

1. Introduction

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 around the world apply mainly static models on large scale road networks for this purpose, although exceptions exist (e.g., the government in the Netherlands applies a quasi-dynamic traffic assignment model in regional and national transport modelling studies, and the San Francisco government in the US recently implemented a dynamic traffic assignment model for assessing transport projects in the city). In the last decade, there has been an increased interest in dynamic traffic assignment models as they have clear advantages with respect to realism. Nowadays, such dynamic models have become feasible on relatively large scale transportation networks due to significant increases in computation power and memory as well as the development of more efficient algorithms for solving these models. As is well-known, traditional static models suffer from several serious limitations, such as traffic flows exceeding link capacities and queues not being explicitly described. As a result travel time delays appear in the wrong locations. Despite this obvious lack of realism, such static models remain popular because they are mathematically rigorous, typically guarantee the existence and uniqueness of an equilibrium solution, can be efficiently solved using fast converging algorithms, are relatively easy to understand and very robust in applications. Dynamic models are significantly more complex, often do not guarantee existence or uniqueness of an equilibrium solution, require much longer run times and are significantly more difficult to calibrate than static models, requiring a large investment in time and money when moving towards dynamic models in practice.

While governments use predominantly static models for long term transport planning, dynamic models are widely applied for traffic operations analysis and evaluation of traffic management measures. In particular microscopic simulation models have become hugely popular. However, almost without exception, governments struggle with the inconsistencies between the static models used for transport planning and the dynamic models for traffic operations. If infrastructure decisions are based on outcomes of a static model (as in most cases), it is problematic if the predicted travel time savings cannot be reproduced in a dynamic model (and in real life). For example, consider a simple corridor with two bottlenecks according to a traditional static model. Expanding the capacity of the second bottleneck may result in significant travel time benefits in the static model, while it may not result in any travel time benefits in the dynamic model because traffic flow is reduced by the first bottleneck and the second bottleneck poses no longer a problem in the dynamic case (and in reality). Such inconsistencies exist because static and dynamic models are built on different underlying principles. The theoretical framework and the models presented in this paper remove these inconsistencies.

Besides inconsistencies being problematic in decision making, they also have impacts on model calibration. The current state of practice is that static models, specifically their networks and origin-destination (OD) matrix, are used as a basis for developing any dynamic model (also in the case of a much smaller sub-network in micro-simulation). However, due to inconsistencies between static and dynamic models, calibrated static models are a very undesirable starting point for creating a dynamic model for two main reasons. Firstly, link capacities in static models do not indicate maximum flow but rather influence the travel time via link performance functions. In dynamic models the link capacities indicate physical maximum flow. A dynamic model is very sensitive to these capacities and their values should be set with much more care. Secondly, the OD matrix has been calibrated to empirical traffic flow data such as link counts. Since traditional static models do not constrain traffic flows to link capacities, the OD matrix is calibrated in such a way that it becomes not very useful for dynamic models that include such capacity constraints. As a result, it becomes a time consuming, complex and often underestimated task to re-calibrate the OD matrix for use in dynamic models. The class of static

models derived in this paper adopts the traffic flow principles from dynamic models, such that a static OD matrix calibrated for this new class of static models is a much better starting point for building dynamic models.

In the past decade, several extensions to static models have been proposed to decrease the present huge gap between static and dynamic traffic assignment models. In particular, novel static (often referred to as quasi-dynamic) models have been proposed as a replacement of traditional unconstrained static models, such that traffic flows and travel times are more in line with dynamic models. As such, these new models have the potential to take away most of the inconsistencies between static and dynamic models. In this paper, instead of trying to extend currently existing static models, we will show how to derive static (and quasi-dynamic) traffic assignment models directly from any first order dynamic traffic assignment model. As such, maximum consistency is guaranteed. In particular, we aim to maintain consistency in spatial interactions described by the route choice model, the link model (described by a fundamental diagram), and the node model (described by turn flow restrictions) by only making different assumptions on how the time dimension is handled (i.e., aggregated) within the model. We will show how to derive first order static models and quasi-dynamic models that realistically describe queues and spillback, based on a generalised first order macroscopic dynamic model with any fundamental diagram, any turn flow restrictions, and route choice behaviour. It should be noted that also for microscopic and mesoscopic dynamic models such a derivation can be made, as long as the underlying (first order) spatial interaction mechanisms can be translated into a specific fundamental diagram, specific turn flow restrictions, and a specific route choice model. For example, the mesoscopic models in AIMSUN and Dynameq consider a triangular fundamental diagram (per lane) and simplified intersection rules that can be directly used to derive a consistent static (or quasi-dynamic) deterministic or stochastic traffic assignment model.

This paper presents several contributions. (i) We present a class of generalised first order dynamic traffic assignment models by combining existing state-of-the-art first order node and link models with a rigorous stochastic user equilibrium route choice problem formulation. (ii) We derive static and quasi-dynamic model classes directly from the generalised dynamic model. Although it seems natural to consider static models as a special case of dynamic models, as far as we are aware we are the first to perform such a derivation. (iii) We explicitly state the ‘static’ assumptions (which we call temporal interaction assumptions) that have been made implicitly in static models in the past (often without realising). In addition, we explicitly state the traffic behavioural assumptions (which we call spatial interaction assumptions) that have been made implicitly in static models (again, often without realising), which include assumptions with respect to the fundamental diagram, turn flow restrictions, and route choice behaviour. Understanding these assumptions provide important insights into the model properties. (iv) By relaxing these spatial interaction assumptions and thereby adopting more mechanisms from dynamic models, this leads to novel more realistic static models. Furthermore, our framework also allows relaxing one or more temporal interaction assumptions, leading to novel semi-dynamic models. (v) We demonstrate that static and quasi-dynamic models earlier proposed in the literature can be seen as special cases by making different temporal and spatial interaction assumptions. Our unified framework therefore describes the whole spectrum of all first order traffic assignment models, ranging from static to dynamic, and enables one to create time-aggregated models consistent with any first order dynamic model.

2. Static, quasi-dynamic and dynamic traffic assignment models

Many traffic assignment models have been proposed in the literature, ranging from static macroscopic to dynamic microscopic and anything in between. Several have been implemented in commercially available software. All traffic assignment models consist of two main

components, namely (i) a route choice model that determines how the travel demand is distributed over available routes in the network, and (ii) a network loading model that propagates (simulates) the traffic on the network. The network loading model is often described in terms of a link model (that describes the traffic flows along a link) and a node model (that describes turn flow restrictions through each node). In this paper we focus on macroscopic models that represent traffic as continuous flows instead of discrete individual vehicles, since we aim to derive static and quasi-dynamic model classes from a dynamic model.

2.1 Static traffic assignment models

Static traffic assignment models are well established. They were mainly developed in the 1950s with the work of Wardrop (1952) and Beckmann et al. (1956), initially for solving for a deterministic user equilibrium solution and later extended to a stochastic user equilibrium (Daganzo and Sheffi, 1977; Fisk, 1980). Route choice is the main part of these static models, while the network loading component is not capacity constrained by a turn flow restrictions and does not allow for residual traffic (e.g., queues). Travel times for traversing links are calculated using link performance functions.

Static traffic assignment model theory has progressed slowly over time. There have been some attempts to include capacity constraints into static models (e.g., Smith, 1987; Yang and Yagar, 1994, 1995; Larsson and Patriksson, 1995; Bell, 1995) or residual queues (Bifulco and Crisalli, 1998; Nesterov and De Palma, 1998, 2000, 2003; Lam and Zhang, 2000; Smith, 2013). None of these static models take spillback into account.

Although some of the static models mentioned in the previous paragraph have been termed quasi-dynamic, there is no explicit modelling of the time dimension in the route choice nor in the network loading model and therefore they are essentially static models. Some authors use the term quasi-dynamic to indicate models that produce results more similar to dynamic models (such as residual queues and spillback). Operational quasi-dynamic algorithms have been proposed in order to take queuing and spillback into account. Examples are QBLOK (4Cast, 2009) and a model in VISUM described in Bundschuh et al. (2006). We would like to point out that these models are merely described as algorithms without explicitly specifying the mathematical problem that is being solved. The lack of a rigorous problem formulation prohibits analysis of the properties of these models. The first quasi-dynamic model that describes spillback in a rigorous problem formulation has recently been proposed by Smith et al. (2013). Since the term quasi-dynamic does not express any specific properties of the model besides outputs that may be more in line with dynamic models, we refer to these quasi-dynamic models as static models due to the absence of any time dynamics.

As far as we are aware, Although the above mentioned residual queuing models are an improvement over the traditional static traffic assignment models, they predict queues inside the bottleneck link and not upstream the bottleneck link because none of these models include a proper (first order) node model. Bliemer et al. (2013) present a novel static model that for the first time includes a proper node model in static traffic assignment, such that residual queues appear upstream bottleneck links.

2.2 Semi-dynamic traffic assignment models

Models in which the time dimension explicitly enters the model in one way or the other we call semi-dynamic. For example, there can be some dynamics in the network loading model, or there can be dynamics across time periods in the route choice. The most important feature of semi-dynamic models is the ability to transfer residual traffic and queues from one time period to the other.

The oldest semi-dynamic traffic assignment model that we are aware of is SATURN (Van Vliet, 1982), which explicitly considers dynamics between time periods (by transferring queues) and

limited dynamics within the network loading. SATURN is again an operational model without a rigorous problem formulation. More recently, Davidson et al. (2011) proposed another operational model that they termed clocktime assignment, in which they propagate the travel demand over paths using one minute time steps, and keep track of traffic that is still on the network at the end of each ten minute time period (in the form of a ‘too late trip matrix’).

Another approach is taken by Brederode et al. (2010) and Bliemer et al. (2012) by combining a static model with a dynamic model. The static model results in vertical queues, while the dynamic model describes the queue dynamics by tracing backward and forward shockwaves resulting in physical queues and spillback. While this approach has both static and dynamic elements, we will not refer to this approach as semi-dynamic, but rather call it a hybrid model.

2.3 Dynamic traffic assignment models

Dynamic traffic assignment models have gained in popularity since the 1990s. Two different approaches can be distinguished in the literature, which we loosely refer to as the (strategic) transport planning approach and the (operational) traffic management approach, referring to the two application areas mentioned in the introduction. In the transport planning approach, network modellers have proposed direct extensions to the static model, maintaining the strong route choice component but adopting a relatively weak network loading model that is not consistent with traffic flow theory. In the traffic management approach, traffic flow theorists have proposed more complex network loading models consistent with traffic flow theory, but these were mostly limited to small networks or unable to find a user equilibrium solution. In the past five years, the two approaches seem to have converged to models that describe both route choice behaviour and traffic flow theory in a realistic, rigorous, and practically feasible way. This is an important observation in itself.

In the transport planning approach, essentially a time index was added to all variables, and traffic flow propagation was based on link travel times described by link performance functions. Such models are often referred to as whole link models in which the travel time is a function of the link flow or density. Examples are Janson (1991), Astarita (1996), Ran and Boyce (1996), Wu et al. (1998), Chen and Hsueh (1998), Xu et al. (1999), Chabini (2001) and Bliemer and Bovy (2003). Flow propagation using link performance functions is not consistent with traffic flow theory (Bliemer, 2007), and due to the lack of a node model, link flows are not capacity constrained and no physical queues are described.

In the traffic management approach, traffic flow theory based on kinematic waves introduced by Lighthill and Whitham (1955) and Richards (1956) is adopted to conduct the network loading. This theory describes traffic flow operations much more realistically, but typically requires solving complex partial differential equations. First order vehicle dynamics can be described by a given fundamental diagram for each link that describes the relationship between flow and density. In these first order models, vehicles are assumed to be able to instantaneously change speed, such that trajectories are piecewise linear. Second order models make the extension to speed dynamics, leading to more realistic smooth trajectories, however they require solving complex second order partial differential equations. Messmer and Papageorgiou (1990) developed METANET, which is based on a simplification of these differential equations and computes an approximate solution. Although second order models can describe some specific phenomena such as capacity drop and hysteresis, first order models already capture most important traffic dynamics, such as queue formation, spillback and shockwaves. Furthermore, Daganzo (1995a) has argued that (approximations of) second order models can have certain inherent limitations and can be problematic. Hence, in this paper we restrict ourselves to first order models. Daganzo (1994, 1995b) proposed the cell-transmission model (CTM) in which the road infrastructure is discretised into smaller cells as to provide an approximate solution to the partial differential equations. Simple node models describing turn flow restrictions for merges and diverges are included as part of the network loading. Newell (1993) describes a

simplified theory of kinematic waves assuming a triangular fundamental diagram, which is an exact method that does not require solving any partial differential equations. Yperman (2007) extended this link model to a network level and named it the link transmission model (LTM). Yperman further extended Newell's theory to a piecewise linear fundamental diagram. Gentile (2010) proposed the generalised link transmission model (G-LTM) in which a fundamental diagram of any concave shape can be taken into account. In many of these models derived from traffic flow theory, route choice is described by splitting rates at nodes. This approach may be suitable for short term forecasting, but for long term planning in which a user equilibrium solution is required, splitting rates are not suitable because they are not consistent with the notion of such an equilibrium. For all models derived from traffic flow theory holds that travel times can be calculated as a post-processing step after the network loading from the cumulative inflows and outflows assuming first-in-first-out (FIFO), instead of relying on link performance functions. More recently, node models have been recognised to play an important role in traffic assignment, and new macroscopic node models have been formulated and investigated (Flötteröd and Rohde, 2011; Gibb, 2011; Tampère et al., 2011; Jin, 2012a; Jin, 2012b).

2.4 From dynamic to static

This paper derives static traffic assignment models from a generalised first order dynamic (G-FOD) traffic assignment model under certain temporal interaction assumptions for handling the time dimension. In this paper we will use the term 'generalised' to indicate that it describes a class of models with any given concave fundamental diagram, any turn flow restrictions, and stochastic or deterministic route choice behaviour. The general equilibrium route choice problem will be formulated as a variational inequality problem. Since the (first order) G-LTM link model in combination with a first order node model presents an exact, general, and analytically rigorous G-FOD traffic assignment model, we can – without loss of generality – use this network loading model as the basis for our derivations. By deriving models explicitly from a generic first order dynamic model, the resulting classes of models inherit many properties from dynamic models, including possible queue formation, spillback, and the transfer of residual traffic to other time periods. The class of models derived from the G-FOD model in which time dynamics do not play any role we refer to as generalised first order static (G-FOS) models, while models in which time dynamics play some role, we refer to as generalised first order semi-dynamic (G-FOSD) models. Quasi-dynamic models are classified as static models, given that time dynamics play no role. Given certain spatial interaction assumptions, which include route choice behaviour, a certain fundamental diagram (note that a link performance function can also be rewritten as a fundamental diagram), and turn flow restrictions, a specific model results. Figure 1 illustrates our proposed theoretical framework. Most traffic assignment models proposed in the literature can be derived from the G-FOD model under certain temporal and spatial interaction assumptions. More importantly, novel static (quasi-dynamic) and semi-dynamic models can be derived with more realistic spatial interactions consistent with dynamic models. We believe that this theoretical framework provides a significant contribution to the literature.

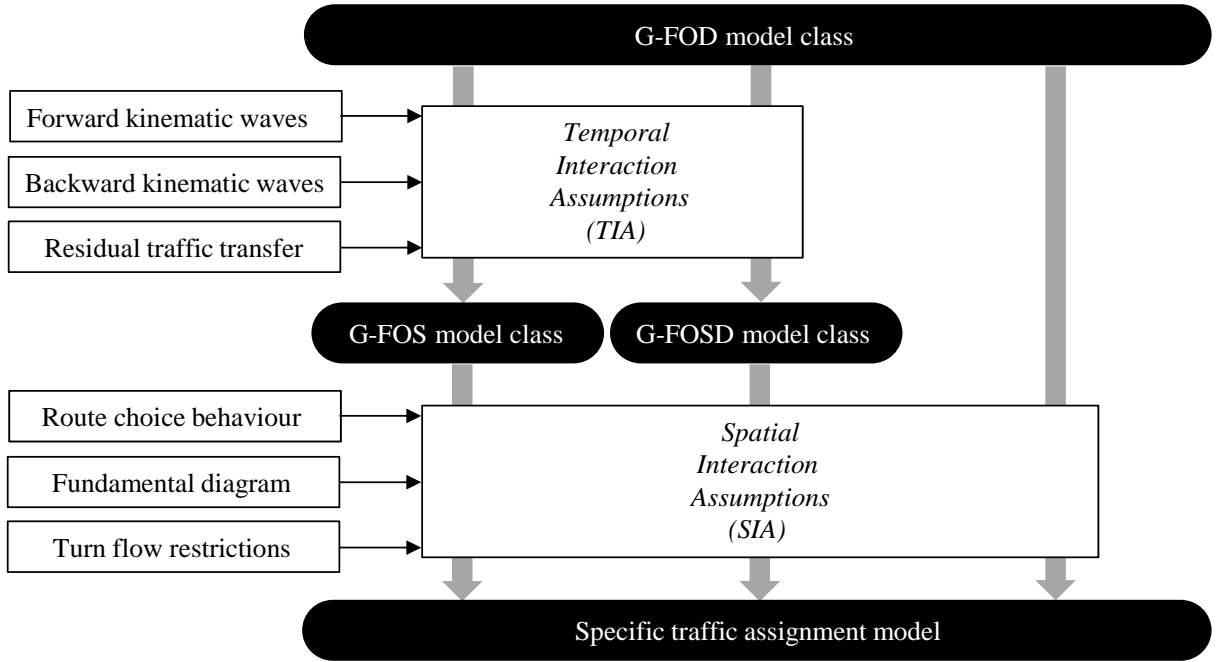


Figure 1 - Theoretical framework for deriving static and semi-dynamic traffic assignment models.

3. Generalised first order dynamic (G-FOD) traffic assignment model

In this section we formulate the generalised (single-class) macroscopic G-FOD traffic assignment model that we will use as the basis for deriving static models. First, we will introduce some notation regarding the input data describing travel demand and infrastructure supply. Consider a general network $G = (N, A)$ where N denotes the set of nodes and A denotes the set of directed links. Each link $a \in A$ is characterised by a length L_a (km) and a fundamental diagram (see also Section 3.2). Let $O \subset N$ and $D \subset N$ be the set of origins and destinations, respectively. Let P^{od} be the set of paths for origin-destination (OD) pair (o, d) , where $o \in O$ and $d \in D$, and let P denote the set of all paths, which is the union of all sets P^{od} . A path $p \in P$ is represented by a series of consecutive links. Each OD pair (o, d) has an associated travel demand rate $M^{od}(t)$ (veh/h) for each departure time instant $t \in [\underline{T}, \bar{T}]$, where \underline{T} and \bar{T} indicate the start and end time instants of the travel demand period, respectively. Links connecting an origin or destination to the physical road network are assumed to have sufficient capacity to accommodate the travel demand. Traffic assignment deals with the problem of assigning the travel demand to available routes given certain route choice behaviour. Since it is practically infeasible to describe route choice in continuous time (although the network loading problem can be solved in continuous time under certain assumptions, the travel demand period $[\underline{T}, \bar{T}]$ is typically split into I route choice time periods, where each time period i corresponds to time interval $[t_i, t_{i+1})$ with a length of $T_i = t_{i+1} - t_i$, $i = 1, \dots, I$. These time periods do not necessarily have to be equal in length. All vehicles departing in a certain route choice time period are assigned simultaneously to routes (i.e., route split proportions can be assumed stationary over each time period). In dynamic traffic assignment models, these periods typically

range from one to 15 minutes¹, and in static models these periods range from one to several hours. The average travel demand rate over route choice time period i , denoted by $M^{od}(i)$, is equal to $M^{od}(i) = (1/T_i) \int_{t_i}^{t_{i+1}} M^{od}(t) dt$. For notational convenience, we will omit the superscript od from path variables (as a path implies the origin and destination), unless strictly necessary.

The G-FOD model consists of a route choice model, a network loading model, and a travel time calculation component that feeds the travel time resulting from the network loading model back into the route choice model. The network loading model consists of a general first order link model and a general first order node model. Each component will be described separately in this section.

3.1 Route choice

Let $\mathbf{f} \equiv [f_p(i)]_{p \in P, i=1, \dots, I}$ (veh/h) denote the departing flow rates on each path p for each time period i . Let all travellers minimise their perceived route travel times. The stochastic route travel times are given by the sum of the actual route travel time and an error term. In this paper we assume that the error term is extreme value type I distributed with a variance of $\pi^2/6\theta^2$, leading to a logit-type route choice model with scale parameter θ . Let $c_p(i)$ be the average travel time in equilibrium on path p for vehicles departing during time period i . Then the perceived route travel time can be written as $c_p(i) + \frac{1}{\theta} \log f_p(i)$ (see e.g., Chen, 1999). Zhou et al. (2012) have formulated a variational inequality (VI) problem for finding a dynamic stochastic user equilibrium solution \mathbf{f}^* based on the C-logit model. Adapting their formulation and extending it to include multiple route choice periods (see e.g., Bliemer and Bovy, 2003), we can formulate the following VI problem:

$$\sum_{(o,d)} \sum_{p \in P^{od}} \sum_i \left(c_p^*(i) + \frac{1}{\theta} \log f_p^*(i) + \beta \log \Psi_p^{od} \right) (f_p(i) - f_p^*(i)) \geq 0, \quad \forall \mathbf{f} \in \Omega, \quad (1)$$

where Ψ_p^{od} is a path commonality factor (Cascetta et al., 1996) leading to the C-logit model with $\beta > 0$ or a path-size factor (Ben-Akiva and Bierlaire, 1999) or corrected path-size factor (Bovy et al., 2008) leading to the PS-logit or PSC-logit models, respectively, with $\beta < 0$. The resulting C-logit or path-size(-correction) logit models aim to take path overlap into account, although we would like to point out that none of these factors can handle path overlap in route choice perfectly (Bliemer and Bovy, 2008). Although VI problem (1) cannot describe other route choice models (e.g., more sophisticated logit or probit models), we believe that this formulation is sufficiently general to include the majority of the (large scale) equilibrium models proposed in the literature and implemented in practice.

Further, Ω is the set of feasible path flow rates given by the following flow conservation and non-negativity constraints that bound the flows from above and below:

$$\sum_{p \in P^{od}} f_p(i) = M^{od}(i), \quad \forall (o,d), \forall i, \quad (2)$$

$$f_p(i) \geq 0, \quad \forall p \in P, \forall i. \quad (3)$$

¹This time period should not be confused with the simulation time step size/period in dynamic network loading models, which are often in the range of 0.1 to 2 seconds.

The average route travel times $c_p(i)$ are computed as an implicit function of the route flows after they have been loaded onto the network. This network loading model is described next.

3.2 Network loading

In the network loading model, path flows \mathbf{f} are simulated on the network using flow conservation and flow propagation equations. Let $u_{ap}(t)$ (veh/h) and $v_{ap}(t)$ (veh/h) denote the path-specific link inflow rate into link a and the outflow rate out of link a at time instant t following path p , respectively. Flow along a path is always conserved, as expressed by the following equation:

$$u_{bp}(t) = \begin{cases} f_p(i), & \text{if link } b \text{ is the first link on path } p, \\ v_{ap}(t), & \text{if link } b \text{ follows link } a \text{ on path } p. \end{cases} \quad (4)$$

The path-specific inflow and outflow rates can be summed over all paths to obtain the total link inflow and outflow rates,

$$u_a(t) = \sum_{p \in P} u_{ap}(t), \quad \text{and} \quad v_a(t) = \sum_{p \in P} v_{ap}(t). \quad (5)$$

Constraints (4) are general flow conservation constraints that are present in all network loading models. For the first link, the path-specific link inflow rate is equal to the average path flow rate in the corresponding time period. For all consecutive links on the path, the path-specific link inflow rates are equal to the path-specific outflow rates of the previous link on the path.

We further introduce the following definitions of cumulative flows. Let

$$U_a(t) = \int_0^t u_a(w)dw, \quad \text{and} \quad V_a(t) = \int_0^t v_a(w)dw, \quad (6)$$

be the total (cumulative) number of vehicles that have entered and exited link a up till time instant t , respectively, and let the corresponding path-specific cumulative flows be defined as

$$U_{ap}(t) = \int_0^t u_{ap}(w)dw, \quad \text{and} \quad V_{ap}(t) = \int_0^t v_{ap}(w)dw. \quad (7)$$

The outflow rates are the result of flow propagation constraints that involve a link and a node model. It is mainly in the choice of the link and node models that (static and dynamic) models proposed in the literature differ. In the next subsections we will describe the flow propagation constraints for general first order macroscopic models.

3.2.1 First order link model

The link model describes the sending flow that would like to leave a link at a certain time instant, as well as the receiving flow that could potentially flow into a link (i.e., can be accommodated) at a certain time instant. In this section we take concepts of the discrete-time LTM model (Yperman, 2007) and the continuous-time G-LTM model (Gentile, 2010) as the basis for formulating our first order link model, which we believe is the most elegant, efficient, exact and general formulation of a first order link model that currently exists. Due to space limitations, we refer to Yperman (2007) and Gentile (2010) for more detailed descriptions of concepts, definitions, and derivations. In this paper we aim to use notation that is common for analytical dynamic traffic assignment models in the literature, which is different from the notation often adopted in traffic flow theory used in Yperman (2007) and Gentile (2010).

Furthermore, we put the G-LTM model into a rigorous route based framework, while Yperman (2007) ignored routes completely and Gentile (2010) adopted split proportions at nodes.

The G-LTM model provides an exact formulation of the flow propagation along a link consistent with first order traffic flow theory and takes any concave fundamental diagram as input. Let the fundamental diagram be described by function $q = \Phi(k)$, which describes the relationship between flow q and density k , see Figure 2. Any point on this curve represents a specific traffic state on a cross-section of a link.

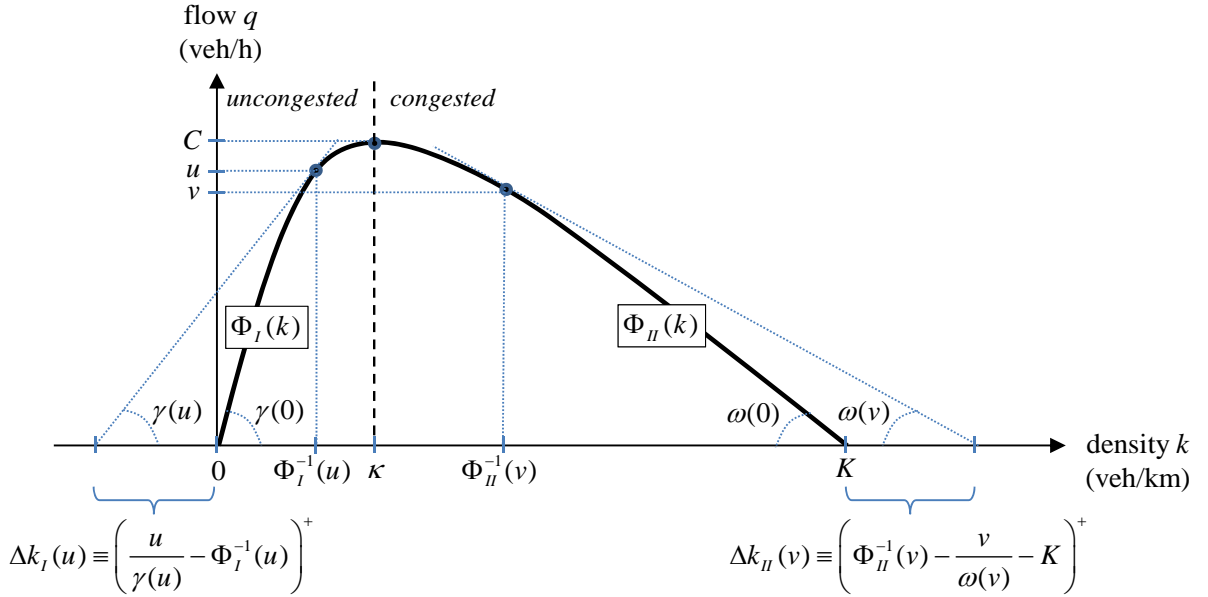


Figure 2: General concave fundamental diagram (figure adapted from Gentile, 2010).

Specific link characteristics that can be obtained from the fundamental diagram are the free-flow (wave) speed $\gamma_a(0)$ (km/h), the jam wave speed $\omega_a(0)$ (km/h), the physical capacity C_a (veh/h), the jam density K_a (veh/km), and a critical density κ_a (veh/km). For densities smaller than the critical density, traffic is uncongested and vehicles are not queuing. This uncongested part of the fundamental diagram is indicated by $\Phi_I(k)$. Densities larger than the critical density are associated with congested conditions and queues. The congested part of the fundamental diagram is indicated by $\Phi_{II}(k)$. Each stable traffic state propagates through the network with the speed of the corresponding kinematic wave, which is given by the slope of the fundamental diagram at that traffic state. In the uncongested part a forward kinematic wave belonging to traffic flow u has a speed of $\gamma(u)$, note that this is usually not the same as the vehicle speed related to this u . In the congested part a backward kinematic wave belonging to traffic flow v has a (negative) speed of $\omega(v)$, again observe that this kinematic wave speed always differs from the vehicle speed related to the traffic flow v . In the special case of a triangular fundamental diagram as proposed in Newell (1993) and used by Yperman (2007), all forward waves in the uncongested part have free-flow speed $\gamma_a(0)$, and all backward waves in the congested part all have (negative) jam wave speed $\omega_a(0)$.

Using the original ideas of Newell (1993), it is possible to directly relate the traffic states at the beginning and at the end of a link without the need to know exactly where vehicles are on a link (and therefore there is no need to split the link into smaller cells like the cell transmission model

does). As such, all information that is needed is contained in the cumulative flows $U_a(t)$ and $V_a(t)$. FIFO of link flows is an important property of this model that we will exploit later.

Let $S_a(t, t + \Delta t)$ (veh) be the sending flow of link a during simulation time period $(t, t + \Delta t)$, where the simulation time step Δt is usually somewhere between 0.1 and 10 seconds. The sending flow is the number of vehicles that would like to exit the link during this time period (the node model described later determines how many vehicles can actually flow out due to possible receiving flow constraints). The sending flows can be computed using cumulative inflows and outflows, and forward kinematic wave speeds:

$$S_a(t, t + \Delta t) = \min \left\{ \min_{u \in \Theta_a^l(t)} \left\{ U_a \left(t + \Delta t - \frac{L_a}{\gamma_a(u)} \right) - V_a(t) + \Delta k_l(u) L_a \right\}, C_a \Delta t \right\}, \quad (8)$$

where $\Theta_a^l(t)$ is the set of all past realistic uncongested traffic inflow states up to time instant t , and operator $(x)^+ \equiv \max\{0, x\}$ is used in Figure 2 to define $\Delta k_l(u)$ as the non-negative distance². The minimum operators in Equation (8) are needed because of the Newell-Luke Minimum Principle, which states that the minimum cumulative flow dominates the others. This equation can be derived as a direct extension of the formulation in Yperman (2007) for piecewise linear fundamental diagrams, or can be derived using the continuous-time formulation in Gentile (2010). The outer minimum returns the first argument if the link is uncongested (i.e., no queue), while the second argument is returned when it is in congestion.

Let the inner minimum be attained at inflow rate $u_a^*(t)$.³ The average sending flow rate can simply be computed as $s_a(t, t + \Delta t) = S_a(t, t + \Delta t) / \Delta t$ (veh/h). Letting $\Delta t \rightarrow 0$ yields the following continuous-time formulation:

$$s_a(t) = \begin{cases} C_a, & \text{if } U_a \left(t - \frac{L_a}{\gamma_a(u_a^*(t))} \right) - V_a(t) + \Delta k_l(u_a^*(t)) L_a > 0, \\ u_a \left(t - \frac{L_a}{\gamma_a(u_a^*(t))} \right), & \text{otherwise,} \end{cases} \quad (9)$$

where $U_a \left(t - \frac{L_a}{\gamma_a(u_a^*(t))} \right) - V_a(t) + \Delta k_l(u_a^*(t)) L_a = 0$ indicates that the sending flow at time instant t did not encounter any congestion. In the same way, let $R_a(t, t + \Delta t)$ (veh) be the receiving flow of link a during simulation time period $(t, t + \Delta t)$. The receiving flow is the number of vehicles that could enter the link during this time period (the node model described later determines how many vehicles will actually flow in due to possible sending flow constraints). The receiving flows can be computed using cumulative inflows and outflows, and backward kinematic wave speeds:

² While in dynamic network models the term cannot be negative due to concavity of the fundamental diagram, when deriving static and quasi-dynamic network loading models in Sections 5 and 6 we will require this property.

³ Note that it is possible that multiple forward shockwaves arrive at exactly the same time at the end of a link. While the minimum in Equation (10) is unique, there may be multiple traffic states (inflow rates) for which this minimum is attained. In that case, the Newell-Luke Minimum Principle prescribes that the minimum inflow rate should be chosen, which due to concavity of the fundamental diagram corresponds to the most recent traffic state in set $\Phi_a^l(t)$.

$$R_a(t, t + \Delta t) = \min \left\{ \min_{v \in \Theta_a^H(t)} \left\{ V_a \left(t + \Delta t + \frac{L_a}{\omega_a(v)} \right) - U_a(t) + K_a L_a + \Delta k_{II}(v) L_a \right\}, C_a \Delta t \right\}, \quad (10)$$

where $\Theta_a^H(t)$ is the set of all past realistic congested traffic outflow states up to time instant t , and $\Delta k_{II}(v)$ is the non-negative distance indicated in Figure 2. The outer minimum returns the first argument if the link is in a spillback state, while the second argument is returned when there is no spillback. Let the inner minimum be attained at outflow rate $v_a^*(t)$.⁴ The average sending flow rate can be computed as $r_a(t, t + \Delta t) = R_a(t, t + \Delta t) / \Delta t$ (veh/h). Letting $\Delta t \rightarrow 0$ yields the following continuous-time formulation:

$$r_a(t) = \begin{cases} C_a, & \text{if } V_a \left(t + \frac{L_a}{\omega_a(v_a^*(t))} \right) - U_a(t) + K_a L_a + \Delta k_{II}(v_a^*(t)) L_a > 0, \\ v_a \left(t + \frac{L_a}{\omega_a(v_a^*(t))} \right), & \text{otherwise,} \end{cases} \quad (11)$$

where $V_a \left(t + \frac{L_a}{\omega_a(v_a^*(t))} \right) - U_a(t) + K_a L_a + \Delta k_{II}(v_a^*(t)) L_a = 0$ indicates that a queue has reached the beginning of the link and is spilling back to upstream links.

3.2.2 First order node model

The node model determines how much of the sending flows can actually leave each link given the receiving flow restrictions on consecutive links on the paths. In other words, the node model describes the turn flow restrictions through each node, possibly leading to restrictions on link outflows. It is one of the most important components of a network model, as all queues originate at nodes and the node model therefore significantly influences delays and travel times. While link models have received significant attention by researchers in the past, only in the last decade node model have been recognised to be of significant importance in order to describe realistic traffic flows, in particular in an urban setting.

The general first order node model discussed in this section is assumed to follow the requirements provided by Tampère et al. (2011) for first order macroscopic node models. To summarise, these requirements are (i) general applicability (i.e., can handle cross-nodes with any number of incoming and outgoing links), (ii) maximizing flows, (iii) non-negativity, (iv) conservation of vehicles, (v) satisfying demand and supply constraints, (vi) obeying conservation of turning fractions (CTF), and (vii) satisfaction of the invariance principle (Lebacque and Khoshyaran, 2005). First-in-first-out (FIFO) diverging rules (Daganzo, 1995b) satisfy requirement (vi), while fair merging rules (Jin and Zhang, 2003) satisfy requirement (vii).

Several node models have been proposed in the last decade (Jin and Zhang, 2003; Jin and Zhang, 2004; Bliemer, 2007; Jin, 2012a; Jin, 2012b) but do not satisfy all of the above requirements. Recently, some first order node models have been proposed that satisfy all the

⁴Similar to Footnote 2, the Newell-Luke Minimum Principle dictates that in case multiple backward kinematic waves reach the beginning of the link at the same time, the one corresponding to the minimum outflow rate should be taken, which is the most recent one (given that the fundamental diagram is concave).

conditions (Tampère et al., 2011; Flötteröd and Rohde, 2011; Gibb, 2011). In this paper we consider any first order node model that satisfies the requirements in Tampère et al. (2011).

The sending flows $s_a(t)$ describe the demand constraints and the receiving flows $r_a(t)$ describe the supply constraints in the node model. To be more precise, we need the sending flows in each direction. Consider a certain node $n \in N$, and let $A_n^{\text{in}} \subset A$ and $A_n^{\text{out}} \subset A$ denote the sets of incoming and outgoing links of node n , respectively. Let $s_{ab}(t)$ (veh/h) be the directional sending flow rate from link $a \in A_n^{\text{in}}$ at time instant t that is destined for link $b \in A_n^{\text{out}}$. We also call (a, b) a turn (in the most general sense, not just left and right but also straight). Further, let $\varphi_{ab}(t)$ be the split proportion of the sending flow $s_a(t)$ heading for link b , such that we can write

$$s_{ab}(t) = \varphi_{ab}(t)s_a(t). \quad (12)$$

Since the FOD link model satisfies the link FIFO rule, we can determine the split proportions at link exit from the flow proportions at link entrance. If $\theta_a^-(t)$ denotes the time of link entrance of a vehicle that reaches the end of the link at time instant t , i.e. $U_a(\theta_a^-(t)) = V_a(t)$, then the split proportions can be computed as follows:

$$\varphi_{ab}(t) = \sum_{p \in P} \delta_{bp} \xi_{ap}(\theta_a^-(t)), \quad (13)$$

where δ_{ap} is a link-path incidence indicator that is equal to one if link a is on path p and zero otherwise, and the path-specific flow proportion at any given time instant t is given by

$$\xi_{ap}(t) = \frac{u_{ap}(t)}{u_a(t)}. \quad (14)$$

The actual directional outflow rates are computed by multiplying the (desired) directional sending flows with certain reduction factors $\alpha_a(t)$ for each link,

$$v_{ab}(t) = \alpha_a(t)s_{ab}(t), \quad (15)$$

Note that there is only a single reduction factor for all turns (a, b) due to the fact that the FIFO property is assumed on each link, which essentially assumes a single queue for all turn directions. Such a FIFO diverging rule is assumed in most (but not all) node models. If $\alpha_a(t) = 1$, then all sending flow can exit the link, i.e. no turn flow restrictions are imposed. However, if $\alpha_a(t) < 1$, some of the sending flow cannot exit and will be added to a queue on link a . These reduction factors are determined by turn flow restrictions described by a first order node model for each node $n \in N$, which we implicitly indicate by function $\Gamma_n(\cdot)$:

$$[\alpha_a(t)]_{a \in A_n^{\text{in}}} = \Gamma_n(s_{a'b'}(t), r_{b'}(t), \forall a' \in A_n^{\text{in}}, \forall b' \in A_n^{\text{out}}). \quad (16)$$

It should be noted that function $\Gamma_n(\cdot)$ in Equation (16) can be fairly complex and often not closed form, as the outflow rate for a specific link a typically depends on sending flows from all incoming links and receiving flows from all outgoing links. We would like to refer to Tampère et al. (2011) for a detailed description of these relationships.

The link outflow rate $v_a(t)$ can now be computed as

$$v_a(t) = \sum_{b \in A_n^{\text{out}}} v_{ab}(t), \quad (17)$$

such that the path-specific flow rates, which feed back into the flow conservation constraints stated in Equation (4), is given by

$$v_{ap}(t) = \xi_{ap}(\theta_a^-(t))v_a(t). \quad (18)$$

3.3 Travel time calculation

In G-FOD models, travel time calculations are done after flow propagation as a post-processing step. Because of the FIFO link property, travel times can easily be derived from the (piecewise linear) cumulative inflows and outflows, see e.g., Newell (1993) or Long et al. (2011). Because of the FIFO property, the x^{th} vehicle enters link a at time instant $U_a^{-1}(x)$ and exits the link at time instant $V_a^{-1}(x)$, hence the link travel time for this vehicle is $V_a^{-1}(x) - U_a^{-1}(x)$ (h). We are interested in computing the average path travel time for vehicles departing during a certain time period i . Since the FIFO property also holds on a path level, the path travel time for the x^{th} vehicle entering path p can be calculated as $V_{\bar{a}_p}^{-1}(x) - U_{\underline{a}_p}^{-1}(x)$, where \underline{a}_p and \bar{a}_p are the first and last link on path p , respectively. Alternatively, this path travel time can also be computed as the sum of consecutive link travel times, $\sum_a \delta_{ap}(V_{ap}^{-1}(x) - U_{ap}^{-1}(x))$. Noting that there are $U_{\underline{a}_p}(t_{i+1}) - U_{\underline{a}_p}(t_i) = f_p(i)T_i$ vehicles departing along path p during time period i , the average path travel time is given by

$$c_p(i) = \frac{1}{f_p(i)T_i} \int_{U_{\underline{a}_p}(t_i)}^{U_{\underline{a}_p}(t_{i+1})} (V_{\bar{a}_p}^{-1}(x) - U_{\underline{a}_p}^{-1}(x)) dx. \quad (19)$$

This integral, which represents the total travel time spent by all vehicles departing during time period i , is graphically illustrated in Figure 4 for a simple example in which there is only one path (so we have omitted the path subscript) consisting of four links with lanes and capacities given in Figure 3. Three one-hour time periods (i.e., $T_i = 1$ for all i) of travel demand are considered, with average path flow rates also given in Figure 3. Figure 4 shows possible cumulative link inflows and outflows. Congestion appears on links 2 and 3, where we for simplicity have assumed that the links are long enough to prevent spillback. The total travel time spent by all vehicles in all time periods is indicated by the grey shaded area. For each time period i , we are only interested in the area enclosed from the left and right by $U_1(t)$ and $V_4(t)$, and from the bottom and top by $U_1(t_i)$ and $U_1(t_{i+1})$, respectively. Dividing this area by the total number of vehicles that entered yields the average path travel time. The average path travel times are input into VI problem (1).

While the difference between $V_a^{-1}(t)$ and $U_a^{-1}(t)$, which is the horizontal distance between the cumulative flow curves, indicates the travel time for a vehicle entering the link at time a , the vertical distance between $U_a(t)$ and $V_a(t)$ indicates the number of vehicles on the link at time instant a (see also the dark black vertical lines in Figure 4). Knowing this number of vehicles will become useful in the derivations that follow, hence we let $Q_a(t)$ (veh) define the number of vehicles on link a at time instant t as

$$Q_a(t) = U_a(t) - V_a(t). \quad (20)$$

The number of vehicles on a link at the beginning of each time period, $Q_a(t_i)$, we refer to as residual traffic from earlier time periods.

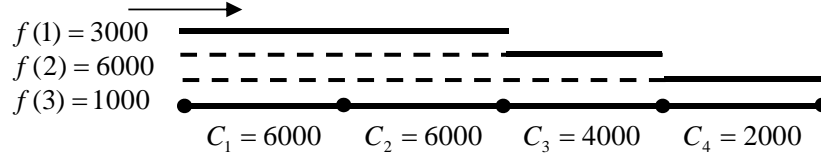


Figure 3 - Example corridor network.

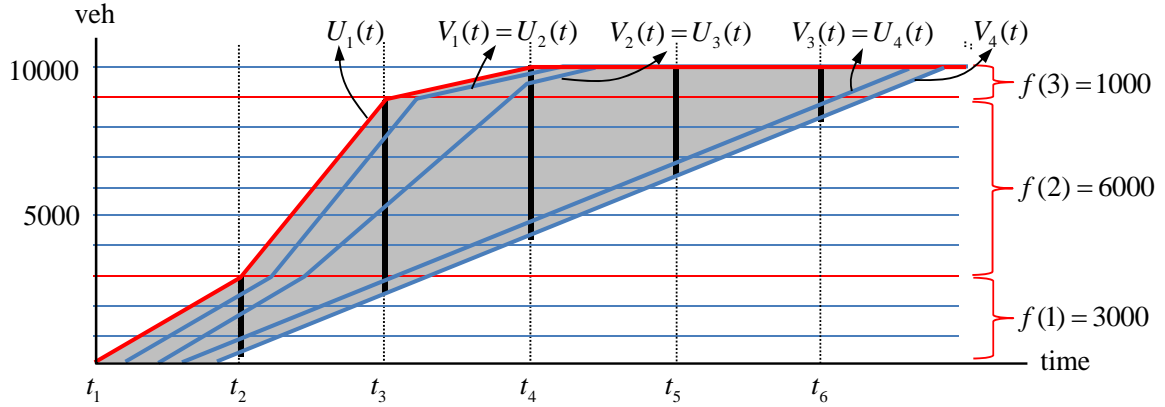


Figure 4 - Using cumulative link inflows and outflows to determine travel times.

4. Temporal interaction assumptions

Now that we have defined a general class of first order dynamic traffic assignment models, in this section we formulate ‘static assumptions’ that eliminate the time dimension in one or more parts of the model, and from which we can derive static and quasi-dynamic models. We will refer to these assumptions as Temporal Interaction Assumptions (TIA), and are related to forward and backward kinematic waves, and to how residual traffic is dealt with. Each of these assumptions aims to remove or simplify any time dynamics within the model, in particular in the network loading model.

4.1 Forward kinematic waves

In the G-FOD model, the forward kinematic waves are responsible for the propagation of the vehicles in the network. In the dynamic model, the speeds of forward kinematic waves, $\gamma_a(u)$, are equal to the slopes of the fundamental diagram in the uncongested branch, $\Phi_l(k)$. This presents two complexities in dynamic models. First, we need to trace each forward kinematic wave over time. Secondly, since the traffic conditions vary over time, the forward kinematic wave speeds also vary over time. The second complexity can be overcome by assuming a finite stationary forward kinematic wave speed over time, but this does not solve the first complexity. However, assuming infinite forward kinematic wave speeds, i.e., $\gamma_a(u) = \infty$, removes the need to explicitly trace the kinematic waves by essentially propagating the vehicles instantaneously through the network. This is the underlying assumption that is implicitly made in all static

A unified framework for traffic assignment: deriving static and quasi-dynamic models consistent with general first order dynamic traffic assignment models

Bliemer, Raadsen, Brederode, Bell and Wismans

models. Note that assuming $\gamma_a(u)=0$ is not meaningful, as vehicles would not propagate forward and stay at the origin.

It is important to realise that in large networks, not all vehicles may be able to reach their destination within the considered time period (resulting in residual traffic). However, in static models this is mostly ignored, such that for example a vehicle with a three hour travel time will reach its destination during a two hour time period. Hence, the instantaneous forward propagation assumption should preferably be restricted by the time period under consideration.

These possible assumptions are summarised in Table 1, where ‘S’ refers to an assumption that removes the time dimension completely and hence results in a static model, while ‘D’ refers to an assumption that is underlying a dynamic model.

Table 1 - Temporal interaction assumptions.

<i>Assumptions on forward kinematic waves</i>	
S	Instantaneous propagation of forward kinematic waves with infinite speeds;
D	Time varying forward kinematic waves with speeds equal to the slope in the uncongested branch of the fundamental diagram;
<i>Assumptions on backward kinematic waves</i>	
S(v)	No propagation of backward kinematic waves;
S(h)	Instantaneous propagation of backward kinematic waves with infinite speeds and finite link storage capacity;
D	Time varying backward kinematic waves with speeds equal to the slope in the congested branch of the fundamental diagram;
<i>Assumptions on residual traffic transfer</i>	
S	No transfer of residual traffic to next time periods;
D	Transfer of residual traffic to next time periods;

4.2 Backward kinematic waves

In the G-FOD model, the backward kinematic waves are responsible for possible spillback of queues to upstream links. Similar to forward kinematic waves, in the dynamic model the speeds of backward kinematic waves, $\omega_a(v)$, are equal to the slopes of the fundamental diagram in the uncongested branch, $\Phi_{II}(k)$. Again, this requires tracing each backward kinematic wave over time, while the wave speeds vary depending on traffic conditions. The simplest assumption to remove any dynamics in backward kinematic waves is to set $\omega_a(v)=0$, i.e. not having any kinematic waves that move backward. This essentially means that queues form as (vertical) point queues at the end of a link and never cause any spillback effects. The same model would result if we would assume infinite link storage (denoted by the maximum number of vehicles that could be present on a link), i.e. $K_a L_a = \infty$. In case of such infinite storage, the backward kinematic waves will never have an impact on upstream links and hence does not cause spillback. More interesting is the assumption of infinite backward kinematic wave speeds, $\omega_a(v) = -\infty$, combined with finite link storage. In this case, queues consume limited physical

storage space and spillback occurs if the finite storage space is exceeded. The upstream influence of the backward kinematic waves will not propagate indefinitely but are automatically restricted by the finite link storage capacities during each time period.

Table 1 summarises the above mentioned assumptions on the backward kinematic waves. Assumption ‘S(v)’ results in a static model without spillback (e.g., vertical queues), while assumption ‘S(h)’ yields a static model with horizontal queues and spillback. Assumption ‘D’ is the underlying assumption in a dynamic model.

4.3 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. These residual vehicles are either (i) in congestion in a residual queue near a bottleneck, or (ii) are not in congestion but simply are not able to reach their final destination because the travel time to reach the destination is longer than the considered time period (see also Section 4.1). This residual traffic influences the traffic flows and travel times in the next time period. If there is residual traffic on a certain link a at the beginning of time period i , then $Q_a(t_i) > 0$. This dependency of traffic across time periods can be eliminated by assuming that any residual traffic has exited the network at the beginning of each time period i , i.e., $Q_a(t_i) = 0$ for all links a .

Hence, we can consider two assumptions on residual traffic transfer as summarised in Table XT. In case no residual traffic transfer occurs, the time periods become completely independent, leading to a static model.

4.4 Combinations of temporal interaction assumptions

Different combinations of the TIAs in Table 1 result in different model classes. However, not every combination is useful or even meaningful. If all assumptions are static (S), then a G-FOS model results in which the time dimension is completely eliminated, both within and across time periods. Depending on whether assumption ‘S(v)’ or ‘S(h)’ is adopted for backward kinematic waves, a static model with vertical or horizontal queues will result, respectively. In case of vertical queues, there will be no spillback, while assuming horizontal queues yields the possibility of having spillback. We will further illustrate these classes of G-FOS models in Section 6. If all assumptions are dynamic (D), then a G-FOD model results as described in Section 3. If the TIAs are a mix of static (S) and dynamic (D), then a G-FOSD model results in which the time dimension is only eliminated in certain parts of the model. In particular assuming ‘D’ in residual traffic transfer leads to interesting semi-dynamic models in which residual traffic is transferred between time periods (like in SATURN).

In this paper we focus on deriving G-FOS models and leave deriving the significantly more complex semi-dynamic models for future research.

Regarding the assumptions on the forward and backward kinematic waves, they are summarised graphically in Figure 5 for the different model types. The G-FOS assumes infinite forward wave speeds and zero or infinite backward wave speeds, as indicated by the horizontal and vertical lines in Figure 5(b). In the G-FOD model in Figure 5(d), these wave speeds are equal to the slopes of the fundamental diagram at the corresponding densities. For the G-FOSD models, the meaningful combinations of the slopes are indicated in Figure 5(c). Figure 5(a) illustrates the case for the traditional static traffic assignment model and will be further explained in Section 6.4.

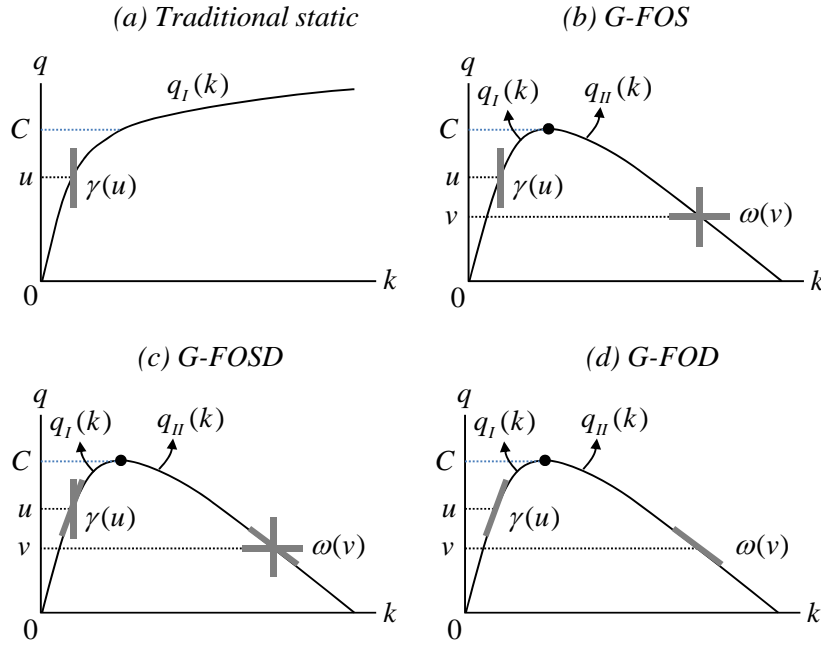


Figure 5 - Fundamental diagrams and assumptions on wave speeds.

We would like to point out that each of the above model classes are generic first order models in the sense that they can take any fundamental diagram, any turn flow restrictions, and deterministic or stochastic route choice into account. We refer to these choices as spatial interaction assumptions, which will be discussed next.

5. Spatial interaction assumptions

While the TIAs describe how the time dimension is handled in the model, the spatial interaction assumptions (SIA) describe how traffic flows spatially interact and directly impact the realism of the model. These spatial interactions are a combination of assumptions on route choice behaviour, the fundamental diagram, and turn flow restrictions (yielding flow reduction factors). As pointed out in the introduction, maintaining consistency in spatial interactions is important when moving between static and dynamic models. These spatial interactions have been extensively analysed theoretically and have been estimated empirically. Specific choices for the SIAs in a class of first order models lead to a specific traffic assignment model.

5.1 Route choice behaviour

Regarding route choice behaviour, often it is assumed that travellers have perfect knowledge regarding travel times, such that $\theta=0$ and $\beta=0$, leading to a deterministic model. In case perception errors are introduced in which users base their route choice decisions on perceived travel times, $\theta>0$ and a (logit-based) stochastic traffic assignment model results. Hence, we can define the following assumptions. In case of such a stochastic model, path overlap is taken into account if $\beta \neq 0$. This can be either done via a path-size factor (leading to a path-size logit model) or a commonality factor (leading to the C-logit model), denoted by Ψ_p^{od} . If no path overlap is taken into account, the model simplifies to the well-known conditional (also referred to as multinomial) logit model. The different SIAs regarding route choice behaviour are summarised in Table 2.

Table 2 - Spatial interaction assumptions.

<i>Assumptions on route choice behaviour</i>	
D	Deterministic route choice behaviour;
S(m)	Stochastic route choice behaviour using the multinomial logit model, ignoring path overlap;
S(c)	Stochastic route choice behaviour using the C-logit model, taking overlap into account using a commonality factor;
S(p)	Stochastic route choice behaviour using the path size logit model, taking overlap into account using a path size factor;
<i>Assumptions on the fundamental diagram</i>	
L-X	Linear uncongested branch, no congested branch;
L-H	Linear uncongested branch, horizontal congested branch;
G-X	General concave uncongested branch, no congested branch;
L-L	Triangular fundamental diagram;
Q-Q	Quadratic fundamental diagram;
Q-L	Quadratic uncongested branch, linear congested branch;
G-G	General concave uncongested branch, general concave congested branch;
<i>Assumptions on the turn flow restrictions</i>	
U	No turn flow restrictions specified, leading to unconstrained flows;
S	Separable link exit capacity constraints;
G	General non-separable turn flow restrictions;

5.2 Fundamental diagram

The shape of the fundamental diagram, which relates traffic flow and density, plays an important role in traffic flow theory and different shapes lead to different spatial traffic patterns in the model. The fundamental diagram can be directly estimated from cross-sectional traffic counts and speeds, and consists in general of an increasing concave uncongested branch, and a decreasing concave congested branch. The first fundamental diagram is described in Greenshields (1935). He proposed a linear relationship between speed and density, which results in a quadratic fundamental diagram. Such a symmetric fundamental diagram may describe uncongested flows pretty accurately, but performs poorly for congested states. A popular choice in traffic flow theory, mainly because of its simplicity, has been an asymmetric triangular fundamental diagram. While a linear relationship in the congested branch seems realistic, a linear relationship in the uncongested branch is less realistic, as this would mean that the free-flow speed is equal to the speed at capacity. Bliemer et al. (2013) therefore propose a quadratic-linear fundamental diagram as a good compromise, in which the uncongested branch of the

fundamental diagram is quadratic, while the congested branch is linear. While fundamental diagrams have been used extensively in dynamic traffic assignment models, static traffic assignment models have mainly relied on link performance functions (also called volume-delay functions), which describe the relationship between link travel time and link flow (volume) or between the speed and flow. The most well-known link performance function is the BPR link performance function (Bureau of Public Road, 1964). Rewriting the travel time in terms of speed and re-arranging terms, the BPR function can be written in form of a fundamental diagram with the following (inverse) density-flow relationship:

$$\Phi_{l,a}^{-1}(u_a(i)) = \frac{u_a(i)}{\gamma_a(0)} \left(1 + \rho_{a,1} \left(\frac{u_a(i)}{C_a} \right)^{\rho_{a,2}} \right), \quad (21)$$

where $\rho_{a,1}$ and $\rho_{a,2}$ are link type specific parameters (with typical values for a motorway of 0.15 and 4, respectively). This function is plotted in Figure 5(a). Two things can be immediately observed from this figure. First, the BPR function only contains the uncongested branch of the fundamental diagram and ignores the congested branch. In fact, any link performance function only describes the uncongested branch. Secondly, the uncongested branch is also increasing beyond the capacity flow rate. Davidson (1966) proposed an alternative to the BPR function in which the travel times goes to infinite as the flow approaches capacity, such that the uncongested branch of the fundamental diagram has a horizontal asymptote at capacity. However, this provides computational problems when flow reaches capacity, and while others have discussed modifications to eliminate these problems (e.g., Daganzo, 1977; Taylor, 1984), it is clear that link performance functions do not realistically describe the relationship between traffic flow and density. In the G-FOS models derived in this paper we therefore propose to use realistic fundamental diagrams as used in dynamic models, and derive travel times using principles from traffic flow theory.

Table 2 summarises the most common assumptions on the fundamental diagram that can be distinguished, where the first letter indicates the shape of the uncongested branch of the fundamental diagram, while the second letter is associated with the congested branch. For example ‘L-L’ indicates a triangular fundamental diagram in which both the congested and uncongested branch are linear. When using a link performance function, one implicitly assumes that no congested branch exists (denoted with ‘X’).

5.3 Turn flow restrictions

While link models and fundamental diagrams have been researched extensively for many decades, node models seem to only have received significant interest in the last decade. Given that queues and travel time delays are mainly determined by such node models, it is surprising to see that essentially all static traffic assignment models lack a node model description. In the absence of a node model, it is implicitly assumed that all flows are unconstrained and can enter and exit all links, i.e. $\alpha_a = 1$ for all links a . Under this assumption, no residual queues form. In case $\alpha_a < 1$ for a certain link a , turn flow restrictions are in place and a residual queue forms at the end of link a . However, depending on the specific formulation of such turn flow restrictions, this residual queue may be (incorrectly) inside the bottleneck link, or (correctly) upstream the bottleneck. A common choice (see e.g., Bifulco and Crisalli, 1998; Smith et al., 2013) is to constrain the outflow rate only by the link capacity. Although previous authors who used this constraint made no reference to a node model, we see this constraint as primitive node model. In this case, the turn flow restrictions are formulated as $\alpha_a(t) = C_a / s_a(t)$, such that the outflow is restricted by the physical link (exit) capacity. Note that these reduction factors are separable, i.e., only depend on the flow on the link itself. However, link inflow may exceed capacity and therefore the residual queue will be (incorrectly) in the bottleneck link.

First order node models describe reduction factors more realistically as in Equation (16), but depend on all sending and receiving flows from all links connecting the node and are therefore non-separable, such that also the link travel times are non-separable. To illustrate, consider a node in which links a and b merge into link c . Then according to the often used (capacity proportional) fair merging principle (see Jin and Zhang, 2003), $\alpha_a(t) = \min\{1, (r_c(t)/s_a(t))(C_a/(C_a + C_b))\}$ and $\alpha_b(t) = \min\{1, (r_c(t)/s_b(t))(C_b/(C_a + C_b))\}$. In case of a node in which link a diverges into links b and c , the reduction factors are often supposed to follow the (demand proportional) FIFO diverging rule (see Daganzo, 1995b), yielding $\alpha_a(t) = \min\{1, r_b(t)/s_{ab}(t), r_c(t)/s_{ac}(t)\}$. For cross nodes with multiple inlinks and outlinks, such explicit analytical expressions may not exist and one has to rely on implicit functions that need to be solved iteratively (Tampère et al., 2011). In such non-separable cases, the variational inequality problem (1) cannot be simplified into a mathematical optimisation problem (see e.g., Dafermos, 1980).

Table 2 summarises the different types of choices for the turn flow restrictions, where ‘X’ represents the absence of any turn flow restrictions such that flows are unconstrained, ‘S’ represents separable exit capacity restrictions, and ‘G’ refers to general node models with non-separable turn flow restrictions, such as the class of first order node models described in Tampère et al. (2011).

6. Generalised First Order Static (G-FOS) traffic assignment model

In this section we will mathematically derive the G-FOS models resulting from the static TIAs stated in Section 4. Further, we show that many existing static models proposed in the literature are a special case of the G-FOS model class under specific SIAs (see Section 5).

6.1 Route choice

Due to the static assumption on residual traffic in which results in an empty network at the beginning of each time period i (i.e., $Q_a(t_i) = 0$ for all links a), traffic flows and travel times are independent across time periods. Hence, we can model each time period i separately. VI problem (1) for finding a dynamic logit-based stochastic user equilibrium solution can therefore be written as a series of independent VI problems (one for each time period i) for finding static logit-based stochastic user equilibrium solutions $\mathbf{f}(i) \equiv [f_p(i)]_{p \in P}$ for each period i :

$$\sum_{(o,d)} \sum_{p \in P^{od}} \left(c_p^*(i) + \frac{1}{\theta} \log f_p^*(i) + \beta \log \Psi_p^{od} \right) (f_p(i) - f_p^*(i)) \geq 0, \quad \forall \mathbf{f}(i) \in \Omega(i), \quad (22)$$

where the set of feasible static path flow rates for time period i , $\Omega(i)$, is determined by the following constraints:

$$\sum_{p \in P^{od}} f_p(i) = M^{od}(i), \quad \forall (o,d), \quad (23)$$

$$f_p(i) \geq 0, \quad \forall p \in P. \quad (24)$$

It is clear that if inequality (22) holds for each time period i , then also inequality (1) holds as it is simply the sum of inequalities (22). In other words, the dynamic problem (1) is split into several simpler independent static sub-problems (22).

6.2 Network loading

In the network loading, we are looking for average link inflow and outflow rates as a result of the static assumptions of (unrestricted) propagation of forward shockwaves with infinite speeds and no propagation of backward shockwaves. The flow conservation constraints in Equation (4) hold for each time instant t , such that we can simply write

$$u_{bp}(i) = \begin{cases} f_p(i), & \text{if link } b \text{ is the first link on path } p, \\ v_{ap}(i), & \text{if link } b \text{ follows link } a \text{ on path } p, \end{cases} \quad (25)$$

where we provide variables that are stationary with an index i to represent all time instants in period $[t_i, t_{i+1})$. These average path-specific link inflow and outflow rates, $u_{ap}(i)$ and $v_{ap}(i)$, can be used to obtain link inflow and outflow rates, $u_a(i) = \sum_p u_{ap}(i)$, and $v_a(i) = \sum_p v_{ap}(i)$.

Now we look in more detail at the flow propagation constraints. Because we are only interested in finding an average inflow rate and an average outflow rate for each link, minimisation over realised inflow and outflow states in Equations (8) and (10) is trivial as there is only a single average inflow rate and a single average outflow rate. Define sending flows $S_a(i) \equiv S_a(t_i, t_{i+1})$ and receiving flows $R_a(i) \equiv R_a(t_i, t_{i+1})$ for the whole of time period i . We will first derive a model with vertical queues (and no spillback), before we are deriving a model with horizontal queues (and spillback).

6.2.1 Horizontal queues and spillback

Letting $\gamma_a(u) \rightarrow \infty$ (following the static assumption for forward shockwaves), $\omega_a(v) \rightarrow -\infty$ and (following the static assumption for backward kinematic waves ‘S(h)’), $Q_a(t_i) = 0$ (no residual traffic), and using $\Delta t = T_i$, Equations (8) and (10) simplify to

$$\begin{aligned} S_a(i) &= \min \left\{ U_a(t_{i+1}) - V_a(t_i) + \left(-\Phi_{I,a}^{-1}(u_a(i)) \right)^+ L_a, C_a T_i \right\} \\ &= \min \left\{ U_a(t_i) + u_a(i) T_i - V_a(t_i), C_a T_i \right\} \\ &= \min \left\{ Q_a(t_i) + u_a(i) T_i, C_a T_i \right\} \\ &= u_a(i) T_i, \end{aligned} \quad (26)$$

and

$$\begin{aligned} R_a(i) &= \min \left\{ V_a(t_{i+1}) - U_a(t_i) + K_a L_a + \left(\Phi_{II,a}^{-1}(v_a(i)) - K_a \right)^+ L_a, C_a T_i \right\} \\ &= \min \left\{ V_a(t_i) + v_a(i) T_i - U_a(t_i) + K_a L_a, C_a T_i \right\} \\ &= \min \left\{ v_a(i) T_i - Q_a(t_i) + K_a L_a, C_a T_i \right\} \\ &= \min \left\{ v_a(i) T_i + K_a L_a, C_a T_i \right\}. \end{aligned} \quad (27)$$

Hence, the average sending and receiving flow rates for time period i are

$$s_a(i) = u_a(i), \quad \text{and} \quad (28)$$

$$r_a(i) = \min \left\{ v_a(i) + K_a L_a / T_i, C_a \right\}. \quad (29)$$

To understand the meaning of the receiving flow, we can also state that $r_a(i) = C_a$ if $(C_a - v_a(i))T_i < K_a L_a$. In other words, the receiving flow is equal to capacity if the number of vehicles that remain on the link is smaller than the maximum number of vehicles that fit on the link. It should be noted that we can also replace K_a with the true queuing density available from the fundamental diagram via $\Phi_{H,a}^{-1}(v_a(i))$ in order to calculate the queue lengths and spillback more accurately.

Using Equations (12) and (13), the directional sending flows $s_{ab}(i)$ can hence be determined as

$$s_{ab}(i) = s_a(i) \sum_{p \in P} \delta_{bp} \frac{u_{ap}(i)}{u_a(i)} = \sum_{p \in P} \delta_{bp} u_{ap}(i). \quad (30)$$

The node model in Equation (16) can now be written as

$$[\alpha_a(i)]_{a \in A_n^{\text{in}}} = \Gamma_n(s_{a'b'}(i), r_b(i), \forall a' \in A_n^{\text{in}}, \forall b' \in A_n^{\text{out}}). \quad (31)$$

The directional outflow rates can then be computed as $v_{ab}(i) = \alpha_a(i) s_{ab}(i)$ (see Equation (15)), and hence the path-specific outflow rates can be written as $v_{ap}(i) = \alpha_a(i) u_{ap}(i)$. Because of Equation (25) we have that $u_{bp}(i) = v_{ap}(i) = \alpha_a(i) u_{ap}(i)$, where b is the link downstream link a on path p . Recursively applying Equation (25) yields

$$u_{ap}(i) = \delta_{ap} f_p(i) \prod_{a' \in \eta_{ap}} \alpha_{a'}(i), \quad (32)$$

where η_{ap} consists of the set of links from the origin of path p to (but not including) link a . Hence, the path-specific inflow into link a can be determined by multiplying the route flow with reduction factors (resulting from the node model) along the path. Combining Equations (30) and (32) yields

$$s_{ab}(i) = \sum_{p \in P} \delta_{ap} \delta_{bp} f_p(i) \prod_{a' \in \eta_{ap}} \alpha_{a'}(i). \quad (33)$$

Further, the receiving flows $r_b(i)$ depend on the link outflows $v_b(i)$ (see Equation (29)), which can be written as

$$v_b(i) = \sum_p v_{bp}(i) = \sum_p u_{ap}(i), \quad (34)$$

where a is the link upstream link b on path p . Hence, combining Equations (29), (32) and (34) gives

$$r_b(i) = \min \left\{ \frac{K_b L_b}{T_i} + \sum_{p \in P} \delta_{bp} f_p(i) \prod_{b' \in \eta_{bp}} \alpha_{b'}(i), C_b \right\}. \quad (35)$$

Noting that the sending flows $\mathbf{s}(i) \equiv [s_{ab}(i)]$ in Equation (33) can be expressed merely in terms of $\boldsymbol{\alpha}(i) \equiv [\alpha_a(i)]$ (and a given vector of path flows that are input to the network loading problem), that the receiving flows $\mathbf{r}(i) \equiv [r_b(i)]$ in Equation (35) can also be expressed in terms of $\boldsymbol{\alpha}(i)$ (given the network characteristics), and that the node model determines reduction

factors $\alpha(i)$ as a function of $\mathbf{s}(i)$ and $\mathbf{r}(i)$, the network loading problem can be written as a fixed point problem. The result will be fixed point α^* and resulting link inflows $u_a(i) = \sum_p \delta_{ap} f_p(i) \prod_{a' \in \eta_{ap}} \alpha_{a'}^*(i)$ and link outflows $v_a(t) = \alpha_a^*(i) u_a(i)$. The fact that the network loading model requires solving a fixed point problem is a direct consequence from the static assumption of instantaneous propagation of forward and backward kinematic waves. In the dynamic case, the network loading problem is typically solved by traffic simulation in which the link and node models are solved consecutively over time. In the static model there are instantaneous dependencies between the link and node models, which result in the fixed point problem.

6.2.2 Vertical queues and no spillback

Letting $\gamma_a(u) \rightarrow \infty$ (following the static assumption for forward shockwaves), $\omega_a(v) \uparrow 0$ (following the static assumption ‘S(v)’ for backward kinematic waves, where we note that the speeds approach zero from below as the backward kinematic waves have negative speeds), $Q_a(t_i) = 0$ (no residual traffic), and using $\Delta t = T_i$, Equations (8) and (10) simplify to

$$\begin{aligned} S_a(i) &= \min \left\{ U_a(t_{i+1}) - V_a(t_i) + \left(-\Phi_{l,a}^{-1}(u_a(i)) \right)^+ L_a, C_a T_i \right\} \\ &= \min \left\{ U_a(t_i) + u_a(i) T_i - V_a(t_i), C_a T_i \right\} \\ &= \min \left\{ Q_a(t_i) + u_a(i) T_i, C_a T_i \right\} \\ &= u_a(i) T_i, \end{aligned} \quad (36)$$

and

$$\begin{aligned} R_a(i) &= \min \left\{ V_a(t_{i+1} - \infty) - U_a(t_i) + K_a L_a + \left(\Phi_{ll,a}^{-1}(v_a(i)) + \infty \right)^+ L_a, C_a T_i \right\} \\ &= \min \left\{ 0 - U_a(t_i) + \infty, C_a T_i \right\} \\ &= C_a T_i. \end{aligned} \quad (37)$$

Hence, the average sending and receiving flow rates for time period i are simply $s_a(i) = u_a(i)$ and $r_a(i) = C_a$. The resulting model is therefore only different in the receiving flows, which are fixed to capacity, such that spillback never occurs. All equations from Section 6.2.1 are the same, except that $r_b(i) = C_b$.

6.3 Travel time calculation

Deriving travel times in a model in which the time dimension is eliminated is not trivial. Travel times can no longer be directly derived from the cumulative flow curves as in Equation (19) for dynamic models due to the fact that there is no explicit time dimension and flows are instantaneously propagated. However, we can compute the route travel times as a summation of uncongested (free-flow) travel times and delays in queues, i.e.,

$$c_p(i) = c_p^{\text{ff}}(i) + c_p^{\text{queue}}(i), \quad (38)$$

where $c_p^{\text{ff}}(i)$ is the travel time in uncongested states, and $c_p^{\text{queue}}(i)$ is the possible additional delay spent in one or more queues.

The delay in queues can be determined by the integral defined in Equation (19), noting that in a static assignment the cumulative inflow and outflow curves are ‘squeezed’ together, reflecting instantaneous flow propagation. In order to calculate travel times for all departing vehicles, we

have to let all travel demand flow out of the network. Consider again the simple corridor network in Figure 3, The grey shaded areas in Figure 6 indicate the total delay faced by departing vehicles in each of the three time periods. Figure 6(a) indicates the cumulative inflows and outflows in case of vertical queues, while Figure 6(b) show the case of horizontal queues. These figures will be further explained below. We note that in our G-FOS models, the path-specific inflow rate into the first link is $f_p(i)$, while the path-specific outflow rate out of the last link is $f_p(i) \prod_{a \in p} \alpha_a(i)$ according to Equation (32). Hence, we can write $U_{a,p}^{-1}(x) = x / f_p(i)$ and $V_{a,p}^{-1}(x) = x / f_p(i) \prod_{a \in p} \alpha_a(i)$, such that the average delay for vehicles departing during time period i on path p is

$$\begin{aligned} c_p^{\text{queue}}(i) &= \frac{1}{f_p(i)T_i} \int_{U_{a,p}^{-1}(x)}^{U_{a,p}^{-1}(x) + f_p(i)T_i} (V_{a,p}^{-1}(x) - U_{a,p}^{-1}(x)) dx \\ &= \frac{1}{f_p(i)T_i} \left[\frac{x^2}{2f_p(i) \prod_{a \in p} \alpha_a(i)} - \frac{x^2}{2f_p(i)} \right]_{x=0}^{f_p(i)T_i} \\ &= \frac{T_i}{2} \left(\frac{1}{\prod_{a \in p} \alpha_a(i)} - 1 \right). \end{aligned} \quad (39)$$

The delay is therefore a direct function of the reduction factors. If all reduction factors are equal to one, i.e. there are no bottlenecks or queues encountered on the path, then the delay is zero. This delay function can be applied both for vertical and horizontal queues.

The free-flow travel times $c_p^{\text{ff}}(i)$ calculated differently for vertical and horizontal queues. First, consider the case of vertical queues. The queues are assumed to be point queues without any physical length, hence the uncongested travel times can simply be computed dividing the link length by the link speed. The link speed is given by $u_a(i) / \Phi_{I,a}^{-1}(u_a(i))$, hence the uncongested travel time is simply

$$c_p^{\text{ff}}(i) = \sum_{a \in p} \frac{L_a \Phi_{a,I}^{-1}(u_a(i))}{u_a(i)}. \quad (40)$$

In case of horizontal queues, the queues will have a certain physical length, such that vehicles on congested links can only drive in free-flow on part of the link. To this end, we calculate the queue length by dividing the vehicles in the queue by the jam density, i.e., $Q_a(t_{i+1}) / K_a$, where $Q_a(t_{i+1}) = (u_a(i) - v_a(i))T_i = u_a(i)(1 - \alpha_a(i))T_i$. Instead of taking the jam density, it is also possible to more accurately determine the queuing density from the fundamental diagram, $\Phi_{II,a}^{-1}(v_a(i))$. by using the queuing density resulting from the congested branch of the fundamental diagram, yielding a queue length of $Q_a(t_{i+1}) / q_{II,a}^{-1}(v_a(i))$. This leads to a the following uncongested route travel time,

$$c_p^{\text{ff}}(i) = \sum_{a \in p} \left(L_a - \frac{u_a(i)(1 - \alpha_a(i))T_i}{\Phi_{II,a}^{-1}(\alpha_a(i)u_a(i))} \right) \frac{\Phi_{I,a}^{-1}(u_a(i))}{u_a(i)}, \quad (41)$$

As far as we are aware, this travel time calculation includes for the first time the complete fundamental diagram in a static model.

To illustrate Equation (39), first consider the example corridor in Figure 3 with vertical queues. In the first time period, only link 4 is a bottleneck and a queue will form at the end of link 3. It can be easily shown that $\alpha_1(1) = \alpha_2(1) = \alpha_4(1) = 1$, and $\alpha_3(1) = 2000/3000 = 0.667$. Hence, the average delay according to Equation (39) is $\frac{1}{2}(1/(1 \cdot 1 \cdot 0.667 \cdot 1) - 1) = 0.25$ h. This can be confirmed by looking at the queues, namely there are 1000 vehicles waiting after one hour at the end of link 3. The last vehicle will therefore have a delay of $1000/2000 = 0.5$ h, while the first vehicle faces no delay, hence the average delay is 0.25 h. In the second period, all vehicles from period 1 are assumed to have flown out and the travel demand of 6000 veh/h during one hour, such that links 3 and 4 will both be bottlenecks and a queue will appear at the end of links 2 and 3. The network loading problem will yield that $\alpha_1(2) = \alpha_4(2) = 1$, $\alpha_2(2) = 4000/6000 = 0.667$, and $\alpha_3(2) = 2000/4000 = 0.5$. Hence, the average delay will be $\frac{1}{2}((1/(1 \cdot 0.667 \cdot 0.5 \cdot 1) - 1)) = 1$ h. This can be confirmed by looking at the delays on links 2 and 3. There will be a queue of 2000 vehicles on link 2, which has an outflow rate of 4000 veh/h, hence the last vehicle will have a delay of $2000/4000 = 0.5$ h. There will also be a queue of 2000 vehicles on link 3, which has an outflow rate of 2000 veh/h, hence the last vehicle in this queue will see a delay of $2000/2000 = 1$ h. However, note that all travel demand has to pass this queue, also the vehicles on link 2. If it takes 1 h to let all 4000 vehicles flow out of link 3 during one hour, it will take $(6000/4000) \cdot (2000/2000) = 1.5$ h for all 6000 vehicles to pass link 3. Hence, the last vehicle faces a delay of $0.5 + 1.5 = 2$ h, such that the average delay is indeed 1 h. The scaling factor of $6000/4000$ is often forgotten, but is necessary in order to compute travel times for all departing vehicles. In time period 3 the travel demand is low, such that no queues appear.

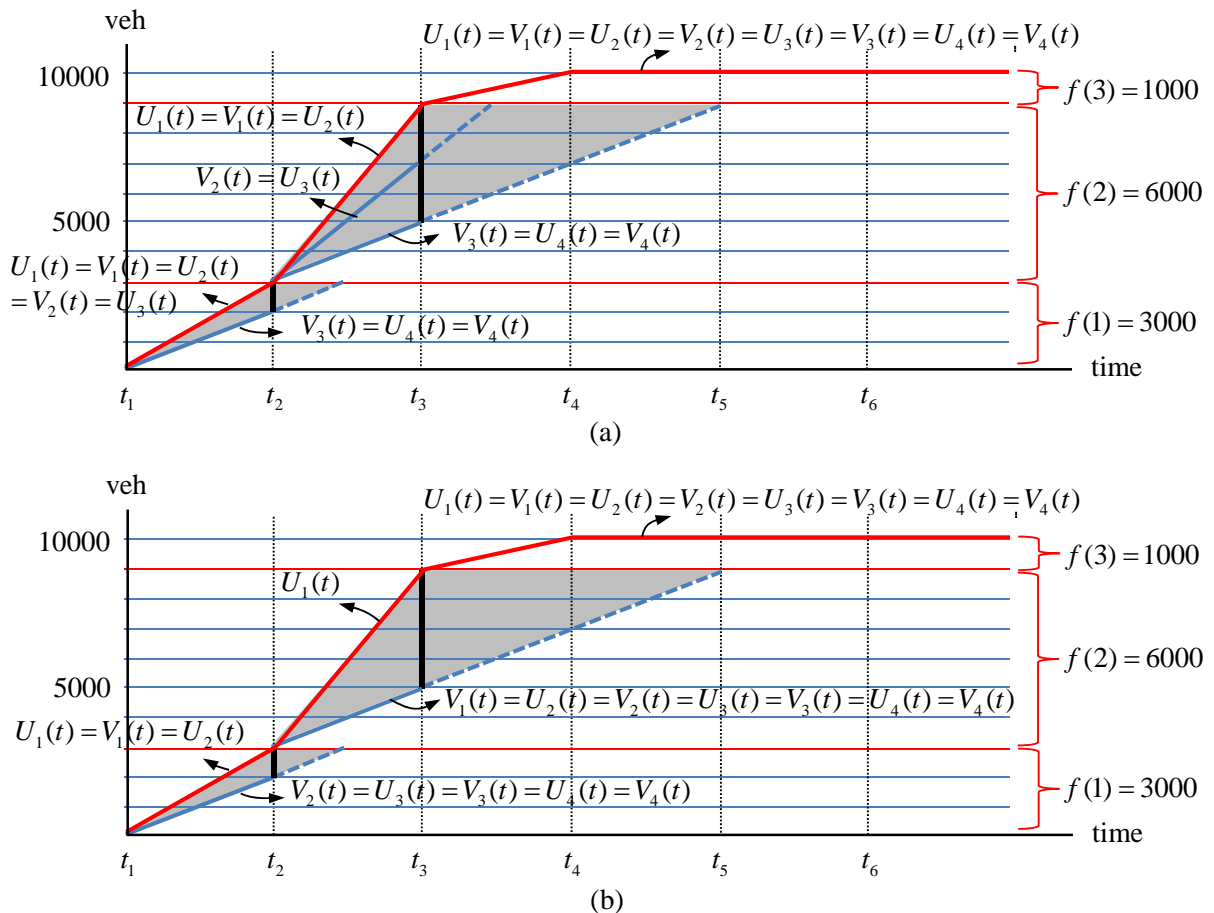


Figure 6- Using cumulative link inflows and outflows to determine delays in static assignment with vertical queues (a) or horizontal queues (b).

Now let us consider the second time period of the same corridor but now with horizontal queues. Assume that all links have a length of 3 km each, and for convenience we will use the following jam densities as queue densities, $K_1 = K_2 = 600$, $K_3 = 400$, and $K_4 = 200$. It can easily be shown that the solving the network loading problem yields $\alpha_1(2) = 2000/6000 = 0.333$, $\alpha_2(2) = \alpha_3(2) = \alpha_4(2) = 1$. In other words, the queue on link 3 is spilling back upstream and reaches link 1, such that all links have an outflow rate of 2000 veh/h. In total there are 4000 vehicles in the queue, of which $400 \cdot 3 = 1200$ vehicles on link 3, $600 \cdot 3 = 1800$ on link 2, and the remaining 1000 on link 1. According to Equation (39) the average delay is $\frac{1}{2} \left((1/(1 \cdot 0.333 \cdot 1 \cdot 1)) - 1 \right) = 1$ h, the same as in the case with vertical queues. Note that in this simple corridor case, it does not matter for the delay whether we consider horizontal or vertical queues, as all that matters is the outflow rate out of the last link. This does not mean that it does not matter for path travel times whether queues or horizontal or vertical. On the contrary, if there was another route passing through link 1 and not through the other links, the path travel time of this route would be affected with an additional delay on link 1 in case of horizontal queues, while in case of vertical queues this delay on link 1 does not exist.

It is important to note the main differences between the delays in Figure 4 and Figure 6. The assumption of no residual traffic transfer leads to empty networks at the beginning of each time period. Clearly, this can reduce the modelled total delay experienced by travellers significantly, as for example can be seen when comparing the delays for time period 3. Without considering residual traffic at the beginning of each time period, there are no delays predicted in the static model, while there are significant delays predicted in the dynamic model. This illustrates that the assumption of no residual traffic transfer is a very strong assumption that preferably should be relaxed, leading to semi-dynamic models.

6.4 Specific first order static models

Most static (and quasi-dynamic) models proposed in the literature are a special case of the class of G-FOS models that we derived. Different spatial interaction assumptions lead to different models.

The model proposed by Smith (2013) can be seen as a special case of the G-FOS model with horizontal queues, in which route choice is assumed deterministic, the uncongested branch of the fundamental diagram is assumed concave, and turn restrictions are defined by simple link exit capacities. Hence, the G-FOS model with horizontal queues as derived in this paper therefore extends the model of Smith (2013) to any fundamental diagram, to any first order node model that describes the turn restrictions, and to stochastic route choice.

Bliemer et al. (2013) proposed a model with vertical queues in which the turn restrictions are describes by the node model of Tampère et al. (2011), while route choice based on the multinomial logit model and a quadratic-linear fundamental diagram are assumed, and hence is a special case of the G-FOS model with vertical queues. Bliemer et al. (2013) have further shown that the fixed point problem for network loading as derived in this paper can be efficiently solved for large scale networks.

The most well-known static traffic assignment model is the deterministic model (i.e., $\theta = \infty$, $\beta = 0$) proposed by Beckmann et al. (1956) using link performance functions, hence it is implicitly assumed there is no congested branch of the fundamental diagram. Further, no turn flow restrictions are considered in this formulation, hence $\alpha_a = 1$ for all links, yielding unrestricted flow propagation along each route. This strong assumption yields that Equation (32) simplifies to $u_{ap}(i) = \delta_{ap} f_p(i)$, hence the fixed point problem in Section 6.2 simplifies to the simple relationship $u_a(i) = \sum_p \delta_{ap} f_p(i)$, which is the familiar constraint in the traditional static traffic assignment model. Hence, also the traditional static traffic assignment model is a special

case of the G-FOS model. Due to the fact that only the uncongested branch of the fundamental diagram is described and no turn flow restrictions are considered, this model is essentially only applicable to uncongested networks, which questions its suitability for use in heavily congested areas.

Extensions to the deterministic route choice model have been proposed by Fisk (1980) and Zhou et al. (2012) by including perception error. Fisk used the multinomial logit model for route choice while Zhou et al. adopts the C-logit route choice model. Both stochastic traffic assignment models use the simple network loading relationship from Beckmann et al. (1956).

Bifulco and Crisalli (1998) were the first to formulate the static network loading problem as a fixed point problem and use a formulation similar to Equation (32). However, they did not include a proper node model but instead assumed that reduction factors are derived from link exit capacities, such that $\alpha_a(i) = C_a / u_a(i)$. This means that outflow is constrained to capacity, leading to residual (vertical) queues typically inside the bottleneck link. Others have implicitly or explicitly also made this assumption, i.e. Lam and Zhang (2000) and Smith et al. (2013).

There are other static models proposed in the literature that cannot be derived from the G-FOD model. For example, Yang and Yagar (1994), Larsson and Patriksson (1995) and Bell (1995) propose capacity constrained static traffic assignment models in which the link inflow may not exceed the link capacity, but no explicit residual queues are formed. Instead, Lagrange multipliers of the link capacity constraints are computed to represent link delays. These delays re-route path flows in order to ensure that no link flows exceed link capacity (assuming there is sufficient capacity in the network). It is not possible to formulate a fundamental diagram, turn flow restrictions and specify route choice behaviour that results in these capacity constrained static models, and hence these models are not first order static models.

7. Conclusions and discussion

In this paper we have presented a theoretical framework in which we can derive static (or quasi-dynamic) and semi-dynamic traffic assignment models from a general first order dynamic traffic assignment model using temporal interaction assumptions on forward and backward kinematic waves and on the transfer of residual traffic. Further, the framework is very general in that it allows to take into account any fundamental diagram (i.e., a link model), any turn flow restrictions (i.e., a node model), and deterministic or stochastic route choice behaviour. This theoretical framework is then used to derive novel static (quasi-dynamic) traffic assignment models that take queuing and spillback into account. We believe this is the first time that static models have actually been rigorously derived from a dynamic model. The advantage of such a derivation is that it maintains maximum consistency with a dynamic model. The resulting static models with vertical and horizontal queues are essentially general forms of the models presented by Bliemer et al. (2013) and Smith et al. (2013), respectively.

One of the main contributions of this paper is that we have explicitly formulated all underlying assumptions made in static (quasi-dynamic) traffic assignment models proposed in the literature. One of the assumptions implicitly made in static models is the absence of residual traffic transfer by assuming that the network is empty at the beginning of each time period under consideration. While this may not be a large problem if the period under consideration in static assignment is sufficiently large (i.e., covers the entire peak period from the moment of queue build up till the moment that all queues have disappeared), it will lead to significant underestimations of travel times in case of successive static traffic assignment on smaller time period (i.e., hourly static traffic assignment while the peak period is several hours long). In the latter case, transfer of residual traffic seems very important. Our framework allows for deriving models that consider residual traffic transfer, leading to what we call semi-dynamic models. These models are significantly more complex and will be mathematically derived in future research.

As a special case, the traditional static traffic assignment model proposed by Beckmann et al. (1956) can be derived by assuming (i) infinite forward kinematic wave speeds, (ii) no backward kinematic wave speeds, (iii) no residual traffic transfer, (iv) a concave function for the uncongested branch of the fundamental diagram while no function is specified for the congested branch, (v) no turn flow restrictions, and (vi) deterministic route choice. Several of these assumptions are very strong, and essentially limit the applicability of this model in practice. The novel static models derived in this paper relax several of these strong assumptions, making the forecasts more realistic and in line with dynamic models.

Our derivations are based on the general link transmission model as an exact and general formulation of a first order dynamic traffic assignment model. A limitation in our derivations is that it only holds for a single vehicle class and cannot handle multiclass traffic. We are not aware of any other first order dynamic model that provides an explicit, exact, and analytical expression of flow rates, which is needed for our derivations. This does not mean we cannot formulate multiclass static models. In fact, this can be done in a relatively straightforward fashion by performing a single class network loading using passenger car units, but in the uncongested part of the travel time calculation use for each vehicle class a different function for the uncongested branch of the fundamental diagram. Queuing delays may be assumed equal for all vehicle classes. This would provide a pragmatic way of including multiple vehicle classes, without explicitly deriving a multiclass static model from a multiclass dynamic model.

References

- 4Cast (2009). Qblok_2004: toedelingsprocedure LMS/NRM (in Dutch). Working report (the Netherlands)
- Astarita, V. (1996). A continuous time link model for dynamic network loading based on travel time function. *Transportation and Traffic Theory. Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pp. 79–101.
- Bakker, D., P.H. Mijjer, and F. Hofman (1994) QBLOK: An assignment technique for modeling the dependency between bottlenecks and the prediction of ‘grid lock’. *Proceedings of Colloquium Vervoersplanologisch Speurwerk, Delft*, pp. 313-332.
- Beckmann, M.J., C.B. McGuire, and C.B. Winsten (1956) *Studies in economics of transportation*, Yale University Press, New Haven CT, USA.
- Bell, M.G.H. (1995) Stochastic user equilibrium in networks with queues. *Transportation Research Part B*, Vol. 29, No. 2, pp. 115–112.
- Ben-Akiva, M. and M. Bierlaire (1999) Discrete Choice Methods and Their Applications to Short Term Travel Decisions. In: R. Hall (ed.) *Handbook of Transportation Science*, Kluwer, Dordrecht, pp. 5-34.
- Bifulco, G., and Crisalli, U. (1998). Stochastic user equilibrium and link capacity constraints: formulation and theoretical evidencies. In *Proceedings of the European Transport Conference*. pp. 85–96.
- Bliemer, M.C.J. and P.H.L. Bovy (2003) Quasi-variational inequality formulation of the multiclassdynamic traffic assignment problem. *Transportation Research Part B*, Vol. 37, pp. 501-519.

A unified framework for traffic assignment: deriving static and quasi-dynamic models consistent with general first order dynamic traffic assignment models

Bliemer, Raadsen, Brederode, Bell and Wismans

Bliemer, M.C.J. (2007) Dynamic queuing and spillback in an analytical multiclass dynamic network loading model. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2029, pp. 14-21.

Bliemer, M.C.J., and P.H.L. Bovy (2008) Impact of route choice set on route choice probabilities. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2076, pp. 10-19.

Bliemer, M.C.J., Raadsen M.P.H., Smits E.S., Zhou B., Bell M.G.H. (2013) Capacity constrained stochastic static traffic assignment with residual point queues incorporating a proper node model. ITLS Working Paper.

Bliemer, M.C.J., Brederode, L., Wismans, L., and Smits, E.S. (2012). Quasi-dynamic network loading: adding queuing and spillback to static traffic assignment. *Annual Transportation Research Board Meeting 2012*.

Bovy, P, S. Bekhor, and C.G. Prato (2008). The path size factor revisited: an alternative derivation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2076, pp. 132-140.

Brederode, L., Bliemer, M. C. J., and Wismans, L. (2010). STAQ: Static traffic assignment with queuing. Presented at: European Transport Conference 2010.

Bureau of Public Roads (1964) *Traffic Assignment Manual*. U.S. Dept. of Commerce, Urban Planning Division, Washington D.C., USA.

Bundschuh, M., Vortisch, P. and T. Van Vuuren (2006) Modelling queues in static traffic assignment. *Proceedings of the European Transport Conference 2006*.

Cascetta, E., Nuzzolo, A., Russo, F., and Vitetta, A. (1996) A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks, in Lesort, J. B. (ed.), *Proceedings of the International Symposium on Transportation and Traffic Theory*. Elsevier, New York, pp. 697–711.

Chabini, I. (2001). Analytical dynamic network loading problem formulation, solution algorithms, and computer implementations. *Transportation Research Record Vol. 1771*, *Transportation Research Board*, Washington, D.C., pp. 191–200.

Chen, H.K., Hsueh, C.F. (1998). A model and an algorithm for the dynamic user optimal route choice problem. *Transportation Research Part B. Methodological*. Vol. 32, No. 3, pp 219-234.

Chen, H.-K. (1999) *Dynamic travel choice models: a variational inequality approach*. Springer-Verlag Berlin-Heidelberg.

Dafermos, S.C. (1980) Traffic equilibrium and variational inequalities. *Transportation Science*, Vol. 14, No. 1, pp. 42-54.

Daganzo, C. and Sheffi Y. (1977) On stochastic models of traffic assignment. *Transportation Science* Vol. 11, No 3, pp 253-274.

Daganzo, C.F. (1977) On the traffic assignment problem with flow dependent costs—II. *Transportation Research*, Vol. 11, pp. 439–441.

**A unified framework for traffic assignment: deriving static and quasi-dynamic models
consistent with general first order dynamic traffic assignment models**

Bliemer, Raadsen, Brederode, Bell and Wismans

Daganzo, C. (1994) The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, *Transportation Research Part B*, Vol. 28, No. 4, pp. 269–87.

Daganzo, C. (1995a) Requiem for second-order fluid approximations of traffic flow, *Transportation Research Part B*, Vol. 29, No. 4, pp. 277–286.

Daganzo, C.F. (1995b) The cell transmission model II: network traffic. *Transportation Research Part B*, Vol. 29, No. 2, pp. 79–93.

Davidson, K.B. (1966) A flow travel-time relationship for use in transport planning. *Proceedings of the 3rd Australian Road Research Board Conference*, Vol. 8, No. 1, pp. 32-35.

Davidson, P., Thomas, A., & Teye-Ali, C. (2011). Clocktime assignment: a new mesoscopic junction delay highway assignment approach to continuously assign traffic over the whole day. In *European Transport Conference Proceedings*.

De Palma, A., Nesterov, Y. (1998). Optimization formulations and static equilibrium in congested transportation networks. 8th WCTR-Conference, pp 1–34.

Fisk, C. (1980). Some developments in equilibrium traffic assignment. *Transportation Research Part B: Methodological*, Vol. 14 No. 3, pp. 243–255.

Flötteröd, G., and J. Rohde (2011) Operational macroscopic modeling of complex urban road intersections. *Transportation Research Part B*, Vol. 45, No. 6, pp. 903-922.

Gentile G. (2010) The General Link Transmission Model for Dynamic Network Loading and a comparison with the DUE algorithm, in L.G.H. Immers, C.M.J. Tampère, and F. Viti (eds) *New developments in transport planning: advances in Dynamic Traffic Assignment*. Transport Economics, Management and Policy Series, Edward Elgar Publishing, MA, USA.

Gibb, J. (2011) Model of traffic flow capacity constraint through nodes for dynamic network loading with queue spillback. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2263, pp. 113-122.

Greenshields, B.D. (1935) A study of traffic capacity. *Highway Research Board Proceedings*, Vol. 14, pp. 448–477.

Janson, B.N. (1991) Dynamic traffic assignment for urban road networks. *Transportation Research Part B*, Vol. 25, No. 2/3, pp. 143-161.

Jin, W.-L. (2012a) The traffic statics problem in a road network. *Transportation Research Part B*, Vol. 46, pp. 1360-1373.

Jin, W.-L. (2012b) A kinematic wave theory of multi-commodity network traffic flow. *Transportation Research Part B*, Vol. 46, pp. 1000-1022.

Jin, W.-L., Zhang, H.M. (2003) On the distribution schemes for determining flows through a merge. *Transportation Research Part B*, Vol. 37, No. 6, pp. 521–540.

Jin, W.L. and H.M. Zhang (2004) Multicommodity kinematic wave simulation model for network traffic flow. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1883, pp. 59-67.

A unified framework for traffic assignment: deriving static and quasi-dynamic models consistent with general first order dynamic traffic assignment models

Bliemer, Raadsen, Brederode, Bell and Wismans

Lam, W., and Zhang, Y. (2000). Capacity-constrained traffic assignment in networks with residual queues. *Journal of Transportation Engineering*, (April), 121–128.

Larsson, T., and M. Patriksson (1995) An augmented Lagrangean dual algorithm for link capacity side constrained traffic assignment problems. *Transportation Research Part B*, Vol. 29, pp. 433-455.

Lebacque, J.P., and M.M. Khoshyaran (2005) First-order macroscopic traffic flow models: Intersection modeling, network modeling. *Proceedings of the 16th International Symposium on Transportation and Traffic Theory (ISTTT)*, pp. 365-386.

Lighthill, M. J., and Whitham, G. B. (1955). On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 229, pp. 317–345.

Long, J., Z. Gao, and W.Y. Szeto (2011) Discretised link travel time models based on cumulative flows: formulations and properties. *Transportation Research Part B*, Vol. 45, No. 1, pp. 232-254.

Messmer A., Papageorgiou M., (1990). METANET: a macroscopic simulation program for motorway networks, *Traffic Engineering and Control*. Volume 31 issue 9, 1990, pp466–470.

Nesterov, Y., De Palma A. (2000) Stable dynamics transportation systems. *Proceedings of the European Transport Conference*.

Nesterov, Y., de Palma, A. (2003). Stationary dynamic solutions in congested transportation networks: summary and perspectives. *Networks and Spatial Economics*, 371–395.

Newell, G.F. (1993) A simplified theory of kinematic waves in highway traffic, Part I: General theory, Part II: Queuing at freeway bottlenecks, Part III: Multi-destination flows, *Transportation Research Part B*, Vol. 27, pp. 281-313.

Ran, B., and Boyce, D. (1996) Modeling dynamic transportation networks: an intelligent transportation system oriented approach. Springer-Verlag, Heidelberg.

Richards, P. I. (1956). Shock Waves on the Highway. *Operations Research*, Vol. 4, pp. 42–51.

Smith, M. J. (1987). Traffic control and traffic assignment in a signal-controlled network with queueing. In N. H. Gartner & Wilson M. (Eds.), *Proceedings of the Tenth International Symposium on Transportation and Traffic Theory* (pp. 61–68). Elsevier.

Smith, M. J. (2013). A link-based elastic demand equilibrium model with capacity constraints and queueing delays. *Transportation Research Part C: Emerging Technologies*, 29, 131–147.

Smith, M. J., Huang, W., and Viti, F. (2013) Equilibrium in capacitated network models with queueing delays, queue-storage, blocking back and control. *Procedia – Social and Behavioral Sciences*, Vol. 80, pp. 860-879.

Tampère, C.M.J. , Corthout, R., D. Cattrysse and L.H. Immers (2011) A generic class of first order node models for dynamic macroscopic simulation of traffic flows, *Transportation Research Part B*, Vol. 45, pp. 289-309.

Taylor, M.A.P. (1984) A note on using Davidson’s function in equilibrium assignment. *Transportation Research Part B*, Vol. 18, No. 3, pp. 181-199.

Van Vliet, D. (1982). SATURN: a modern assignment model. *Traffic Engineering and Control*, 23, 575–581.

Wardrop, J.G. (1952) Some theoretical aspects of road traffic reseach. *Proceedings of the Institute of Civil Engineers, Part II*, pp. 325-378.

Wu, J., Chen, Y., & Florian, M. (1998). The continuous dynamic network loading problem: a mathematical formulation and solution method. *Transportation Research Part B: Vol. 32 No.3*, pp. 173–187.

Xu, Y.W., H. Wu, M. Florian, P. Marcotte and D.L. Zhu (1999). Advances in the continuous dynamic network loading problem. *Transportation Science*, Vol. 33, No. 4, pp. 341-353.

Yang, H., and S. Yagar (1994) Traffic assignment and traffic control in general freeway arterial. *Transportation Research Part B*, Vol. 28, No. 4, pp. 463-486.

Yperman, I. (2007) *The Link Transmission Model for Dynamic Network Loading*. PhD Thesis, Katholieke Universiteit Leuven, Belgium.

Zhou, Z., A. Chen, and S. Bekhor (2012) C-Logit Stochastic User Equilibrium Model: Formulations and Solution Algorithm. *Transportmetrica*, Vol. 8, No. 1, pp. 17–41.