Congestion Effects in Transport Modelling and Forecasting

Jens West

Licentiate Thesis in Transport Science 2015
KTH Royal Institute of Technology
Stockholm, Sweden
Akademisk avhandling som med tillstånd av KTH i Stockholm framläggs till offentlig granskning för avläggande av teknisk licentiatexamen måndagen den 8 juni kl. 13:00 i sal Nash/Wardrop, KTH, Teknikringen 10, Stockholm.
Acknowledgements

I would like to thank my supervisors Maria Börjesson, Jonas Eliasson and Oded Cats and my mentors and colleagues Leonid Engelsson, Ida Kristoffersson, Daniel Jonsson, Svante Berglund, Karin Brundell-Freij and Gunnar Flötteröd for their contributions to this thesis and Trafikverket, Vinnova and SLL for financial support.
Abstract

Transport investments and policies are increasingly turned towards dealing with transport congestion rather than with shortening the potential free flow travel time. However, appraisal methodologies for projects meant to reduce congestion are relatively less well developed compared to methodologies for projects aiming to reduce travel times. Static assignment models are for instance incapable of predicting the build-up and dissipation of traffic queues and capturing the experienced crowding caused by uneven on-board passenger loads. Despite of the availability of dynamic traffic assignment and despite of fairly concrete ideas of how integration with demand models could take place, only few model systems have been developed for real applications.

The predicted reduction of traffic volume across the Gothenburg congestion charge cordon in the peak, 11%, turned out to be an accurate estimate of the observed reduction, 12%. The reduction in the off-peak, however, was overpredicted, as it was also in the Stockholm case. To analyse congestion charges in Stockholm it is necessary and fully possible to integrate DTA with the demand model. In the performed tests it could be seen that both tested models had problems replicating the flow on the main bypass early in the morning but otherwise performed well. A case study of a metro extension in Stockholm demonstrated that congestion effects constitute more than half of the total benefits and that these effects are excessively underestimated by a conventional static model. Effects of various operational measures can be analysed with BusMezzo and the results have been validated against observed data. The findings indicate that all three tested measures in a case study (boarding through all doors, headway-based holding and bus lanes) had an overall positive impact on service performance and that there are synergetic effects.
Using a continuous VTT distribution and hierarchal route choice was demonstrated as a successful method of modelling the multi-passage rule implemented in Gothenburg congestion charges and was shown to give realistic predictions of route choice effects. First results from integration of DTA with a travel demand model for the Stockholm region show that even without systematic calibration the DTA is in reasonable agreement with observed traffic counts and travel times. The presented experiments did not reveal a striking difference between using a macroscopic and a microscopic assignment package. While travel time savings are often the only benefit included in public transport project appraisals, the best practice assigns weighted value of time to average load/capacity measures. However, failure to represent dynamic congestion effects may lead to substantial underestimation of the benefits of projects primarily designed to increase capacity rather than reduce travel times. The impact of small operational measures should not be underestimated. These measures are relatively cheap compared to investments in new transit infrastructure and large societal gains can therefore be achieved by their implementation.
Sammanfattning

Investeringar och åtgärdsprogram inom transportområdet handlar allt oftare om att råda bot på trängsel snarare än att minska potentiella friflödesrestider. Utvärderingsmetoder för projekt som syftar till att minska trängsel är dock mindre utvecklade än metoder för projekt som syftar till att minska restider. Statiska nätutläggningsmodeller är till exempel inte lämpade för att prognosera uppbyggnad och avveckling av körer eller för att fånga upplevd trängsel orsakad av ojämn belastning på kollektivtrafikfordon. Trots förekomsten av dynamisk trafikutläggning och trots tämligen konkret kunskap om hur integration med efterfrågemodeller kunde gå till, finns det endast ett fåtal modellsystem utvecklade för verklig tillämpning.

Den predikterade trafikflödesminskningen på 11 % över trängselskattesnittet i Göteborg visade sig stämma väl överens med den observerade minskningen på 12 %. Minskningen under lågtrafik visade sig dock vara överskattad, precis som i Stockholmsfallet. För att analysera trängseleffekter i Stockholm är det nödvändigt och fullt möjligt att integrera en DTA-modell med efterfrågemodellen. I de utförda testerna observerades att båda de testade modellerna hade problem med att återskapa korrekta flöden på Essingeleden under morgonens maxtimme, men i övrigt gav god överensstämmelse. I en fallstudie av blå tunnelbanelinjes förlängning i Stockholm visade det sig att trängseleffekter utgjorde mer än hälften av de totala nyttorna och att dessa effekter är kraftigt underskattade i en konventionell statisk modell. Effekter av olika operationella åtgärder kan analyseras med BusMezzo och resultaten har blivit validerade mot observerade data. Resultaten tyder på att alla tre testade åtgärder i en fallstudie (ombordstigning i alla dörrar, turtäthetsbaserad reglering och busskörfält) hade en positiv inverkan på servicenivån och att det förekommer synergieffekter dem emellan.
**Introduction**

For a long time, the growth of cities went hand in hand with the invention of faster modes of transport which gradually made longer commuting distances possible. However, in contemporary cities, vehicle top speed is rarely the delimiter of city growth. As cities grow and sprawl, traffic congestion effectively and inevitably lowers the actual travel speed and comfort. Hence in growing cities of today, transport investments and policies are turned towards dealing with transport congestion rather than with shortening the potential free flow travel time. However, appraisal methodologies for projects meant to reduce congestion are relatively less well developed compared to methodologies for projects aiming to reduce travel times.

Integrated four-stage travel demand models and static equilibrium assignment procedures with level-of-service feedback have been used for urban and regional forecasts for many years (Vovsha et al. 2004, Wegener 2004). The drawbacks of this approach – in particular misspecification of congestion phenomena (Verhoef 1999) and coarse aggregation of travel demand (Ben-Akiva and Bowman 1998) – have been long recognized as hindrances in the investigation and evaluation of congestion mitigating measures such as road charges and public transport capacity expansions.

Take, for instance, the analysis of a time-dependent congestion charging scheme. Peak hour and off-peak traffic both offer their own difficulties to the modeller, where peak hour modelling is focused on sufficient capturing of network congestion effects while the difficulties in off-peak modelling mainly lie in predicting demand for trips other than work trips. A static assignment package is inadequate for predicting the build-up and dissipation of queues around the to-be-established toll zone during the rush hour (Engelson and van Amelsfort 2011). Despite of the availability of dynamic traffic assignment (DTA) and despite of fairly concrete ideas of how integration with demand models...
could take place (Flötteröd and Bierlaire 2012, Nagel and Flötteröd 2012), only few model systems have been developed for real applications.

The counterpart to car congestion in public transport is a whole array of weak system constraints, ranging from a lack of available seats on-board the bus to track capacity problems. For simplification, capacity constraints and discomfort factors are often ignored in public transport analysis. Travel time metrics obtained from static transport assignment models – which is commonly used as input for cost-benefit analysis to evaluate long-term planning decisions - is expressed in terms of average performance measures (Prud’homme et al. 2012, Kroes et al. 2013, Pel et al. 2014). Static assignment models are thus incapable of capturing the impact of capacity increases on the distribution of passengers over different vehicles and hence their implications on experienced crowding.

In the context of general traffic operations, simulation models have evolved as the primary tool for evaluation at the operational level. Transit simulations may serve several interests (Meignan et al. 2007): observation of network dynamics and design; evaluation and control of dynamic processes, and; evaluation of network performance under alternative designs. Transit simulation models may therefore be instrumental in testing the implications of various operational measures prior to their implementation. Most of the previous transit simulation studies were conducted by adjusting traffic simulation models that do not represent transit operations or enhancing existing simulation models by extending their capabilities for specific applications (Abdelghany et al. 2006, Ding et al. 2001, Chang et al. 2003, Cortes et al. 2005). This prohibits the analysis of measures that may have effects that extend beyond a single segment and may even influence other lines.
Theoretical models

Value of time
In order to prioritize among candidate transport investment projects through social cost-benefit analyses (which normally include factors such as investment cost and effects on emissions, noise and accident rates, but where travel time savings are often the largest benefit), it is important to be able to translate travel time to money (Oort 1969). This is complicated by the fact that travel time is not directly sold or bought on any market.

The value of travel time (VTT) is essential in both prediction and evaluation. VTT, which in public transport modelling is tightly connected to the value of comfort and crowding, is a key determinant in destination choice, route choice and the choice whether to board a full bus or wait for the next one (Li and Hensher 2011). The VTT applied in traffic route choice is crucial for prediction of the effect of congestion pricing, including revenues, but has still received surprisingly little attention in the literature, possibly due to the few congestion pricing system implemented around the world.

In reality, VTT varies both with socioeconomic factors and trip-related characteristics (e.g., trip purpose and trip comfort). The VTT is the sum of the opportunity cost of time and the direct (marginal) disutility of spending time in a particular situation relative to some reference situation. The opportunity cost of time will be equal in all situations for a given individual, assuming that individuals allocate their time optimally across potential activities. The direct disutility, however, will be different in different situations. It depends on characteristics of the specific situation such as comfort and possibilities to use time productively. Crowding will hence increase the direct disutility of time spent in vehicles and potentially on platforms, and the direct disutility of spending time waiting on a platform will usually be higher than spending time in the vehicle.
A common method to estimate VTT is by applying a logit model on stated preference (SP) survey data (Börjesson and Eliasson 2014). Recent advances in estimation techniques (e.g., Fosgerau 2007) have enhanced the possibilities to capture the full distribution of VTT. However, there are several difficulties in translating the hypothetical choices in an SP study to reasonable VTT.

Assigning different VTT to different traveller groups and transport modes is certainly not uncontroversial, even if the purpose would be merely to make a better representation of reality. Many authors (e.g., Mackie et al. 2001) argue that the effect of income differences on VTT should be removed. As different transport modes attract different socio-economic groups, using mode specific VTT also needs to be carefully motivated. However, according to Börjesson and Eliasson (2014), the impact of income on VTT is small compared to other variables.

**Congestion and crowding in public transport**

In research on car traffic, the term congestion is rather well defined. According to Wikipedia, “traffic congestion is a condition on road networks that occurs as use increases, and is characterized by slower speeds, longer trip times, and increased vehicular queuing.” The concept of car congestion has been studied extensively and is possible to express in rather simple equations (e.g., volume-delay functions, speed-density curves) or in agent based simulation models.

In research and debate on public transport however, there are several layers of crowding and congestion related problems that can be brought up;

1. Crowding in the vehicles increases the value of time of passengers and hence their generalized travel cost, i.e., it is less convenient to travel when the vehicle is full and the passengers might need to stand.
2. If a vehicle gets full, some passengers will be denied boarding and have to wait for the next vehicle. This effectively increases waiting times. Waiting due to denied boarding imposes a higher disutility per minute for passengers than normal waiting times since they are unpredictable, and moreover cannot be partly spent at home (or similar) as normal waiting times can.

3. Boarding takes time, so high demand tends to cause “bunching” of vehicles when long boarding times cause some vehicles to fall behind their schedule and other vehicles to catch up. This is a vicious circle: when a vehicle falls behind schedule, the number of passengers waiting to board that vehicle on the next stop will be even higher, causing it to fall even further behind. Conversely, when a vehicle begins to catch up the vehicle before it, fewer passengers will be waiting to board at the next stop, possibly causing it to catch up even further. This increases average waiting times since vehicles are not evenly spaced, increases average crowding since passengers are not evenly distributed across vehicles, and increase unreliability since vehicles cannot keep their schedule.

4. Boarding and alighting time increases as the vehicle becomes more crowded. This could truly be described as congestion.

5. If the bus lane or rail segment gets full, the public transport vehicles will cause delays to each other, just like car traffic congestion.

6. Car traffic congestion can cause delays in public transport if the public transport is not entirely separated from car traffic.

7. Crowding and congestion can occur at stops and transfer nodes, causing both discomfort and actual delays.

The very influential paper on public transport modelling by Spiess and Florian, Optimal strategies: A new assignment model for
transit networks (1989) intentionally uses the term “congestion” to mean anything caused by increased travel demand that would make the service less attractive to the travellers, mentioning on-board crowding as an example and longer vehicle travel time as another. This is because their model did not have an explicit representation of travellers or vehicles, as it is a network flow model. In this context, the term congestion is borrowed directly from literature on car congestion modelling, which was already quite mature then.

Subsequently it seems that this convention spread in the public transport modelling society, which more and more started to use the word congestion as synonym for crowding. As an example, Kroes et al. (2013) in the background to their research topic say that “the reduction in congestion levels can also be forecasted by more advanced traffic models”. It is not easy to say what is meant by congestion in this sentence, but the following paragraphs as well as the title of the paper (“The value of crowding on public transport in Île-de-France”) reveal the meaning. Nuzzolo et al. (2012) stated that “in the context of transit networks, congestion usually refers to the decrease in on-board comfort as the on-board load increases up to a maximum threshold (vehicle capacity), after which users are not allowed to board (oversaturation) and have to wait for next arriving vehicle.” In their paper, the word crowding is not used at all. However, in papers written by native English speaking authors, the term crowding is generally used instead of congestion.

**Congestion pricing**

Congestion pricing is the use of time-variable fees to charge the users of public goods for the negative externalities generated by peak demand. To relieve traffic congestion within a city centre the fee can be a cordon-based toll paid by car drivers to enter or leave the area.
Dynamic traffic assignment

Static assignment models distribute OD-flows over routes in the network for an analysis period assuming a volume-delay function for each link. The drivers are assumed to choose the cheapest route where the cost of the route is a combination of travel time and monetary cost. Representation of travel time in static models suggests a one-to-one relationship between volume (demand for trips through the road link) and travel time on the link, i.e. a separable volume-delay function. In real world situations with congestion, there is no such separable relationship, and the distribution of travel times for a given demand depends on demand for trips on other links and at time proceeding the analysis period. Therefore the calculation of travel time based on separable volume-delay functions is misleading in such situations.

Even if the volume-delay functions and the demand matrices can be calibrated so that the static assignment results in plausible traffic volumes and travel times in a baseline situation, the change in travel times due to changes in travel demand will usually be underestimated in a static model because the positive and non-linear (convex) relationships to other links and periods are not taken into account. Truly dynamic traffic assignment (DTA) models, on the other hand, capture the spatio-temporal dynamics of vehicles, vehicle packages or vehicle flows. This typically implies a temporal model resolution in the order of seconds, and a spatial resolution in the order of vehicle lengths.

Approximately dynamic (semi-dynamic) models typically replace exact spatio-temporal vehicle dynamics by loosely coupled assignments per time slice. They allow for a much coarser, say hourly, time resolution. Due to the coarser representation of traffic dynamics, the usage of semi-dynamic models might be easier. In particular, the semi-dynamic models typically require fewer parameters, less calibration effort and shorter calculation times. However the semi-dynamic models suffer from inconsistencies that result from their inherent approximations; queue build-up
and dissipation processes are not exactly captured and the representation of trips that span multiple time periods is problematic.

In a view of model systems used in strategic planning and cost-benefit analysis, possibility of coupling with demand models is an important criterion for the choice of representation of traffic dynamics. Given that a major concern in the coupling is the maintenance of consistent data associations (e.g. between persons, trips, and cars), a truly dynamic model is highly preferable over an approximation thereof. Any approximation of the dynamics in the network loading procedure is likely to lead to a further degradation of the representable data associations. Also, recent algorithmic and computational advances (event-based simulation, multi-threaded simulation) render the truly dynamic simulation even of large metropolitan regions feasible.

Moreover, currently there are no models in planning practice that consistently incorporate travel demand modelling with truly dynamic assignment models. The main obstacle is requirement of detailed calibration of behaviour at specific road links and intersections. The truly dynamic models tend to be very sensitive to details that are not possible to input as they are today and to keep in a future scenario. Indeed many intersections need to and will be modified in future scenarios.

**Results**

**Ex-post evaluation of national transport model – Gothenburg congestion charges application**

The topology of the transport network in Gothenburg implies a large number of OD-relations where the driver has the choice between a faster charged route and an uncharged but slower route. Due to this, the predicted route choice proved to be highly sensitive to the VTT assumed in traffic assignment. On top of this a special
multi-passage rule was applied in Gothenburg, introducing a non-additive component into the route choice utility function. Nevertheless, we found high accuracy of the predicted reduction of traffic volume across the cordon in the peak, 11% compared to the observed 12%. The reduction in the off-peak, however, was overpredicted, as it was also in the Stockholm case. The lower accuracy of the off-peak predictions seems to be driven by the different and possibly more diversified adaptation strategies applied to discretionary trips, whereas virtually all commuters priced off the road switched to public transport.

**Integration of dynamic traffic assignment with a travel demand model for the Stockholm region**

To analyse congestion charges in Stockholm it is necessary, yet fully possible to integrate DTA with the demand model. In the performed tests we could see that both tested models had problems replicating the flow on the main bypass early in the morning. It is likely that a big part of the explanation is that we have used the same departure time profile for all origin-destination-pairs. Traditionally, DTA have been used with fixed demand. This is not a problem for limited policy interventions, or for designing local traffic facilities. Our ambition, however, is to be able to use DTA tools for more strategic policies, where demand effects are not only expected but intended. In such cases we cannot rely on fixed demand estimated to replicate flows, we need a demand model. Furthermore, in some cases, such as time-differentiated congestion pricing, departure time choices will be of crucial importance. But as our results show, the importance of geographically different departure times may actually become an issue even without policies that shift departure times explicitly. A potential disadvantage in using a simulation model is the necessity to perform several replications of the simulation for making a statistically sound comparison between alternatives. This still presents a challenge due to extensive computing time.
Appraisal of increased public transport capacity
Crowding effects are a large majority of the benefits in a typical metro rail project. These effects are not possible to capture in a static assignment model. A case study of a metro extension in Stockholm demonstrated that congestion effects constitute more than half of the total benefits and that these effects are excessively underestimated by a conventional static model. In other words, accounting for the dynamic congestion effects added 120% to the benefits of a conventional static model which essentially only captures travel time savings. Using crowding factors to adjust the travel time in the static model in the case study only adds 3% to the benefit.

Evaluating transit preferential measures - priority lanes, boarding and control strategies
Effects of various operational measures can be analysed with BusMezzo and the results have been validated against empirical data. Our findings indicate that all three tested measures (boarding through all doors, headway-based holding and bus lanes) had an overall positive impact on service performance and that they exercised negative synergy effects with their combined effect being smaller than the sum of their marginal contributions, except for headway-based holding which exercised positive synergy effects with the two other measures. It is therefore advisable to simulate alternative measures prior to their implementation to assess their impacts and refine their design.

Discussion
The theories used in this work are mostly not novel. Analytical transport models have been used for a long time in project appraisal and utility functions including value of time for different user classes and transport modes are usually key features in these. Using a continuous value of time distribution in a route choice
model is certainly not standard, but there is no new theory behind it. No new parameter estimations were done in this work.

Traffic simulation models are not new either. Until recently though, the usage of micro simulation has been limited mainly by computer power, but that constraint is now rapidly becoming less important. In more abstract analytical models many assumptions and theories about variable dependencies are usually needed, often expressed in complicated formulas. A pure micro simulation traffic model on the other hand does not depend on complicated formulas, even if the algorithms can be rather incomprehensible at first look. The concept of agent-based simulation is rather simple and straightforward, at least in theory. Still, the programming and calibration effort of good micro simulation models is a serious hindrance to their further usage. In practice, many promising attempts have halted in their early developing stages due to lack of resources or devotion.

The cases studied in this work are relatively specific to their Swedish circumstances. A Swedish standard for how congestion charges are implemented has been set by the Stockholm and Gothenburg systems and other systems might not necessarily have much in common regarding the design of the system. Crowding problems experienced in Swedish public transport are not even close to the levels that can be experienced in larger cities and the Swedish willingness to pay for rather mild improvements of comfort might not be existent in many other places in the world. Still, the examples in this work can be regarded as good examples of how state of the art and state of the practice can be brought closer to each other when it comes to modelling and evaluating effects of capacity improvement and congestion mitigation.

The simulation models that were chosen in the case studies, namely BusMezzo and TransModeler, both have their strengths and weaknesses. They should not be regarded as ultimate solutions to every problem, but as fine examples of fairly advanced micro
simulation models that have the crucial property that they can be scaled to large applications and handle route choice. The models used are not rigorously estimated, meaning that the error margin of the results is typically large.

What is new in this work is to actually have tried these methods in practice. Most studies stay at a highly exploratory level, with simplified models and hypothetical cases. In this work, all cases are full scale demonstrations with the most ambitious modelling approach found.

**Conclusions**

Using a continuous VTT distribution and hierarchal route choice was demonstrated as a successful method of modelling the multipassage rule implemented in Gothenburg congestion charges and was shown to give realistic predictions of route choice effects. This method can be applied to the Stockholm network and other comparable congestion charging systems as well, which would improve realism in the distribution of choices, especially when there are similar discount structures.

The Stockholm case showed that in highly congested road networks a dynamic simulation model might be crucial for capturing the full benefit of congestion charges. First results from integration of DTA with a travel demand model for the Stockholm region show that even without systematic calibration the DTA is in reasonable agreement with observed traffic counts and travel times. The presented experiments did not reveal a striking difference between using a macroscopic and a microscopic assignment package. However, given the clear trend to microscopic modelling and simulation on the travel demand side, the use of a micro-simulation-based DTA package appears more natural from a system integration perspective.
While travel time savings are often the only benefit included in public transport project appraisals, the best practice assigns weighted value of time to average load/capacity measures. However, failure to represent dynamic congestion effects may substantially underestimate the benefits of projects primarily designed to increase capacity rather than reduce travel times such as the construction of high-capacity public transport, redesigning vehicle capacity or increased service frequency.

The impact of small operational measures should not be underestimated. Our bus line case study showed that measures such as priority lanes headway control and faster boarding can give large gains to the passengers. These measures are relatively cheap compared to investments in new transit infrastructure and large societal gains can therefore be achieved by their implementation. Evaluating such gains requires detailed modelling of public transport supply.

**Included papers**


References


Nagel, Kai, and Gunnar Flötteröd. 2012. “Agent-Based Traffic Assignment: Going from Trips to Behavioral Travelers.” In Travel Behaviour Research in an Evolving World: Select Papers from the 12th International Conference on


Wegener, Michael. 2004. “Overview of Land-Use Transport Models.” In Handbook of Transport Geography and