Genetics of traffic assignment models for strategic transport planning

Bliemer, Michiel C J; Raadsen, MPH; Brederode, Luuk; Bell, MGH; Wismans, Luc J J; Smith, Mike J.

DOI
10.1080/01441647.2016.1207211

Publication date
2016

Document Version
Peer reviewed version

Published in
Transport Reviews: a transnational, transdisciplinary journal

Citation (APA)

Important note
To cite this publication, please use the final published version (if applicable). Please check the document version above.
Genetics of traffic assignment models for strategic transport planning

Michiel Bliemer¹*, Mark Raadsen¹, Luuk Brederode²³, Michael Bell¹, Luc Wismans²⁴, Mike Smith⁵

¹Institute for Transport and Logistics Studies, University of Sydney Business School, Australia
²DAT.Mobility, the Netherlands
³Department of Transport and Planning, Delft University of Technology, the Netherlands
⁴Centre for Transport Studies, University of Twente, the Netherlands
⁵Department of Mathematics, University of York, United Kingdom

*Corresponding author: T +61.2.9114.1840, E michiel.bliemer@sydney.edu.au

22 April 2015

Abstract

This paper presents a review and classification of traffic assignment models for strategic transport planning purposes by using concepts analogous to genetics in biology. Traffic assignment models share the same theoretical framework (DNA), but differ in capability (genes). We argue that all traffic assignment models can be described by three genes. The first gene determines the spatial capability (unrestricted, capacity restrained, capacity constrained, capacity and storage constrained) described by four spatial assumptions (shape of the fundamental diagram, capacity constraints, storage constraints, and turn flow restrictions). The second gene determines the temporal capability (static, semi-dynamic, dynamic) described by three temporal assumptions (wave speeds, vehicle propagation speeds, and residual traffic transfer). The third gene determines the behavioural capability (all-or-nothing, one shot, equilibrium) described by two behavioural assumptions (decision making and travel time consideration). This classification provides a deeper understanding of the often implicit assumptions made in traffic assignment models described in the literature. It further allows for comparing different models in terms of functionality, and paves the way for developing novel traffic assignment models.

Keywords

Traffic assignment, strategic transport planning, spatial assumptions, temporal assumptions, behavioural assumptions, fundamental diagram, model capabilities

Acknowledgment

This work was supported by the Australian Research Council under Grant LP130101048.
1. Introduction

1.1 Background

Traffic assignment models are used all over the world in strategic (long term) transport planning and project appraisal to forecast future traffic flows and travel times. Road authorities typically apply traditional models on large scale road networks for this purpose. These models describe the interaction between road travel demand (in particular passenger cars) and road infrastructure supply and were initially developed in the 1950s. The overall structure as depicted in Figure 1 has not changed much since (although solution algorithms have become more efficient). Traffic assignment models consist of a route choice sub-model that determines path flows and a network loading sub-model that propagates these path flows through the network and yields travel times. The route choice sub-model has a (possibly time-varying) origin-destination travel demand matrix as input, while the network loading sub-model considers infrastructure characteristics including road segment length, number of lanes, maximum speed, and possibly intersection layout and average green times of traffic controls.

Over the past few decades, there have been many new developments (especially in dynamic network loading models) leading to more advanced traffic assignment models that describe flows and travel times more realistically and (in certain ways) enhance their applicability. Such advancements can be categorised as being spatial, temporal, or behavioural in nature. We will refer to models incorporating such advancements as more capable models that have a larger ability to incorporate phenomena observed in reality.

Figure 1: Interaction between travel demand and infrastructure supply

There exists a wide range of traffic assignment models proposed in the literature, ranging from static to dynamic models, ranging from models that consider only free-flow conditions to models that consider congestion with queuing and spillback, and ranging from all-or-nothing assignment to equilibrium models. These models differ in capabilities, each making their own underlying assumptions.

In this paper we aim to disentangle some of the characteristics of traffic assignment models and explicitly state the assumptions underlying these models. Deeper insights in these assumptions allows a better understanding of the capabilities of each model and the circumstances under which models may reasonably be applied, as well as develop new more capable models.
1.2 Scope

In this paper we focus on capabilities of traffic assignment models with a focus on motorised private transport. This means we do not consider public transport or active modes of transport (such as walking and cycling). We would like to point out that “capability” is only one aspect when selecting suitable models for strategic transport planning. There are many other relevant aspects, such as ease of use (i.e., short run times, easy calibration, low input requirements), accountability (convergence of algorithms, existence and uniqueness of solutions, model complexity), and robustness (i.e., does the model generate stable outcomes). It is for example likely that a highly capable model has a higher computational complexity and less favourable solution properties, so a transport planning analyst should always balance these aspects when choosing a suitable model. We refer to Bliemer et al. (2013) for a more general discussion on these requirements for traffic assignment models.

We narrow the scope of this paper further by making the following eight limiting assumptions: (i) macroscopic description of traffic flow, (ii) only first order effects are considered, (iii) only pre-trip route choice is considered, (iv) no day-to-day dynamics are considered, (v) individual travellers are guided by selfish (non-cooperative) behaviour, (vi) inelastic travel demand, (vii) only a single user class is considered, and (viii) only travel time is considered in route choice.

The first five assumptions are made because the focus is on traffic assignment models for strategic transport planning purposes, which in general do not consider mesoscopic or microscopic representations of traffic flows (with possible random components), ignore dynamical second order effects (such as capacity drop, stop-and-go waves, and hysteresis), do not consider en-route travel decisions (which are more relevant for short term traffic operations), do not consider learning processes and disequilibria (partly due to difficulties when comparing base and future scenarios), and does not consider system optimal conditions (which can be useful for network design).

The last three assumptions are made to restrict ourselves to core components of traffic assignment models in which we assume a given travel demand (and do not include departure time choice, mode choice, destination choice, or other travel choices influencing demand) for a single user class (passenger cars) considering only travel time (and do not include tolls, travel time reliability, parking costs, etc.). These last three assumptions can be relaxed and are not strictly necessary for our framework, but they allow a more focussed presentation of the concepts in this paper. For example, one can replace travel time with a generalised cost or (dis)utility function that includes travel times and travel costs. Further, multiple user classes can be taken into account by considering different sensitivities to time and cost in these generalised cost functions (e.g., people with a high or low willingness to pay for travel time savings). Taking different vehicle types into account in a macroscopic model is usually more challenging due to asymmetric interactions between for example cars and trucks (see e.g. Bliemer and Bovy, 2003), which is partly why modellers often choose to convert all vehicle types into passenger car units.

1.3 Genetics

In this paper we describe the ‘genetics’ of traffic assignment models, which allows us to describe and characterise models in a qualitative fashion. Although the various traffic assignment models proposed in the literature may seem very different and sometimes incompatible, they share the same DNA and can be seen as descendants of the same ancestors having different genes.

In biology, DNA is the blueprint of life that consists of instructions that control the functions of cells. Each species (e.g., humans) shares more or less the same DNA. The building blocks of DNA are called nucleotides, which store genetic information. Genes describe basic functions of living organisms and consist of a specific sequence of nucleotides. The genetic code therefore describes all characteristics of the organism. DNA is inherited from parents through recombination, and evolves through mutation (i.e., genetic variation).
Traffic assignment models can be thought of as being characterised by a genetic code containing model assumptions and genes that describe functionality. Each traffic assignment model for strategic transport planning shares the same theoretical framework (namely, DNA). We identify three different genes: (i) a gene that describes spatial interactions, (ii) a gene that describes temporal interactions, and (iii) a gene that describes behaviour. These genes are composed of nucleotides that delineate each individual assumption that impacts on the functional capability of the model. By combining different temporal, spatial, and behavioural assumptions, different traffic assignment models are created.

A very capable organism with many positive characteristics is sometimes said to have ‘good genes’. Advanced traffic assignment models may be thought of as having ‘better’ genes than their simpler traditional counterparts regarding realism. An organism is defined by physical appearance and its behaviour, both defined by genes. In strategic macroscopic models, the network loading sub-model can be seen as a physical process in which traffic flow is modelled as a fluid following hydrodynamic theories. While traffic flow is a result of underlying individual driving behaviour (e.g., speed choice and lane choice), this level of behaviour is not described by macroscopic models; instead it is aggregated to a physical relationship through a cost function or the fundamental diagram of traffic flow (see Sections 2.1 and 3.2). Thus, the network loading sub-model is physical in nature and described by a spatial and temporal gene. In contrast, the route choice sub-model describes a behavioural process and is described by a, third, behavioural gene.

Just like living organisms, traffic assignment models have evolved over time, often by small mutations in one of the underlying assumptions, sometimes by recombination of existing models into a new model. By discovering basic underlying assumptions of each model (genetic code), we can investigate model functionality and limitations, as well as propose improved models. It also allows genetic modifications of existing models to develop novel models.

1.4 Paper outline

In Section 2 we describe the DNA of traffic assignment models, which allows us to classify each traffic assignment model. Section 3 describes the first gene using four nucleotides that represent the spatial assumptions. Section 4 describes the second gene, consisting of two nucleotides that represent the temporal assumptions. Section 5 discusses the third gene, consisting of two nucleotides representing behavioural assumptions. Section 6 establishes the genetic code for a selection of traffic assignment models proposed in the literature based on the spatial, temporal, and behavioural assumptions. Finally, we draw conclusions in Section 7 and state some potential for new model development.

2. DNA of traffic assignment models

In the literature, the main distinction that is often made between models is with respect to temporal assumptions, i.e. whether a model is static or dynamic. Dynamic models are typically seen as superior over static models. However, in terms of spatial interactions, certain static models are capable of accounting for queues and even spillback while certain dynamic models may not. Also, regarding the underlying route choice behaviour, some simple static models may be more advanced than certain dynamic models. We therefore need a more elaborate classification of traffic assignment models that describes their characteristics and capabilities in greater detail.

In this section we propose a unified theoretical framework (DNA) for traffic assignment models. This classification leads to model types and capabilities that result from three different genes that describe spatial, temporal, and behavioural assumptions, see Figure 2. Details of these underlying assumptions will be discussed in Sections 3, 4, and 5.

---

1 Although there is debate in the literature whether behaviour is determined by genes or by the environment (or both), in biology the field of study called behavioural genetics examines the origins of individual differences in behaviour.
Gene 1 describes the assumptions regarding spatial interactions, resulting in four distinct model classes (see Section 2.1). Gene 2 describes the assumptions regarding temporal interactions, resulting in three model classes (see Section 2.2). Finally, Gene 3 describes the behavioural assumptions, leading to three model classes (see Section 2.3). Combining the different model classes, the framework in Figure 2 describes in total 36 different model types, each with their own capabilities. The most capable model type according to this framework is a dynamic capacity and storage constrained equilibrium traffic assignment model, while the least capable model type is a static unrestrained all-or-nothing traffic assignment model. Each less capable model type is a special case of a more capable model type. In other words, less capable models can typically be derived from more capable models by making simplifying assumptions.

Figure 2: DNA of traffic assignment models

2.1 Model classes and capabilities resulting from spatial assumptions

As a result of spatial assumptions (Gene 1), the following model types are distinguished (in increasing order of capability):

- Unrestrained models;
- Capacity restrained models;
- Capacity constrained models;
- Capacity and (queue) storage constrained models.

The most capable traffic assignment models are models that constrain both the capacity (of flow) and the storage (of queues) on road segments. These models ensure that flow does not exceed capacity by diverting traffic to routes with spare capacity or by buffering vehicles in a physical queue. If the length of the queue exceeds the length of the road segment, the queue will spillback to upstream road segments. A capacity constrained model is a special case in which there are no constraints on the (queue) storage and as such spillback does not occur. An even more simplified model class is the capacity restrained model. In this model class, flows can exceed the physical road capacity and therefore queues are not described explicitly. To mimic the effect of queues (in these models) travel
times simply increase with increasing levels of flow. Finally, the simplest and least capable model is an unrestrained model with fixed (usually free flow) travel conditions and travel times.

Capacity restrained models are the most common model class in strategic transport planning, although the use of capacity (and storage) constrained models is gaining in popularity. Unrestrained models are rarely used. Each model class has different capabilities and a particular model should ideally only be used in cases where the underlying spatial assumptions are valid; however, as remarked above, there are many other factors which may influence model choice.

Figure 3 indicates a fundamental diagram describing the theoretical relationship between flow and density that can be empirically observed from traffic counts, and depends, among other things, on the number of lanes, the maximum speed limit, and the road type. Such a fundamental diagram may be assumed to hold for each cross-section on a homogeneous road segment (and is independent of the length of the road segment). Each point on this diagram represents a specific steady-state traffic state. While the diagram only shows flows (veh/h) and densities (veh/km), the speed of a vehicle (km/h) can be determined using the fundamental relationship that (space-mean) speed equals flow divided by density. For low densities (indicated by A and B in Figure 3) there is no congestion and no queues appear. Such traffic states are called hypocritical states (below the critical density) in which flow increases with density (i.e., throughput increases with more vehicles on the road). High densities (indicated by C and D) are a result of congestion and queues on the road. These traffic states are called hypercritical states in which flow decreases with density (i.e., throughput deteriorates with more vehicles on the road). The jam density provides an upper bound on the number of vehicles that can be stored on a certain road segment (assuming zero speed). For more information on the fundamentals of traffic flow theory and the fundamental diagram we refer to e.g. Cascetta (2009).

Figure 3: Spatial assumptions and model capabilities

Unrestrained models are only suitable for light traffic conditions (A) in which flow increases linearly with density, indicating that vehicles drive at maximum speed. Capacity restrained models are only suitable for light to medium traffic conditions (A and B) in which the flow does not exceed capacity, but some slight delays may occur due to increasing density. These models do not describe the hypercritical part of the fundamental diagram. Capacity constrained models are suitable for light to

---

2 In other words, this relationship only describes first order effects and does not explicitly describe transitions between traffic states (which requires explicit modelling of braking and acceleration as second order effects). As mentioned in Section 1.2, second order effects are usually not considered in large scale strategic transport planning for tractability reasons, but also to avoid illogical behaviour such as negative flows and traffic going backwards as outlined by Daganzo (1995b).
heavy traffic conditions (A, B, and C) in which short queues can form. These models cannot describe queues longer than the length of the road. Most capable is a capacity and storage constrained model, which can be applied to all traffic conditions (A, B, C, and D); including very heavy traffic when queues can grow longer than the road length and spillback to upstream road segments occurs.

Section 3 describes the underlying assumptions of these model classes in more detail.

2.2 Model classes and capabilities resulting from temporal assumptions

As a result of temporal assumptions (Gene 2), the following model types can be distinguished (in increasing order of capability):

- Static models;
- Semi-dynamic models;
- Dynamic models.

Dynamic models consider time-varying travel demand and multiple time periods for route choice and within each time period there exist (smaller) time steps for network loading in which flows are propagated through the network. These models explicitly account for variations over time in path flows, link flows, and travel times, and are the most capable models considered. Semi-dynamic models are special cases that only consider part of the dynamics. They often consider only a single time step for network loading within each route choice period, but may propagate traffic flows between route choice periods. Finally, static models are the simplest and least capable models that consider a stationary travel demand and only a single time period (with a specified or unspecified duration) for both route choice and network loading.

Some models are referred to as quasi-dynamic, which can be confusing. Quasi-dynamic models only consider a single time period and do not explicitly model time-varying flows. As such, these models are essentially static; they may be thought of as static models with certain dynamic elements (such as queues), see Miller et al. (1975) and Payne and Thompson (1975). Due to lack of a formal definition, we define quasi-dynamic models as static models that impose capacity and/or storage constraints and thereby can explicitly account for queues (similar to more advanced dynamic models).

Static models are the most common model class adopted for strategic transport planning purposes, although semi-dynamic models are used in some countries. Dynamic models are increasing in popularity, but applications for strategic planning purposes remain relatively rare due to the much higher model complexity and related needed computation times. As before, model classes defined by temporal assumptions have different capabilities and should ideally only be used in cases where these assumptions are valid; however, as remarked above, there are many other factors that may influence model choice.

Figure 4 illustrates how static, semi-dynamic, and dynamic models represent travel demand. The solid red line indicates the actual travel demand for a single origin-destination pair, and the grey bars represent the average demand in the model during each period. The areas of the grey bars (indicating the number of vehicles) are equal to the areas underneath the demand curves.

A static model considers a single time period, typically consisting of an entire peak period (e.g., a three hour period from 6.30am till 9.30am), and assumes that traffic outside this time period does not influence flows or travel times in the considered period. In other words, traffic in different periods can be assigned separately. Route choice proportions are assumed stationary during this period and network loading also considers a single time period in which all traffic reaches the destination and link flows are interpreted as average flows during this period.

---

3 Note that the line that separates traffic conditions C and D in Figure 2 is plotted somewhat arbitrary between the critical density and jam density since it is case specific, i.e. depends on the inflow rate and the link length.
In a semi-dynamic model multiple time periods are considered (e.g., one hour time slices, such as periods 6-7am, 7-8am, 8-9am, and 9-10am). It can be seen as a sequence of static models, however it takes the result from a previous period (such as vehicles in a queue) into account, for example by passing on residual traffic to the next period. As such, semi-dynamic models are more capable in describing travel demand variations as well as interactions of vehicles across time periods. Route choice proportions are assumed stationary during each time period, while network loading within each time period is usually done in a simple fashion similar to a static model. However, this typically does include the limitation that vehicles cannot be propagated for more than the duration of each period. In other words, vehicles that do not reach their destination within a single time period may be transferred to the next time period.

Dynamic models are capable of describing interactions between vehicles within and across each time period. They usually consider many smaller time periods (e.g., time slices of 15 minutes), which allows them to more accurately represent time-varying travel demand. Route choice proportions are typically assumed stationary during each time period. Network loading is much more sophisticated and similar to simulation models, i.e. they typically consider small time steps (e.g., 1 second) in which vehicles are propagated through the network.

Section 4 describes the underlying assumptions of these model classes in more detail.

Figure 4: Temporal interaction assumptions and model capabilities

2.3 Model classes and capabilities resulting from behavioural assumptions

As a result of behavioural assumptions (Gene 3), the following model types can be distinguished (in increasing order of capability):

- All-or-nothing models;
- One shot models;
- Equilibrium models.

Equilibrium models are the most capable models in which travellers consider congested travel times when choosing their route. In an equilibrium state, often referred to as a user equilibrium in which travellers are assumed to be non-cooperative (i.e., exhibit selfish behaviour), no traveller can unilaterally change routes to improve his or her travel time (Wardrop, 1952). This is in contrast to system optimal models that assume travellers cooperate and minimise the total (or average) travel time in the system. In this context, in this paper, only user equilibrium models are considered. One shot models are simplified models in which there is no feedback from previous travel time experience but rather a single network loading is performed based on initial path flow proportions. Such path flow proportions are either pre-determined or based on instantaneous travel times considering current traffic conditions. Finally, the simplest and least capable is an all-or-nothing model that is a special case
of a one shot model in which all travellers follow the fastest route based on given (typically free-flow) travel times.

Each of these model classes can be further differentiated into deterministic and stochastic models. Deterministic models usually assume perfect information, such that travellers base their decisions on actual travel times. In contrast, stochastic models assume imperfect information, such that travellers make decisions based on perceived travel times (Daganzo and Sheffi, 1977).

Equilibrium models are the most widely used model class in strategic transport planning, while system optimal assignments are mainly used to provide a benchmark solution. One shot models are often applied to simulate traffic using a more advanced (dynamic) network loading model based on route choice proportions from a simpler (static) model. All-or-nothing assignments, static or time-dependent, are not that common (anymore), but are often sub-models in equilibrium models.

Section 5 describes the underlying assumptions of these model classes in more detail.

3. Gene 1: Spatial assumptions

The first gene represents the spatial assumptions, which describe how traffic flows in network loading spatially interact and directly impact on the realism of the model (see also Figure 2). These spatial interactions are a combination of assumptions on the link level (shape of the fundamental diagram, capacity and storage constraints), and the node level (turn flow restrictions yielding turn reduction factors). These spatial interactions have been analysed separately or jointly in the literature and can be calibrated empirically.

The four specific assumptions (nucleotides) within this gene are summarised in Table 1 and are discussed in more detail in the following subsections. The nucleotide level refers to the spatial level at which interactions are described. The spatial assumptions of a traffic assignment model can be indicated using a sequence of letters representing the genetic code. For example, the most widely used assignment model for strategic transport planning purposes is a static capacity restrained model with the following code for Gene 1: CN-UU-U-N. The most sophisticated and capable model according to this classification is defined by genetic code CC-CC-C-F.

Table 1: Genetic code for Gene 1 (spatial assumptions)

<table>
<thead>
<tr>
<th>Nucleotide code for Gene 1 (spatial assumptions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide</td>
</tr>
<tr>
<td>Shape of the fundamental diagram</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Capacity constraints</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Storage constraints</td>
</tr>
<tr>
<td>Turn flow restrictions</td>
</tr>
</tbody>
</table>

3.2 Nucleotide 1 – Shape of the fundamental diagram

All traffic assignment models explicitly or implicitly assume a fundamental diagram. The shape of the fundamental diagram plays an important role in traffic flow theory and different shapes lead to different traffic patterns on a link (some more realistic than others). We indicate the maximum flow
through any part of a homogeneous road segment by physical road capacity $C$, also referred to as the saturation flow, which depends among other things on the number of lanes and the speed limit. The inflow and outflow capacity, however, are at best equal to $C$ and in many cases lower. For example, the outflow capacity may be restricted due to traffic controls and competing traffic (e.g., a merge) and the inflow capacity may be restricted due to spillback of a downstream bottleneck. This does not influence the fundamental diagram itself, but rather means that only specific traffic states on the diagram are observed in practice.

Figure 5: Shapes of the fundamental diagram

The fundamental diagram is generally defined by an increasing concave hypocritical branch (for densities lower than the critical density, indicated in blue in Figure 5, consistent with traffic conditions A and B in Figure 3) and a decreasing concave hypercritical branch (for densities higher than the critical density, indicated in red in Figure 5, consistent with traffic conditions C and D in Figure 3). The shape of such a general function can be indicated by CC using the coding from Table 1.

The first fundamental diagram was described by Greenshields (1935). He proposed a linear relationship between speed and density, which results in a quadratic fundamental diagram QQ, see Figure 5(b). Such a symmetric fundamental diagram may describe hypocritical traffic conditions quite accurately, but performs poorly for hypercritical states. A popular choice in traffic flow theory due to computational advantages has been an asymmetric triangular fundamental diagram LL (Newell, 1993) as shown in Figure 5(c). While a linear relationship in the hypercritical branch is often considered sufficiently realistic, a linear relationship in the hypocritical branch is less realistic (since it assumes that the speed at capacity is equal to the maximum speed). Therefore, piecewise linear fundamental diagrams PP as shown in Figure 5(d) have been proposed (e.g., Yperman, 2007), which maintain many of the computational benefits. A special case of such a piecewise linear fundamental diagram is the trapezoidal fundamental diagram (Daganzo, 1994) shown in Figure 5(e).
Diagrams shown in Figure 5(a)-(e) result in models with physical queues since they have a downward sloping hypercritical branch, while the diagrams in Figure 5(f)-(g) do not result in any queues since the hypercritical branch is absent. Other shapes of the hypercritical branch of the fundamental diagram have been proposed that result in specific types of queues. A fundamental diagram with a horizontal hypercritical branch as shown in Figure 5(h) is consistent with a model with vertical (non-spatial) queues, while a vertical hypercritical branch as shown in Figure 5(i) yields a model with horizontal (spatial) queues in which all queues are assumed to have a fixed queuing density, either equal to the jam density (leading to very compact queues) or some other fixed queuing density (Bliemer, 2007).

Fundamental diagrams have been used extensively in more advanced capacity and storage constrained dynamic traffic assignment models; in contrast, static models have mainly relied on link performance functions (also called volume-delay functions or travel time functions or cost-flow functions), which describe the relationship between link travel time and link flow (volume) or between speed and flow. Branston (1976) reviews link performance functions. The most well-known link performance function is the BPR link performance function (Bureau of Public Road, 1964). The corresponding fundamental diagram that is implicitly assumed is plotted in Figure 5(f). Two things can be observed from this CN shape. First, the BPR function gives rise to only the hypocritical branch of the fundamental diagram and ignores the hypercritical branch. Secondly, the hypocritical branch increases beyond the physical road capacity $C$, making it suitable only for capacity restrained models. Another popular choice in capacity restrained models is the conical link performance function proposed by Spiess (1990), which exhibits less rapid increases in link travel times when flows exceed capacity.

Davidson (1966) proposed a specific function in which the travel time goes to infinity as the flow approaches capacity (as suggested by Beckmann et al., 1956). Such a function is called a barrier function and guarantees that flows do not exceed the road capacities, hence this function can be used in a capacity constrained model. The corresponding fundamental diagram is shown in Figure 5(g) in which the hypocritical branch has a horizontal asymptote at capacity. However, this model may give rise to computational problems and perhaps unrealistic travel times when flow approaches capacity. Several others have discussed modifications to eliminate these problems (e.g., Daganzo, 1977; Taylor, 1984; Akçelik, 1991).

Link performance functions have also been used in several dynamic models (e.g., Janson, 1991; Friesz et al., 1993; Ran and Boyce, 1996; Bliemer and Bovy, 2003) in which travel times are calculated for vehicles at the time of link entrance (based on the flow at link entrance or all flows that previously entered or exited the link). These computed travel times, also referred to as predictive travel times, are then used to calculate the link exit times for flow propagation. Such link performance functions cannot realistically describe flows and travel times under (very) heavy traffic conditions (at densities $C$ and $D$ in Figure 3) since these functions do not represent the hypercritical branch of the fundamental diagram and do not explicitly describe queues.

### 3.3 Nucleotide 2 – Capacity constraints

Some models consider capacity constraints, while others assume no upper bounds on traffic flows. In case no constraints on the link entrance and exit flows are assumed, i.e., $UU$ in Table 1, no queues build up. This is consistent with fundamental diagrams of the shape shown in Figure 5(f). When considering both link entrance and exit capacity constraints, i.e. $CC$, these are typically set to the single physical link capacity $C$. In this case, residual queues will form upstream the bottleneck link. Some models consider $UC$, in which only link exit capacities are considered. In other words, flow is not restricted to flow in, but is restricted when flowing out. Such an assumption leads in some situations to queues inside the bottleneck link. Finally, models can also consider $CU$ with link entrance capacity constraints and no explicit outflow constraints.

### 3.4 Nucleotide 3 – Storage constraints

When the number of vehicles in a queue exceeds the available link storage, the queue will spill back to upstream links. The theoretical maximum number of vehicles that can be physically stored on a link...
should be equal to the jam density times the link length, although in moving queues (with a density lower than the jam density) the number of vehicles that can be present on the link is much lower. Some models do not consider spillback, thereby implicitly assuming no storage constraints (U). This essentially means an infinite jam density, which is consistent with the fundamental diagram presented in Figure 5(h). Models that take storage constraints into account (C) have a finite jam density, consistent with the fundamental diagrams in Figures 5(a)-(e) and 5(i).

3.5 Nucleotide 4 – Turn flow restrictions

Given that queues and travel time delays mainly arise due to interactions at the node level (i.e., merges, intersections), it is perhaps surprising to see that many static traffic assignment models and some dynamic models completely lack a node model description. In case there are no capacity constraints on the link entrance or exit flows, queues will never occur and hence a node model can often be omitted (N). In addition to node models (or sometimes instead of node models), junction models can be used to calculate additional delays per turn and may also impose turn capacities as well (based on junction configurations and controls).

In the presence of capacity constraints, node models determine the turn flows at intersections, merges, and diverges. Tampère et al. (2011) describes requirements for a first order node model for a node with any number of incoming and outgoing links. These requirements include flow maximisation, non-negativity, satisfying demand and supply constraints, satisfying the conservation of turn fractions (CTF) and the invariance principle (see Lebacque and Khoshyaran, 2005). Merge constraints that follow the capacity based weighted queueing rule (Ni and Leonard II, 2005) satisfy the invariance principle, in which the outflow rates are capacity proportional in case both in-links are congested. An often used merge constraint that does not satisfy the invariance principle is the fair merging rule in which inflow rates are demand proportional (Jin and Zhang, 2003).

Bliemer (2007) combines a first-in-first-out diverging rule and the fair merging rule into a closed form demand proportional model for general cross nodes. Several node models for general nodes have been proposed in the last decade (e.g., Jin and Zhang, 2004; Jin, 2012a; Jin, 2012b), none of them satisfy both CTF and the invariance principle and are therefore classified under other turn flow restrictions (O). More recently, models have been proposed that satisfy all requirements for first order node models (F), including CTF and the invariance principle, see e.g. Tampère et al. (2011), Flötteröd and Rohde (2011), Gibb (2011), and Smits et al. (2015).

4. Gene 2: Temporal assumptions

In this section we consider temporal assumptions in network loading identified in the second gene. Temporal assumptions determine whether a model is static, semi-dynamic, or dynamic. These assumptions consider interactions within time periods (wave speeds and vehicle propagation speeds) as well as across time periods (residual traffic transfer). They can be used to remove or simplify time dynamics within the model.

The three specific assumptions (nucleotides) within this gene are summarised in Table 2 and are discussed in more detail in the following subsections. Note that the level refers to the temporal level (within-period or across periods) at which the interactions are described. The temporal assumptions for traditional static models can be described by the following code for Gene 2: IN-IN-N. The most capable dynamic model is defined by genetic code KK-VV-T.
Table 2: Genetic code for Gene 2 (temporal interaction assumptions)

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Level</th>
<th>Type</th>
<th>Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave speeds</td>
<td>Within</td>
<td>Hypocritical</td>
<td>K, V, I</td>
<td>Kinematic / Vehicular / Infinite</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypercritical</td>
<td>K, I, Z, N</td>
<td>Kinematic / Infinite / Zero / Not available</td>
</tr>
<tr>
<td>Vehicle propagation</td>
<td>Within</td>
<td>Hypocritical</td>
<td>V, I</td>
<td>Vehicular / Infinite</td>
</tr>
<tr>
<td>speeds</td>
<td></td>
<td>Hypercritical</td>
<td>V, I</td>
<td>Vehicular / Infinite / Not available</td>
</tr>
<tr>
<td>Residual traffic transfer</td>
<td>Across</td>
<td>T, N</td>
<td>Transfer / No transfer</td>
<td></td>
</tr>
</tbody>
</table>

4.1 Nucleotide 5 – Wave speeds

Temporal interactions on a network are described by wave speeds as well as vehicle propagation speeds. Wave speeds are used to propagate traffic states through the network while vehicle propagation speeds describe how vehicles move through the network. Vehicle propagation speeds are discussed in the next nucleotide.

We first consider wave speeds in the hypocritical branch (i.e., forward waves). In the first order kinematic wave model proposed by LWR (Lighthill and Whitham, 1955; Richards, 1956), traffic conditions travel at the kinematic wave speed (K) equal to the slope of the hypocritical branch of the fundamental diagram as shown in Figure 6(a) for traffic flow rate $q$. It is important to realise that the speeds at which traffic states propagate and the speeds at which vehicles are propagated through the network are in general not the same. In case of a concave hypocritical branch, the kinematic wave speed is always smaller than (or equal to) the vehicular speed (V), which is equal to the flow divided by the density and hence equal to the slope of the line connecting the origin to the traffic state as shown in Figure 6(b). Only if the hypocritical branch is linear, these speeds are equal. More recent dynamic models consider kinematic wave speeds, but especially earlier dynamic models and semi-dynamic models consider vehicular speeds.

All static models simplify the within-period interactions by implicitly assuming infinite forward wave speeds (I) in which traffic states instantaneously propagate through the network and reach their destination within the single period. This situation is illustrated in Figure 6(c). This assumption effectively removes the necessity (and possibility) to track traffic states over time.

Figure 6: Speeds in hypocritical branch

Backward waves track how traffic states in the hypercritical branch propagate backwards on a road segment, and are responsible for queue build up and possible spillback to upstream road segments. In the LWR model traffic conditions travel at the (negative) kinematic wave speed (K) equal to the slope of the hypercritical branch of the fundamental diagram as shown in Figure 7(a) for traffic state $q$. 
Similar to forward waves, it requires a dynamic model to explicitly deal with the effects of such backward kinematic waves over time.

An unconstrained static model gives rise to a fundamental diagram, which does not have a hypercritical branch; and so backward wave speeds are not available (N). A capacity constrained static model however does give rise to a hypercritical branch. In these fundamental diagrams two different temporal assumptions regarding backward waves can be made (since the time dimension does not exist in a static model). The most widely adopted assumption is that backward wave speeds are zero (Z) as shown in Figure 7(b). In this case, traffic conditions never move backwards, which usually means vertical non-spatial queues and no spillback. (Note that stationary physical queues are also consistent with zero backward wave speeds.) The zero speed assumption is consistent with fundamental diagrams of the shape shown in Figure 5(h). Another assumption is that there is a (negative) infinite speed (I) as depicted in Figure 7(c); this allows the model to describe spillback when the number of vehicles in the queue exceeds the available link storage. Note that an infinite backward wave speed does not mean that queues build up indefinitely, since the length of the queue is constrained by the number of vehicles in the queue. The fundamental diagram in Figure 5(i) is consistent with the infinite speed assumption.

4.2 Nucleotide 6 – Vehicle propagation speeds

Instead of looking at the speeds at which traffic states propagate, we now look at the assumption on the speed with which vehicles propagates on a road segment. As mentioned in the previous section, traffic states and vehicles in general do not move at the same speed.

In the hypocritical branch, traffic states and vehicles both move forward, but vehicles never move slower than traffic states (see Figure 6). In static models, the vehicle propagation speed is assumed to be infinite (I) such that vehicles move instantaneously through the network within a single time period. Note that although vehicles are propagated instantaneously in static models, this does not mean that the travel times are zero, since travel times are calculated separately from vehicular speeds. In contrast, dynamic models consider finite vehicular speeds (V), such that travel times are consistent with vehicle propagation speeds.

Traffic states in the hypercritical branch (if considered in the model) move upstream (i.e., have a negative speed), while vehicles move downstream (i.e., have a positive speed), see Figure 7. In dynamic models the vehicle propagation speed is assumed to be equal to the finite vehicular speed (V). In static models that do not describe residual queues the vehicle propagation speed is implicitly assumed to be infinite (I), however, in static models that consider residual queues, the vehicle propagation speed is assumed to be finite and set to the vehicular speed (V). Note that this does not make the model dynamic since it only requires applying capacity and storage constraints to traffic flows instead of explicitly tracking vehicles over time.

Figure 7: Speeds in hypercritical branch
4.3 Nucleotide 7 – Residual traffic transfer

Residual traffic at the end of a time period results when vehicles are not able to reach their final destination within the considered time period (or the smaller network loading time step). These residual vehicles are either (i) in a residual queue due to a bottleneck downstream, or (ii) simply are not able to reach their final destination because the travel time to reach the destination is longer than the considered time period. Residual traffic influences traffic flows and travel times in the next time period. This dependency of traffic across time periods can be eliminated by assuming that any residual traffic has no impact on the next time period, in other words, assuming that the network is empty at the beginning of each time period.

Dynamic models transfer all traffic (T), thereby describing the full temporal interactions within and across time periods. Static models have just one (fairly long) time interval and so do not consider residual traffic transfer (N). Thus static models are unsuitable for modelling short time periods in a congested network. The main difference between static and semi-dynamic models is that the latter does assume residual traffic transfer across time periods as discussed in Section 2.2.

5. Gene 3: behavioural assumptions

The third and final gene represents the behavioural assumptions, which describe travellers’ route choice. From biology we know that describing which genes affect behaviour is difficult, since behavioural characteristics are complex and polygenic (i.e., influenced by multiple genes). The same holds for describing route choice behaviour in traffic assignment models, and many types of behaviours have been described in the literature.

In this section we put route choice behaviour into a single gene with two nucleotides as summarised in Table 3 and discussed in more detail in the following subsections. We note that while we try to be as inclusive as possible, this list is not exhaustive and is limited by the scope set out in Section 1.2 (for example, we do not consider day-to-day learning effects). The most capable model considered is a (equilibrium) model with the following code for Gene 3: BI-E, while the simplest model is a (all-or-nothing) model defined by genetic code FP-I.

Table 3: Genetic code for Gene 3 (behavioural assumptions)

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Type</th>
<th>Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision making</td>
<td>Rationality</td>
<td>F, B</td>
<td>Full / Bounded</td>
</tr>
<tr>
<td></td>
<td>Information</td>
<td>P, I</td>
<td>Perfect / Imperfect</td>
</tr>
<tr>
<td>Travel time consideration</td>
<td>I, P, E</td>
<td>Instantaneous / Predictive / Experienced</td>
<td></td>
</tr>
</tbody>
</table>

5.1 Nucleotide 8 – Decision making

Decision making behaviour has many dimensions. We limit ourselves to the ones that have most often been used in the context of route choice, namely rationality, uncertainty, and motivation.

In terms of rationality, most traffic assignment models consider full rationality (F) which assumes that travellers consider all alternatives and eventually all travellers select their own best routes. In reality, travellers are unlikely to behave in such an optimal way due to resistance in change (inertia effects) and the fact that people often minimise effort and time in decision making. Bounded rationality (B) is a term that is often used to describe such decision making behaviour, which includes habitual route choice, or route choice in which travellers expose satisficing behaviour and consider routes with travel times sufficiently close to the fastest route travel time (see e.g., Di et al., 2013; Han et al., 2015).
If travellers have perfect information (P), then decision making can be described by a deterministic process. In contrast, if travellers are considered to have imperfect information (I) with a given level of uncertainty, then decision making is referred to as probabilistic or stochastic. For example Fisk (1980) proposed a stochastic assignment model that adopts a logit model, Zhou et al. (2012) adopt a C-logit model, and Kitthamkesorn and Chen (2013) adopt a path-size weibit model, where the latter two aim to correct the path choice probabilities for path overlap. Deterministic models can be seen as special cases of stochastic models where the level of uncertainty is equal to zero.

Although outside of the scope, we point out that travellers may be driven by different motivations for choosing a certain route. As stated in Section 1.2, here we only consider selfish drivers who minimise their individual travel time leading to a user equilibrium based model. Other models exist in which drivers are guided by different motivations, yet these models are hardly ever used in the context of strategic transport planning.

5.2 Nucleotide 9 – Travel time consideration

In (semi-)dynamic models, different types of path travel times can be considered in route choice, see e.g., Ran and Boyce (1996) and Buisson et al. (1999). Instantaneous path travel times (I) for a certain departure time consider only the traffic states at this time instant and the corresponding link travel times, and hence ignores any changes in traffic conditions while driving. Models that consider instantaneous travel times are often referred to as reactive. Predictive path travel times (P) consider the addition of link travel times based on the traffic conditions at the time of link entrance, hence time-varying traffic conditions along the path are taken into account. Such travel times can be considered as an estimate, since changing traffic conditions while traversing the link are ignored. More recent models calculate experienced travel times (E), which consider the actually experienced link travel times at the time of link exit (instead of link entrance). In static models (in which no such differences in path travel times exist) we assume that travel times are instantaneous.

6. Classification of existing traffic assignment models

Many traffic assignment models have been proposed in the literature that we can classify using the nine nucleotides in the three genes. Table 4 provides a list of some prototypical models described in the literature, which is by no means intended to be complete.

Looking at temporal assumptions, all static models assume infinite wave and vehicle propagation speeds in the hypocritical branch and no residual traffic transfer. In case a hypercritical branch of the fundamental diagram is considered, either zero or infinite backward wave speeds are assumed, and vehicle propagation speeds equal to vehicular speeds or infinity. On the other hand, all dynamic models assume forward wave speeds that are not infinite, i.e. either equal to the vehicular speed or kinematic wave speed. Backward wave speeds are equal to the kinematic wave speeds and follow the shape of the fundamental diagram (and can therefore be equal to zero or infinity if the hypercritical branch of the fundamental diagram is horizontal or vertical, respectively). Vehicle propagation speeds are equal to the vehicular speed in both the hypocritical and the hypercritical branch (if considered). Further, dynamic models assume residual traffic transfer.

Regarding behavioural assumptions, all models in Table 4 are (user) equilibrium models. Exceptions are Bovy (1990) who describes a one shot model for uncongested situations, while Daganzo (1994, 1995a), Yperman et al. (2005), Bliemer (2007) and Gentile (2010) mainly describe the network loading sub-model and omit behavioural route choice information.

Finally, looking at spatial assumptions, many models are capacity restrained using a strictly increasing link performance function, although more recently several capacity constrained models have been proposed that can explicitly account for queues. Relatively few models are storage constrained in which spillback is described. A wide variety of shapes of fundamental diagrams has been used. More advanced models include turn flow restrictions through the incorporation of a node model, which allow more realistic queueing and spillback of traffic.
Semi-dynamic models are neither completely static nor completely dynamic. This means with respect to the temporal assumptions that they typically assume a sequence of connected static models as described in Nakayama and Connors (2014). In such a case, wave speeds and vehicle propagation speeds in the hypocritical branch are infinite. However, vehicle propagation speeds in the hypercritical branch are considered finite and vehicles that reside in a queue at the end of a time period are transferred to the next time period. We have omitted semi-dynamic models from the list in Table 4 because the papers are either in Japanese (Fujita et al., 1988; Fujita et al., 1989; Miyagi and Makimura, 1991; Akamatsu et al., 1998; Nakayama, 2009) or have been described as operational procedures and algorithms rather than mathematical problems (e.g., Van Vliet, 1982; Davidson et al., 2011), which makes them difficult to classify accurately.

7. Discussion and conclusions

In this paper we have presented a theoretical framework, which classifies traffic assignment models for strategic transport planning purposes. This framework is described in terms of a genetic code with three genes and nine nucleotides consisting of four spatial assumptions, three temporal assumptions, and two behavioural assumptions. This framework leads to in total 36 different model types, each with their own underlying assumptions and their own capabilities.

As a special case, the widely applied capacity restrained equilibrium static traffic assignment model can be derived by assuming (i) a concave hypocritical part and no hypercritical part of the fundamental diagram, (ii) no flow capacity constraints, (iii) no storage constraints, (iv) no turn flow restrictions, (v) infinite forward wave speeds and no backward waves, (vi) infinite vehicle propagation speeds, and (vii) no residual traffic transfer, (viii) perfectly rational travellers with full information, and (ix) instantaneous travel time consideration. Such strict assumptions limit the capability and hence realism of this particular model in certain instances. At the same time, we acknowledge that more capable models often have other less favourable characteristics, such as higher computational complexity and possible non-uniqueness of solutions. As a result, transport planners may decide to choose less capable models, but should be aware of model limitations when interpreting outputs.

Capacity constrained (and possibly also storage constrained) models are more capable and can explicitly describe queues (and possibly spillback). Several sophisticated dynamic models exist that are capable of describing flows and travel times under all traffic conditions. Such static models also exist, which extend the capability (realism) of static models in congested situations by sharing the same spatial assumptions made in advanced dynamic models. This opens up possibilities for static models that are derived from advanced dynamic models by simply using static temporal assumptions. Therefore, the framework described in this paper may not only be useful for classifying models, but also for developing new models with new genetic codes by combining different spatial, temporal, and behavioural assumptions (and hence inherit genetic properties from other models).
Table 4: Overview of assumptions made in different traffic assignment models proposed in the literature

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>fundamental diagram</td>
<td>wave speeds</td>
<td>decision making</td>
</tr>
<tr>
<td>capacity constraints</td>
<td>vehicle prop. speeds</td>
<td>travel time consideration</td>
</tr>
<tr>
<td>storage constraints</td>
<td>residual traffic transfer</td>
<td></td>
</tr>
<tr>
<td>turn flow restrictions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Static models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Spatial Assumptions</th>
<th>Temporal Assumptions</th>
<th>Behavioural Assumptions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bovy (1990)</td>
<td>LN</td>
<td>IN</td>
<td>FI</td>
<td>I</td>
</tr>
<tr>
<td>Beckmann et al. (1956)</td>
<td>CN</td>
<td>IN</td>
<td>FP</td>
<td>I</td>
</tr>
<tr>
<td>Irwin et al. (1961)</td>
<td>CN</td>
<td>IN</td>
<td>FP</td>
<td>I</td>
</tr>
<tr>
<td>Fisk (1980)</td>
<td>CN</td>
<td>IN</td>
<td>FI</td>
<td>I</td>
</tr>
<tr>
<td>Smith (1987)</td>
<td>LH</td>
<td>IN</td>
<td>FP</td>
<td>I</td>
</tr>
<tr>
<td>Bell (1995)</td>
<td>LH</td>
<td>IZ</td>
<td>FI</td>
<td>I</td>
</tr>
<tr>
<td>Bifulco and Crisalli (1998)</td>
<td>CH</td>
<td>IZ</td>
<td>FI</td>
<td>I</td>
</tr>
<tr>
<td>Lam and Zhang (2000)</td>
<td>CH</td>
<td>IZ</td>
<td>FI</td>
<td>I</td>
</tr>
<tr>
<td>Zhou et al. (2012)</td>
<td>CN</td>
<td>IZ</td>
<td>FI</td>
<td>I</td>
</tr>
<tr>
<td>Smith (2013)</td>
<td>LH</td>
<td>IZ</td>
<td>FI</td>
<td>I</td>
</tr>
<tr>
<td>Smith et al. (2013)</td>
<td>CV</td>
<td>IZ</td>
<td>FI</td>
<td>I</td>
</tr>
<tr>
<td>Bliemer et al. (2014)</td>
<td>LC</td>
<td>IZ</td>
<td>FI</td>
<td>I</td>
</tr>
</tbody>
</table>

**Dynamic models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Spatial Assumptions</th>
<th>Temporal Assumptions</th>
<th>Behavioural Assumptions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Janson (1991)</td>
<td>CN</td>
<td>VN</td>
<td>FP</td>
<td>I</td>
</tr>
<tr>
<td>Daganzo (1994, 1995a)</td>
<td>PL</td>
<td>KK</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Chen and Hsueh (1998)</td>
<td>CN</td>
<td>VN</td>
<td>FP</td>
<td>P</td>
</tr>
<tr>
<td>Li et al. (2000)</td>
<td>LH</td>
<td>KZ</td>
<td>FP</td>
<td>I</td>
</tr>
<tr>
<td>Chabini (2001)</td>
<td>CN</td>
<td>KN</td>
<td>FP</td>
<td>P</td>
</tr>
<tr>
<td>Bliemer and Bovy (2003)</td>
<td>CN</td>
<td>KN</td>
<td>FP</td>
<td>P</td>
</tr>
<tr>
<td>Yperman et al. (2005)</td>
<td>LL</td>
<td>KK</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Bliemer (2007)</td>
<td>CV</td>
<td>KI</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Gentile (2010)</td>
<td>CC</td>
<td>KK</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Friesz et al. (2013)</td>
<td>CH</td>
<td>KZ</td>
<td>FP</td>
<td>E</td>
</tr>
<tr>
<td>Han et al. (2015)</td>
<td>LL</td>
<td>KK</td>
<td>BP</td>
<td>E</td>
</tr>
</tbody>
</table>
References


