Requirements for traffic assignment models for strategic transport planning: A critical assessment.

By
Michiel Bliemer, Mark Raadsen, Erik de Romph[1] and Erik-Sander Smits[1]

[1] Department of Transport and Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology   The Netherlands

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in

Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

| TITLE: | **Requirements for traffic assignment models for strategic transport planning: A critical Assessment.** |

| ABSTRACT: | Transport planning models are used all over the world to assist in the decision making regarding investments in infrastructure and transport services. Traffic assignment is one of the key components of transport models, which relate travel demand to infrastructure supply, by simulating (future) route choices and network conditions, resulting in traffic flows, congestion, travel times, and emissions. Cost benefit analyses rely on outcomes of such models, and since very large monetary investments are at stake, these outcomes should be as accurate and reliable as possible. However, the vast majority of strategic transport models still use traditional static traffic assignment procedures with travel time functions in which traffic flow can exceed capacity, delays are predicted in the wrong locations, and intersections are not properly handled. On the other hand, microscopic dynamic traffic simulation models can simulate traffic very realistically, but are not able to deal with very large networks and may not have the capability of providing robust results for scenario analysis. In this paper we discuss and identify the important characteristics of traffic assignment models for transport planning. We propose a modelling framework in which the traffic assignment model exhibits a good balance between traffic flow realism, robustness, consistency, accountability, and ease of use. Furthermore, case studies on several large networks of Dutch and Australian cities will be presented. |

| KEY WORDS: | *Traffic assignment, strategic transport planning, quasi-dynamic, model assessment, strict capacity constraints, physical queuing.* |

| AUTHORS: | **Bliemer, Raadsen, de Romph and Smits** |

| CONTACT: | INSTITUTE of TRANSPORT and LOGISTICS STUDIES (C13) The Australian Key Centre in Transport and Logistics Management |
| | The University of Sydney   NSW   2006   Australia |

	Telephone:	+612 9114 1824
	Facsimile:	+612 9114 1722
	E-mail:	business.itlsinfo@sydney.edu.au
	Internet:	http://sydney.edu.au/business/itls

1.   Introduction

Transport planners aim to prepare, assess, and implement different plans and projects in order to improve and manage transport systems, which include

  (i)   road and rail infrastructure (e.g., adding new or expanding existing infrastructure);
  (ii) public transit services (e.g., new bus routes, frequency changes, etc.);
  (iii) demand management policies (e.g., road pricing);
  (iv) traffic management policies (e.g., ramp metering);
  (v)   information strategies (e.g., real-time route information); and
  (vi) land use policies (e.g., new urban developments).

These plans and projects typically involve large amounts of money, and the decisions will usually have long term impacts. Therefore, many governments all over the world use strategic transport models to make forecasts of such impacts and compare different scenarios.

Australia is no different, as indicated in Table 1. A strategic transport model exists for every major metropolitan area and serves as a tool that supports decision making in transport systems. These models can be applied in the preparation phase to do a quick scan of a wide range of possible solutions, later in the assessment phase to compare different alternative solutions in more detail, and finally in the implementation phase to for example look at the consequences of the construction, which may take many years.

Table 1: Main strategic transport models in Australia

Model name	Abbreviation	State	Area
Sydney Strategic Travel Model	STM	NSW	Sydney
Melbourne Integrated Transport Model	MITM	VIC	Melbourne
Canberra Strategic Transport Model	CSTM	ACT	Canberra
Brisbane Strategic Transport Model – Multi Modal	BSTM-MM	QLD	Brisbane
Metropolitan Adelaide Strategic Transport Evaluation Model	MASTEM	SA	Adelaide
Strategic Transport Evaluation Model	STEM	WA	Perth

1.1   Forecasts used in decision making

When making decisions regarding infrastructure investments and transport policies, the advantages and disadvantages of the investment or policy need to be determined. This is often done by economic appraisal, which takes a wide range of costs and benefits into account. The most common are a cost-benefit analysis (CBA) in which all costs and benefits are quantified in monetary terms. In case outputs are difficult to measure in monetary terms, one can apply a cost-effectiveness analysis (CEA). If the benefits are larger than the costs, i.e. if the benefit-cost ratio is larger than 1, the project is said to be economically beneficial. An environmental impact assessment is often part of the economic appraisal.

While the costs are relatively easy to determine (although the costs have a wide range due to risk and uncertainty involved), the benefits are more difficult to establish. There can be a whole wide range of benefits (although they can also be dis-benefits if they make the current status quo worse):

  • Decrease in travel time
  • Increase in travel time reliability
  • Increase in health (measured by emissions of NO_x, PM_{10}, $PM_{2.5}$, etc.)
  • Decrease in climate change (measured by emissions of CO_2)
  • Decrease in noise
  • Increase in employment

- Increase in safety (measured by the number of fatal and non-fatal crashes)
- Other effects (e.g., agglomeration effects, wider economic impacts)

In order to express each of these benefits into a monetary value, certain conversions are required. For example, an hour less travel time can be converted to dollars using the value of travel time savings (VTTS), also often referred to as the value of time (VOT).

Strategic transport models may not provide direct estimates of these benefits, but they can often be derived from model outputs. While travel time savings are a natural and often the most important outcome of strategic transport models, the benefits related to travel time reliability, health, climate change, and noise are often derived from traffic flows, speeds, and distance travelled, which are also outcomes of the model.

1.2   *Model components*

Strategic transport models often adopt a similar structure as outlined by the classical four stage model (Ortúzar and Willumsen, 2011), but exceptions exist. They all comprise of the following two main components: (i) demand model, and (ii) supply model.

The demand model generates the travel demand, and reflects the travel decisions of agents in the transport system. For passenger transport, these decisions include activity choice, trip choice, destination choice, mode choice, departure time choice, and route choice. The result is therefore the travel demand from a certain origin to a certain destination using a certain mode at a specific time on a specific route. Similarly, freight transport can be described by a demand model.

The supply model describes the interaction of the travel demand with the supply of infrastructure and transport services. Like in any economic model, if the demand is larger than the supply, costs will increase (i.e., congestion and delays will occur). Traffic simulations are typical examples of supply models, in which the travel demand and infrastructure supply is input, and the traffic simulation determines the flows, speeds, travel times and delays on road segments. Clearly, there is interaction between the demand and supply models: agents may decide to change their route, departure time, mode, or destination to avoid long delays or high costs, which in return will have an effect on the traffic conditions. The traffic simulations can range from relatively simple static traffic assignment models that consider macroscopic traffic flow rates (i.e., vehicles per hour), to very elaborate microscopic simulation models, in which each vehicle is simulated individually.

1.3   *Paper contributions and outline*

In this paper, we will focus on traffic assignment models of *road* transport for strategic transport planning purposes. In other words, we will concentrate on supply models with the inclusion of route choice behaviour. Further, we will only consider passenger and freight transport, in which we refer to the decision makers as agents, i.e. car drivers and transporters of goods. We will not discuss public transport due to their very specific driving pattern with many stops. We will also not consider taxis, although they could be considered as passenger cars with possibly extra permissions regarding infrastructure use (bus and taxi lanes). Since we are focussing on *strategic* models, we are only looking at models for long term prognoses, not for short term analysis such as incidents or road works.

Many traffic assignment models have been proposed in the literature, and there exists a wide range of commercially available software that can perform traffic assignment, namely static models in general transportation software (such as TransCAD, OmniTRANS, EMME, VISUM, Cube), dynamic models that present flow macroscopically as flow rates using a fundamental diagram (such as StreamLine, INDY), and dynamic models that present flow microscopically as individual vehicles (AIMSUN, VISSIM, PARAMICS). Microscopic models often use car following and lane changing behaviour instead of a fundamental diagram. Some dynamic models are called mesoscopic, as they simulate individual vehicles or packets of vehicles using

a fundamental diagram (INTEGRATION, DYNASMART, Dynameq). More recently, network models have been developed that can mix meso and micro levels on the same network (e.g., Transmodeler, AIMSUN). There also exist models that are somewhere in between static and dynamic (such as SATURN, QBLOK).

For the purpose of strategic transport models, the current state of the practice is the use of static traffic assignment models. However, these models have serious drawbacks and may produce very unrealistic outcomes, which may lead to significant errors in decision making. Dynamic models are able to produce much more realistic traffic conditions, and advances in computing power have made them more feasible for larger areas. But the question is whether these detailed dynamic models are the right tool for strategic transport planning. Dynamic assignment models prove to be rather difficult and cumbersome to operate. Furthermore, feedback loops between the supply and demand models require long running times and the results not stable enough to be practical for scenario analysis. Likely, the 'best' model is therefore somewhere in between the two extremes of static models and dynamic microscopic simulation models.

In this paper we explore a range of desired properties for traffic assignment models to determine the 'best' traffic assignment model for the purpose of *strategic* transport planning. We will argue that traditional static traffic assignment models, which may have some merits and are still widely used by transport planners all over the world, often generate problematic and unrealistic traffic conditions and travel times, and are therefore not the most suitable tool for decision making in transport planning. Further, we will argue that moving towards very detailed microscopic simulation models is not the answer either. Based on a detailed analysis of criteria, we will show that the 'best' model will be a route-based capacity-constrained traffic assignment model that is consistent with a proper link model (consistent with a realistic fundamental diagram) and a proper node model (consistent with conditions stated in Tampère, 2011), which converges to a unique stochastic user-equilibrium. We will then show that the quasi-dynamic approach is the most computational efficient approach which satisfies these criteria. This model will exhibit realistic route choice behaviour and traffic flow characteristics, will yield robust results, will be consistent with dynamic models, will provide reliable accountable results, and will be easy to use. We finally show that such a model is feasible on large realistic networks, including networks of the size currently used in strategic transport models in Australia.

Section 2 will discuss the properties that we believe a good traffic assignment model for strategic transportation planning should have. Section 3 will then assess the impact of these properties on the choices for an appropriate model. Following from these choices, in Section 4 we will present a traffic assignment modelling framework that adheres to these choices as much as possible. Section 5 presents some case studies illustrate the feasibility of our approach, and we conclude and give a final discussion in Section 6.

2. Desired properties for traffic assignment models

In this section we discuss desired properties for traffic assignment models for strategic planning purposes. We distinguish the following properties:

(i) Realism of results
(ii) Robustness of results
(iii) Consistency of results
(iv) Reliability and accountability of results
(v) Ease of use

We will look at each property in more detail, and determine some model criteria. Some criteria may be conflicting, therefore in Section 3 we will have to find a good balance between these criteria.

2.1 Realism of results

2.1.1 Realistic route choice behaviour

Agents base their route decisions on many factors. Car drivers will choose their route based on travel time (both free-flow and congestion delays), travel costs (including tolls and running costs), travel distance, travel time reliability (for example, expressed in terms of a standard deviation), familiarity with the route, the type of road along the route (motorway, urban roads), and the number of intersections encountered (where it may also matter how often one turns left or right). Transporters of freight will aim to optimise their routes for deliveries and typically minimise costs. The routes they can choose may be limited, dependent on the size of their vehicle and their load.

Route choice should at least consider the travel time including congestion and other delays (e.g., at intersections due to traffic lights), as travel time is one of the main determinants in route choice. It also needs to include costs such as running costs and tolls, in order to correctly forecast route changes due to taxation and pricing policies. In the last decade, travel time reliability has been argued as another important factor for route decisions. Depending on the trip purpose, one may choose a more reliable (but potentially longer) route, in order to guarantee being in time at the destination. Including travel time reliability is not trivial, but some studies have showed that it may be possible to express travel time reliability as a function of the travel time (Hellinga et al., 2012).

Familiarity is closely related to habitual behaviour. In strategic models, it is common to just consider a single representative day or peak period, such that only one route choice decision is used. Clearly, travellers do not always take the same route, such that it is important to consider repetitive choice making in which multiple routes are taken into account. Familiarity will have a direct influence on these repetitive choices.

Agents may have different preferences towards the route attributes. These preferences may depend on the trip purpose (i.e., work, education, leisure, shopping), the person type (i.e., gender, income), and many other factors. It is therefore important to take preference heterogeneity into account.

Since we are interested in strategic models for long-term prognoses, in which we aim to compare scenarios or variants, it is common to adopt the concept of a Wardropian user-equilibrium (Wardrop, 1952). A user-equilibrium is a long term prediction of a stable travel situation, which enables comparing different scenarios. This means that the model should be able to generate such a user-equilibrium state and use pre-trip route choice with feedback. In contrast, short term models using en-route route choice without feedback and in which travellers respond for example to incidents, are assumed not to take this information into account the next time they travel, and therefore will not reach a user-equilibrium.

Finally, different vehicle types may have different infrastructure available. For example, trucks may not be allowed to drive on certain urban roads, while dedicated infrastructure may be available for them. Hence, the route choice set may be different across vehicle types, such that different vehicle types have to be considered explicitly.

2.1.2 Realistic traffic flow propagation

Given that all agents have chosen a route, the agents can be simulated on the network in order to assess the efficiency of the transport system in terms of flows, speeds, queues, and travel times.

For analysing where problems occur in the network, it is of utmost importance that bottleneck locations are identified accurately. These are locations where the travel demand exceeds the infrastructure supply, and will be the point from which queues will build upstream and cause congestion on the roads. These queues have a physical length and will spill back to upstream road segments when they exceed the road segment length. Typical locations are lane drops,

merges of motorways, and (non-)signalised intersections. The basic relationships between flow, speed, and density can be described with the fundamental diagram which, together with the conservation of vehicles law, describes stationary flow behaviour. This can predict most essential features of traffic flow, including wave formation and propagation. Other more empirical phenomena in traffic are the so-called capacity or speed drop (which occurs when traffic is near the critical density), hysteresis (which describes different acceleration and deceleration patterns), platoon diffusion, etc. In contrast to stationary flow behaviour, these phenomena cannot be described by basic fundamental traffic flow theory and require more sophisticated relationships (Zhang, 2001). We argue that for strategic transport models, it is important that the model describes stationary flow behaviour (first order effects), but does not necessarily need to be able to describe these additional traffic phenomena (second order effects).

For benefit cost analysis in which usually travel time savings are an important input, the predicted travel times need to be accurate. Such travel times can only be accurately calculated if the length of queues and speed within queues are predicted correctly, including queues that are spilling back over intersections. Even routes without any bottlenecks can be seriously affected by queue spillback, such that the travel times on many routes may increase.

Finally, traffic flow will consist of a mix of different vehicle types that may have very different driving characteristics. In particular, we need at least to distinguish passenger cars and trucks. Trucks impede cars more than vice versa, and trucks drive at lower speeds, resulting in longer travel times.

2.2 *Robustness of results*

Strategic transport models are often used to compare different scenarios or variants, therefore it is important that differences between scenarios can be attributed to the scenarios themselves, and not to unstable model results. Therefore, we require that the model is robust. A model is said to provide robust results if marginally different inputs only lead to marginally different outputs. For example, if a different random seed can lead to substantially different results, comparing scenarios will be problematic.

2.3 *Consistency of results*

Governments often apply several different models. For example, a static macroscopic model is applied for the whole city, a dynamic mesoscopic model is applied on the city centre, and a dynamic microscopic model is applied on a couple of roads and intersections. It is also common that different models are applied in different phases of the project. For example a static model for quick-scan and project appraisal and a microscopic model to investigate details during the implementation phase. Although these models are used for different purposes, it is not beneficial for the decision making process if they give conflicting results.

It is therefore important that the model results across the different models are as consistent as possible. Even though the level of detail may be different in each model, the main mechanisms should be similar or at least use the same underlying principles. Hence, mesoscopic models should be seen as an aggregation of microscopic models, and macroscopic models should be seen as an aggregation of mesoscopic models, such that the underlying principles of micro models transfer to meso and macro models. We will use the microscopic models as the basis, as these are widely used by governments as operational models, and compare other models in terms of consistency with such micro models.

2.4 *Accountability of results*

A model will be more accountable if the model properties are well understood and results can be explained and easily verified. Explainable results are very important in order to convince policy makers and the community.

The model should therefore not be a black box, but rather formulated as a rigorous mathematical problem, such that convergence towards a user-equilibrium can be guaranteed, and such that

existence and possibly uniqueness of solutions can be proved. Such a deeper understanding of the model should prevent unexpected results. Regarding model complexity, the model should be as complex as it needs to be in order to describe the most important transport aspects (realism of results), but not any more complex. For accountability reasons, often a less complex model is preferred over a highly complex model. Or stated differently; it is always better to have a model that is guaranteed to be fairly close to reality, instead of a model that is potentially very realistic, but this level of realism comes without any guarantees.

2.5   *Ease of use*

Last but not least, the model should be user friendly, such that it can timely provide results for decision making. After all, making a model is an iterative process of running the model, tuning parameters and correcting errors in the input.

This means that the model must have relatively short run times. A rule of thumb is that complete scenarios can run overnight, i.e. within 12 hours. Specifically when the traffic assignment is embedded in an iterative demand loop with multiple user classes and day parts, the time to run the assignment should be limited. With computational power increasing, the run times are becoming less of a problem each year.

Preferably a minimum of input data is required by the model to enable easy input and quick calibration. The infrastructure should be described by road segments and intersections. For a strategic model, it suffices to characterise the road segments by length, number of lanes, capacity, maximum speed, and possibly a speed at capacity (critical speed). These attributes are mostly easily obtainable, although the capacity is an important input that requires the most attention, as it will determine the bottleneck locations. Intersections will be defined by allowed turns, settings of traffic controls (i.e., green times for traffic lights). The physical layout of an intersection for a strategic model is often not needed, as long as the capacities in each direction can be properly calculated by the model. An extension would be to add priority rules to intersections in order to be able to compute any additional delays. Note that the supply model is essentially completely determined by the infrastructure, and therefore does not need to contain parameters that require calibration. A proper model should therefore not contain any additional parameters, as such parameters are merely present to correct imperfections of the model (for example, parameters of travel time functions are such parameters, as will be discussed in Section 3).

Since the supply model should not need any further parameters to calibrate, the calibration process comes down to calibrating the origin-destination matrix (or matrices), and any parameters in the route choice model. Often, vehicle counts on road segments are used to calibrate the matrix. Using such counts assumes that the model is at least strictly capacity constrained, i.e. that the flow on a road segment cannot exceed the capacity (consistent with reality). Further, in order to calibrate matrices, the model should be able to easily provide select link information (i.e., routes and origin-destination pairs that pass through a certain link). Route-based models can readily provide this information.

Regarding model outputs, in order to be able to quickly analyse and assess traffic assignment outcomes, the model should provide information on the bottleneck locations, queues, flows, speeds, densities and level of service matrices (i.e., skims with travel times, travel costs, travel distances, etc.).

3. Critical assessment of models and criteria

Given the desired properties and criteria established in Section 2, we will now critically assess the different models and determine which models are most suitable.

3.1 Realism of results

3.1.1 Realistic route choice behaviour

In order to take multiple route attributes into account, it is common to define a so-called generalised cost function, in which each attribute is converted into dollars. For example, if only travel time and cost is considered, then we can write:

$$c_{mnp} = \beta_{mn}\tau_{mp} + \theta_{mp}, \quad \forall p \in P_{mn}^{rs}, \tag{1}$$

where c_{mnp} is the generalised cost (or the systematic utility) of route p for user class n driving vehicle type m, τ_{mp} is the route travel time for vehicle type m, θ_{mp} is the travel cost (e.g., running costs and toll costs) on route p for vehicle type m, β_{mn} is the value-of-time for user class n driving vehicle type m, and P_{mn}^{rs} is the set of relevant routes for vehicle type m and user class n from origin r to destination s. Clearly, the preferences are heterogeneous for different user classes and vehicle types. Furthermore, the route sets can be vehicle type and user class specific.

Since route choice is a repetitive choice, habitual behaviour may exist. For repetitive choices, choice behaviour can be decomposed in an habitual part, and in a variety seeking (backup) part, which leads to a probabilistic choice model. Swait and Bliemer (2013) apply this methodology for mode choice. Swait and Marley (in press) have shown that the probability of choosing a certain alternative (in our case, a route) can be written as the well-known conditional logit model,

$$\pi_{mnp} = \frac{\exp\left(-\mu_{mn}^{rs}c_{mnp}\right)}{\sum_{p' \in P_{mn}^{rs}} \exp\left(-\mu_{mn}^{rs}c_{mnp'}\right)}, \quad \forall p \in P_m^{rs}, \forall(r,s), \forall m, \forall n, \tag{2}$$

with a positive scale parameter that reflects the level of habitual behaviour (i.e., if $\mu_{mn}^{rs} \to \infty$, then a driver of user class n in vehicle type m will due to habit always take the least cost route, while if $\mu_{mn}^{rs} = 0$, this driver is variety seeking and randomly selects a route). Note that we also added superscripts rs to the scale parameter, as the behaviour typically depends on the distance between origin r and destination s. We would like to point out that the generalised cost (or utility) function in Eqn. (1) can be extended with any additional terms, including socio-demographics of the driver n, and is therefore completely flexible.

In traditional static traffic assignment models, and also in several dynamic models, it is assumed that all travellers take the cheapest (in terms of generalised costs) route, which will lead to a so-called deterministic user-equilibrium (Wardrop, 1952). The route choice model in Eqn. (2) means that drivers do not always take the cheapest/fastest route. This is similar to the notion of a so-called stochastic user-equilibrium. Note that deterministic assignment is the limiting case of stochastic assignment in which $\mu_{mn}^{rs} = \infty$ (i.e., all travellers behave in a purely habitual fashion).

Using Eqn. (2) has an important consequence: routes will have to be explicitly generated. While models that search for a deterministic user-equilibrium often do not determine route choice sets, in a stochastic user-equilibrium based on the logit model this is a requirement. The number of

relevant routes will be very large in case of networks with many zones, but Bliemer and Taale (2006) have shown that it is feasible, which we will illustrate in Section 5 when presenting our case studies. Note that only relevant routes are needed, hence we can filter out many route alternatives that are unlikely to be chosen. Furthermore, Bar-Gera (2010) has developed an alternative way of using routes called paired alternative segments (PAS), which significantly reduces the amount of memory required and speeds up convergence.

Explicitly generating routes has more advantages than merely enabling more realistic choice behaviour. It also significantly speeds up convergence to a (stochastic) user-equilibrium. The reason is that traffic flows will be distributed over multiple routes from the first iteration on, and iterative stochastic route choice has a much smoother result than iterative all-or-nothing route choice. Another advantage is that the resulting flows in user-equilibria will be route proportional. Bar-Gera (2010) has shown the importance of this property, and it is particularly important when dealing with intersection delays.

A final remark we have to make is about route overlap. In logit based choice models, such as Eqn. (2), the implicit assumption is that all alternatives are disjoint. However, in practice many routes will be partially overlapping, which distorts the route choice probabilities. Several corrections have been proposed in the literature by using a route commonality factor (Cascetta et al., 1996) or a path size factor (Bierlaire and Ben-Akiva, 1999), which simply adds an overlap term to the route cost functions. We would advise using such an overlap term, although one has to be careful not to include any irrelevant routes in the route set, as this can lead to unexpected results (Bliemer and Bovy, 2008).

3.1.2   *Realistic traffic flow propagation*

Network loading of route flows to the network can be done statically or dynamically. Clearly, traffic is dynamic in nature, and therefore dynamic network loading models (macroscopic models or microscopic simulation models) are clearly superior over static models in terms of realism. Static models basically aim to predict average traffic conditions over a certain time period assuming stationary travel demand and instantaneous flow propagation. The assumption of instantaneous flow propagation is particularly convenient from a computational perspective, but it also assumes that a vehicle is on all parts of the network at the same time.

Besides the above simplifying assumptions, traditional static traffic assignment models are particularly weak in determining bottleneck locations and queue formation to derive proper travel times. Most of these models adopt the original model formulation of Beckmann et al. (1956) and compute the link travel times as a function of the link flow. Well-known travel time functions (more correctly called link performance functions or volume-delay functions) are the BPR (Bureau of Public Roads, 1964) function and the Akçelik (1991) function. Both functions are of the form:

$$\tau_a = \frac{L_a}{\gamma_a} + f_a\left(\frac{q_a}{C_a}\right), \tag{3}$$

where τ_a is the travel time on link a, L_a is the link length, γ_a is the maximum speed, q_a is the link volume (flow), and C_a is the capacity of the link. The first part of this term represents the free-flow travel time, and the second part represents the additional delay, where f_a denotes a certain increasing function of the volume/capacity-ratio with certain parameters that have to be calibrated (see for example the US Highway Capacity Manual). It is important to note that this ratio can be larger than 1, in other words, the link flow is not constrained to capacity. This means that high flows will merely lead to increased delay instead of vehicles queuing in front of bottlenecks. Hence, the travel times are not consistent with traffic flow theory. This means that traffic flows, predicted by a static model, will be rather meaningless and mainly too high, such that bottleneck locations will be wrong, and travel times will be incorrect. Due to these flaws,

that it is impossible to correctly calibrate such models to link counts and measured travel times. We do not believe one should draw too many conclusions based on the outcomes of such static models. It is therefore somewhat worrying that many assessments of large infrastructure projects are based on such model outcomes. There have been extensions of the formulation of Beckmann et al. that add capacity constraints (Larsson and Patriksson, 1999; Nesterov and De Palma, 2000), but these models result often in even more unrealistic traffic conditions by constraining the entire route flow to the smallest capacity on a route. Daganzo (1998) proposed to use a travel time function with an asymptote near the capacity, which aims to prevent the link flow from exceeding the capacity, but cannot guarantee this.

Early dynamic traffic assignment models, such as Janson (1991), were basically a direct extension of static models by introducing a time index in the travel time functions. In the last decade, it has become very clear that realistic dynamic models cannot rely on such travel time functions, but that traffic flow needs to be derived from traffic flow theory. De Romph (1994) therefore introduced the use of speed-density relationships instead of travel time functions. Travel time has to be considered an implicit result of the traffic flow propagation, not an explicit function of the flow (Bliemer, 2007). This insight has led to models where traffic is modelled consistent with fundamental diagrams of macroscopic traffic flow theory. In these models, flow can never exceed capacity, such that queues will build up. The simplest models assumed vertical queues without any physical length (as in the original bottleneck model introduced by Vickrey, 1969), but have recently been replaced by models with simple horizontal queues (e.g., Bliemer, 2007), and more advanced physical queues (Yperman, 2007; Gentile, 2010) in which the queue may move along a road segment depending on the shockwaves. The most widely accepted macroscopic theory is the traffic flow theory based on kinematic waves of Lighthill and Whittam (1955) and Richards (1956), which is able to explain most essential traffic phenomena. Other phenomena mentioned earlier, such as the empirically observed capacity drop and hysteresis, can only be reproduced by more advanced higher order models (Parzani and Buisson, 2012; Zhang, 1999). However, these higher order models are computationally much more complex and may exhibit inconsistencies (Daganzo, 1995). As argued earlier, first order models sufficiently reproduce most relevant traffic phenomena for strategic transport models, including queue formation and spillback. The simplified theory of kinematic waves of Newell (1993) presents a basic but powerful first order model. Therefore, we propose to adopt Newell's model instead of a second or higher order model.

During the development of dynamic models the focus was primarily on the development of link models. This is not surprising since the dynamics of traffic flow occur on roads. As well, static models do not constrain flow to capacity and at nodes no restrictions to flow are imposed. However, the nodes are the locations where queues originate and – moreover – the available supply is distributed over demand. The first used node models are very unrealistic, they block traffic that can pass through or cause alternations (i.e., flip-flops) in simulations. It was not until Tampère (2011) that the importance of proper node model was recognized. They formulate requirements for node models to represent first order phenomena at intersections. The node model determines the severity and direction of congestion and is therefore very important in network models.

Instead of using fundamental diagrams and macroscopic traffic flow theory, others have adopted microscopic traffic flow theory in which all vehicles are considered separately. These microscopic models include mostly car-following behaviour, gap acceptance, speed adaptation, ramp merging, lane-changing, and overtaking behaviour (Olstam and Tapani, 2004). These models require a high level of detail and, when this level of detail is provided, are able to mimic the behaviour of each vehicle, which results at a more aggregate level in macroscopic traffic flows. The aggregate behaviour of microscopic models is likely to be more or less similar to the fundamental diagrams in macroscopic traffic flow theory, but differences will exist.

Mesoscopic models are hybrids of macroscopic and microscopic models. They are based on macroscopic traffic flow theory, but propagate individual vehicles or packets of vehicles.

Mesoscopic models have gained popularity the last few years due to their reliance on robust macroscopic traffic flow theory while at the same time individual information (e.g., route, vehicle class) can be easily tracked.

In practice, the step from static to dynamic models is considerable. In order to fill the gap, so-called quasi-dynamic models have been proposed. They are basically static models that consider a single time period, but constrain the flows to capacity, such that bottlenecks appear and queues build up. Examples are the operational model QBLOK (4Cast, 2009), which has been used in the Dutch national and regional models for many years, and a model described by Bundschuh et al. (2006) which has been implemented in VISUM. Although these models are strict capacity constrained, they do not consider a realistic fundamental diagram nor a proper node model. Brederode et al. (2010) and Bliemer et al. (2011) derive a new quasi-dynamic model from a macroscopic dynamic model assuming stationary flow and instantaneous flow propagation. As such, their quasi-dynamic models inherit the most important properties from macroscopic traffic flow theory. Travel times are derived after the flow propagation using cumulative inflow and outflow curves.

Now consider different user classes and different vehicle types. First, we note that while it makes sense to distinguish different user classes in route choice, it is much less important to distinguish user classes in flow propagation. When there is congestion, the driver will just have to queue and wait, no matter what their socio-demographics are. Different vehicle types, however, do have an important impact. While microscopic models can naturally consider different vehicle types, in macroscopic models this is less obvious and usually requires some assumptions. The first assumption that is often made in macroscopic models, is that every vehicle type is converted into passenger car units (pcu), see e.g., Petigny (1967). For example, a large truck could have the same impact as 2.5 cars, and a small truck can be converted using a pcu value of 1.5 cars. This pcu value is determined by a combination of the space occupied by the value when standing still, and the impedance of the vehicle on other vehicles. We believe that adopting pcu's is a simple and workable way to include multiple vehicle types, even though it does not capture all the nuances that microscopic models could capture. Extra care is needed when calculating the travel times for vehicle types other than the car. In case of free-flow, the car speed will be significantly higher than the truck speed, hence the travel time for the truck needs to be scaled up appropriately. In case of heavy congestion, the car and truck will drive at approximately the same speed, such that the travel time of the truck is about the same as that of the car. Therefore, we propose for vehicle types such as trucks to use a scaling factor for travel time dependent on the traffic conditions. Another option is to simulate different modes on different layers of the network that can interact (for example via a common density). This causes additional problems that are beyond the scope of this paper, but does allow for direct simulation of different modes with different characteristics. The benefit of this approach is the absence of the need for post simulation adjustments to correct the results.

3.2   *Robustness of results*

Traditional static models have proven to be very robust, although it has shown to be important to set the convergence criterion (e.g., the duality gap) relatively small to guarantee proper convergence (Boyce et al., 2004). This property is mainly due to the fact that these models are not strict capacity constrained (i.e., capacity is a soft constraint that merely increases the travel time, not a hard constraint limiting the flow), resulting in smooth functions. In contrast, all strict capacity constrained models with proper queuing behaviour are more sensitive and gridlock of a network can occur in models with spillback. This increase in sensitivity to model inputs does not mean that such models are not robust. All macroscopic and mesoscopic models are usually quite robust against small perturbations due to their fundamentals in macroscopic traffic flow theory.

In contrast, microscopic models are based on car-following, lane-changing and other driving behaviour, and often include random components or probabilities. This means that running a

microscopic model with exactly the same input but only a different random seed may yield different results. This could mean the difference between congestion and no congestion occurring with large impacts on route choice. It is therefore recommended to perform multiple runs of a microsimulation model. Simply averaging the results of multiple runs is not possible, as this may lead to an infeasible result. While microsimulation models are suitable for short term prognoses, due to potential instability (see e.g., Sbayti and Roden, 2010), we believe that microscopic simulation models are typically not suitable for comparing scenarios in strategic transport planning. However, microsimulation models can be complementary in later stages of the planning process.

3.3   Consistency of results

As mentioned, we require that model results should be as much as possible consistent with outcomes from a dynamic microscopic model. Differences in results can be minimised by using the same underlying principles. As mentioned in Section 3.1.2, we propose to derive a quasi-dynamic model based on a dynamic model. By using a proper link model that is consistent with first order traffic flow theory and a proper first order node model, differences will be minimised.

3.4   Accountability of results

The static traffic assignment model has been proposed over 50 years ago and has not changed much since. The properties of this model are well understood. For example, there exists a unique solution to the original model formulation by Beckmann et al. (1956) for a single vehicle type in case the travel time function is strict monotonically increasing (like the BPR function). In the case of multiple vehicle types in which pcu values are used to convert all vehicles into passenger cars, the travel time function is no longer strictly monotone, hence the solution need not be unique. Many algorithms have been proposed to solve for the user-equilibrium solution, of which the Frank-Wolfe method (1956) is the most well-known. More recently origin based models have been proposed with better converging properties, such as Bar-Gera (2002), Florian et al. (2009), and Gentile and Noekel (2009).

In contrast, properties of dynamic models are less well understood. Different dynamic models will have different properties, but it has been shown that for some models existence of a dynamic user-equilibrium solution cannot be guaranteed (Szeto and Lo, 2006), and if it exists, it is not likely to be unique (Bliemer and Bovy, 2003). Models with horizontal queues and spillback in particular do not have elegant theoretical properties (Szeto and Lo, 2006). Models with vertical queues have nicer properties, but are clearly less realistic. Simulation based algorithms will typically find an approximate solution, but it is problematic to compare such approximate user-equilibria in different scenarios.

We will propose a model that is mathematically defined similar to a static traffic assignment model in which we proof existence of a user-equilibrium solution, and under certain assumptions also uniqueness. Further, we will propose a converging algorithm.

3.5   Ease of use

Due to their computational complexity and configuration hungry nature, microscopic dynamic models are, even when feasible, not desirable to use on large scale networks with thousands of traffic analysis zones (TAZ), links, and nodes, millions of routes and possibly millions of vehicles on the road at the same time. Although mesoscopic and macroscopic dynamic models can handle much larger networks, they still have a high computational complexity, require more detailed inputs, and require much more time for calibration compared to static models. Static models on the other hand are very fast and can handle very large networks, but generate very unrealistic traffic conditions. Quasi-dynamic models seem to sit comfortably between the two, which combine a low computational complexity with realistic flow propagation, which is the model type we will adopt.

The input into quasi-dynamic traffic assignment models is actually less than in static traffic assignment models that use travel time functions. Since the supply model is completely determined by the infrastructure, there is no need to include these functions and hence no need to estimate the corresponding parameters. Dynamic models require more input data in the form of time-specific travel demand.

Calibration of this time-dependent travel demand in dynamic models is not an easy task, as the time dimension adds significant complexity. Not only are there many more input variables that require calibration, each calibration run will require simulations and therefore long computation times. Static and quasi-dynamic models only consider a single origin-destination matrix for the whole time period, which makes calibration of the model easier more feasible task. Calibration towards link counts is essentially impossible in static traffic assignment models due to the fact that the flows on the network are not strict capacity constrained and as such are actually desired flows, not realised flows. Quasi-dynamic models that are strict capacity constrained yield flows that can directly be compared with the link counts. Furthermore, by adopting a route-based approach, there will be an immediate mapping from links to routes, which assists in select link analysis and matrix calibration.

Therefore, quasi-dynamic models are superior in ease of use, even compared to static traffic assignment models. They come with an additional bonus that they can show bottlenecks and queues, which are model outcomes that are easily understood.

4.   Proposed quasi-dynamic traffic assignment methodology

Analysis of the desired properties and how the different models score are summarised in Table 2. This table draws a clear picture. Static models are robust, reliable, and easy to use, but lack realism and consistency. In the other side, dynamic models are realistic and consistent, but are less easy to use, less reliable, and less robust. In between are the quasi-dynamic models, which are designed to be sufficiently realistic, robust, consistent, reliable, and easy to use. Therefore, quasi-dynamic traffic assignment models seem to be very suitable for strategic transport planning purposes in which realistic aggregate results are required without too many details.

Table 2: Models and scoring on desired properties

Property	static macro	quasi-dyn. macro	dynamic macro	dynamic meso	dynamic micro
Realistic results	--	+	++	++	++
- stochastic route-choice	-	+	+	+	+
- multiple vehicle types	+	+	+	++	++
- multiple user classes	+	+	+	++	++
- strict capacity constrained	--	++	++	++	++
- queue spillback	--	+	++	++	++
- realistic link model	--	+	++	++	++
- realistic node model	--	+	+	++	++
Robust results	++	++	+	+	-
- stable outcomes	++	++	+	+	-
Consistent results	--	+	++	++	++
- consistent with dyn. micro model	--	+	++	++	++
Accountable results	++	++	+	+	-
- convergence to equilibrium	++	+	+	+	-
- existence and uniqueness	++	++	-	-	--
- low model complexity	++	+	-	-	--
Ease of use	+	++	-	-	--
- short run times	++	+	-	-	--
- little input required	++	++	-	-	--
- easy of calibration	+	++	-	-	--

Literature on quasi-dynamic traffic assignment models is scarce. Bakker et al. (1994) were pioneers in this field, and developed QBLOK, which added capacity constraints, horizontal queues and spillback to static traffic assignment. However, this model is not route-based and uses split proportions to propagate traffic through the network, which leads to inconsistent results when blocking flow. Bundschuh et al. (2006) proposed a similar approach, which is implemented in VISUM. Both models suffer from the lack of a realistic underlying fundamental diagram and a proper node model. Furthermore, the queues build up inside bottlenecks, while realistic queues form upstream of bottlenecks. 4Cast (2009) makes several changes to the QBLOK model, among other queue formation upstream bottlenecks. Brederode et al. (2010) and Bliemer et al. (2011) developed a novel route-based approach derived from a macroscopic dynamic model (to be more specific, the link transmission model, see Yperman, 2007, which is based on the simplified theory of Newell, 1993), which includes a simplified but sufficiently realistic triangular fundamental diagram and a demand proportional node model, resulting in queues upstream bottlenecks. Furthermore, they formulated their quasi-dynamic model in a much more rigorous fashion, which allows inspection of the model properties (important for accountability of model results).

Tampère et al. (2011) investigated node models and came to the conclusion that most existing (demand proportional) node models are flawed and distribute capacities and flows in an unrealistic way. They formulated a set of conditions for proper node models and formulated a capacity proportional node model that satisfies all conditions. Realistic distribution of capacities to different turns at nodes is essential for any strict capacity constrained flow propagation model. In this paper we adopt the approach of Brederode et al. and Bliemer et al. and derive a quasi-dynamic model from the macroscopic dynamic link transmission model assuming Newell's triangular fundamental diagram (Newell, 1993) and the capacity proportional node model proposed by Tampère et al. (2011). Adding this node model is by no means a trivial exercise. The flow propagation model has to be solved in an entirely different way than proposed by Brederode et al. and Bliemer et al. in order to guarantee consistency with this node model.

Our proposed model consists of four components:

    (i)   Route generation submodel;
    (ii) Route choice submodel;
    (iii) Strict capacity constrained static network loading submodel;
    (iv) Dynamic physical queuing submodel;
    (v) Travel cost submodel.

The first component first finds all the relevant and likely routes between each origin-destination pair. The second component determines for each origin-destination pair the route flows depending on the generalised costs (utilities) of each route. The third component instantaneously propagates the route flows through the network, in which turn capacities are determined by a first order node model, and flows are strictly constrained to these capacities. This will yield vertical queues upstream bottlenecks. The fourth component dynamically applies a first order link model to convert these vertical queues into horizontal physical queues with realistic shockwaves and spillback. Finally, the fifth component calculates the link and route travel costs.

It is clear that this model is a hybrid between a static model and a dynamic model, hence the term quasi-dynamic. In the following subsections we will elaborate on each of these five components. We will assume that the total travel demand over a certain time period $[0,T]$ is given by vehicle type and user class specific origin-destination (r-s) trip matrices $D_{mn} \equiv [D_{mn}^{rs}]$. Further, we assume that the network is given by a directed graph $G = (N, A)$, where N is the set of nodes and A is the set of links. Each link $a \in A$ has an associated link length L_a (in km),

a capacity C_a (in pcu/T), a vehicle type specific maximum speed γ_{ma} (in km/hour), and a jam density K_a (in pcu/km).

4.1   *Route generation submodel*

Different methods exist for generating route choice sets. They can basically be split into stochastic methods and deterministic methods. Stochastic methods (e.g., Fiorenzo-Catalano et al., 2004) iteratively generate cheapest cost routes by randomising the link travel costs. Deterministic methods (e.g., Prato and Bekhor, 2006) find all routes that satisfy a set of given constraints. Both methods are able to generate relevant routes depending on the network. More routes will automatically be generated if many similar alternatives exist (e.g., in a grid network), while less routes will be generated when only few relevant alternatives exist (e.g., in a network with motorways). The notion of relevancy is important, as indicated by Bliemer and Bovy (2008). A route is irrelevant if it is unlikely that travellers will choose that route. Examples are routes with long detours, or routes that include off-ramp on-ramp behaviour. In deterministic methods, such irrelevant routes can be ruled out by setting specific constraints. We will adopt a combination of a stochastic and deterministic method by adopting a stochastic generation method followed by a deterministic route filtering method that excludes irrelevant routes.

It is important to note that while route set generation can be time-consuming, once we have generated such a route set, we can re-use this route set in subsequent model runs, thereby avoiding expensive shortest path computations while running the model. This significantly speeds up running the assignment model. However, by generating the routes in advance, we cannot guarantee that each relevant route is included in the route set in order to find a user equilibrium. Therefore, it may be wise to search for new routes at the end of the traffic assignment run based on the current travel costs and include any newly found routes in the route set.

The outputs of this submodel are route sets P_{mn}^{rs} for each origin-destination pair (r,s) for each vehicle type *m* and user class *n*.

4.2   *Route choice submodel*

In the proposed model, we aim to find a stochastic user equilibrium. In general, the vehicle type and user class specific route flows $\mathbf{f}^* \equiv [f_{mnp}^*]$ corresponding to a stochastic user equilibrium can be found by solving the following variational inequality (VI) problem (for more details, see Nagurney, 1993, for a derivation in the case of a single vehicle type and user class):

$$\sum_m \sum_n \sum_{(r,s)} \sum_{p \in P_{mn}^{rs}} \left[\left(f_{mnp} - \pi_{mnp} D_{mn}^{rs} \right) \frac{\partial c_{mnp}}{\partial f_{mnp}} \right] \left(f_{mnp} - f_{mnp}^* \right) \geq 0, \quad \forall \mathbf{f} \in \Omega \tag{4}$$

where π_{mnp} are the logit route choice probabilities defined in Eqn. (2), and Ω is the set of feasible route flows defined by the following flow conservation and non-negativity constraints:

$$\sum_{p \in P_{mn}^{rs}} f_{mnp} = D_{mn}^{rs}, \quad \forall (r,s), \forall m, \forall n, \tag{5}$$

$$f_{mnp} \geq 0, \quad \forall p \in P_m^{rs}, \forall (r,s), \forall m, \forall n. \tag{6}$$

Different iterative schemes could be used to solve this route-based stochastic user equilibrium problem (Nagurney, 1993). There are some simple strategies that could be adopted, such as the well-known method of successive averages (MSA, see e.g., Sheffi and Powell, 1982). Liu et al. (2009) propose some variations of MSA that may be more efficient.

It is important to note that the original optimisation problem formulation introduced by Beckmann et al. (1956) can no longer be used, as the resulting link performance functions are no longer separable (Dafermos and Sparrow, 1969).

The outputs of each iteration of this submodel are route flows f_{mnp} for each route $p \in P_{mn}^{rs}$ for each origin-destination pair (r,s) for each vehicle type m and user class n.

4.3 Strict capacity constrained static network loading submodel

The main difference between static models and quasi-dynamic models is the propagation of traffic flow and computation of the resulting route costs c_{mnp}. The first step in the flow propagation is the strict capacity constrained network loading, which is later followed by dynamic physical queuing. Our strict capacity constrained static network loading model will move all stationary traffic flow instantaneously through the network (consistent with static assumptions), in which traffic flows are capped at turn capacities, which are outcomes of the node model specified in Tampère et al. (2011). Our novel model formulation consists of the following set of equations:

$$q_{ap} = \sum_m \sum_n \left(\prod_{b \in p_a} \alpha_b \right) \rho_m f_{mnp}, \quad \forall a, \forall p \in P^{rs}, \forall (r,s), \tag{7}$$

$$\boldsymbol{\alpha} = \Psi(\mathbf{q} \mid \mathbf{C}), \tag{8}$$

where $\mathbf{q} \equiv [q_{ap}]$ denotes the vector of flows on link a following a certain path p, p_a denotes the set of links on route p from the origin up to the link previous to a, $\boldsymbol{\alpha} = [\alpha_a]$ is the vector of outflow reduction factors at the end of link a due to capacity restrictions, $\mathbf{C} \equiv [C_a]$ is the vector of link capacities, ρ_m is the pcu-value of vehicle type m, and $\Psi(\cdot)$ is a mapping from desired flows at each node to reduction factors, which is described by the node model specified in Tampère et al. (2011). The reduction factors range from 0 to 1, where 0 means that no flow leaves the link (complete halt), while a reduction factor equal to 1 means that all flow can leave the link and no queue forms. In traditional static traffic assignment models, $\alpha_a = 1$ for all links, i.e. no strict capacity constraints. In our case, the flow along a route will decrease each time it is constrained to capacity. Note that in Eqn. (7) all flows over user classes n can be summed. The flows over different vehicle types are also summed taking pcu-factors ρ_m into account.

When inspecting these two equations, it can be seen that the flows $\mathbf{q}$ depend on $\boldsymbol{\alpha}$ given path flows $\mathbf{f}$, while $\boldsymbol{\alpha}$ depends on flows $\mathbf{q}$ given capacities $\mathbf{C}$. Writing Eqn. (7) into the form $\mathbf{q} = \Gamma(\boldsymbol{\alpha} \mid \mathbf{C})$, where $\Gamma(\cdot)$ is the function that performs the strict capacity constrained network loading, we can rewrite Eqns. (7) and (8) into the following fixed point (FP) problem:

$$\mathbf{q} = \Gamma\big(\Psi(\mathbf{q} \mid \mathbf{C}) \mid \mathbf{f}\big), \tag{9}$$

The vector of reduction factors $\mathbf{q}^*$ that satisfies $\mathbf{q}^* = g(\mathbf{q}^* \mid \mathbf{f}, \mathbf{C})$, where $g = \Gamma \circ \Psi$ is the composite function, is called a fixed point solution. It can be shown that under some mild conditions, this FP solution exists and is unique. Function $g(\cdot)$ is a non-expansive mapping and under mild conditions, it is a contraction mapping such that this FP solution can be found by iteratively solving the strict capacity constrained network loading and the reduction factor (node) model. Faster accelerated iterative schemes have been proposed to solve fixed point problems, such as Polyak iterations (Polyak, 1990; Bottom and Chabini, 2001) or Anderson acceleration (Anderson, 1965; Walker and Ni, 2011).

We can further define the link flows q_a, which are merely a sum of path-specific flows q_{ap} that pass through link a, i.e.,

$$q_a = \sum_{(r,s)} \sum_p \delta_{ap} q_{ap}, \quad \forall a, \tag{10}$$

where δ_{ap} is an indicator that equals one if link a is on route p, and zero otherwise (also known as the assignment map).

The cumulative link inflow U_a and outflow V_a (both in pcu) over time period $[0,T]$ can be written as $U_a = q_a$ and $V_a = \alpha_a q_a$, respectively. The (vertical) queues (expressed in pcu) at time instant T can be computed as

$$Q_a = (1-\alpha_a)q_a. \tag{11}$$

The outputs of this submodel are therefore link flows q_a (in pcu) and vehicles waiting in the queue Q_a (in pcu) for each link $a \in A$.

4.4 Dynamic physical queuing submodel

While the strict capacity constrained model presented in the previous subsection is a major improvement over traditional static traffic assignment models and the resulting travel times will be much more realistic than the travel times generated by simple link travel time functions like the BPR function, the vertical queues are not very realistic and ignore the fact that bottlenecks have consequences for a much larger part of the network due to spillback effects. In order to overcome this problem, in this second submodel we will propagate the shockwaves realistically through the network consistent with the dynamic link transmission model (Yperman, 2007). Bliemer et al. (2011) have proposed an event-based procedure for queues and spillback of the queues. All shockwaves through the network are simulated for the whole simulation time period $[0,T]$. While this may sound computationally intensive, it is in fact a very fast and efficient procedure. The first reason is, that it is an event-based procedure in which only changes in the flow rates are of interest. Since in a static traffic assignment model the input is a stationary flow, the flow rates only change when forward or backward shockwaves hit the other end of the link. Hence, the number of events will be limited. Secondly, the dynamic model only has to compute in the local area around the bottlenecks, not the entire network. Details of this event-based procedure are beyond the scope of this paper. We refer to to Bliemer et al. (2011) for details. The only difference with Bliemer et al. is that we are using the capacity proportional node model proposed by Tampère et al. (2011) instead of the demand proportional node model as proposed in, e.g., Bliemer (2007).

It should be pointed out that this submodel is optional. Although the physical queues determined in this submodel are more realistic, resulting in more appropriate travel times, one could opt to only determine vertical queues (output of the previous submodel) for the first or all of the route choice iterations.

The outputs of this submodel are dynamic cumulative inflow and outflows $U_a(t)$ and $V_a(t)$, where t is a certain time instant in the entire simulation period.

4.5 Travel cost submodel

The route costs consist of travel time and additional costs, see Eqn. (1). The route travel time is the summation of travel times on links along the route,

$$\tau_{mp} = \sum_a \delta_{ap} \tau_{ma}, \quad \forall p \in P_{mn}^{rs}, \forall (r,s), \forall m, \forall n. \tag{12}$$

The link travel times are determined differently, depending on whether one runs the optional dynamic physical queuing submodel or not.

First, suppose that one does not run the dynamic submodel. The queue at $t = 0$ is zero, and the queue at $t = T$ is Q_a, such that the average queue length is $\frac{1}{2}Q_a$. The outflow rate for each link a is given by $\alpha_a q_a / T$, such that the delay as a result of waiting in the queue is given by $\frac{1}{2}Q_a /(\alpha_a q_a / T) = (1 - \alpha_a) q_a /(\alpha_a q_a / T) = T(1 - \alpha_a)/(2\alpha_a)$. Hence, the link travel time is given by

$$\tau_{ma} = \frac{L_a}{\gamma_{ma}} + \frac{1 - \alpha_a}{2\alpha_a}T. \tag{13}$$

Clearly, the free-flow travel times L_a / γ_{ma} are vehicle type specific, but the average delay $T(1 - \alpha_a)/(2\alpha_a)$ is vehicle type independent. If the link flow does not exceed capacity, then $\alpha_a = 1$, and the resulting travel times are free-flow. The more the outflow is reduced due to capacity constraints, the longer the delay.

Now, suppose that we do a run with the dynamic submodel. Then we can use the dynamic cumulative inflow and outflows $U_a(t)$ and $V_a(t)$ to determine the link travel time, see e.g., Carey (2004). Then the average travel time over time period $[0, T]$ on link a for vehicle type m is given by:

$$\tau_{ma} = \frac{\xi_m}{q_a} \int_{t=0}^{T} \left(V_a^{-1}(U_a(t)) - t \right) dt, \tag{14}$$

where ξ_m is a correction factor depending on vehicle type m. If m is a passenger car, $\xi_m = 1$, while if m is a truck, then $\xi_m \geq 1$ due to possible lower maximum speed γ_{ma}.

The outputs of this submodel are route travel costs c_{mnp} that can be used in the next iteration of the route choice submodel.

5.   Case studies

In order to test our proposed quasi-dynamic traffic assignment model, we have implemented submodels (iii)-(v) in the StreamLine modelling framework in the OmniTRANS software package. Streamline is a modular framework, initially designed to run dynamic traffic assignment models, see Raadsen et al. (2010). StreamLine included implementations for submodels (i) and (ii), can calculate turn capacities for junctions, and has graphical capabilities. This enabled us to implement our model in a relatively short amount of time.

For our initial feasibility tests presented in this paper, we make a series of simplifying assumptions. While these assumptions may result in shorter computation times, they are by no means too restrictive. Relaxing these assumptions is possible without great sacrifices. The following simplifying assumptions are made. We consider a single vehicle type and a single user class. We adopt Polyak iterative averaging for solving the fixed point problem in submodel (iii) and a simple MSA scheme for solving for a stochastic user equilibrium solution in submodel (iv). We assume that the travel demand in all test cases considers a peak period of 2 hours. In order to speed up computations for Eqn. (13), instead of calculating the average travel time we compute the median travel time for a vehicle leaving at $t = \frac{1}{2}T = 1$ hour. The paired combinatorial logit model (Prashker and Bekhor, 1998; Gliebe et al., 1999) is available in Streamline, which is able to handle the route overlap earlier discussed. However, we have adopted the simpler conditional logit model as stated in Eqn. (2) for reasons of computational

speed. We further assume that the travel costs only consist of travel times. Finally, we have not used green times at signalised intersections to provide more detailed information on turn capacities, although StreamLine and the framework we have developed can handle it.

We point out that we have not calibrated the models in any way, and did not compare the outcomes of the models to real measured link counts travel times. Therefore, the purpose of these initial case studies are to solely to show the computational feasibility of our quasi-dynamic approach on large networks. We leave the validation of our results for further research. We again note that, since our model is derived from a first order dynamic model, results will be much closer to outcomes of a dynamic model than to the results of a traditional static model.

Figure 1 shows the four road networks that we will use for the case studies. The first two networks (Amsterdam and Rotterdam) are from the Netherlands, and the next two networks (Gold Coast and Sydney) are from Australia. Table 3 summarises the dimensions of the case studies. The networks vary in size, and the number of traffic analysis zones (TAZ) varies as well. Clearly, the number of TAZs has clearly a significant impact on the number of routes generated by submodel (i). The CPU time per iteration is the summation of computation time for submodels (ii) through (v). We have generated route choice sets prior to running the models and did not update the route choice sets at the end of the traffic assignment runs. This means that the stochastic user equilibria that we converge towards are conditional on the route choice sets generated at the beginning.

Table 3: Network data, computation time, and memory use

Network	Number of TAZs	Number of links	Number of nodes	Number of routes	Number of OD pairs	Number of vehicles	CPU time per iteration[3]
Amsterdam[1]	418	9,408	4,281	266,505	275,722	135,886	9 sec.
Rotterdam[1]	1,744	17,187	6,422	1,394,853	737,415	130,162	58 sec.
Gold Coast[2]	1,067	9,565	2,987	1,221,524	592,856	121,919	78 sec.
Sydney[2]	3,264	75,379	30,573	2,394,496	1,045,156	784,849	227 sec.

[1] Network and OD matrix kindly provided by Goudappel Coffeng BV, The Netherlands

[2] Network and OD matrix kindly provided by Veitch Lister Consulting Pty Ltd, Australia

[3] Using a notebook computer with Intel Core i7 @ 2.80Ghz running Windows 7

The CPU times reported in Table 3 are per route choice iteration, including calculating the route choice proportions for over 1 million origin-destination pairs, solving the fixed point problem in the strictly capacity constrained traffic assignment submodel (69 fixed point iterations were required in the first iteration), and performing the event-based dynamic physical queuing model (1,124,381 events were generated in the first iteration). As a result of the first iteration, 1,333 nodes were blocked, yielding 1,799,407 blocked routes. The maximum number of blocked turns on a single route is 152. In total 9 per cent of all links were in a congested state. All these computations for the first iteration were done within 4 minutes (on a single core) and required 3.5GB of RAM. Note that these results are based on our initial implementation. We can like achieve efficiency gains that will bring down the required CPU time and memory usage.

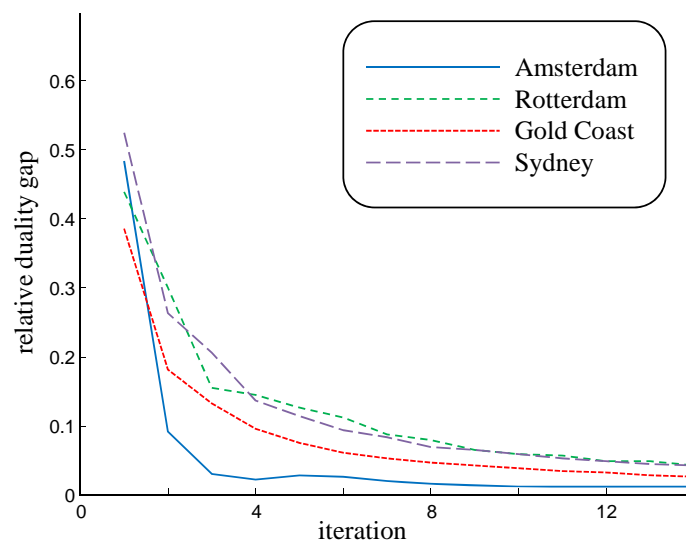*Figure 1: Road networks for the different case studies*

Figure 2: Convergence towards a stochastic user equilibrium solution

Figure 2 shows convergence over multiple route choice iterations in terms of the relative duality gap. For a deterministic user equilibrium, this gap will go to zero. However, for a stochastic user equilibrium, it will stabilise at a certain value larger than zero. As can be observed from the figure, for all case studies this gap seems to stabilise quite quickly, which is typical for models with stochastic route choice, as they are able to distribute the route flows over the network quite quickly.

To illustrate the difference between the queues of submodel (iii) and submodel (iv), Figure 3(a) shows the bottleneck locations and the vertical queues after running the strict capacity constrained traffic assignment submodel, while Figure 3(b) shows the horizontal queues after running the dynamic physical queuing submodel. Figure 3(a) is very useful to get an insight into the bottleneck locations, whereas Figure 3(b) shows how the queues spill back upstream. Both provide very powerful visualisations that are easy to interpret.
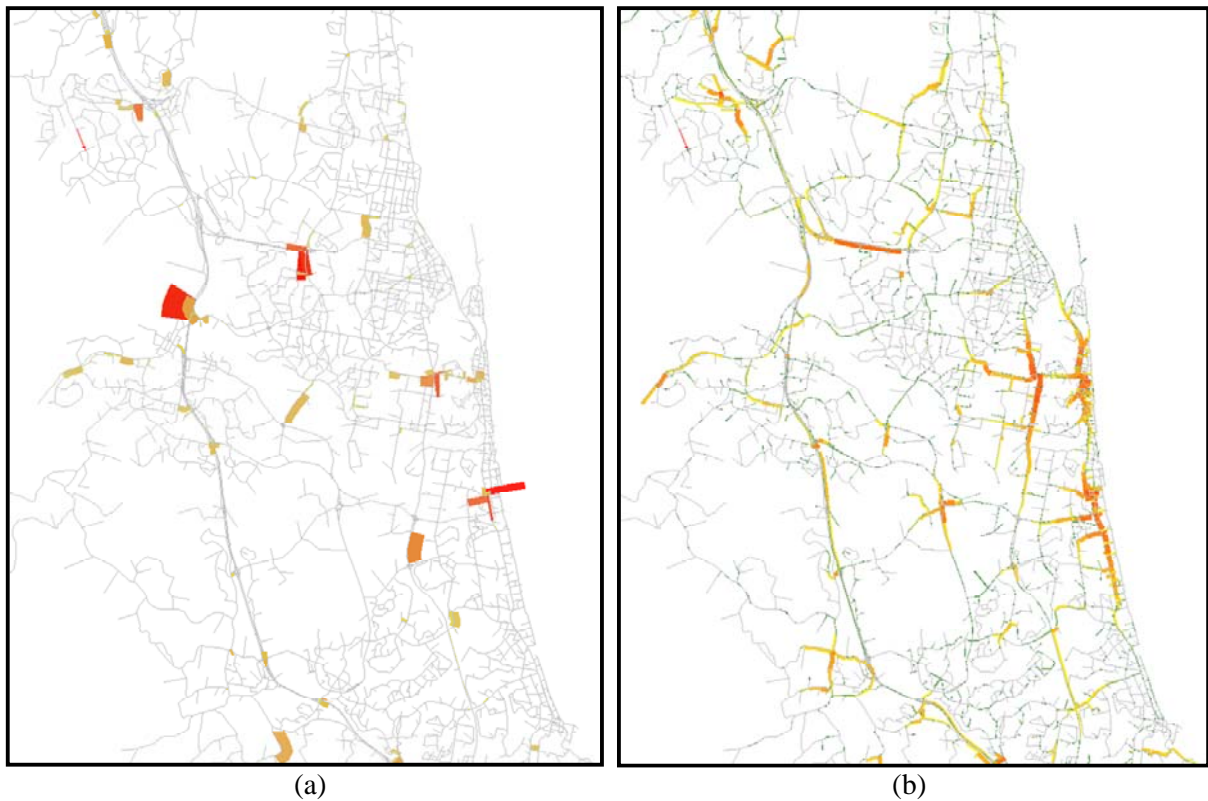

(a)   (b)

Figure 3: Gold Coast case study results: (a) vertical queues, and (b) horizontal physical queues

6.   Conclusions and discussion

In this paper we have considered traffic assignment models for strategic planning purposes. In a critical assessment based on realistic outcomes, robust results, consistent results, accountable results, and ease of use, we argue that traditional static models yield unrealistic results and outcomes that are inconsistent with dynamic microscopic models. At the other hand, dynamic models are less robust, less accountable, and not easy to use. Therefore, for the purpose of long term strategic transport planning, in which we would like to compare scenarios and run models on large scale networks, we propose to use a hybrid approach which is sometimes termed quasi-dynamic. This approach maintains the realism of dynamic models, but adopts the more rigorous mathematical foundations and computational efficiency of static models.

Our hybrid modelling framework is basically a special case of a first order dynamic model with some 'static' assumptions. It includes a proper node model to determine turn capacities, and a strict capacity constrained network loading model, such that link flows do not exceed the

capacity. Furthermore, physical queues, shockwaves, and spillback can be achieved by adopting an event-based dynamic physical queuing model.

We believe that the model approach formulated in this paper is a major step forward in traffic assignment for strategic planning. We have shown that the computational complexity of these models is feasible for large scale networks.

Although the formulated model is the result of about two years of research, there are still several research steps to be made. First of all, we are currently investigating all the mathematical properties of the model and providing rigorous proofs, such that the exact conditions for existence, uniqueness and convergence are known. Second, while we have already implemented more detailed junction models for signalised and non-signalised intersections, they still need to be tested. Third, the simple iterative schemes that we have adopted are by no means optimal; therefore we expect to increase the computational efficiency of our algorithms by smarter iterative schemes. Fourth, we so far have considered only Newell's triangular fundamental diagram in the original link transmission model. Gentile (2010) generalised the link transmission model to include any fundamental diagram. We have already developed a new event-based dynamic physical queuing model that assumes a quadratic shape of the fundamental diagram in the free-flowing part, in contrast to a linear shape. This has the advantage that the free-flow speed gradually reduces from the maximum speed to the critical speed, instead of always assuming the maximum speed in free-flow conditions. This new event-based model, however, still requires more testing, but it seems that the event-based algorithm can be extended to a fundamental diagram of any shape. Finally, and perhaps more importantly, we want to compare the outcomes of our new model to actual traffic data and thereby empirically validate our approach. In the coming year(s) we will address the above mentioned points.

We hope that with this research, we are able to move away from the current strategic (mostly static) traffic assignment models that rely on theories developed in the 1950s, and enter a new era of traffic assignment models.

References

4Cast (2009) *Qblok_2004: toedelingsprocedure LMS/NRM* (in Dutch), working report, the Netherlands.

Akçelik, R. (1991) Travel time functions for transport planning purposes: Davidson's function, its time-dependent form and an alternative travel time function. *Australian Road Research*, Vol. 21, No. 3, pp. 49–59.

Anderson, D.G. (1965) Iterative procedures for nonlinear integral equations. *Journal of the Association for Computing Machinery*, Vol. 12, pp. 547–560.

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'. *Presented at Colloguium Vervoersplanologisch Speurwerk*, The Netherlands, pp. 313-332.

Bar-Gera H. (2002) Origin-based algorithm for the transportation assignment problem. *Transportation Science* 36, pp. 398-417.

Bar-Gera, H. (2010) Traffic assignment by paired alternative segments. *Transportation Research Part B*, Vol. 44, No. 8/9, pp. 1022-1046.

Beckmann, M.J., C.B. McGuire, and C.B. Winsten (1956) *Studies in economics of transportation*, Yale University Press, New Haven CT, USA.

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.

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 (2003) Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem. *Transportation Research Part B*, Vol. 37, pp. 501-519.

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., L. Brederode, L. Wismans, and E.-S. Smits (2011) Quasi-dynamic traffic assignment: static traffic assignment with queuing and spillback. *Presented at the Annual Meeting of the Transportation Research Board*, Washington DC, USA.

Bliemer, M.C.J. and H. Taale (2006) Route Generation and Dynamic Traffic Assignment for Large Networks. *Proceedings of the 1ˢᵗ International Symposium on Dynamic Traffic Assignment*, Leeds, UK.

Bottom, J. and I. Chabini (2001) Accelerated averaging methods for fixed point problems in transportation analysis and planning. *Proceedings of the Triennial Symposium on Transportation Analysis (TRISTAN IV)*, São Miguel, Azores, pp. 69–74.

Boyce, D., B. Ralevic-Dekic, and H. Bar-Gera (2004) Convergence of traffic assignments: how much is enough? Technical Report Number 155, National Institute of Statistical Sciences, USA.

Brederode, L., M. Bliemer, and L. Wismans (2010) Static traffic assignment with queuing (STAQ): Static assignment revisited. *Presented at the European Transport Conference*, Glasgow, UK.

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

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

Carey, M. (2004) Link Travel Times I: Desirable Properties. *Networks and Spatial Economics*, Vol. 4, pp. 257-268.

Cascetta, E., A. Nuzzolo, F. Russo, and A. Vitetta (1996) A Modified Logit Route Choice Model Overcoming Path Overlapping Problems: Specification and Some Calibration Results for Interurban Networks. In: J.B. Lesort (ed.) *Proceedings of the International Symposium on Transportation and Traffic Theory*, Lyon, pp. 697-711.

Dafermos, S.C. and F.T. Sparrow (1969) The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards*, Vol. 73B, pp. 91-118.

Daganzo, C.F. (1995) Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B*, Vol. 29, pp. 277-286.

Daganzo, C. (1998) Queue spillovers in transportation networks with a route choice. *Transportation Science*, Vol. 32, No. 1, pp. 3-11.

De Romph, E. (1994) *A dynamic traffic assignment model: theory and applications*. PhD Thesis, Delft University of Technology, the Netherlands.

Fiorenzo-Catalano, S., R. van Nes, and P.H.L. Bovy (2004) Choice set generation for multi-modal travel analysis. *European Journal of Transport and Infrastructure Research*, Vol. 4, No. 2, pp. 195-209.

Florian M.A, I. Constantin, and D. Florian (2009) A new look at projected gradient method for equilibrium assignment. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2090, pp. 10-16.

Frank, M. and P. Wolfe (1956) An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, Vol. 3, pp. 95-110.

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.

Gentile G., and K. Noekel (2009) Linear User Cost Equilibrium: the new algorithm for traffic assignment in VISUM. *Proceedings of European Transport Conference 2009*, The Netherlands.

Gliebe, J.P., F.S. Koppelman and A. Ziliaskopoulos (1999) Route choice using paired combinatorial logit model. *Presented at 78th Annual Meeting of the Transportation Research Board*, Washington, DC.

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

Hellinga, B., H. van Lint, and F. Hofman (2012) Use of exogenously defined standard deviation versus mean travel time relationships for estimating the impact of policy measures on reliability. *Presented at the Annual Meeting of the Transportation Research Board*, Washington, DC, USA.

Larsson, T. and M. Patriksson (1999) Side constrained traffic equilibrium models – analysis, computation and applications. *Transportation Research Part B*, Vol. 33, pp. 233-264.

Lighthill, M.J., and G.B. Whittam (1955) On kinematic waves: II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society*, London, Ser. A 229(1178), pp. 317-345.

Liu, H.X., X. He and B. He (2009) Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. *Networks and Spatial Economics*, Vol. 9, pp. 485-503.

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

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.

Olstam, J.J. and A. Tapani (2004) Comparison of car following models. VTI Meddelande 960A, Swedish National Road and Transport Research Institute, Sweden.

Ortúzar, J.de.D. and L.G. Willumsen (2011) *Modelling transport*. Wiley, 4th edition.

Parzani, C., and C. Buisson (2012) Second-order model and capacity drop at merge. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2315, Washington DC, USA, pp. 25-34.

Petigny, B. (1967) Reflexions sur l'equivalent d'un vehicule en unites de vehicules particuliers (U.V.P.) (in French). *Transportation Research Part A*, Vol. 1, No. 4, pp. 339-348.

Polyak, B.T. (1990) New method of stochastic approximation type. *Automation and Remote Control*, Vol. 51, No. 7, pp. 937–946.

Prashker, J.N. and S. Bekhor (1998) Investigation of stochastic network loading procedures. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1645, Washington, DC, pp. 94–102.

Prato, C.G., and S. Bekhor (2006) Applying branch-and-bound technique to route choice set generation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1985, Washington DC, USA, pp. 19-28.

Raadsen, M.P.H., H.E. Mein, M.P. Schilpzand, and F. Brandt (2010) Implementation of a single dynamic traffic assignment model on mixed urban and highway transport networks including junction modelling. *Proceedings of the 3rd International Conference on Dynamic Traffic Assignment*, Takayama, Japan.

Richards, P.I. (1956) Shock waves on the highway. *Operations Research*. Vol. 4, 42-51.

Sbayti, H. and D. Roden (2010) Best practices in the use of micro simulation models. AECOM report prepared for the American Association of State Highway and Transportation Officials (AASHTO), Standing Committee on Planning, USA.

Sheffi, Y. and W.B. Powell (1982) An algorithm for the equilibrium assignment problem with random link times. *Networks*, Vol. 12, No. 2, pp. 191–207.

Swait, J., and M.C.J. Bliemer (2013) A strategic perspective on work mode choice behaviour: The concept of Commuting Plans. *Presented at the International Choice Modelling Conference*, Sydney, Australia.

Swait, J. and A.A.J. Marley (in press) Probabilistic Choice (Models) as a Result of Balancing Multiple Goals, *Journal of Mathematical Psychology*, forthcoming.

Szeto, W.Y., and H.K. Lo. (2006) Dynamic traffic assignment: Properties and extensions. *Transportmetrica*, Vol. 2, No. 1, pp. 31-52.

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.*

Vickrey, W.S. (1969) Congestion theory and transport investment *American Economic Review*, Vol. 59, No. 2, pp. 251–261.

Walker, H.F. and P. Ni (2011) Anderson acceleration for fixed-point iterations. *SIAM Journal of Numerical Analysis*, Vol. 49, No. 4, pp. 1715-1735.

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

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

Zhang, H.M. (1999) A mathematical theory of traffic hysteresis. *Transportation Research Part B*, Vol. 33, pp. 1-23.

Zhang, H.M. (2001) New perspectives on continuum traffic flow models. *Networks and Spatial Economics*, Vol. 1, pp. 9-33.