Delft University of Technology

Fac. of Information Technology and Systems **Control Systems Engineering**

Technical report bds:01-11

Models for traffic control*

T. Bellemans, B. De Schutter, and B. De Moor

If you want to cite this report, please use the following reference instead:
T. Bellemans, B. De Schutter, and B. De Moor, "Models for traffic control," *Journal A*, vol. 43, no. 3–4, pp. 13–22, 2002.

Control Systems Engineering
Faculty of Information Technology and Systems
Delft University of Technology
Delft, The Netherlands
phone: +31-15-278.51.19 (secretary)
fax: +31-15-278.66.79
Current URL: http://www.dcsc.tudelft.nl

*This report can also be downloaded via http://pub.deschutter.info/abs/01_11.html

Models for traffic control

Tom Bellemans[*], Bart De Schutter[†], Bart De Moor[‡]

Abstract

Traffic models play an important role in both today's traffic research and in many traffic applications such as traffic flow prediction, incident detection and traffic control. In this paper we concentrate on motorway traffic models for traffic control. First, we discuss why traffic control is important for both practice and research. We also argue why a model-based approach to traffic control is used and how it is implemented in traffic applications. Every traffic application requires specific features of the traffic model. After classifying the models based on these features, we look in detail at three macroscopic models since these models are best suited for implementation in traffic control systems. We conclude with an illustrative example that illustrates a model-based predictive control approach for ramp metering.

Keywords: traffic, control, simulation, macroscopic models

1 Introduction

1.1 Traffic congestion and traffic control

Traffic congestion on motorways is becoming an ever more pressing problem in countries all over the world. One approach to tackle congestion could be the construction of new roads to enlarge the capacity of the traffic infrastructure. This approach is very costly and it is often not possible due to environmental or societal constraints. In addition, it can only be executed on the longer term. So there is a need for another, short-term solution. This short-term solution exists of controlling traffic in such a way that congestion is solved, reduced, or at least postponed.

In most countries traffic operators in traffic control centers monitor the traffic situation based on video images and measurements from e.g. loop detectors[1]. Especially during rush hours the interpretation by experts of information about traffic intensities, average velocities, weather conditions, incidents, ... on the motorways leads to advice for the travelers and/or actions in order to solve, to prevent or to postpone congestion. Traffic models allow for a simulation of future traffic densities and average velocities. These predictions can help traffic operators to determine which control measures to take. Another advantage of simulation models is that they allow for a prediction of the traffic state that takes exceptional situations such as (partially) blocked roads into account.

[*]Tom Bellemans is Research Assistant with the K.U.Leuven. ESAT/SISTA, K.U.Leuven, Kasteelpark Arenberg 10, 3001 Heverlee, Belgium. Email: Tom.Bellemans@esat.kuleuven.ac.be

[†]Bart De Schutter, Control Systems Engineering, Fac. Information Technology and Systems, Delft University of Technology, PO Box 5031, 2600 GA Delft, The Netherlands. Email: b.deschutter@its.tudelft.nl

[‡]Bart De Moor is Full Professor at K.U. Leuven. Email: Bart.DeMoor@esat.kuleuven.ac.be

[1]Loop detectors are devices that count the number of vehicles passing on a freeway. They consist of a conducting loop in the road surface and an electronic device that monitors the changes of inductance of the loop as vehicles are passing over it.

Figure 1: Picture of a highway with ramp metering in Delft, The Netherlands.

A more advanced approach towards tackling the congestion problem is the application of a traffic control system that functions as a decision support tool for the traffic operators in the traffic control center by presenting them with a list of the most adequate control measures and their predicted performance or results. State-of-the-art traffic models are one of the core components of such a system.

In the remainder of this section we discuss how model-based traffic control systems operate.

1.2 Goals

One of the goals of traffic control is to solve congestion. This goal is translated into a cost function expressed in terms of the traffic states: traffic flows, traffic densities, average speeds, … As the name cost function states, the cost associated with an undesirable traffic state with congestion needs to be higher than the cost associated with a traffic state with less congestion. This way, finding a traffic state without congestion or where the congestion is as small as possible corresponds to looking for the traffic states with the lowest value of the cost function.

Typically, the total time spent by all the vehicles in the traffic network is included in the cost function. The total time spent by all the vehicles in the network can be calculated from the traffic density and the dimensions of the traffic network. Congestion results in a higher traffic density (and lower speeds) during a longer time and thus the total time spent will increase with an increasing level of congestion.

Besides the total time spent in the network, which is directly related with congestion, also other terms can be added to the cost function such as terms expressing environmental costs (e.g. air pollution, fuel consumption), social costs (e.g. stress) and economic costs (e.g. delayed deliveries). In this case, the traffic regime corresponding to a minimal overall cost will be a trade off between the costs described by the terms in the cost function. By weighing the terms we can put more emphasis on certain costs. The definition of a sustainable cost function is a research area of its own [Vickrey, 1963], [Calthrop, 1998], [Proost, 2001].

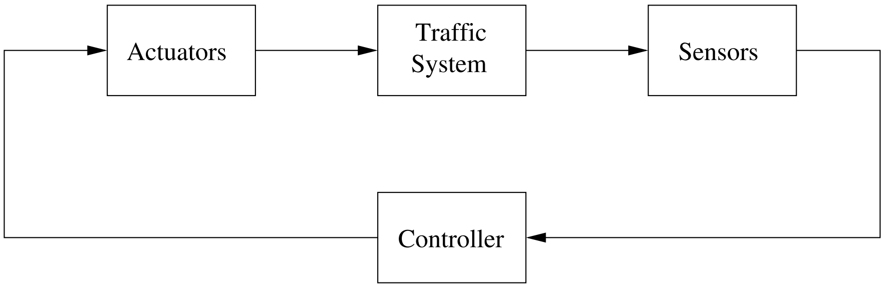

Figure 2: Schematic representation of an advanced traffic management system consisting of the traffic system, actuators, sensors and the controller.

1.3 Control problem

The traffic control problem can be identified as a traditional control problem with a plant and feedback control as shown in Figure 2. The plant (i.e. the traffic system) has actuators such as traffic signals, dynamic route information panels, traffic information, ramp meters, variable speed signs, etc., which are able to interact with and influence the state of the system. The goal of traffic control is to find actuator signals that minimize the cost function. In an automatic procedure this is done by optimizing the cost function subject to the traffic model and the physical constraints of the system, resulting in automatically generated advice and/or control actions for the traffic control signs.

Due to the dynamic nature of traffic on motorways, the control actions need to be updated regularly to account for changing traffic situations. This leads to an adaptive control scheme.

The model-based approach described above requires prediction of the future traffic situation for different control strategies followed by selection of the control strategy resulting in the lowest cost. In scope of this approach, simulation of traffic situations using traffic models shows some important advantages:

- Simulation of traffic models allows for fast, reproducible and cost effective experiments. Using currently available traffic models and computers it is possible to simulate some traffic situations faster than real-time. This allows for a real-time evaluation (optimization) of different alternative control strategies that would not be possible otherwise.
- Since the safety of the travelers cannot be compromised at any time, models are an interesting alternative for experiments in extreme traffic regimes which could lead to potentially dangerous situations and which can therefore not be invoked intentionally. E.g. traffic operation near congestion with shock waves or with large speed differences between adjacent lanes can safely be simulated and reproduced using models.

In this paper we will concentrate on models that can be used to describe traffic flows since such models are — in combination with conventional or advanced control design techniques such as PID control, model predictive control, adaptive control, fuzzy control, etc. — the basic ingredients of advanced traffic control design methods.

2 Traffic model classification

A wide variety of traffic models exists. These models can be classified based on their properties. In this section we illustrate four ways to classify traffic models from the point of view of a (traffic) control engineer:

- Physical interpretation
- Level of detail
- Discrete versus continuous
- Deterministic versus stochastic

2.1 Physical interpretation

According to system theory [Ljung, 1987], there are three major approaches towards modeling: white box, black box, and grey box modeling. We will now explain each of them in more detail and discuss their application to traffic models.

- White box modeling or the deductive approach. In the deductive approach physical equations describe the relationships between different states of the traffic system. This approach is used for instance to describe a mechanical system by Newton's laws and is called white box modeling (first principles modeling). The relationship between the states of the system is described by properties that can be measured (e.g. mass in a mechanical system).
- An alternative approach is inductive, or black box, modeling where input and output data of the system are recorded and a generic parameterized model is fitted to the data. For a traffic system, the input/output data typically consist of the evolution of traffic flows, traffic densities and measured speeds over time. With the black box method there is in general no physical relationship between the modeled traffic situation and the model structure. An example of black box traffic modeling is the application of a neural network that is trained to mimic the behavior of the traffic system by e.g. modeling the relationship between the future density as a function of the current density and speed, and the on-ramp and off-ramp flows [Ho and Ioannou, 1996].
- Finally, we can distinguish an intermediate method between black and white box modeling, known as gray box modeling. Grey box modeling is a combination of the inductive and the deductive approach. In the deductive phase, parameterized equations between the states of the motorway system are written down. During the inductive phase, the parameters in the model are tuned by fitting the input-output relation of the traffic model to input-output measurements of the traffic system. As an example of this intermediate approach, we mention the traffic models of Lighthill, Whitham [1955] and Richards [1956] and Payne [1971], which are discussed in more detail in the next section. The equations in these models have a physical interpretation but contain parameters, which need to be fitted using input/output data from the traffic system.

2.2 Level of detail

Based on the level of detail, we distinguish two types of traffic models in this paper: microscopic and macroscopic models. For a more detailed classification based on the level of detail we refer the interested reader to Hoogendoorn and Bovy [2000].

- **Microscopic** traffic models are models in which all vehicles or 'particles' in the system are described individually (Figure 3). A model for the behavior of every single vehicle is defined. These vehicle models include e.g. the interaction between the vehicles or between the vehicles and the motorway. During simulation, all the individual cars and their interactions are simulated using the models. Combination of all simulation results of the vehicle models leads to an image or snapshot of the traffic situation.

Microscopic traffic simulators can be implemented based on different microscopic traffic models. Some microscopic simulators start from a description of the motorway network we want to simulate where the network is characterized by the speed limit, the number of lanes, overtaking prohibitions, ... for all motorway sections. Every vehicle in the network is characterized by some parameters describing its origin, its destination, the desired velocity, acceleration and deceleration abilities, vehicle type, driver's patience, aggressiveness and so on. During simulation, a flow of vehicles is entered into the network. The vehicle parameters (mass, acceleration, ...) and the driver parameters (aggressiveness, reaction time, ...) are sampled from a stochastic distribution function, which is determined based on measurements of real-life traffic. The vehicles each attempt to reach their desired destination by traveling through the network. On their way from their origin to their destination, the vehicles interact with each other (e.g. by overtaking slower vehicles) and with the motorway network (e.g. by obeying speed limits) according to their microscopic vehicle models.

The two most important microscopic models are:

1. The car following model, which describes how a vehicle follows preceding vehicles. The car following model describes e.g. the headway a driver preserves between himself and the preceding vehicle, how the driver reacts on acceleration or deceleration of the vehicle in front of him, etc. The car following model uses vehicle and motorway specific parameters. The acceleration ability, the aggressiveness of the driver, the mass of the vehicle are vehicle specific parameters that can be used in the car following model. The maximal speed limit and the number of lanes are motorway specific parameters influencing the way a car follows its predecessor.
2. The overtaking model describes how a driver decides whether or not to overtake its predecessor. Vehicle and driver properties that are important in the overtaking model are e.g. the desired speed of the driver and acceleration abilities of the vehicle. The number of lanes is a motorway property that influences overtaking behavior. Vehicle interaction is also important. A driver will decide to overtake based on the speed difference between his lane and the adjacent lane and on the available gap on the other lane.

Microscopic models can also be implemented as cellular automaton models [Nagel, 1996]. A cellular automaton model describes the motorway as a network of connected 'cells' (Figure 4). The behavior of a vehicle in the network is described by the way the vehicles are 'hopping' from one cell to the next. Each cell is approximately of vehicle dimensions (typically 7.5 meter) and can be empty or contain exactly one vehicle. The microscopic model consists of a set of rules describing how vehicles hop from one cell to another. These rules describe how adjacent cells influence each other e.g. no vehicle can hop from a cell to the next if the next cell is occupied (a cell can hold only one vehicle).

- **Macroscopic** traffic models are models that use aggregate variables[2] in order to describe the traffic situation. Typically, a macroscopic model defines a relation between the traffic density,

[2]An aggregate traffic variable is a variable that summarizes information about multiple vehicles. E.g. the average speed contains information on the speed of all the vehicles present in a given section of the road.

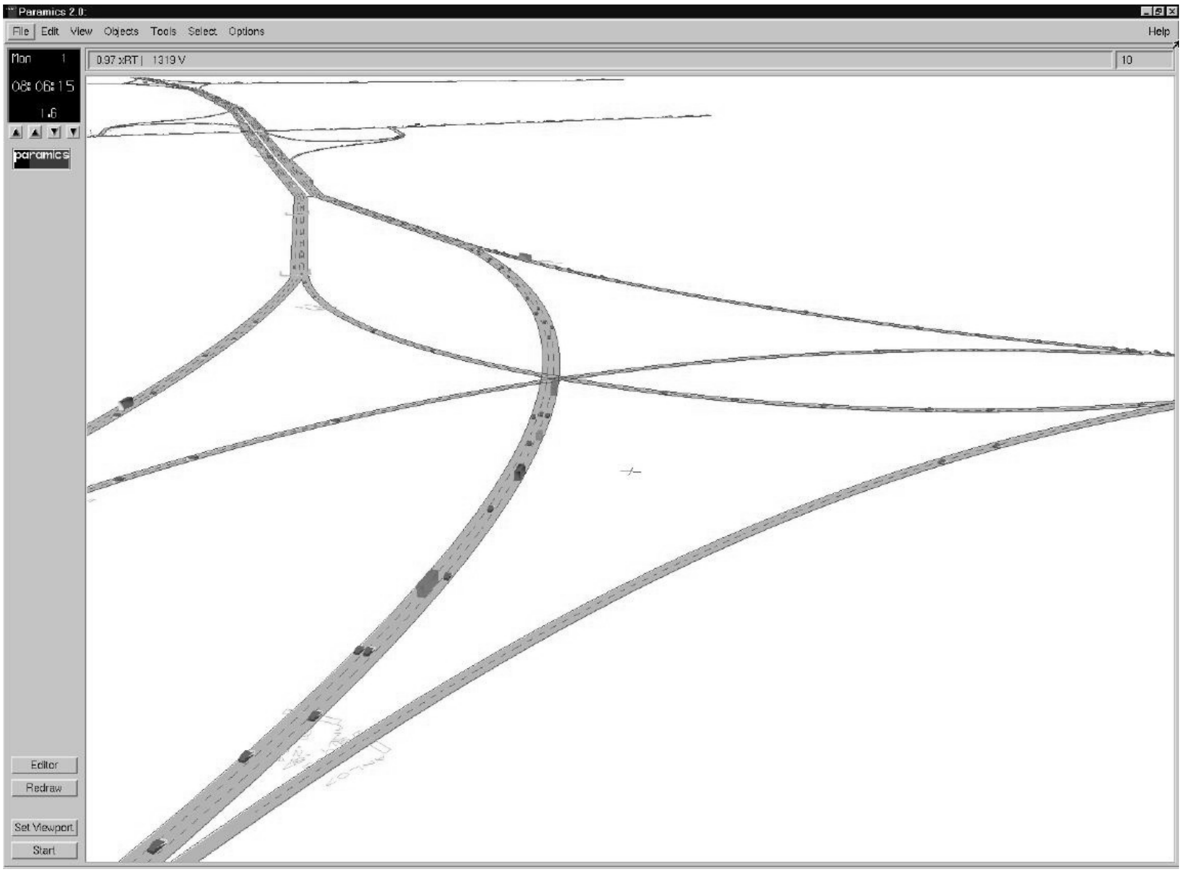

Figure 3: Screenshot of Paramics microscopic simulation software during simulation of part of the E17 motorway Ghent - Antwerp (Belgium). Every vehicle is simulated and visualized individually. Note the different types of vehicles (cars, lorries, ...) present in the simulation.

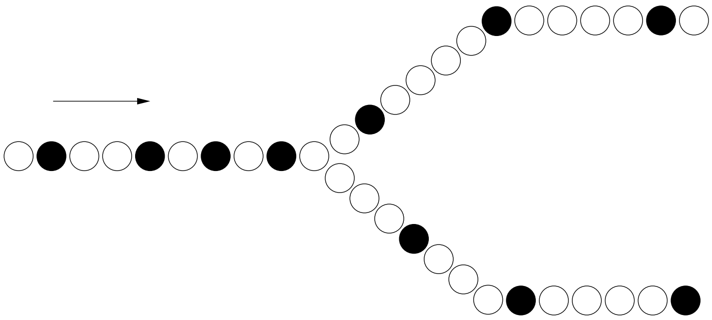

Figure 4: Schematic representation of a cellular automaton model for the simulation of a motorway network. The motorway network consists of a bifurcation of a single lane highway into two single lane highways. A white circle represents an empty cell and a black circle a cell containing a vehicle. During simulation, vehicles can hop from one cell to the next depending on the state of the neighboring cells. The arrow denotes the traffic driving direction.

the average velocity and the traffic flow. The traffic density is defined as the number of vehicles per kilometer and per lane, while the traffic flow or the traffic intensity is defined as the number of vehicles passing a certain point per hour. Within the class of macroscopic models, a classification based on the order of the models can be made. The oldest model was proposed by Lighthill and Whitham in 1955 and independently by Richards in 1956 and is of first order. The only state variable of this model is the traffic density. The Lighthill, Whitham, Richards model was extended later on in order to be able to cope with shock waves and stop-and-go traffic in congested traffic situations [Newell, 1993]. The model described by Payne [1971] is of second order since it has two state variables: traffic density and average velocity. Helbing [1996] proposed a third order macroscopic traffic model with as state variables the traffic density, the average velocity and the variance on the velocity.

Since macroscopic traffic models only work with aggregate variables and do not describe the traffic situation on the level of independent vehicles, they are less computationally intensive than microscopic models. This allows a fast simulation of the studied motorway network that is imperative for on-line predictive control. Often, the cost function is expressed in terms of the aggregate variables, which makes it easy to calculate the cost corresponding to a model state. Due to the fact that a macroscopic traffic model has fewer parameters to estimate than a microscopic model, it is easier to identify and to tune a macroscopic model. Macroscopic models are written as state space models, which are linked to control theory. As a result of the above, macroscopic models are very well suited for optimal predictive and adaptive control. We will discuss some macroscopic traffic models in more detail later on.

2.3 Deterministic versus stochastic

In a deterministic traffic model there is a deterministic relation between the input, the states and the output of the model. If we simulate a traffic situation twice, starting from the same initial conditions and with the same inputs and boundary conditions, the outputs of the model will be the same. Payne's macroscopic traffic simulation model, discussed later on, is an example of a deterministic traffic model.

A stochastic traffic model contains at least one stochastic variable. This implies that two simulations of the same model starting from the same initial conditions, the same boundary conditions and the same inputs may give different results, depending on the value of the stochastic variable during each simulation. The stochastic variable is characterized by a distribution function or a histogram. For example, in the microscopic simulation package Paramics, a distribution of the level of patience over different drivers needs to be defined. During simulation, this distribution is sampled to determine a level of patience for every simulated driver in the network. It is clear that a second simulation, and thus a new sampling of the distribution, will result in drivers with other levels of patience. Therefore, stochastic models need to be simulated repeatedly and an average over the results needs to be taken in order to be able to draw conclusions. If we aim at on-line traffic simulation for predictive control purposes, this requires a lot of computational power, especially if we consider large traffic networks. As such, this is one of the major drawbacks of stochastic simulation models.

2.4 Discrete versus continuous

A motorway traffic model describes the evolution of the state variables of the traffic network over time. This means that there will be two independent variables, namely space and time. These independent variables can be considered to be either continuous or discretised. Since the continuous traffic models are generally too complicated to solve analytically, especially if the size of the considered traffic network is large, they are discretised in time and space in order to simulate their behavior using a computer [Lebacque, 1996]. Issues concerning discretisation techniques and sampling frequencies (aliasing) are beyond the scope of this text. Payne's traffic model is from origin a continuous model which is discretised in space (with motorway stretches of typically 500 meter) and time (with time intervals of typically 15 seconds) for implementation and simulation on a computer [Papageorgiou, 1990a].

In the remainder of this paper, we will more extensively look at macroscopic traffic models, since this kind of model is better suited for model-based control design methods than microscopic traffic flow models. But first we discuss one of the basic concepts in traffic flow theory: the fundamental diagram.

3 The fundamental diagram

The fundamental diagram is a basic tool in understanding the behavior of traffic systems: it relates the traffic flow and the traffic density at a given location or section of the motorway. When observing traffic on a motorway and plotting the traffic flow [vehicles/hour] versus the traffic density [vehicles/kilometer/lane] for a location along the motorway, a curve as in Figure 5 occurs. This curve has a characteristic shape that is the same for every motorway section and is known in the traffic literature as the "fundamental diagram". The fundamental diagram was already described by Greenshields in 1935.

Let us discuss Figure 5 in detail now. For zero traffic density (no cars on the motorway), the traffic flow is obviously zero. With increasing traffic density (more vehicles on the motorway), the traffic flow also increases. If the traffic density increases further, the increase of the traffic flow will slow down. At a certain point, the flow reaches a maximum for a density that we call the critical density C_{cr}. The traffic flow on the motorway decreases with increasing density if the density is greater than the critical density. Eventually the flow becomes zero again for a density we define to be the jam density C_{jam}.

We can interpret the behavior of traffic on a motorway and the fundamental diagram as follows: when the traffic density is zero, the traffic flow will obviously be zero as well. When the traffic density starts to increase (more cars are travelling on the motorway) the traffic flow increases. For low traffic densities, the relationship between flow and density is almost linear due to the small interaction between the vehicles if the traffic density is small. As the traffic density increases further, this interaction between the vehicles increases rapidly. E.g. fast cars will have to slow down while waiting to overtake a slower car. Due to the decreasing average speed with increasing density, the traffic flow increases at a slower and slower rate. The traffic flow reaches a maximum for the critical density C_{cr} after which the flow decreases again with further increasing traffic density. In high-density traffic situations, drivers tend to slow down, resulting in a drop of traffic flow, even though the traffic density is high. The maximal possible traffic density on the motorway is C_{jam}, the traffic density at which the vehicles come to a halt: if a vehicle stops, the upstream vehicles are forced to stop as well.

For a real-life motorway, the fundamental diagram can be fitted to measurement data in order to determine the parameters.

In summary, two important properties of traffic flow-density relationship of traffic on motorways can be distinguished:

- The traffic flow reaches a maximum at the critical density C_{cr}.

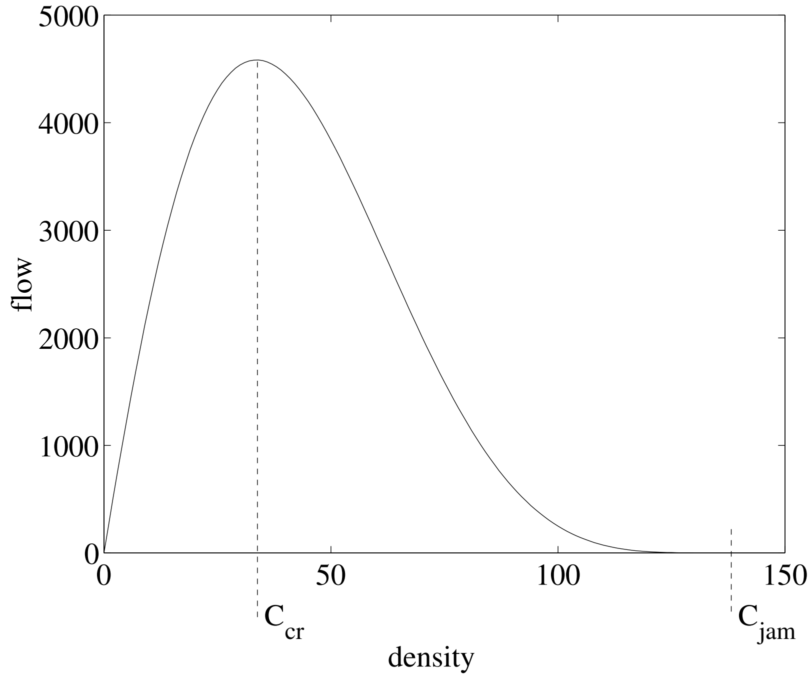

Figure 5: The Fundamental Diagram. A plot of traffic flow versus traffic density on a motorway results in the figure above. The density at which the traffic flow is maximal is called the critical density C_{cr}. The (non-zero) traffic density at which the flow equals zero (i.e. the vehicles come to a full stop) is called the jam density C_{jam}.

- With increasing traffic density on the motorway, traffic will eventually come to a halt (average speed equal to zero) when the density reaches the maximal possible value C_{jam}.

We will further illustrate the fundamental diagram using measurements of the traffic situation on a stretch of the three-lane motorway E17 Gent-Antwerp in Belgium. A plot of the traffic measurements of a day in the week (Thursday March 2, 2000) looks like Figure 6. The upper figure shows the evolution of the total traffic flow (veh/h) over time and the lower figure shows the evolution of the average speed (km/h).

In Figure 6 we see an increase of the traffic flow in the early morning (5-6 a.m.). During this period the average speed on the motorway remains nearly constant. As the morning rush hour starts (around 6.30 a.m.), the traffic flow continues to increase, but suddenly the increasing flow breaks down (7 a.m.) as the critical density of the motorway is exceeded. Congestion has set in. The breakdown of the traffic flow is accompanied by a drastic reduction of the average speed (50km/h as can be read on the lower graph). At the end of the rush hour (9.30 a.m.), the traffic density on the motorway decreases and the average speed on the motorway is restored. This real life example illustrates the fundamental diagram, i.e. the relations between the measured values of traffic density, average velocity and traffic flow on a motorway stretch.

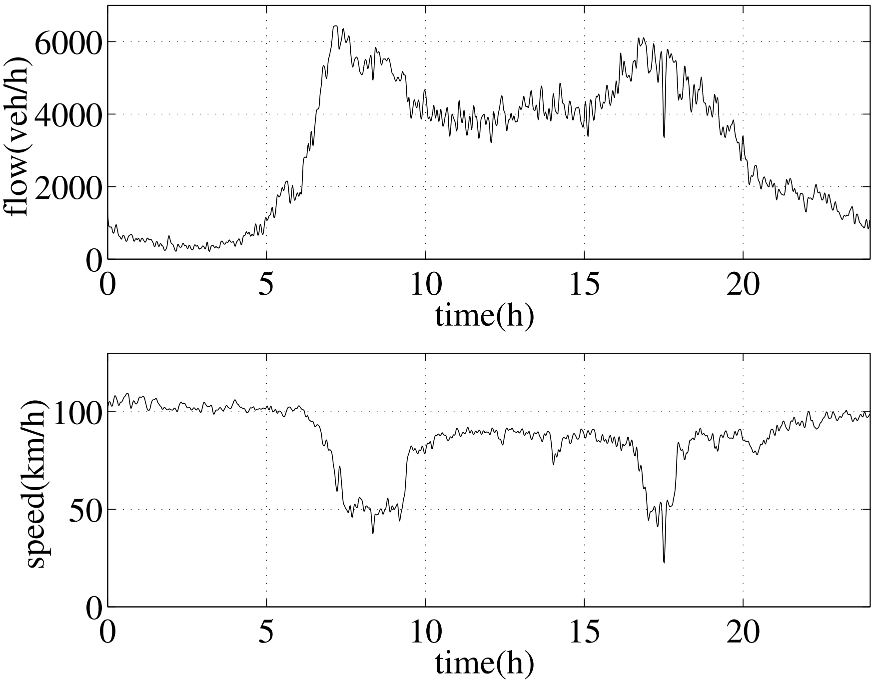

Figure 6: Evolution of traffic flow (upper graph) and average speed (lower graph) in time for the three-lane motorway E17 Gent-Antwerp on Thursday 02 March 2000.

4 A closer look at macroscopic traffic models

Since macroscopic traffic flow models are better suited for model-based control design techniques, we will now discuss this type of models in more detail. Although there are many different macroscopic models we will limit ourselves here to the three most important models: the Lighthill-Whitham-Richards model, the Payne model, and the model of Papageorgiou, since these are most often used in practice. We will write down the model equations for each type and give an intuitive interpretation of their meaning.

4.1 The Lighthill, Whitham and Richards (LWR) model

Macroscopic models use aggregated variables to describe the behaviour of traffic. Macroscopic traffic models are often derived using the analogy between traffic flows and the fluid flows. As well as there exists a law of conservation of mass in fluid dynamics, we can write down a law of conservation of vehicles in the traffic context:

$$\frac{n(x)\partial C(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \ , \tag{1}$$

where $C(x,t)$ denotes the traffic density in vehicles per lane per kilometer at location x and at time t, with $n(x)$ the number of lanes at position x and with $q(x,t)$ the traffic flow in vehicles per hour at location x at time t, sometimes also called the traffic intensity. The aggregated variables $C(x,t)$ and $q(x,t)$ are continuous functions of time and space, although the system they describe is intrinsically discrete (the cars on the motorway are discrete entities). Equation (1) is a 'physical law' in traffic

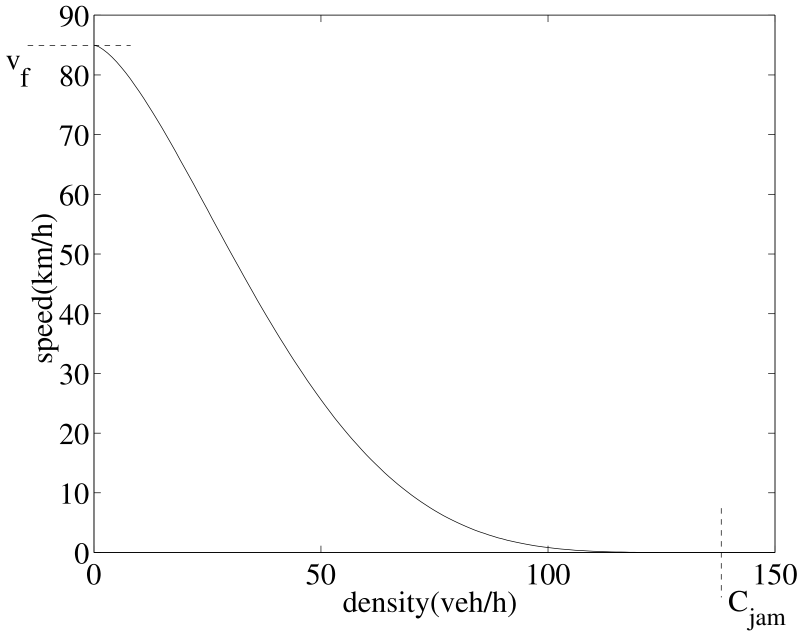

Figure 7: Plot of the empirical speed-density relation given in expression (5). The free flow speed is 85 km/h and the jam density is 140 veh/km.

theory stating that no cars can vanish nor appear out of the blue. By consequence, equation (1) holds exactly for every possible traffic regime and at all times. The traffic flow from equation (1) can be expressed in terms of the traffic density and the traffic speed $v(x,t)$:

$$q(x,t) = C(x,t)\, v(x,t)\, n(x) \quad . \tag{2}$$

Lighthill and Whitham [1955] and also Richards [1956] observed that the average equilibrium speed of the vehicles is a function of the traffic density:

$$v(x,t) = F(C(x,t)) \quad . \tag{3}$$

The general equation (3) can be substituted by one of the empirical expressions proposed in the literature [May, 1990]. In this paper we use expression (5) below, which is plotted in Figure 7. This relationship between the average speed and the density on a motorway can intuitively be understood: when the road becomes crowded, the average speed will drop untill eventually congestion occurs.

The Lighthill-Whitham-Richards (LWR) model consists of the equations (1)–(3). This model is a continuum model in both time and space. An analytical solution of these equations for a motorway network is hard or even impossible to find. In practice the model is discretised in time and in space to allow for a computer simulation of the difference equations describing the behaviour of the motorway network. This is illustrated in Figure 8. The discretization in time is done by considering time steps Δt (a typical value is 15 s); the discretization in space is obtained by dividing the motorway in sections (a typical length is 500–1000 m). For stability reasons, the discretisation needs to be such that $\Delta x > v\Delta t$ for all sections in the network [Papageorgiou, 1990a].

Now we derive a difference equation that describes the Lighthill-Whitham-Richards model. Dis-

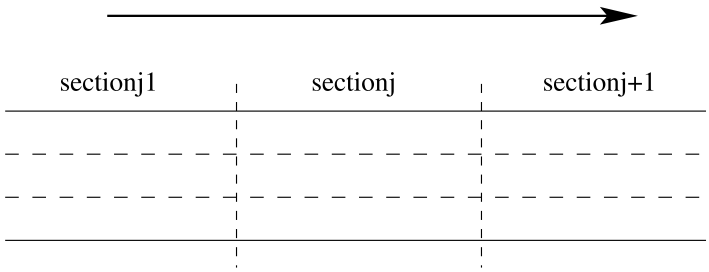

Figure 8: Discretisation of a stretch of motorway in three connected sections ($j-1$, j and $j+1$). The arrow denotes the direction of traffic flow.

cretisation of equation (1) with time step Δt results in

$$C_j(k+1) = C_j(k) + \frac{\Delta t}{l_j n_j}\left[q_{\mathrm{in},j}(k) - q_{\mathrm{out},j}(k)\right] \quad , \tag{4}$$

where $C_j(k)$ is the average traffic density in section j and in period k, l_j is the length of the section and n_j the number of lanes. The net inflow in section j is defined as the difference between the inflow $q_{\mathrm{in},j}$ in section j minus the outflow $q_{\mathrm{out},j}$ out of section j. This equation states that the traffic density in a section over a certain time interval $k+1$ equals the traffic density during the previous time interval k plus a change in density due to the net inflow in the section.

One empirical formula for the average speed proposed in the literature [May, 1990] is

$$v_j(k) = V^{\mathrm{e}}(C_j(k)) \triangleq v_{\mathrm{f}}\left[1 - \left[\frac{C_j(k)}{C_{\mathrm{jam}}}\right]^{\alpha}\right]^{\beta} \quad , \tag{5}$$

where v_{f} is the free flow speed, i.e., the average speed of the vehicles in section j assuming the vehicles do not influence each other (e.g. if a vehicle is alone in section j), α and β are parameters, and C_{jam} is the density at which the average speed of the vehicles becomes zero. The parameters of the empirical relationship (5) need to be fitted on traffic data. Data acquisition, pre-processing and parameter fitting are beyond the scope of this text. We refer the interested reader to [Bellemans, De Schutter, and De Moor, 2000a, 2000b] for more information.

The flow in each section can be written as the average speed in that section times the traffic density times the number of lanes in the section:

$$q_j(k) = C_j(k)\, v_j(k)\, n_j \quad . \tag{6}$$

By dividing a motorway network into several sections and writing down equations (4), (5) and (6) for every section we can implement a simulation model. In the LWR model the different sections are coupled by defining the flow from section $j-1$ to section j to be the minimum of the outflow of section $j-1$ (the supply) and the maximal possible inflow of section j (the demand). This simulation model allows to simulate the traffic density, the average speed and the traffic intensity in every section. These parameters are important when optimizing traffic control strategies.

4.2 Payne

The second macroscopic traffic model we discuss in detail is the Payne model. Where the first order LWR-model contained only one partial differential equation, the second order Payne model contains two partial differential equations.

The first differential equation (1) describes the conservation of vehicles and is the same as in the LWR model.

Observations of traffic show that the average speed in a section is not only dependent on the traffic density in that section but also on the state of the neighboring sections and on the dynamics of the average speed itself. Three major mechanisms that influence the average speed can be distinguished: relaxation, convection and anticipation. As a consequence, equation (5) is replaced by

$$\frac{\partial v}{\partial t} + \underbrace{v\frac{\partial v}{\partial x}}_{1} = \underbrace{\frac{V^{\mathrm{e}}(C) - v}{T}}_{2} - \underbrace{\frac{c_0^2}{C}\frac{\partial C}{\partial x}}_{3} \ , \tag{7}$$

where V^{e} is the average speed in equilibrium, c_0 is the anticipation constant and T is the relaxation constant. Equation (7) is derived from a simple car following model that describes the behavior and interaction of the vehicles as they follow a leading vehicle on the road [Payne, 1971; May, 1990]. It describes the change in average speed in a motorway section over time due to convection (term 1), relaxation (term 2) and anticipation (term 3).

Just like the LWR model the Payne model is discretised for implementation on a computer. Discretisation of equation (7) yields [Papageorgiou, 1990a]:

$$v_j(k+1) = v_j(k) + \underbrace{\frac{\Delta t}{l_j} v_j(k)[v_{j-1}(k) - v_j(k)]}_{1} + \underbrace{\frac{\Delta t}{\tau}[V(C_j(k)) - v_j(k)]}_{2} - \underbrace{\frac{\nu_{\mathrm{a}}\Delta t[C_{j+1}(k) - C_j(k)]}{\tau l_j [C_j(k) + \kappa]}}_{3} \ , \tag{8}$$

where τ is a time constant, ν_{a} the anticipation constant, κ a tuning parameter and $V(C_j(k))$ the average speed in equilibrium. For a schematic representation of the discretisation in space we refer to Figure 8. The average speed in equilibrium is an empirical function of traffic density and needs to be calibrated. Different formulas are proposed in the literature [May, 1990; Kotsialos,1999]. In this paper we use equation (5).

According to equation (8) the speed in section j at time $k+1$ equals the speed in section j at time k plus a correction for convection, relaxation and anticipation:

- **Convection** (term 1 in (7) and (8)): Cars travelling from section $j-1$ to section j do not adapt their speed instantaneously. For instance, a driver travelling at high speed through section $j-1$ will gradually slow down to the speed he feels comfortable with in section j. During the braking process, the driver is actually going faster than his desired equilibrium speed in j. This way, the higher average speed in section $j-1$ increases the average speed in section j. The opposite is also true. Cars travelling at low speed in section $j-1$ need time to accelerate towards the higher average speed in section j. During this time, they are travelling at a lower speed, thus pulling the average speed down. Convection is proportional to the difference in speed between the two sections since the higher the speed difference, the longer the vehicles will need to accelerate or decelerate and the greater the impact on the average speed in the section. The effect of convection is also proportional to the average speed in the section. The higher the speed, the longer it takes to brake or to accelerate due to the increased kinetic energy of the vehicle and thus the greater the effect of convection on the average speed. The impact of convection decreases with the length of the section. The longer the section, the longer the distance the vehicles drive at their desired speed. Note that it is assumed here that convection is symmetrical, meaning that drivers will accelerate and decelerate at the same rate (on the average).

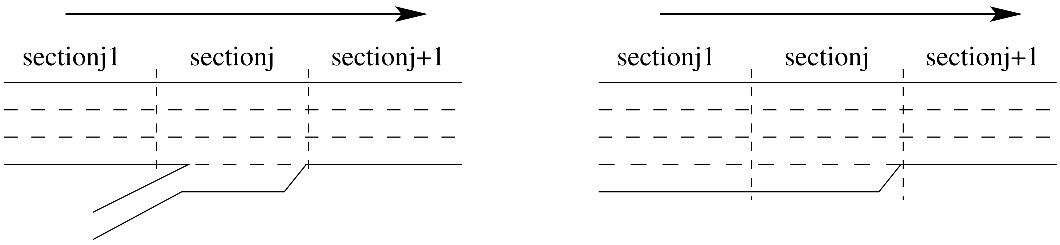

Figure 9: Merging (left) and weaving (right) configurations on motorways. Note the similarity between the two configurations.

- **Relaxation** (term 2 in (7) and (8)): The relaxation term describes the fact that all drivers tend to reach their desired speed. If the actual speed is lower (e.g. because there is/was an obstacle, congestion, ...) the driver will accelerate, if the actual speed is higher (e.g. because the desired speed is dropping rapidly due to onset of congestion) the driver will apply the brakes. When we average the behavior over all drivers, we find a density dependent desired speed (5). If the average actual speed differs from the desired average speed, the average speed will evolve towards the desired average speed. The larger the difference between actual and desired speed, the larger the drivers' action and the larger the relaxation term. The time constant τ depends on the drivers' swiftness. The larger τ, the slower the drivers will react to a deviating speed and the smaller the relaxation term.

- **Anticipation** (term 3 in (7) and (8)): The anticipation term takes into account that drivers are looking ahead. If a driver sees that the traffic density ahead is higher, he will anticipate and slow down (note the sign!). Vice versa, if a driver sees that the traffic density ahead is lower than the current density, he will accelerate. The anticipation constant is proportional to the relative density difference between sections j and $j+1$. Note that we are working with the downstream section $j+1$, resulting in an information stream opposite to the driving direction. The time constant τ describes the swiftness of the drivers' response as in the relaxation term. The parameters ν_{a} and κ allow for tuning of the anticipation term.

The combination of equations (4), (5), (6) and (8) yields a complete second order motorway traffic model. The major difference with the LWR model is the dynamic equation for the speed.

4.3 Improvements to the Payne model by Papageorgiou

Papageorgiou [1990a] proposed to add two extra terms to (8) in order to take additional dynamic phenomena in traffic flows into account.

When driving in intense traffic on a motorway, one often notices that traffic is slowing down in the vicinity of on-ramps (Figure 9). This dropping of the speed can be partially explained by the increased traffic density at and after the on-ramp. However, it is observed that the speed drop is larger than can be accounted for by the increased density. The vehicles on the on-ramp need to merge into the traffic flow on the mainstream. This involves lane changing and disturbs the traffic flow, resulting in a decrease of the average speed. Adding equation (9) as an extra merging term in (8) accounts for this additional decrease in average speed:

$$-\left(\frac{\delta\Delta t}{l_j n_j}\right)\frac{q_\mu(k)v_j(k)}{C_j(k)+L} \quad , \tag{9}$$

where δ and L are tuning parameters for the merging term. The decrease in average speed is proportional to the flow on the on-ramp $q_\mu(k)$. Since merging becomes harder as the speed on the mainstream increases, the impact of the merging (slower) cars becomes larger as the speed $v_j(k)$ increases. This can intuitively be understood as follows: in a high-speed, low-traffic-density situation, the impact on the average speed of a fast car that is slowed down due to a lane changing maneuver of a merging car will be larger than in a low-speed, high-traffic-density situation where the drop in speed due to the lane changing maneuver will be smaller. The merging phenomenon is inversely proportional to the traffic density on the mainstream. At first sight, this seems like a contradiction. One might expect that the higher the traffic density is, the harder it gets to merge into the traffic stream and the higher the impact of lane changing effects. However, due to the lower mainstream speed at higher densities, the impact of lane changing on the average speed becomes smaller.

A phenomenon that is very similar to merging is weaving (Figure 9). When the number of lanes on a motorway drops, vehicles driving on the disappearing lanes need to merge into the other lanes. The lane changing process and its impact is nearly the same as with merging described above. Equation (10) is the term to be added to equation (8) in order to take a reduction of lanes between section j and $j+1$ into account.

$$-\left(\frac{\phi \Delta t}{l_j n_j}\right)\left(\frac{\Delta n_j C_j(k)}{C_{\mathrm{jam},j}}\right) v_j^2(k) \quad , \tag{10}$$

where ϕ is a tuning parameter and $\Delta n_j = n_j - n_{j+1}$ is the number of lanes dropped. We now explain the meaning of (10) by emphasizing the relation with (9).

The vehicles travelling on the disappearing lane(s) can be viewed as a flow that needs to merge into the flow on the non-disappearing lanes through lane changing. The merging process causes a drop of the average velocity as described by (9). Assuming that the traffic flow of the disappearing lane(s) is divided equally over the non-disappearing lanes and by using (6) we find that $q_\mu(k)$ in (9) can be approximated by

$$\frac{\Delta n_j}{n_j} C_j(k) v_j(k) n_j \quad .$$

The weaving phenomenon differs from merging in the sense that there is a fraction of cars that will anticipate the lane drop by changing lanes early. This causes the density on the remaining lanes to become higher than on the disappearing lanes. We discuss the effect of early changing both in the low and the high traffic density case:

- During low-density operation, early anticipation of the lane drop by the drivers leads to a smoother lane changing. Therefore, the impact on the average speed of lane changing due to merging is in general smaller than the impact accounted for by (9). Therefore, we replace $C_j(k)+L$ in (9) by the larger $C_{\mathrm{jam},j}$ in (10). This takes the smaller disturbance of traffic due to anticipation in low traffic density operation into account.

- During high-density operation, the increase of the traffic density on the remaining lanes, caused by anticipation, results in an additional decrease of the average velocity in the motorway section as compared to (9). Replacing $C_j(k)+L$ in (9) by $C_{\mathrm{jam},j}$ in (10) takes this into account. During high-density operation $C_{\mathrm{jam},j}$ becomes smaller than $C_j(k)+L$, resulting in a larger correction term for the average speed.

Equation (10) expresses that the higher the density on the motorway is, the more intense and more disturbing the weaving process is. Weaving influences on the average speed in a section are proportional to the square of the average speed, expressing that drivers will slow down significantly when weaving occurs in high speed circumstances.

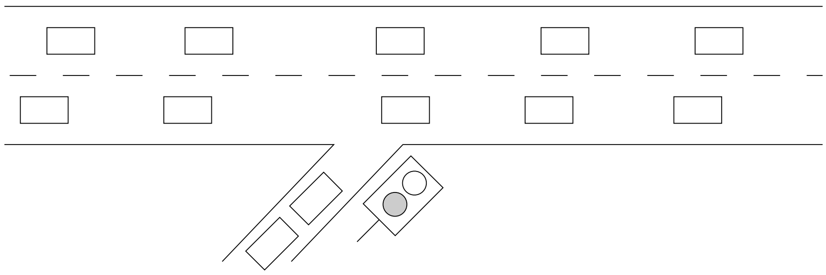

Figure 10: A motorway with ramp metering.

4.4 Higher order models

The second-order model of Payne can be further extended towards a third-order model as proposed by Helbing [1996]. The model proposed by Helbing consists of three states: the vehicle density, the average speed in the section and the variance on the average speed in the section. The meaning of the additional state can be illustrated as follows. Imagine two traffic situations with identical traffic density and identical average traffic speed in the section but with a different variance on the speed. If the variance on the speed is small, all cars adopt more or less the same speed, resulting in smooth traffic. If the variance on the speed is high there are vehicles driving significantly faster than others, resulting in lane changing and overtaking. Traffic with a high speed variance is much more chaotic than traffic with small speed variance and will respond in a different way to disturbances than smooth traffic. The Helbing model describes how the chaotic nature of the traffic on the motorway evolves in time and through space. The dynamics of the traffic density and the average velocity are considered to be dependent on the speed variance. A detailed description of the equations of the model proposed by Helbing goes beyond the scope of this text.

5 Simulation example

In this section we illustrate the application of a second order traffic flow model for optimization of ramp metering. In order to do so, we consider a traffic network consisting of a motorway with two lanes and an on-ramp as shown in Figure 10. The motorway section upstream of the on-ramp is discretised in two segments of 500 meters as well as the downstream section. At the on-ramp we implement ramp metering by installing a traffic light. By altering the duration of the green phase of the traffic light, the number of cars that can enter the highway — the metering rate — is controlled. The arrival rate of the queue that can form at the traffic light is the traffic demand of the on ramp while the service rate of the queue is determined by the metering rate.

The goal of optimal traffic control is to determine the metering rate that minimizes a given cost criterion. In order to illustrate the improvement of the traffic situation that can be achieved using ramp metering, we compare the no-control case with a situation with a model predictive control (MPC) based controller. We use the discrete second order Payne model as described earlier. The motorway consists of four 500 meter sections with two lanes each. The maximal density on the highway is 180 vehicles per kilometer and the critical density is 33.5 vehicles per kilometer. The free flow speed is 102 kilometer per hour. We simulate three and a half hours and the discrete time step is chosen to be 10 seconds. The traffic demand on the motorway during the simulation is plotted in Figure 11. The demand on the mainline is constant at 3400 vehicles per hour while the demand on the on-ramp starts at 500 vehicles an hour and increases temporarily to 1500 vehicles per hour.

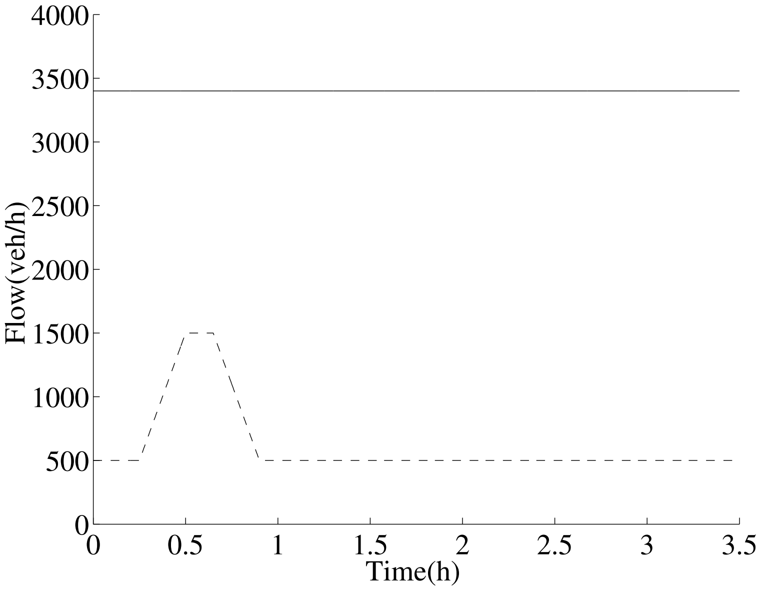

Figure 11: The evolution of traffic demand during simulation. The solid line represents the demand on the mainline while the dashed line represents the demand on the on-ramp.

In Figure 12 we see the simulation results of the no-control case. As the demand on the on-ramp increases, the density of the traffic on the motorway increases. This increase is accompanied with a decrease of the speed on the motorway. As can be seen in Figure 12, the increased demand on the on-ramp causes the traffic flow on the mainline to drop resulting in a queue at the entrance of the mainline. The criterion we use here to assess the cost, or the performance, of the state of the network is the total time spent by all the vehicles in the network (TTS). This includes the vehicles in the queues as well as the vehicles on the motorway.

The performance of the no-control case is 1267 vehicle hours spent in the network during the simulated period of 3.5 hour.

In the control case we use MPC to optimize the metering rate such that the cost function is minimal. This cost function consists of the TTS plus an additional term to penalize a change of the control signal. This penalization term on the change of the metering rate is added in order to limit the metering rate variations as much as possible. A maximal queue length of 100 vehicles at the on-ramp is added as an additional constraint. This constraint is typically imposed on the system in order to prevent the queue from blocking traffic on the roads in the vicinity of the on-ramp. The prediction horizon is 8 minutes and the control horizon is 2 minutes. The metering rate is kept constant over a minute and, by consequence, the metering rate can only change twice during the control horizon.

Figure 13 shows the results of the simulation using MPC. We see the most obvious difference with the no-control case when looking at the queue lengths. The metering rate decreases after approximately half an hour (Figure 14) in order to limit the traffic density on the motorway resulting in a queue at the on-ramp. After a while, the metering rate has to increase in order to avoid the queue at the on-ramp to grow over 100 vehicles. This results in an increased density in the mainline and eventually in a queue at the mainline entrance. When looking at the flow, we see that the metering rate avoids a collapse of the flow in the mainline.

The advantage of ramp metering becomes clear when comparing the performance of the MPC controlled case with the performance of the no-control case. The performance of the MPC controlled

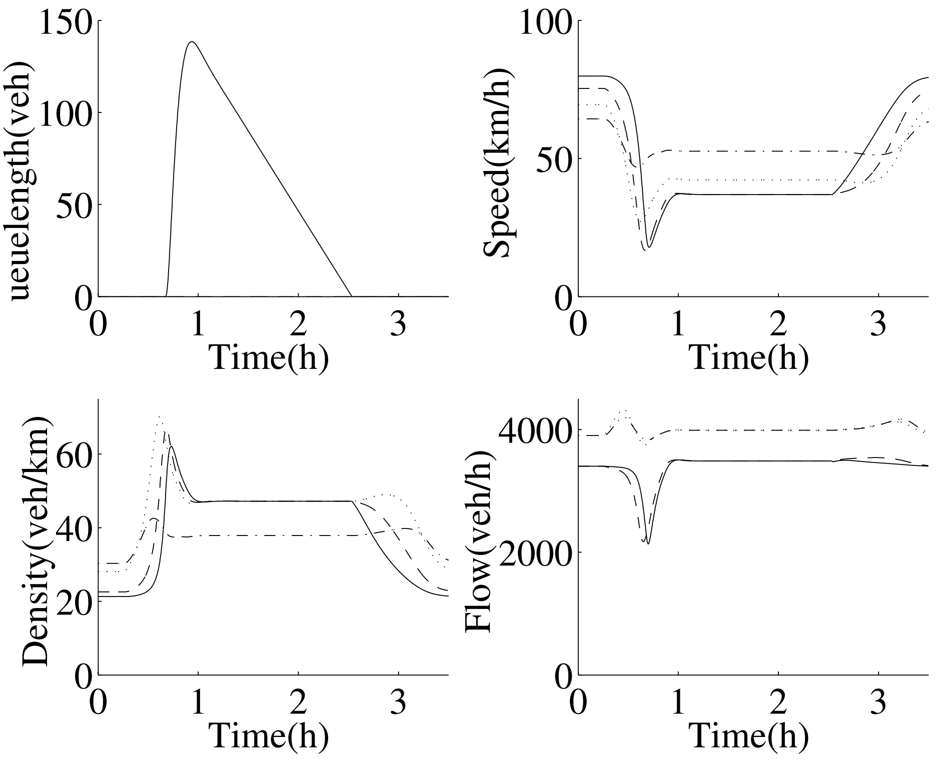

Figure 12: Simulation results of the no-control case. Plot of the queue lengths and the states of the motorway. The solid line represents the mainline queue while the dashed line represents the queue at the on-ramp. The motorway segments in downstream order are represented by a solid, a dashed, a dotted and a dashed-dotted line respectively.

case is 1183 vehicle hours over a simulation period of 3.5 hours. Thus, the MPC controller combined with ramp metering achieves a reduction of 84 vehicle hours or 3.5 vehicle days over a simulation period of 3.5 hours!

6 Conclusions

In this paper we have presented a brief overview of traffic flow models from the point of view of a control engineer. These models can be used to develop traffic controllers using model-based control system design procedures. We have presented some model classifications based on the following properties: physical interpretation of the model, level of detail, discrete or continuous models, and deterministic versus stochastic models. Since macroscopic traffic flow models are best suited for traffic control purposes, we have discussed the most important representatives of this class: the Lighthill-Whitham-Richards model, the Payne model, and an improved Payne model developed by Papageorgiou. For each of these models we have given an intuitive interpretation of the equations that describe the model.

In addition, we have also presented the fundamental diagram, which describes the behavior of traffic in aggregated variables (average density, average flow and average speed), since it is one of the basic concepts in traffic flow theory and since it provides more insight in the nature of traffic flows. So in this paper we have given an overview of traffic models, which are a basic ingredient

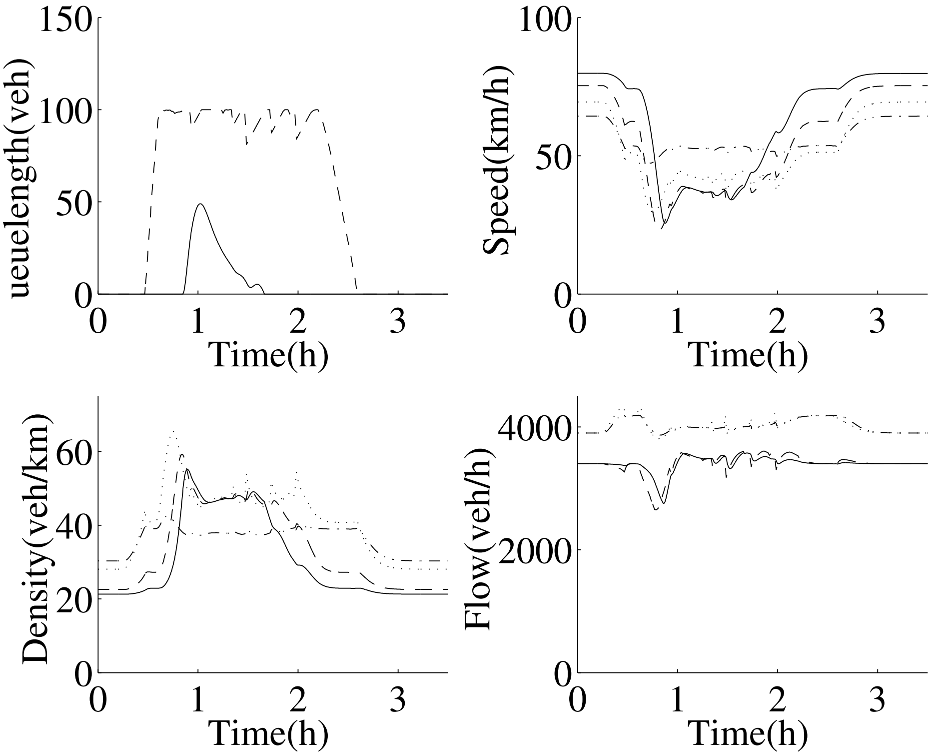

Figure 13: Simulation results of the MPC case. Plot of the queue lengths and the states of the motorway. The solid line represents the mainline queue while the dashed line represents the queue at the on-ramp. The motorway segments in downstream order are represented by a solid, a dashed, a dotted and a dashed-dotted line respectively.

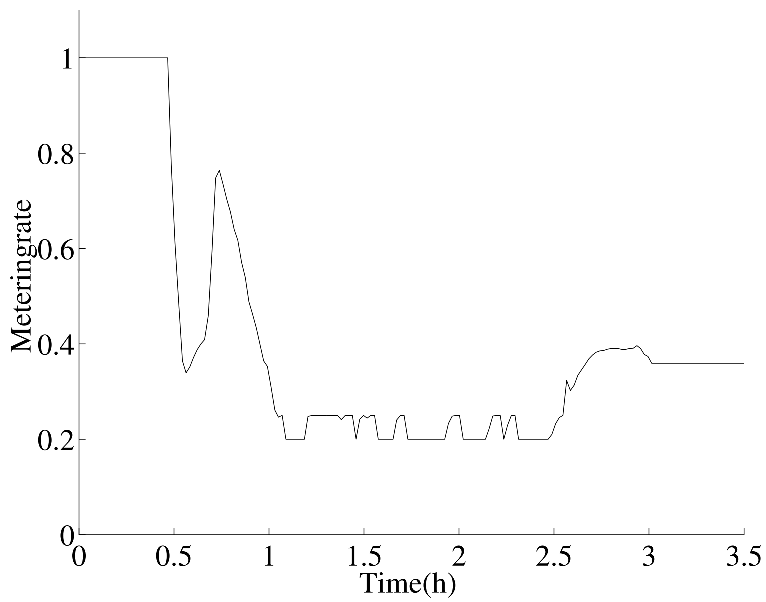

Figure 14: Evolution of the optimal metering rate over time.

of model-based control design procedures for traffic flow control. Today — and even more in the future — there is a growing need for such advanced traffic control procedures, since they are one of the most promising methods to effectively deal with the negative effects of traffic congestion on the environment, economy and society as a whole, and to guarantee a sustained mobility.

List of symbols

$C(x,t)$	: vehicle density on the motorway at point x and at time t. (veh/km/lane)
$C_j(k)$	: vehicle density in motorway section j during the time interval $[k\Delta t,(k+1)\Delta t)$. (veh/km/lane)
$C_{\mathrm{cr},j}$	: critical density in section j, the traffic density corresponding to the maximal flow in the motorway section. (veh/km/lane)
$C_{\mathrm{jam},j}$	: jam density in section j, the density at which the vehicles come to a halt. This is the maximal possible density in the motorway section. (veh/km/lane)
l_j	: length of motorway section j. (km)
n_j	: number of lanes in section j.
$q(x,t)$	: vehicle flow on the motorway at point x and at time t. (veh/h)
$q_j(k)$	: vehicle flow in motorway section j during the time interval $[k\Delta t,(k+1)\Delta t)$. (veh/h)
$v(x,t)$	: average speed on the motorway at point x and at time t. (km/h)
$v_j(k)$	: average speed in section j of the motorway during the time interval $[k\Delta t,(k+1)\Delta t)$. (km/h)
$v_{\mathrm{f},j}$	: free flow speed of section j, the speed that an average vehicle obtains in section j when there is no interaction with other vehicles. (km/h)

Acknowledgements

This work is supported by several institutions: the Flemish Government: Research Council K.U.Leuven: Concerted Research Action Mefisto-666, the FWO projects G.0240.99, G.0256.97, and Research Communities: ICCoS and ANMMM, IWT projects: EUREKA 2063-IMPACT, STWW; the Belgian State, Prime Minister's Office - OSTC -: IUAP P4-02 (1997-2001) and IUAP P4-24 (1997-2001), Sustainable Mobility Programme - Project MD/01/24 (1997-2000); the European Commission: TMR Networks: ALAPEDES and System Identification, Brite/Euram Thematic Network: NICONET; Industrial Contract Research: ISMC, Electrabel, Laborelec, Verhaert, Europay). The scientific responsibility is assumed by its authors.

Bibliography

Bellemans T., De Schutter B., and De Moor B. (2000a). "Data acquisition, interfacing and preprocessing of highway traffic data." In: *Proceedings of Telematics Automotive 2000*, Birmingham, UK, vol. 1, pp. 4/1–4/7.

Bellemans T., De Schutter B., and De Moor B. (2000b). "On data acquisition, modeling and simulation of highway traffic." In: *Proceedings of the 9th IFAC Symposium on Control in Transportation Systems 2000*, Braunschweig, Germany, pp. 22–27.

Calthrop E. and Proost S. (1998). "Transport and Environment - interaction between theory and empirical research" In: *Environmental and Resource Economics*, vol. 11, pp. 3–4.

Greenshields B.D. (1935). "A study in Highway Capacity". In: *Highway Research Board, Proceedings*, vol. 14, pp. 458.

Helbing D. (1996). “Gas-kinetic derivation of Navier-Stokes-like traffic equations.” In: *Physical Review*, vol. E53, no. 3, pp. 2266–2381.

Ho F.S. and Ioannou P. (1996). “Traffic flow modeling and control using artificial neural networks.” In: *IEEE Control Systems Magazine*, vol. 16, no. 5, pp. 16–26.

Hoogendoorn S.P. and Bovy P.H.L. (2000). “State-of-the-art of vehicular traffic flow modelling.” Tech. rep., Transportation and Traffic Engineering Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands. Accepted for publication in the Special Issue on Road Traffic Modelling and Control of the *Journal of Systems and Control Engineering — Proceedings of the Institution of Mechanical Engineers, Part I.*

Hoogendoorn S.P. and Bovy P.H.L. (2000b). “Modelling multiple user-class traffic flow.” In: *Transportation Research*, vol. B34, no. 2, pp. 123–146.

Kotsialos A., Papageorgiou M., and Messmer A. (1999). “Optimal co-ordinated and integrated control motorway traffic control.” In: *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, Jerusalem.

Lebacque J.P. (1996). “The Godunov scheme and what it means for first order traffic flow models”. In: Lesort, J.B. (ed.), *Proceedings of the 13th International Symposium of Transportation and Traffic Theory*, Lyon, pp. 647–677.

Lighthill M. J. and Whitham G.B. (1955). “On kinematic waves II. A theory of traffic flow on long crowded roads.” In: *Proceedings of the Royal Society of London*, vol. 229A, pp. 317–345.

Ljung L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, New Jersey.

May A.D. (1990). *Traffic Flow Fundamentals*. Prentice-Hall, Englewood Cliffs, New Jersey.

Messmer A. and Papageorgiou M. (1990). “METANET: A macroscopic simulation program for motorway networks.” In: *Traffic Engineering & Control*, vol. 31, no. 8/9, pp. 466–470.

Nagel K. (1996). “Particle hopping models and traffic flow theory” In: *Physical Review*, vol. E53, pp. 4655–4672.

Newell G. (1993), “A simplified theory of kinematic waves in highway traffic, Part 1: General theory.” In: *Transportation Research B*, vol. 27B, no. 4, pp. 281–287.

Papageorgiou M., Blosseville J. M., and Hadj-Salem H. (1990a). “Modelling and real-time control of traffic flow on the southern part of boulevard périphérique in Paris: Part I: Modelling.” In: *Transportation Research A*, vol. 24A, no. 5, pp. 345–359.

Papageorgiou M. (1990b). “Dynamic modeling, assignment, and route guidance in traffic networks.” In: *Transportation Research B*, vol. 24B, no. 6, pp. 471–495.

Payne H.J. (1971). “Models of freeway traffic and control.” In: *Mathematical Models of Public Systems*, (G.A. Bekey, ed.), vol. 1, no. 1 of *Simulation Council Proceedings Series*, pp. 51–61, La Jolia, California.

Proost S., Van Dender K. (2001). “Welfare impacts of alternative policies to address atmospheric pollution in urban road transport” In: *Regional Science and Urban Economics*, forthcoming June 2001.

Richards P.I. (1956). “Shockwaves on the highway” In: *Operations Research*, vol. 4, pp. 42–51.

Vickrey W. S. (1963). “Pricing in urban and suburban transport.” In: *American Economic Review*, vol. 53, pp. 452–465.