016/j.jtte.2017.07.008

1. Introduction

Macroscopic modelling has long been able to accurately model traffic flow and capture many traffic related phenomena. However, the capacity drop phenomenon in traffic flow is an important phenomenon that initially was not able to be reproduced in early first order traffic models. And still today, even with growing consensus on the causes of the capacity drop, there remains no one accepted cause or modelling method. This paper focuses on developing a yet limited investigated idea in modelling the capacity drop, which is caused, in part, by the reaction times of drivers to the movement of their predecessors (Kesting and Treiber, 2008). The idea that differences in acceleration and deceleration by drivers can cause the capacity drop is well known, but has not been widely investigated as a function of the reaction of drivers in congestion. This approach is further considered and developed in this paper and implemented in a state of the art first order macroscopic model with consideration of individual microscopic car following characteristics and stochastics. The model is practise ready and the difficulty of including stochastic reaction times in a discrete formulation is addressed.

The capacity drop is defined as the difference between the breakdown capacity and the discharge capacity on a section of road and can be frequently observed after traffic breakdown between observations in a critical under-saturated traffic state and an over-saturated traffic state. The occurrence of the capacity drop was originally signalled by Hall and Agyemang-Duah (1991) and Banks (1991) and was generally attributed to the so called hysteresis effect (Banks, 1991; Hall and Agyemang-Duah, 1991). The hysteresis effect occurs in part due to differing driving behaviour as vehicles enter and exit congested traffic states and is most commonly captured in macroscopic models in second order formulations. In these models, an additional equation is given that describes the dynamics of vehicle flow. There have also been attempts to include the capacity drop in first order model, such as by Laval (2004). The drop in capacity after a breakdown event is obviously detrimental to traffic flow and network performance as it reduces the potential throughput of traffic. For this reason, the prevention of the capacity drop or reduction of its effects is often targeted as an important method to improve traffic throughput within areas such as traffic management. From empirical research, it has become apparent that the capacity drop does not have a single value, but can vary from almost non-existent to values up to 30%, while values ranging from 3% to 15% appear to be most common (Zhang and Levinson, 2004). It has also been demonstrated and argued that the capacity drop, just as capacities themselves, do not hold to a static value, but are also stochastic entities (Calvert et al., 2015b; Lorenz and Elefteriadou, 2001).

Traditionally first order macroscopic models consider traffic flow based on principles laid in the LWR model, first described as a compressible fluid (Richards, 1956; Whitham, 1955), which allows general traffic flow features to be described. The LWR model has long been applied due it is the eloquent description and easy implementation in describing macroscopic traffic flow. Despite their popularity in traffic modelling, basic first order models, do not capture many of the detailed dynamics of traffic flow, such as the interaction between vehicles in various traffic states and therefore do not sufficiently describe phenomena such as kinematic waves (Kerner, 1999) and the capacity drop. This has previously been described in detail and led to the development of second order models (Aw and Rascle, 2000; Daganzo, 1995; Lebacque et al., 2007). Second order macroscopic models do allow the capacity drop to be captured, however often at a cost, such as higher calculation time, or with simplifications, such as presuming homogeneity in traffic flow. Daganzo (1995), among others, described certain flaws of second order models. Furthermore, the simplicity of first order models offers a major advantage over second order models, which has led to the proposal of many extensions for first order models to help capture more traffic dynamics while retaining much of their simplicity (Leclercq, 2007a).

In this contribution, an extension to a first order model is proposed that allows the capacity drop phenomenon to be modelled based on delays in driver reaction times in heterogeneous traffic. In heterogeneous traffic, the aggressiveness of a driver is considered by a drivers' willingness or ability to accelerate at different rates to maintain a headway, which is described in the FOMSA model, a Lagrangian kinematic wave model. In this contribution, we first give a description of the dynamics that lead to the capacity drop and the ways that this has been modelled. In Section 3, we describe the applied model and give a description of the capacity drop inducing components. In Section 4, a demonstration of the method is given, with a discussion and the conclusions given in Sections 5 and 6.

2. Capacity drop

2.1. *Dynamics of the capacity drop*

In traffic flow theory and modelling there are few variables that are as fundamental as road capacity. Road capacity is applied in modelling for the likes of infrastructure planning and the evaluation of traffic measures. Various interpretations of capacity exist; traditionally the capacity of a road is defined as "the maximum traffic flow on a section of road under fluent traffic conditions". In more recent decades, capacity is seen increasingly in relation to the likelihood of a flow value being achieved as "the maximum flow rate that can reasonably be expected to traverse a uniform segment of road under prevailing roadway, traffic and control conditions", as defined in the Highway Capacity Manual. It is well accepted that the capacity of a road can be stochastic and Lorenz and Elefteriadou (2001) define capacity as "the rate of flow along a uniform freeway segment corresponding to the expected probability of breakdown deemed acceptable under prevailing traffic and roadway conditions in a specific direction". These definitions all relate to the breakdown capacity, often also referred to as the free-flow capacity, which indicates the traffic flow in under-saturated conditions. The discovery of a discharge capacity for over-saturated traffic flow by Hall and Agyemang-Duah (1991) and Banks (1991) and later consolidated by empirical evidence by Cassidy and Bertini (1999) among

others, led to the term capacity drop coming into existence. The term discharge refers to the maximum rate that traffic can flow out of, or discharge, from an area with congested over-saturated traffic. The capacity drop is considered as the difference between the under-saturated capacity value and the over-saturated discharge capacity value. At this point it should be evident that the capacity drop can, and maybe should, also be considered as a stochastic variable. If capacity, either under-saturated or discharged, can be stochastic, then the same must also hold for the difference between the two (Calvert et al., 2015b; Cassidy and Bertini, 1999; Srivastava and Geroliminis, 2013; Zhang and Levinson, 2004).

Throughout the start of the new millennium, researchers have attempted to describe and give causes for the capacity drop. Differences in acceleration rates between vehicles and a limitation of acceleration (bounded acceleration) have been widely considered as a possible cause (Coifman and Kim, 2011; Hall and Agyemang-Duah, 1991; Leclercq et al., 2016), while merging at low speeds with increased lane changing has also been considered as a likely cause for many locations (Treiber et al., 2006). General heterogeneity in traffic flow, especially in saturated traffic states is another reason that can be found from literature (Chen et al., 2014; Chung and Cassidy, 2004). A lack of coherent and timely reaction of drivers to traffic conditions has also been mentioned (Kesting and Treiber, 2008), in which drivers are slower to react to accelerating vehicles leaving congestion resulting in a reduction in flow. It is suggested that for most bottlenecks, the cause of the capacity drop can be attributed to a number of various mechanisms, which will vary depending on the type and characteristics of the bottleneck. As there appear to be multiple mechanisms, it is not easy to point specifically to one definitive cause.

2.2. *Previous work on capacity drop modelling*

The complex dynamics that lead to the capacity drop form a distinctive challenge to those aiming to model the phenomenon. As these dynamics and exact interplay between various possible causes of the capacity drop are yet unknown, it is common for models to represent the capacity drop through an exogenous or generalised approach. In first order models, adjustments to the fundamental diagram are an easy and effective way to introduce capacity drop characteristics into traffic flow, as the fundamental diagram directly relates to the flow in both under- and over-saturated traffic states and therefore the difference between these states. Early adaptions of the fundamental diagram can be found in the use of the inverse lambda fundamental diagram. Other shapes have been proposed, such as various discontinuous and bivariate fundamental diagrams. A number of approaches consider the traffic behaviour in merges, such as at ramps and weaving sections, and lane drops where there is a significant increase in lane changes (Chung et al., 2007; Laval and Daganzo, 2006; Leclercq et al., 2011). Within such approaches, the effect of slowing and accelerating traffic can be represented. As traffic re-accelerates after decelerating, it is known that this often occurs at a slower rate than the deceleration. Bounded acceleration was previously described (Calvert et al., 2015a; Lebacque, 2003; Leclercq, 2007b) as an example how to represent this. In the basic kinematic wave models, vehicles may accelerate at an unrealistic speed, which is restricted by bounding the maximum acceleration. The effect is especially visible for acceleration of vehicles from low speeds such as out of congestion. Further approaches make use of second order descriptions of traffic flow that allows the influence of speed dynamics in various traffic states to be described (Siebel et al., 2009). In this paper, we describe the dynamics leading to the capacity drop as a function of the reaction time of drivers. This takes the premise of delayed or limited acceleration to the level of the driver. It is not reasonable to presume that a limitation of acceleration is only due to vehicle or comfort constraints, rather that the inability of drivers to sufficiently and timely react to traffic conditions, especially when discharging from congestion. The influence of reaction times of drivers was previously described by (Kesting and Treiber, 2008) in relation to traffic flow and is investigated as an approach to describe the capacity drop in the following sections of this paper.

3. Model description

3.1. *FOMSA model*

3.1.1. *Lagrangian kinematic wave model as basis for FOMSA*

The applied model in this contribution is the first order model with stochastic advection (FOMSA), which is a Lagrangian formulation of the kinematic wave model (LKWM), and includes a heterogeneous stochastic invariant component that describes vehicle specific differences, such as desired headways for different vehicles (Calvert et al., 2015c). The kinematic wave model (KWM) finds its origins in the work performed by Whitham (1955) and Richards (1956), and their formulation of the so-called LWR model. In the KWM, construction of the kinematic waves is achieved through use of the fundamental relationship of traffic flow. The model further relies on the conservation equation and initial boundary conditions. In this contribution, the FOMSA model is applied, which uses the Lagrangian coordinate system that describes coordinates of particles in a flow explicitly as a function of their speed. Jin et al. (2014) previously argued that Lagrangian coordinates can be incorporated into continuum traffic flow models by either establishing moving boundary conditions for Euler formulations or through the application of hydrodynamic flow. Leclercq et al. (2007) showed that hydrodynamic flow is able to be derived using a space function based on variational theory. More recently Laval and Leclercq (2013) further applied the theory of Hamilton-Jacobi to KWM in which the theory is applied to three two-dimensional coordinate systems, which included Eulerian and Lagrangian systems.

In traditional macroscopic modelling, Eulerian coordinates are usually applied, which state that for a specific time and location, a flow, such as traffic, will pass with certain characteristics (Helbing and Treiber, 1999; van Wageningen-Kessels et al., 2010). In this case, it is the flow which moves in relation to the coordinate system. Lagrangian coordinates in contrast are not fixed in space, but can transform with the resulting flow. The time coordinates on the y-plain are set at

static intervals, while the flowing particles are explicitly considered in individual consecutive cells and are assigned a location depending on the speed of traffic flow. Therefore, the coordinates follow the flow rather than the flow following the coordinates. When describing the KWM in Lagrangian coordinates, one still has the same underlying principles that obey the conservation of vehicles and fundamental relations, denoted as a functional of the traffic speed. The conservation equation is given by

$$\partial_t s + \partial_n v = 0 \tag{1}$$

where s denotes the mean space headway of vehicles in a single cell, v denotes the mean speed of vehicles, while n is the vehicles number, which decreases in the driving direction. The fundamental relation in Lagrangian coordinates makes use of the speed v in relation to the density ρ, which is derived from the mean headway spacing $s = 1/\rho$. The fundamental relation is given as

$$v = V(s) \tag{2}$$

where V is the shape of the fundamental diagram.

The use of Lagrangian coordinates has been proven to lead to more accurate results as the error due to numerical diffusion is much less than in the traditionally applied Eulerian coordinates (Leclercq, 2007b; van Wageningen-Kessels et al., 2010). This can be clearly seen from an example of the same flow depicted in Fig. 1 for Eulerian (on the right) and Lagrangian (on the left). The fundamental relation in Lagrangian coordinates makes use of the speed v, in relation to the density ρ, derived from mean space headway $s = 1/\rho$. In Lagrangian coordinates, it is possible to explicitly define the number of vehicles per time-space cell. As the KWM is a macroscopic model, this can be any number of aggregated vehicles.

3.1.2. *Stochastic advection in FOMSA*

The use of Lagrangian coordinates is a game changer in the respect that the numerical scheme follows vehicles rather than time, and allows specific characteristics of these vehicles to propagate with the flow. Propagation of information with a vehicle (group) is described as an advection invariant term. Leclercq (2007b) introduced a generic invariant term which allows numerous descriptive variables to be propagated with traffic flow in a second order macroscopic model. In the FOMSA, an invariant term is introduced as a first order Lagrangian model, which retains the relative simplicity of first order modelling approaches. The invariant can describe an arbitrary characteristic of traffic flow and the drivers' specific behaviour in this model, referred to here as the vehicle specific invariant I. It is applied with a conservation equation and in the fundamental relation

$$\partial_t \rho I + \partial_x(\rho v I) = 0 \tag{3}$$

$$v = V(\rho, I) \tag{4}$$

where the invariant, I, is the vehicle specific invariant, a term that denotes a vehicle dependent adjustment factor that directly influences the critical density, ρ, for each vehicle or group of vehicles depending on the level of discretization.

$$\rho_{\text{crit}} = I\rho_{\text{crit.0}} \tag{5}$$

$$\rho_{\max} = I\rho_{\max.0} \tag{6}$$

where ρ_{crit} is the critical density, $\rho_{\text{crit.0}}$ is the deterministic critical density, $\rho_{\max}$ is the jam density, and $\rho_{\max.0}$ is the deterministic jam density.

The values for I are sampled independently from a uniform distribution, which bounds the extremity of possible values of the invariant:

$$I \sim U([1-\alpha, 1+\alpha])$$

where I is a random number between $1-\alpha$ and $1+\alpha$, α is the stochastic boundary parameter which indicates the maximum extent of the stochastic influence. The vehicle specific invariant, I, is assigned to each vehicle or platoon at the entrance of a network. More detailed construction of the invariant term have been proposed in Calvert et al. (2015c), but are omitted here, as they are not of relevance for the question at hand.

3.2. *Capacity drop modelling*

The capacity drop is induced here through the application of the reaction time (T_r) of drivers to downstream speed increases in combination with heterogeneous traffic. As discussed previously, there are many possible causes for the capacity drop. The reaction time of drivers in acceleration is one such cause for the capacity drop (Kesting and Treiber, 2008) and the one that is modelled here. In the discrete model, speeds v are updated at each discrete time step. However, reaction times are generally much shorter than a time step, usually in the vicinity of 1.0 s, and furthermore are not consistent between drivers. The reaction time T_r, is included as the delayed time required to induce a speed altering action. To include the T_r within a time step Δt and allow an update of the speed v and location x, an updated

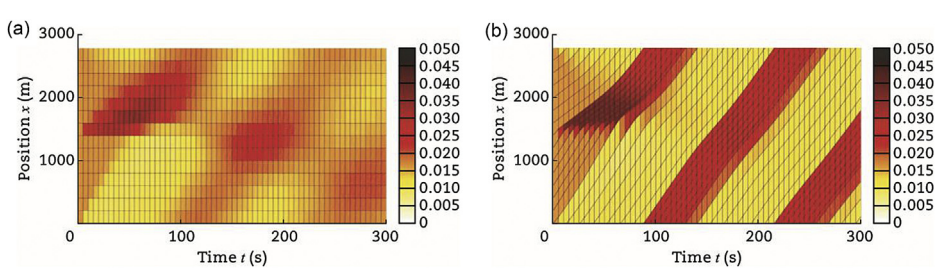


Fig. 1 – Example of the same flow depicted in different systems. (a) Lagrangian system. (b) Eulerian system (van Wageningen-Kessels et al., 2010).

location of the following vehicle is calculated based on the location of a vehicle without reaction time, x^*. The principle is shown in Fig. 2, in which the reactive influence of the car following model is ignored for clarity, $(n+1)^*$ denotes the location of the following vehicle group $(n+1)$ if the reaction time is not considered.

The new location x of a vehicle n, in a time-step t is dependent on the speed of the vehicle as a consequence of the space headway s to the predecessor n. As the space headway is directly correlated to the density of traffic, $s = \frac{1}{\rho}$, the speed of vehicle n is calculated from the prevailing fundamental relation. This is described by

$$v_t^{(n)} = V\left(s_{t-1}^{(n)}, I^{(n)}\right) \tag{7}$$

$$x_t^{(n)} = x_{t-1}^{(n)} + v_t^{(n)}\Delta t \tag{8}$$

where $x_t^{(n)}$ is the location of vehicle (group) n at t, $v_t^{(n)}$ is the speed of vehicle (group) n at t, $I^{(n)}$ is the relative vehicle specific invariant term of vehicle (group) n.

Eq. (7) includes the vehicle specific variant from the applied FOMSA model. For the explanation of the inclusion of the reaction time, this term is not required and is omitted for ease of understanding, to give

$$v_t^{(n)} = V\left(s_{t-1}^{(n)}\right) \tag{9}$$

In the model, speeds are updated using the space headway in Eq. (9). We define an uncorrected space headway without reaction time as s^*, and the corrected space headway with reaction time as s. These are calculated from the difference in locations, with x^* of x respectively, of a vehicle n to its predecessor $n+1$.

$$s_t^{*(n+1)} = x_t^{(n)} - x_t^{*(n+1)} \tag{10}$$

$$s_t^{(n+1)} = x_t^{(n)} - x_t^{(n+1)} \tag{11}$$

when we consider the influence of the reaction time T_r for a single vehicle n, the error made compared by not considering the reaction time in terms of location $x_t^{(n)} - x_t^{*(n)}$, and spacing $s_t^{(n)} - s_t^{*(n)}$, of vehicles is caused by the reaction time during which a vehicle travelled at a higher speed too early in reaction to the leading vehicle $n+1$. The distance covered due to this difference in indicated in Fig. 2 by the black arrow.

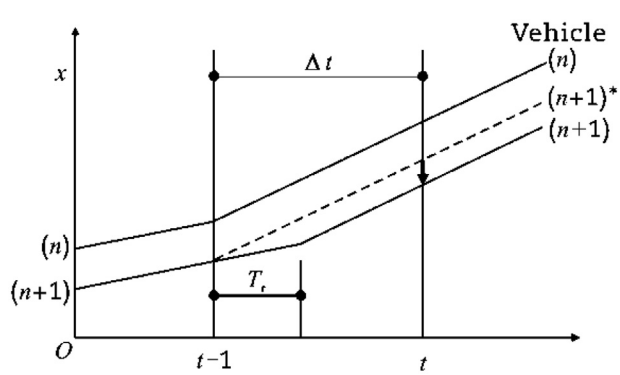


Fig. 2 – Method principle with modified vehicle location and spacing.

Therefore, the correction is applied as the difference in distance that vehicle n should of travelled, given by

$$x_t^{(n+1)} - x_t^{*(n+1)} = s_t^{(n+1)} - s_t^{*(n+1)} = \frac{T_r}{\Delta n}\left(v_{t-1}^{(n)} - v_{t-1}^{(n+1)}\right) \tag{12}$$

where Δn is the vehicle group size.

The correction of the reaction time is only required when the following vehicle has a speed higher than its predecessor. For the case in which the speed of the leading vehicle is higher than the follower, this is a different case, which is not relevant for the capacity drop and for which the model presumes no reaction time. Eq. (12) can be rewritten to give the corrected space headway, which is applied for the following time step, according to Eq. (7).

$$s_t^{(n+1)} = s_t^{*(n+1)} - \frac{T_r}{\Delta n}\left(v_{t-1}^{(n)} - v_{t-1}^{(n+1)}\right) \quad v_t^{(n+1)} > v_t^{(n)} \tag{13}$$

where $s_t^{(n+1)}$ is the space headway of vehicle group $(n+1)$ at t.

This process is repeated over all vehicles within a time step and over all time steps to give the locations, speeds and related traffic metrics of all vehicles in each time step considering bounded traffic flow of the vehicle dynamics with the inclusion of vehicles specific stochastics and reaction times to vehicle speed differences. It should be noted that in a continuous formulation, this would not be a major issue, however the discrete formulation is much easier to apply and is therefore very relevant for use in practise. Furthermore, it is necessary to note that the calculations are performed for a set vehicle group size. The values for the reaction time T_r correspond to the aggregated reaction time of the vehicle group, dependant on the vehicle group size.

4. Method demonstration

4.1. *Case setup*

The effect of T_r is shown on an arbitrary 10 km single bottleneck corridor with a capacity at a critical density of 25 veh/km corresponding to a capacity flow of 2125 veh/h. The applied fundamental diagram is triangular and has the jam density set at 140 veh/h. The bottleneck has a capacity flow set at 70% of the road capacity at a value of 1488 veh/h and is present on the road between 6800 and 8800 m from the start of the corridor. Traffic demand is gradually increased to a value of 2400 veh/h, which is above the bottleneck capacity, and later decreased to show the effect of congestion onset and dissipation and the effect of the capacity drop. The applied time step of the model is 3 s with a vehicle group size of $3\frac{1}{3}$ vehicles. The capacity drop is measured using flows downstream of the bottleneck with a 5 min moving average aggregation. Aggregation is required for the heterogeneous cases to give a good representative quantification of the overall traffic flow values and avoid capturing outliers. T_r values of 0, 0.5, 1.0, and 1.5 s are applied as test scenarios. Traffic heterogeneity is set at 0, 10%, and 20% deviation for the vehicle specific invariant, I, for time headways. The main goal of the case is to demonstrate the ability of the proposed addition to model to capture the capacity drop and to evaluate the sensitivity of the applied parameters in the process.

4.2. Results

4.2.1. Congestion onset and propagation

Each scenario, a combination of a reaction time and traffic heterogeneity, is compared against the reference scenario in which vehicles have no reaction time and are in perfectly homogeneous traffic. The trajectories of this reference scenario are shown in the time–space plot in Fig. 3(a). The plot clearly shows the trajectories of traffic as it approaches the bottleneck location and the triggering of congestion once capacity is reached. The darker colours in the plot show a higher density value, with black showing congestion. Congestion builds up slowly, but rapidly disappears once the traffic demand drops. In contrast, Fig. 3(b) shows the case for a traffic heterogeneity of 10% and a driver reaction time of 1.0 s with all other variables remaining identical to the reference case. The plot shows that congestion occurs sooner than in the reference case, builds up quicker and lasts longer once the traffic demand is reduced. Also notice that the heterogeneity of traffic can be seen from the figure, by the varying darkness of the trajectory lines, which indicate differences in the traffic density and therefore in the following distance of vehicles, which was defined as part of the level of aggressiveness between vehicles.

When we zoom in on the point of traffic breakdown and the onset of congestion (Fig. 4), the differences between the reference and the case with heterogeneity and reaction time delay quickly become apparent. In the reference case (Fig. 4(a)), it is clear that there is a longer period of time in which traffic flow slightly at the bottleneck, before being able to accelerate away from the bottleneck. Later in time, the reduction in speed becomes such that the vehicles pass the point of flow saturation and congestion sets in. Fig. 5(a) shows this process in the fundamental diagram, in which the reduction of speed in the under-saturated region is clear leading up to the point of capacity. From the fundamental diagram, the path back to under-saturated flow can also be

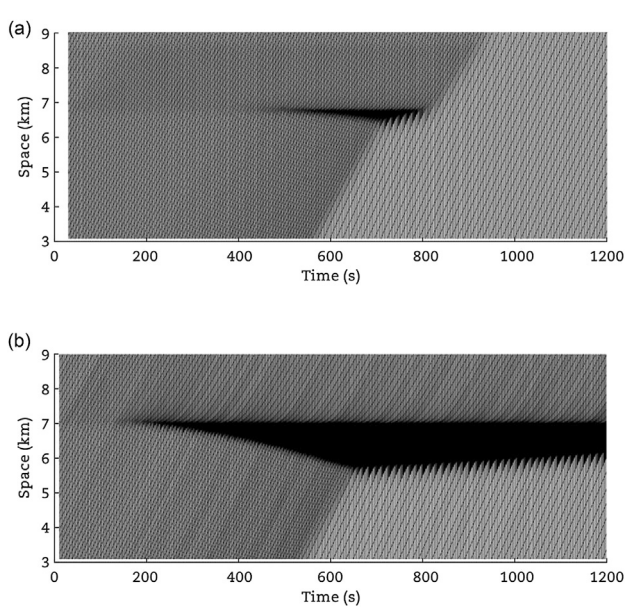


Fig. 3 – Trajectories. (a) The reference case. (b) Case with $I = 10\%$, $T_r = 1.0$ s.

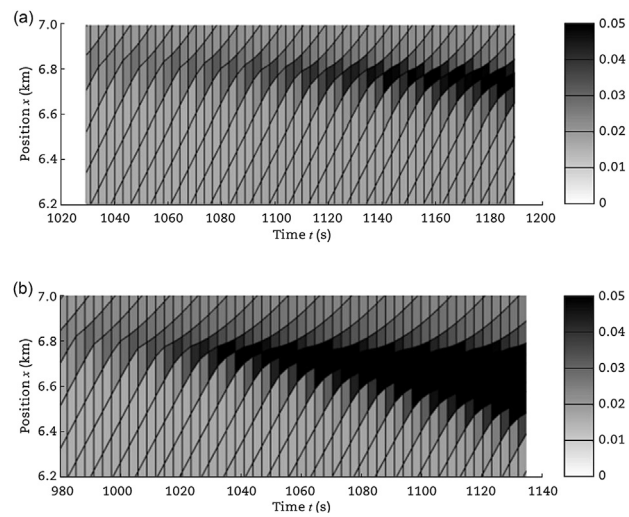


Fig. 4 – Onset of congestion with density (veh/m) (Lagrangian coordinates). (a) The reference case. (b) Case with $I = 10\%$, $T_r = 1.0$ s.

easily observed, with congestion only present for a short time. In Fig. 4(b), the onset of congestion for the 10% heterogeneity and 1.0 s reaction time is shown. There, we see that the time in which traffic starts to slow at the bottleneck to the point at which congestion sets in is much shorter. This is also clear from the fundamental diagram, where a lower capacity level is achieved (Fig. 5(c)). The presence of heterogeneity plays a major role in the quicker onset of congestion as in the lower capacity value. The effect of the different reaction times between the reference and scenario cases is also visible. In Fig. 4(a), vehicles accelerate at a much greater rate out of congestion, therefore allowing following vehicles to do the same, which reduces the severity of congestion and the rate that traffic flow can recover. However, with reaction times (Fig. 4(b)), the vehicles take longer to accelerate out of congestion and therefore cause greater delays upstream.

Further observations from the fundamental diagrams can be made in relation to the capacity drop. In the reference case (Fig. 5(a)), there is no notable capacity drop. Traffic flow quickly decreases and returns to under-saturated flow conditions upon lowering of traffic demand. For the two fundamental diagrams with reaction times of 0.5 and 1.0 s, a clear increase in capacity drop can be observed with increasing reaction time values. Again, once the inflow of traffic is lowered, traffic starts to return to under-saturated conditions, only this takes longer as more congestion has accumulated and the operational capacity is lower. The return to under-saturated conditions is not visible in Fig. 5(c), as congestion lasted until the end of the simulation time, however this is not important for the analysis, as the capacity and discharge flow can already be observed.

4.2.2. Capacity drop

The capacity drop values for the considered T_r and heterogeneity values are shown in Table 1. To accommodate different

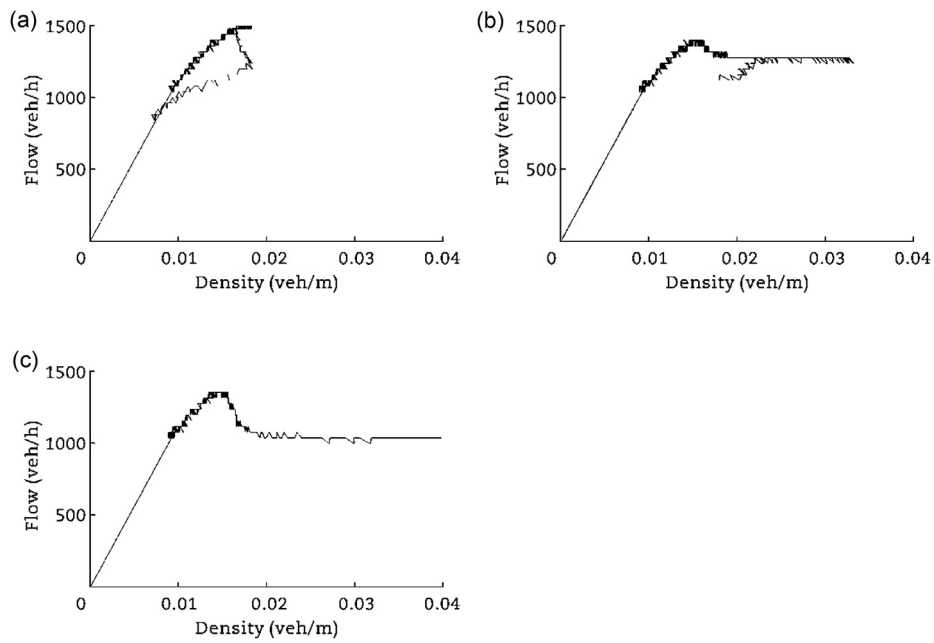


Fig. 5 – Fundamental diagrams. (a) The reference case. (b) $I = 10\%$, $T_r = 0.5$ s case. (c) $I = 10\%$, $T_r = 1.0$ s case.

definitions of the capacity drop and give insight into the results, two different capacity drop values are given for each scenario. The first is the capacity drop values given against the highest flow pre-breakdown for each heterogeneity level. For example, the reaction times for $I = 10\%$ are all compared to the highest flow value (which is the capacity) for the zero reaction time case for $I = 10\%$. For the second capacity drop definition, shown in brackets, a comparison is made compared to the reference case with no heterogeneity or reaction time. In this second approach, the capacity value may be seen as a hypothetical maximum flow that can only be achieved under perfect homogeneous traffic conditions, but is rarely seen when traffic becomes heterogeneous. That is also the reason why the values in brackets are higher, as the comparison is against a higher breakdown capacity value. The difference in breakdown capacity between the scenarios is also visible from Fig. 5.

The results from Table 1 show that higher reaction times give increasingly higher capacity drop values, as is expected, and therefore demonstrate the ability of the proposed method to capture the capacity drop. This trend is the case for all applied levels of traffic heterogeneity. Increased heterogeneity in traffic flow does not show any substantial increase in capacity drop compared to the capacity of the same heterogeneity. An increase in heterogeneity leads to lower capacities for higher levels of traffic heterogeneity. Therefore, when the discharge capacity is compared to the homogeneous case, then higher capacity drops are found for higher levels of heterogeneity, as shown in the brackets in Table 1.

Table 1 – Capacity drop results per case – compared to case capacity and compared to homogeneous maximum capacity in brackets.

Traffic heterogeneity I (%)	Capacity drop (%)			
	$T_r = 0$ s	$T_r = 0.5$ s	$T_r = 1.0$ s	$T_r = 1.5$ s
0	0	9 (14)	23 (27)	40 (43)
10	0	9 (14)	24 (30)	41 (46)
20	0	6 (16)	24 (32)	41 (49)

5. Discussion

The results from the case demonstrate that the application of reaction times to replicate the capacity drop succeeded and the trends that were found are reasonable. The influence of traffic heterogeneity yielded a limited effect on the capacity drop. Traffic heterogeneity did lead to an earlier onset of congestion and therefore a lower breakdown capacity. The results also showed that a larger reaction time led to higher capacity drops. The absolute values that were found for the capacity drop can be seen on the high side, however there is insufficient evidence to state what a correct reaction time should be and therefore further research may need to be performed to indicate if indeed a reaction time for these driving actions is to be expected nearer 0.5 s or 1.0 s. A reaction time of 1.5 s should not be expected to be a realistic values, however was still used to give insights into the effectiveness of the method in extreme cases.

The causes and mechanisms behind the capacity drop have shown in previous work not to be generic to one specific causality and modelling approach. Lane changing, merging at low speeds and heterogeneous lane behaviour have all been mentioned and shown to contribute to some sort of capacity

drop. In this work, we have demonstrated the ability to model the capacity drop due to reaction time delays of drivers. Ideally, research should aim to correctly identify each type of capacity drop and apply an approach that can capture and model each in a single framework. This research gives a contribution to how that may be performed for reaction time delays.

The demonstrated approach may be further improved through improvements in the dynamics behind the application of heterogeneity. At present, the approach is limited to a single variable, but can and maybe should be expanded with a greater degree of behavioural aspects. This can be achieved in part through further collaboration with human factors and empirical research of microscopic driving behaviour. Further improvements may be found in the analysis of reaction times and delays of drivers for various tasks. This may allow the construction of more complex reaction time distributions and in turn allow more accurate modelling of the considered phenomenon.

6. Conclusions

In this paper, a method is proposed to capture the capacity drop in a first order model using driver reaction time in heterogeneous traffic. The model is a Lagrangian formulation of the kinematic wave model with vehicle specific invariant to capture heterogeneity in traffic flow. The capacity drop phenomenon has been shown to have various causes, which have been modelled by various researchers. This paper is one of the first to model the capacity drop based on explicit consideration of reaction time delays in heterogeneous traffic in such a model. The applied method overcomes difficulties in including reaction times in a discrete time model through relaxation in the formulation of the updating process in the discretization, correcting discrete reaction time inaccuracies. The method was demonstrated in a case study to successfully produce the capacity drop phenomenon for different levels of heterogeneity and reaction times. This showed that traffic heterogeneity has a large influence on the time of traffic onset and level of breakdown capacity, while the discharge capacity and therefore the capacity drop are largely affected by the reaction time.

REFERENCES

Aw, A., Rascle, M., 2000. Resurrection of second order models of traffic flow. Siam Journal on Applied Mathematics 60 (3), 916–938.

Banks, J., 1991. The two-capacity phenomenon: some theoretical issues. Transportation Research Record 1320, 83–90.

Calvert, S.C., Snelder, M., Taale, H., et al., 2015a. Bounded acceleration capacity drop in a Lagrangian formulation of the kinematic wave model with vehicle characteristics and unconstrained overtaking. In: IEEE 18th International Conference on Intelligent Transportation Systems, Santa Catalina, 2015.

Calvert, S.C., Taale, H., Hoogendoorn, S.P., 2015b. Quantification of motorway capacity variation: influence of day type specific variation and capacity drop. Journal of Advanced Transportation 50 (4), 570–588.

Calvert, S.C., Taale, H., Snelder, M., et al., 2015c. Vehicle specific behaviour in macroscopic traffic modelling through stochastic advection invariant. Transportation Research Procedia 10, 71–81.

Cassidy, M.J., Bertini, R.L., 1999. Some traffic features at freeway bottlenecks. Transportation Research Part B: Methodological 33 (1), 25–42.

Chen, D., Ahn, S., Laval, J., et al., 2014. On the periodicity of traffic oscillations and capacity drop: the role of driver characteristics. Transportation Research Part B: Methodological 59, 117–136.

Chung, K., Cassidy, M., 2004. Test of theory of driver behavior on homogeneous freeways. Transportation Research Record 1883, 14–20.

Chung, K., Rudjanakanoknad, J., Cassidy, M.J., 2007. Relation between traffic density and capacity drop at three freeway bottlenecks. Transportation Research Part B: Methodological 41 (1), 82–95.

Coifman, B., Kim, S., 2011. Extended bottlenecks, the fundamental relationship, and capacity drop on freeways. Transportation Research Part A: Policy and Practice 45 (9), 980–991.

Daganzo, C.F., 1995. Requiem for second-order fluid approximations of traffic flow. Transportation Research Part B: Methodological 29 (4), 277–286.

Hall, F.L., Agyemang-Duah, K., 1991. Freeway capacity drop and the definition of capacity. Transportation Research Record 1320, 91–99.

Helbing, D., Treiber, M., 1999. Numerical simulation of macroscopic traffic equations. Computing in Science & Engineering 1 (5), 89–99.

Jin, P.J., Han, K., Ran, B., 2014. Some theoretical and practical perspectives of the travel time kinematic wave model: generalized solution, applications, and limitations. In: Transportation Research Board 93rd Annual Meeting, Washington DC, 2014.

Kerner, B.S., 1999. The physics of traffic. Physics World 12 (8), 25–25.

Kesting, A., Treiber, M., 2008. How reaction time, update time, and adaptation time influence the stability of traffic flow. Computer-Aided Civil and Infrastructure Engineering 23 (2), 125–137.

Laval, J.A., 2004. Hybrid Models of Traffic Flow: Impacts of Bounded Vehicle Accelerations (PhD thesis). University of California, Berkeley.

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

Laval, J.A., Leclercq, L., 2013. The Hamilton–Jacobi partial differential equation and the three representations of traffic flow. Transportation Research Part B: Methodological 52, 17–30.

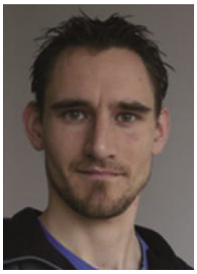
Lebacque, J., 2003. Two-phase bounded-acceleration traffic flow model: analytical solutions and applications. Transportation Research Record 1852, 220–230.

Lebacque, J., Mammar, S., Salem, H.H., 2007. Generic second order traffic flow modelling. In: Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17, London, 2007.

Leclercq, L., 2007a. Bounded acceleration close to fixed and moving bottlenecks. Transportation Research Part B: Methodological 41 (3), 309–319.

Leclercq, L., 2007b. A new numerical scheme for bounding acceleration in the LWR model. In: 4th IMA International Conference on Mathematics in Transport, London, 2007.

Leclercq, L., Laval, J.A., Chevallier, E., 2007. The Lagrangian coordinates and what it means for first order traffic flow models. In: Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17, London, 2007.
Leclercq, L., Laval, J.A., Chiabaut, N., 2011. Capacity drops at merges: an endogenous model. Transportation Research Part B: Methodological 45 (9), 1302–1313.
Leclercq, L., Knoop, V.L., Marczak, F., et al., 2016. Capacity drops at merges: new analytical investigations. Transportation Research Part C: Emerging Technologies 62, 171–181.
Lorenz, M.R., Elefteriadou, L., 2001. Defining freeway capacity as function of breakdown probability. Transportation Research Record 1776, 43–51.
Richards, P.I., 1956. Shock waves on the highway. Operations Research 4 (1), 42–51.
Siebel, F., Mauser, W., Moutari, S., et al., 2009. Balanced vehicular traffic at a bottleneck. Mathematical and Computer Modelling 49 (3), 689–702.
Srivastava, A., Geroliminis, N., 2013. Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model. Transportation Research Part C: Emerging Technologies 30, 161–177.
Treiber, M., Kesting, A., Helbing, D., 2006. Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps. Physical Review E 74 (1), 114–123.
van Wageningen-Kessels, F., van Lint, H., Hoogendoorn, S., et al., 2010. Lagrangian formulation of multiclass kinematic wave model. Transportation Research Record 2188, 29–36.
Whitham, G.B., 1955. On kinematic waves II. A theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences 229 (1178), 317–345.
Zhang, L., Levinson, D., 2004. Some properties of flows at freeway bottlenecks. Transportation Research Record 1883, 122–131.



Simeon C. Calvert, PhD, is a research fellow at Delft University of Technology in the Department of Transport & Planning, the Netherlands. Previously, he worked as a research scientist in the research institute TNO, Netherlands organization for scientific applied research. His main research activities involve traffic flow theory, traffic flow modelling, effects of automated vehicles on traffic, traffic management, and Intelligent Transportation Systems (ITS).

Femke L.M. van Wageningen-Kessels, PhD, is an independent scholar and visiting researcher at Delft University of Technology in the Department of Transport & Planning, the Netherlands. She has her doctoral degree from Delft University of Technology, and now works as independent scholar and author, with her main areas of expertise in applied mathematics, modelling and simulation, and traffic and crowds.

Prof. Serge P. Hoogendoorn, PhD, is the Chair of the Traffic Management, Transport & Planning of the Delft University of Technology and professor of Traffic Flow Theory and Simulation, and staff member of the TRAIL Research School on transport and logistics. His research has focused on traffic flow theory, pedestrian flow, traffic and pedestrian modelling, and simulation of traffic and transportation networks.