WORKING PAPER

ITLS-WP-16-01

Genetics of traffic assignment models for strategic transport planning

By
Michiel Bliemer, Mark Raadsen, Luuk Brederode³, Michael Bell, Luc Wismans⁵ and Mike Smith⁶

³ Delft University of Technology, the Netherlands
⁵ University of York, United Kingdom

January 2016

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council’s Key Centre Program.
Genetics of traffic assignment models for strategic transport planning

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 functionality (genes). We argue that all traffic assignment models can be described by two genes. The first gene determines the spatial functionality (unrestricted, capacity restrained, capacity constrained, capacity and storage constrained) described by five spatial interaction assumptions, while the second gene determines the temporal functionality (static, semi-dynamic, dynamic) described by two temporal interaction assumptions. This classification provides a deeper understanding of the often implicit assumptions made in traffic assignment models described in the literature, particularly with respect to networking loading where the largest differences occur. It further allows for comparing different models in terms of functionality, and opens the way for developing novel traffic assignment models.

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

Michiel Bliemer, Mark Raadsen, Luuk Brederode, Michael Bell, Luc Wismans and Mike Smith

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

INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H73)

The University of Sydney NSW 2006 Australia
Telephone: +61 9114 1824
E-mail: business.itlsinfo@sydney.edu.au
Internet: http://sydney.edu.au/business/itls

January 2016
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 travel demand and infrastructure supply and were initially developed in the 1950s. The overall “equilibrium” 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 model that propagates these path flows through the network and yields travel times.

Over the past few decades, there have been several new developments (especially in dynamic network loading models) leading to more advanced traffic assignment models that describe flows and travel times more realistically and dynamically. However, these sophisticated models need significantly more calculation time and memory to run and are generally much more difficult to program and to calibrate.

Figure 1: Equilibrium 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 and ranging from models that consider free-flow conditions to models that consider congestion with queuing and spillback. These models differ in computational complexity and realism, each making their own simplifying assumptions. However, these assumptions are often implicit. In this paper we aim to disentangle some of the characteristics of traffic assignment models and explicitly state the assumptions underlying these models, with a particular focus on network loading in which most models differ. 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 and realistic models.

1.2 Scope

We narrow the scope of this paper by making the following limiting assumptions: (i) macroscopic description of traffic flow, (ii) within-day equilibrium, no day-to-day dynamics, (iii) only first order effects are considered, (iv) inelastic travel demand, (v) only a single user class is considered, and (vi) only travel time is considered in route choice.
The first three 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), assume a user-equilibrium solution in order to compare scenarios, and ignore dynamical and second order effects (such as capacity drop, stop-and-go waves, and hysteresis). The last three assumptions are made to restrict ourselves to core components of traffic assignment models in which we aim to find a route choice equilibrium with 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, or passenger car equivalents) considering only travel time (and do not include other generalised cost components). 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.

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 children of the same ancestors having different genes.

In biology, DNA is a 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 (DNA). We identify two different genes, namely one gene that describes spatial interactions, and one gene that describes temporal interactions. These genes are composed of nucleotides that delineate each individual assumption that impacts the functional capability of the model. By combining different temporal and spatial interaction assumptions, different traffic assignment models can be 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. 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. Sections 3 describes the first gene using five nucleotides that represent the spatial interaction assumptions. Section 4 describe the second gene, consisting of two nucleotides that represent the temporal interaction assumptions. Section 5 touches upon travel time calculation as an important component of any traffic assignment model. Section 6 classifies a selection of traffic assignment models proposed in the literature based on the spatial and temporal interaction assumptions. Finally, we draw conclusions in Section 7 and state the 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 interaction 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. We therefore need a more elaborate classification of traffic assignment models that describe their characteristics and capabilities.

In this section we propose a unified theoretical framework (DNA) for traffic assignment models.

2.1 Model types as a result of two genes

Traffic assignment models can be classified according to different model types that result from two different genes, namely spatial and temporal interaction assumptions, see Figure 2. Details of these assumptions will be discussed in Sections 3 and 4.

As a result of spatial interaction assumptions (Gene 1), the following model types are distinguished:

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

In unrestrained models, traffic flow is not interrupted and all travellers choose the route with the (perceived) lowest free-flow travel time.

In all other model types, travellers may experience delays due to other vehicles on the road, which affects route choice. These latter models adopt the notion of some sort of user equilibrium (e.g., Wardrop, 1952; Daganzo and Sheffi, 1977) in which individual travellers aim to minimise their (expected) travel time. Beckmann et al. (1956) formulated mathematical conditions for finding such user equilibria in general transportation networks in a deterministic setting.

In capacity restrained models, queues do not exist and traffic flows can exceed the physical road capacity, although link travel times may rise steeply for flows beyond the capacity.

Capacity constrained models impose constraints on the traffic flows, such that traffic flows never exceed road capacities. Instead, queues appear or traffic flow is diverted to other routes with spare capacity. However, this model type does not put a constraint on the number of vehicles than can be ‘stored’ on the road and therefore cannot describe queue spillback.

The most capable traffic assignment models are models that are both capacity and storage constrained, because they take both physical constraints (capacity and storage) into account. Capabilities of models due to spatial interaction assumptions are further discussed in Section 2.2.

As a result of temporal interaction assumptions (Gene 2), the following model types can be distinguished:

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

Static models consider a single time period (having either a finite or an infinite duration) for both route choice and network loading. Travel demand and traffic flows are considered to be stationary or assumed to represent average traffic flows.

Dynamic models consider multiple time periods for travel demand and route choice and within each of these periods consider multiple (smaller) time periods for network loading in which flows are propagated.
from period to period. Therefore, dynamic models explicitly account for variations over time in path flows and link flows.

Semi-dynamic models consider more than one time period for route choice, but are not completely dynamic. They often only consider a single time period for network loading within each route choice period, but may propagate traffic between route choice periods. Capabilities of models due to temporal interaction assumptions are further discussed in Section 2.3.

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 dynamic models with certain stationary elements (such as demand or link flows), see Miller et al. (1975) and Payne and Thompson (1975). Static models are called ‘quasi-dynamic’ if they impose capacity and/or storage constraints and thereby explicitly account for queues (similar to advanced dynamic models).

![Figure 2: DNA of traffic assignment models](image)

Combinations of spatial and temporal assumptions lead to twelve specific model classes. The simplest model class is a static unrestricted traffic assignment model. This model essentially performs an all-or-nothing assignment. The most elaborate model class is a dynamic capacity and storage constrained traffic assignment model. Capabilities of the model increases from top left to bottom right in Figure 2.

### 2.2 Spatial interaction assumptions and model capabilities

The model types associated with different spatial interaction assumptions discussed in Section 2.1 have different capabilities and should ideally only be used in cases where these assumptions are valid. Figure 3 indicates a fundamental diagram describing the relationship between flow and density that can be empirically observed from traffic counts on a road segment. For low densities (indicated by A and B), flow increases with density and no queues appear. High densities (indicated by C and D) are a result of congestion in which flow decreases with density. Traffic states A and B are called hypocritical states and traffic states C and D are called hypercritical states (Cascetta, 2009).
Flow is constrained by capacity and density is constrained by the jam density. The jam density provides an upper bound on the number of vehicles that can be stored on a certain road segment (assuming zero speed).

Unrestricted models are only suitable for light traffic conditions (A) in which flow increasing 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. These models cannot describe the hypercritical part of the fundamental diagram, such that the flow will continue to increase with increasing density. Capacity constrained models are suitable for light to 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). In very heavy traffic, queues can grow longer than the road length and spillback to upstream road segments.

2.3 Temporal interaction assumptions and model capabilities

The model types associated with different temporal interaction assumptions discussed in Section 2.1 have different capabilities and should ideally only be used in cases where these assumptions are valid. 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). It is assumed that traffic outside this time period does not influence flows or travel times in the peak period. Therefore, the time period needs to be sufficiently large such that it contains the entire period during which the demand is above capacity, but also sufficiently small in order to avoid spreading the travel demand too evenly over time (e.g., peak demand may then fall below capacity). Route proportions are assumed stationary during this period. Network loading also considers a single time period in which all traffic reaches the destination and all link flows should be interpreted as average flows during this period.

1 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 somehow 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 better capable of describing variations in travel demand and the interactions of vehicles across time periods. Route proportions are assumed stationary during each time period. Network loading within each time period is usually done in a simple fashion similar to a static model with the limitation that vehicles cannot be propagated more than the duration of each period. In other words, vehicles may not reach their destination within a single time period and may be transferred to the next time period. SATURN was one of the first semi-dynamic models; see Van Vliet (1982).

Dynamic models are the most capable models and usually consider many smaller time periods (e.g., time slices of 15 minutes). This allows dynamic models to more accurately represent time-varying travel demand. Route proportions are again assumed stationary during each time period. Network loading is much more sophisticated and similar to simulation models they typically consider small time steps (e.g., 1 second) in which vehicles are propagated on the network. Compared to static and semi-dynamic models, dynamic models are able to describe more detailed interactions between vehicles within each time period.

3. Gene 1: Spatial interaction assumptions

The first gene represents the spatial interaction assumptions, which describe how traffic flows spatially interact and directly impact the realism of the model (see also Figure 2). These spatial interactions are a combination of assumptions on the path level (route choice behaviour), 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 five 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 interaction 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 restraint model with the following code for Gene 1: FP-CN-UU-U-N. The most sophisticated and capable model according to this classification is defined by genetic code BI-CC-CC-C-F.
Table 1: Genetic code for Gene 1 (spatial 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>Route choice behaviour</td>
<td>Path</td>
<td>Rationality</td>
<td>F, B</td>
<td>Full / Bounded</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Information</td>
<td>P, I</td>
<td>Perfect / Imperfect</td>
</tr>
<tr>
<td>Shape of the fundamental diagram</td>
<td>Link</td>
<td>Hypocritical branch</td>
<td>L, P, Q, C</td>
<td>Linear / Piecewise linear / Quadratic / Concave</td>
</tr>
<tr>
<td></td>
<td>Link</td>
<td>Hypercritical branch</td>
<td>L, P, Q, C, H, V, N</td>
<td>Linear / Piecewise linear / Quadratic / Concave / Horizontal / Vertical / Not available</td>
</tr>
<tr>
<td>Capacity constraints</td>
<td>Link</td>
<td>Inflow</td>
<td>U, C</td>
<td>Unconstrained / Constrained</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Outflow</td>
<td>U, C</td>
<td>Unconstrained / Constrained</td>
</tr>
<tr>
<td>Storage constraints</td>
<td>Link</td>
<td>U, C</td>
<td></td>
<td>Unconstrained / Constrained</td>
</tr>
<tr>
<td>Turn flow restrictions</td>
<td>Node</td>
<td>F, O, N</td>
<td></td>
<td>First order / Other / No restrictions</td>
</tr>
</tbody>
</table>

3.1 Nucleotide 1 – Route choice behaviour

Each traffic assignment model makes assumptions regarding route choice. Often it is assumed that travellers have perfect knowledge regarding travel times and are fully rational, i.e. FP route choice, leading to an assignment model that aims to find a deterministic user equilibrium. If travellers are assumed to have perception errors when making route choice decisions (i.e., imperfect knowledge), a stochastic user equilibrium can be adopted. For example Fisk (1980) proposed an assignment model that adopts a logit model, Zhou et al. (2012) adopts a C-logit model, and Kitthamkesorn and Chen (2013) adopt a path-size weibit model in order to describe FI route choice, where the latter two aim to correct the path choice probabilities for path overlap. Others assume that travellers are boundedly rational, in which drivers are assumed not to change route until the cost savings exceed a certain threshold (Han et al., 2015), resulting in a BP route choice model.

Another distinction can be made regarding the type of path travel time considered in route choice, see e.g., Ran and Boyce (1996). Under the assumption of instantaneous travel times, travellers consider only the traffic conditions at the time of departure. Clearly, traffic conditions may change over time.

3.2 Nucleotide 2 – 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). A fundamental diagram is often directly estimated using cross-sectional traffic counts and space-mean speeds. 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 others 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 downstream bottleneck. This does not influence the fundamental diagram itself, but rather means that some traffic states on the diagram may be observed in practice.
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(a), 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(a), 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 4(b). Such a symmetric fundamental diagram may describe hypocritical traffic conditions quite accurately, but performs poorly for hypercritical states because the range of densities observed part of hypocritical traffic conditions is much larger. 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 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)-(c) 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 dynamic traffic assignment models, however, static models have mainly relied on link performance functions (also called volume-delay functions or simply travel time functions), which describe the relationship between link travel time and link flow (volume) or between the 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 only contains 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 for capacity restraint models. Another popular choice in capacity restraint models is the conical link performance function proposed by Spiess (1990), which exhibits less rapid increases in link travel times when flows exceed capacity.

Beckmann et al. (1956) was the first to describe a relationship between flow and travel time in which the travel time goes to infinity when flow reaches capacity, although they did not propose a specific functional form. In operations research, such a function is called a barrier function and guarantees that flows do not exceed the road capacities. Davidson (1966) proposed a specific link performance function in which the travel time goes to infinity as the flow approaches capacity (as suggested by Beckmann et al., 1956). 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 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 (see Section 5.2), 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. Models that more realistically describe (very) heavy traffic conditions first propagate the flow (using the fundamental diagram and first order traffic flow theory or queuing theory) and calculate travel times afterwards, also referred to as experienced travel times (see Section 5.2). More recent dynamic models adopt this strategy (e.g., Bliemer, 2007; Yperman, 2007; Gentile, 2010).

### 3.3 Nucleotide 3 – 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., $U$, 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. $C$, 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 $U$, 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 only consider $C$, which consider link entrance capacities and no explicit outflow constraints.
3.4 Nucleotide 4 – 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 diagrams 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 5 – Turn flow restrictions

Given that queues and travel time delays are mainly determined by node models, it is perhaps surprising to see that almost all 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 describe turn flow interactions. Junction models typically 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 queuing 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. Often used merge constraints that satisfy the fair merging rule (Jin and Zhang, 2003) in which inflow rates are demand proportional, do not satisfy the invariance principle.

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).

Recently, models have been proposed that satisfy all requirements for first order node models, including CTF and the invariance principle, see e.g. Tampère et al. (2011), Flötteröd and Rohde (2011) and Gibb (2011). These models mainly differ in the merge constraints. Recently, Smits et al. (2015) described a family of macroscopic first order node models (F), which includes the three node models mentioned above.

4. Gene 2: Temporal interaction assumptions

In this section we consider temporal interaction assumptions identified in the second gene that determines whether a model is static, semi-dynamic, or dynamic. These assumptions consider interactions within time periods (forward and backward wave speeds and flow propagation speeds) as well as across time periods (residual traffic transfer). Each of these assumptions aims to remove or simplify time dynamics within the model, in particular in the network loading model.

The three specific assumptions (nucleotides) within this gene are summarised in Table 2 and are discussed in more detail in the following subsections, where level refers to the temporal level (within-period or across periods) at which the interactions are described. The temporal interaction assumptions for traditional static traffic assignment models can be described by the following code for Gene 2: IN-I-N. The most capable dynamic model is defined by genetic code KK-V-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>Forward</td>
<td>K, V, I</td>
<td>Kinematic / Vehicular / Infinite</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Backward</td>
<td>K, I, Z, N</td>
<td>Kinematic / Infinite / Zero / Not available</td>
</tr>
<tr>
<td>Flow propagation speeds</td>
<td>Within</td>
<td></td>
<td>V, I</td>
<td>Vehicular / Infinite</td>
</tr>
<tr>
<td>Residual traffic transfer</td>
<td>Across</td>
<td></td>
<td>T, N</td>
<td>Transfer / No transfer</td>
</tr>
</tbody>
</table>

4.1 Nucleotide 6 – Wave speeds

Temporal interactions on a network are described by forward and backward wave speeds as well as flow propagation speeds. Wave speeds are used to propagate traffic states through the network while flow propagation speeds describes how vehicles move through the network. Flow propagation speeds are discussed in the next nucleotide.

We first consider forward wave speeds. In the first order LWR model (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 use 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 to track traffic states or vehicles over time.

![Figure 6: Forward wave speeds](image-url)
Backward waves track how traffic states propagate backwards on a road segment, and are responsible for queue build up and possible spillback to upstream links. 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.

Most static models do not consider a hypercritical branch in the fundamental diagram such that backward wave speeds are not available (N). Only quasi-dynamic models with capacity constraints can describe queues in static models by implicitly assuming the presence of a hypercritical branch. Two different temporal assumptions regarding backward waves can be made in order to remove the time dimension in a static model. The most widely adopted implicit assumption in quasi-dynamic models is setting backward wave speeds equal to 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; although stationary physical queues is also consistent with zero backward wave speeds. The zero speed assumption is consistent with fundamental diagrams of the shape shown in Figure 5(h). The other assumption that removes time from the model is considering a (negative) infinite speed depicted in Figure 7(c), which allows the model to describe spillback when the number of vehicles in the queue exceeds the available link storage. Note that an infinite 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.

**Figure 7: Backward wave speeds**

4.2 **Nucleotide 7 – Flow propagation speeds**

Instead of looking at the speeds at which traffic states propagate, we now look at the assumption on the speed with which traffic flow propagates in case there are no interruptions by queues (i.e., only considering the hypocritical branch of the fundamental diagram). In dynamic models this speed simply equals the vehicular speed as indicated in Figure 6(b). In contrast, static models move traffic flows instantaneously through the network within the single period, as shown in Figure 6(c).

4.3 **Nucleotide 8 – 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. This 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 from period to period (T), thereby describing the full temporal interactions within route choice intervals and across time periods. Static models consider no residual
traffic transfer (N), such that each time period is modelled independently. Especially when modelling short time periods in a congested network, this independence assumption will not be valid. 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 2.3.

5. Travel time calculations

Travel times are output of the network loading sub-model and are important inputs into the route choice sub-model. Network loading is completely defined by the temporal and spatial interaction assumptions, and as such travel times should be a direct output of the network loading sub-model. However, the way travel times are calculated varies in the literature, and this section attempts to provide a brief overview.

5.1 Link travel times

Link travel times can be decomposed into three components, namely,

1. Free-flow travel time (low traffic, no vehicle impedances);
2. Additional travel time due to hypo-congestion (medium traffic, vehicle impedances);
3. Additional travel time due to hyper-congestion ((very) heavy traffic, queuing delay).

The first travel time component ($\tau_1$) is constant and can simply be computed as the division of the link length and the maximum vehicle speed. The second component ($\tau_2$) depends on flows on the link itself, and the third component ($\tau_3$) depends in general on all network flows since queues appear due to vehicles competing for road capacity from different routes.

It is important to note that in most traffic assignment models only the total link travel time $\tau_1 + \tau_2 + \tau_3$ is of interest and not the exact time spent driving in free-flow and in a queue (unless they are weighted differently in a generalised cost function). There exist different ways to compute total link travel times that produce the same result as shown in Figure 8.

Newell (1993) illustrates that (in case of a single vehicle class) total link travel times can be derived from the cumulative inflow and outflow curves without knowledge of the traffic conditions inside the road segment, which is illustrated in Figure 8. In this example, the stationary link inflow rate is higher than the stationary outflow rate, such that a queue builds up inside the road segment. We are interested in determining the total link travel time for a vehicle entering the link at time instant $t$.

![Figure 8: Link travel time calculation using cumulative link inflow and outflow curves](image)

In Figure 8(a) we assume vertical queues while in Figure 8(b) we assume horizontal queues. In case of vertical queues, $\tau_1 + \tau_2$ is the same for all departing vehicles (since the flow rate is stationary), and $\tau_3$ can be calculated by dividing the number of vehicles in the vertical queue by the outflow rate following Payne and Thompson (1975). In case of horizontal queues, one can explicitly account for the length of
the queue by making $\tau_1 + \tau_2$ proportionally smaller, while making $\tau_3$ larger (since the tail of the queue is moving backwards and hence more vehicles end up in the queue). However, the resulting $\tau_1 + \tau_2 + \tau_3$ is the same in both cases.

Smith et al. (2013) describe a quasi-dynamic spatial queuing model and point out that in such a model it is necessary to adjust the free-flow travel times to account for the queue length in order to avoid double-counting. However, as argued above, no double counting occurs in a spatial queuing model if one calculates travel times consistently, considering either vertical or horizontal queues. It is easy to see that assuming vertical queues leads to simpler calculations as there is no need to determine the location of the tail of the queue.

5.2 Path travel times

Different types of path travel times can be considered, see e.g., Ran and Boyce (1996) and Buisson et al. (1999), depending on how link travel times are added along a path:

1. Instantaneous travel time;
2. Predictive travel time;
3. Experienced travel time.

Note that this distinction is not relevant for static models.

Instantaneous path travel times are calculated for a certain departure time considering only the traffic states at this time instant and the corresponding link travel times, and hence ignores any changes in traffic conditions while driving. This type of traffic assignment is often referred to as reactive.

Predictive path travel times 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 (simulation-based) models calculate experienced travel times, which consider the actually experienced link travel times at the time of link exit (instead of link entrance).

6. Classification of existing traffic assignment models

Many traffic assignment models have been proposed in the literature that we can classify using the eight nucleotides in the two genes. Table 3 provides a list of models, although it should be stressed that this list is by no means intended to be complete, but rather to provide typical examples of certain types of traffic assignment models.

6.1 Static models

All static models assume the following temporal interaction assumptions: infinite flow propagation speeds (I), infinite forward wave speeds (I), and no residual traffic transfer (N). In case a hypercritical branch of the fundamental diagram is considered, either zero (Z) or infinite (I) backward wave speeds are assumed. The models therefore mainly differ in their spatial interaction assumptions.

6.1.1 Unrestrained models

Unrestrained models are not widely used given their limited applicability. Models for uninterrupted traffic conditions are described in Bovy (1990). These models assume fully rational travellers with imperfect information (FI) in route choice and a linear hypocritical branch of the fundamental diagram, while the hypercritical branch is absent (LN). No capacity (UU) or storage (U) constraints nor turn flow restrictions (N) are considered.
6.1.2 Capacity restrained models

The capacity restrained traffic assignment model with deterministic route choice (FP) using a strictly increasing link performance function is perhaps the most widely adopted static model formulation, however the assumed spatial interactions are only realistic in low to medium traffic conditions.

One of the first applications of such a model is described in Irwin et al. (1961). They consider a piecewise linear link performance function, which translates into a concave hypocritical branch of the fundamental diagram while the hypercritical branch is missing (CN). Further, no capacity constraints (UU) or storage constraints (U) nor turn flow restrictions (N) are considered, resulting in link traffic flows exceeding physical link capacities. In other words, this model is essentially only appropriate for uncongested conditions without queues building up.

Fisk (1980) extended this model to include stochastic route choice (FI) based on the multinomial logit model, assuming imperfect information in route choice (I). Zhou et al. (2012) further extended this model to taken path overlap into account consistent with the C-logit model.

6.1.3 Capacity constrained models

Beckmann et al. (1956) proposed a model in which the implicitly assumed fundamental diagram has a horizontal asymptote at capacity (CN), see Figure 5(g), which ensures that flows do not exceed capacity. However, in such a model no queues appear.

Miller et al. (1975), Payne and Thompson (1975) and Smith (1987) introduced explicit queues into steady state assignment models with deterministic route choice (FP) and capacity constraints (UC), while Bell (1995) extended this model to include logit based route choice (FI).

Bifulco and Crisalli (1998) were among the first to propose a model that explicitly describes residual queues which may not be constant. Stochastic logit based route choice (FI) is considered. They implicitly assume a fundamental diagram (CH) as in Figure 5(h) in which fixed link exit capacity constraints are taken into account. Moreover, neither link storage constraints (U) nor turn flow restrictions (N) are considered. Backward kinematic wave speeds are implicitly assumed to be zero (Z), such that the resulting model describes vertical point queues without spillback. Implicitly taking the CTF node constraints, they propose outflows equal to capacity divided by the sending flow (i.e. the flow that could potentially flow out if there were no constraints). Similar models that explicitly describe residual queues have been proposed by Lam and Zhang (2000) and Smith (2013) in which deterministic route choice (FP) is assumed. None of these models consider specific turn flow restrictions (N).

Bliemer et al. (2014) are the first to include a first order node model (F) that explicitly describes turn flow restrictions into a static traffic assignment model with residual queues for general transport networks. Their model can be seen as an extension of the model of Bifulco and Crisalli (1998). Bliemer et al. (2014) further derive a path travel time function consistent with queuing theory and travel time calculations in dynamic models.

6.1.4 Capacity and storage constrained models

The first static traffic assignment model to consider a finite link storage constraints (C) is proposed by Smith et al. (2013). Similar to previous models with residual queues, only link exit capacity constraints (C) are considered in an implicit node model (O), while link entrance constraints and other turn restrictions are mostly ignored (although merges and diverges are briefly discussed). They implicitly assume a fundamental diagram (CV) similar to Figure 5(i), which leads to compact horizontal queues (with densities equal to the jam density) and spillback when the number of vehicles in the queue exceeds the available link storage. Their model is restricted to a specific hypothetical network and has not been shown to be generally applicable to transport networks that include more complex merges and diverges.
6.2 Dynamic models

All dynamic models assume forward wave speeds that are not infinite, i.e. either equal to the vehicular speed (V) or kinematic wave speed (K). 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). Flow propagation speeds are equal to the vehicular speed (V). Further, dynamic models assume residual traffic transfer (T). Again, the models differ mainly in their spatial interaction assumptions.

6.2.1 Unrestrained models

We were not able to find any dynamic unrestrained models proposed in the literature.

6.2.2 Capacity restrained models

One of the earlier dynamic traffic assignment models is described in Janson (1991), which can be seen as a direct extension of the traditional static traffic assignment model. The underlying spatial interaction assumptions are fully rational travellers with perfect knowledge (FP), a fundamental diagram (CN) as in Figure 5(f), without considering a hypercritical branch, no capacity constraints (UU), no storage constraints (U), and no turn flow restrictions (N). Link travel times are a function of only the link flows in the current time period. Hence, travel times are separable and the problem can be formulated as an optimisation problem. Chen and Hsueh (1998) consider a similar model, but the link travel times are considered to be a function of all previous link inflows, leading to a non-separable link travel time function with asymmetric Jacobians, and resulting in a variational inequality problem formulation. Both models assume that forward wave speeds are equal to the vehicular speeds (V).

As Ran and Boyce (1996) point out, forward wave speeds equal to vehicular speeds assumes stationary travel times and ignores platoon dispersion or concentration. Astarita (1996) derives a more general relationship that is consistent with kinematic wave theory, i.e. forward wave speeds are assumed to equal kinematic wave speeds (K), which has also been used in Li et al. (2000) and Bliemer and Bovy (2003), and is further generalised in Chabini (2001).

6.2.3 Capacity constrained models

Li et al. (2000) propose a deterministic model (FP) that considers a fixed free-flow travel time and point queues based on a fixed outflow rate, hence assuming a linear hypocritical branch and a horizontal hypercritical branch of the fundamental diagram (LH). Instantaneous travel times are computed, leading to a reactive assignment. Han (2003) proposed a similar model, but includes logit route choice (FI) resulting in a stochastic dynamic user equilibrium. Outflow constraints are considered (UC) but no storage constraints (U) or turn flow restrictions (N).

Friesz et al. (2013) derives a hydrodynamic model with a concave hypocritical and a horizontal hypercritical branch of the fundamental diagram (CH), see Figure 5(h). They assume kinematic forward wave speeds (K), and zero backward wave speeds (Z), resulting in point queues. Further, CTF constraints are considered, but no explicit turn flow restrictions (N).

6.2.4 Capacity and storage constrained models

Bliemer (2007) proposed an analytical dynamic traffic assignment models that includes out capacity constraints (UC), link storage constraints (C), and turn flow restrictions. The implicitly assumed fundamental diagram (CV) is consistent with Figure 5(i) with infinite backward wave speeds (I) at a queuing density lower than the jam density. Queues spill back to upstream links if a fixed and finite link storage capacity is exceeded. CTF constraints and the fair merging rule was used, resulting in demand proportional turn flows (O) not consistent with a first order node model (since it violates the invariance principle). Yperman et al. (2005) proposed a first order link model with physical spatial queues, under the strict assumption of a triangular fundamental diagram (see Figure 5(c)). The node model describing turn flow restrictions is similar as in Bliemer (2007). Gentile (2010) relaxed the assumption of a
triangular fundamental diagram to allow for any concave fundamental diagram (CC), see Figure 5(a). The models of Yperman et al. (2005) and Gentile (2010) allow proper forward and backward kinematic wave speeds (KK) consistent with first order traffic flow theory. Yperman et al. (2005) and Gentile (2010) focus on the dynamic network loading model, while mostly ignoring the route choice aspect. Han et al. (2015) rigorously formulate a dynamic traffic assignment model under the assumption of boundedly rational users (BP). In the numerical example in their paper they use the model of Yperman et al. (2005) for network loading.

6.3 Semi-dynamic models

Semi-dynamic models are neither completely static nor completely dynamic as semi-dynamic. This means with respect to the temporal interaction assumptions that they typically assume a sequence of connected static models as described in Nakayama and Connors (2014). In such a case, forward wave speeds and flow propagation speeds are infinite (I). However, vehicles that reside in a queue at the end of a time period are transferred to the next time period (T).

A couple of analytical semi-dynamic traffic assignment models have been proposed in the literature, namely Fujita et al. (1988), Fujita et al. (1989), Miyagi and Makimura (1991), Akamatsu et al. (1998), and Nakayama (2009). Unfortunately, these publications are all in Japanese and hence we will not discuss them there.

Other semi-dynamic models have been described as operational procedures and algorithms rather than mathematical problems, e.g., Van Vliet (1982) and Davidson et al. (2011), which makes them difficult to accurately classify, and hence are also omitted in Table 3.

7. Discussion and conclusions

In this paper we have presented a theoretical framework with which we can classify traffic assignment models for strategic transport planning. This framework is described in terms of DNA with two genes consisting of five spatial interaction assumptions and two spatial interaction assumptions.

As a special case, the widely applied capacity restrained static traffic assignment model can be derived by assuming (i) rational travellers with perfect knowledge, (ii) a concave hypocritical part and no hypercritical part of the fundamental diagram, (iii) no flow capacity constraints, (iv) no storage constraints, (v) no turn flow restrictions, (vi) infinite forward wave speeds and no backward waves, (vii) infinite flow propagation speeds, and (viii) no residual traffic transfer. These very strict assumptions limit the applicability of this particular model. Although this model type is applied to heavily congested networks all around the world, this model is actually only appropriate for uncongested traffic conditions (A and B as indicated in Figure 3).

We have also shown that many dynamic models that have been proposed in the literature make similar spatial interaction assumptions and are therefore also not appropriate for (very) heavy traffic with queues and spillback.

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 can realistically describe traffic under all traffic conditions. Such static models also exist, referred to as quasi-dynamic models, which extend the capabilities and realism of static models in congested situations by sharing the same spatial interaction assumptions made in advanced dynamic models. This opens up possibilities for quasi-dynamic models that are derived from advanced dynamic models by simply using static temporal interactions assumptions. Therefore, the framework described in this paper may not only useful for classifying models, but also for developing new models.
Table 3: Overview of assumptions made in different traffic assignment models proposed in the literature

<table>
<thead>
<tr>
<th>Static models</th>
<th>Gene 1: Spatial interaction assumptions</th>
<th>Gene 2: Temporal interaction assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>route choice behaviour</td>
<td>fundamental diagram</td>
</tr>
<tr>
<td>Bovy (1990)</td>
<td>FI</td>
<td>LN</td>
</tr>
<tr>
<td>Beckmann et al. (1956)</td>
<td>FP</td>
<td>CN</td>
</tr>
<tr>
<td>Irwin et al. (1961)</td>
<td>FP</td>
<td>CN</td>
</tr>
<tr>
<td>Fisk (1980)</td>
<td>FI</td>
<td>CN</td>
</tr>
<tr>
<td>Smith (1987)</td>
<td>FP</td>
<td>LH</td>
</tr>
<tr>
<td>Bell (1995)</td>
<td>FI</td>
<td>LH</td>
</tr>
<tr>
<td>Bifulco and Crisalli (1998)</td>
<td>FI</td>
<td>CH</td>
</tr>
<tr>
<td>Lam and Zhang (2000)</td>
<td>FP</td>
<td>CH</td>
</tr>
<tr>
<td>Zhou et al. (2012)</td>
<td>FI</td>
<td>CN</td>
</tr>
<tr>
<td>Smith (2013)</td>
<td>FP</td>
<td>LH</td>
</tr>
<tr>
<td>Smith et al. (2013)</td>
<td>FP</td>
<td>CV</td>
</tr>
<tr>
<td>Bliemer et al. (2014)</td>
<td>FI</td>
<td>LC</td>
</tr>
<tr>
<td>Dynamic models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Janson (1991)</td>
<td>FP</td>
<td>CN</td>
</tr>
<tr>
<td>Chen and Hsueh (1998)</td>
<td>FP</td>
<td>CN</td>
</tr>
<tr>
<td>Li et al. (2000)</td>
<td>FP</td>
<td>LH</td>
</tr>
<tr>
<td>Chabini (2001)</td>
<td>FP</td>
<td>CN</td>
</tr>
<tr>
<td>Bliemer and Bovy (2003)</td>
<td>FP</td>
<td>CN</td>
</tr>
<tr>
<td>Han (2003)</td>
<td>FI</td>
<td>LH</td>
</tr>
<tr>
<td>Yperman et al. (2005)</td>
<td>--</td>
<td>LL</td>
</tr>
<tr>
<td>Bliemer (2007)</td>
<td>--</td>
<td>CV</td>
</tr>
<tr>
<td>Gentile (2010)</td>
<td>FP</td>
<td>CC</td>
</tr>
<tr>
<td>Friesz et al. (2013)</td>
<td>FP</td>
<td>CH</td>
</tr>
<tr>
<td>Han et al. (2015)</td>
<td>BP</td>
<td>LL</td>
</tr>
</tbody>
</table>
References


