COMPARISON OF REGIONAL CONGESTION METRICS WITH STATIC AND DYNAMIC ASSIGNMENT MODELS

Norman L. Marshall
Smart Mobility, Inc.
205 Billings Farm Rd. Unit 2-E
White River Jct., VT 05001
Tel: 802-649-5422 Email: nmarshall@smartmobility.com

Word count: 4,824 words text + 10 tables/figures x 250 words (each) = 7,324 words

Submission Date: 7/29/15
ABSTRACT
Most regional travel demand models assign traffic to roadway networks using static traffic assignment (STA) models. However, it is well known that STA does not model congested networks well because each roadway segment in the model is treated independently. Dynamic traffic assignment (DTA) models represent congestion much better. Tests done with STA and DTA for the same congested networks show that: 1) freeway and ramp assignments are higher in STA than DTA and can exceed capacity, 2) vehicle hours of delay (VHD) is significantly greater in DTA relative to STA because of the delays associated with queuing, and 3) STA performance metrics are unreliable in congested networks, particularly for future conditions where many regional STA models show extreme levels of congestion. It is recommended that DTA be used instead.

Keywords: Regional Planning, Travel Demand Modeling, Dynamic Traffic Assignment, Congestion, Delay
INTRODUCTION
The regional travel demand models developed and maintained by Metropolitan Planning Organizations (MPOs) are used to evaluate transportation alternatives in regional transportation plans (RTP) and for major transportation projects in Environmental Impact Statements (EIS). Congestion and delay measures derived from these models are critical performance measures.
Most, and possibly all, of the MPO travel demand models assign traffic to roadway networks using static traffic assignment (STA) models. This is true even for the new generation of Activity-Based Models (ABMs). While the long-term goal is to combine the ABM demand model with microsimulation, this has not yet been accomplished due to the much greater computer resources required (1). STA models treat each roadway segment as independent. STA models have no queues and no spillback.

In a static model, inflow to a link is always equal to the outflow: the travel time simply increases as the inflow and outflow (volume) increases. The volume on a link may increase indefinitely and exceed the physical capacity ... as represented by a volume-to-capacity (V/C) ratio > 1... The drawback of using V/C is that it does not directly correlate with any physical measure describing congestion (e.g., speed, density, or queue (2).

Dynamic traffic assignment (DTA) models have been developed that have intermediate computer processing requirements between STA and microsimulation. A 2012 reference on modeling practice states: “The DTA methodology offers a number of advantages relative to the STA methodology, including the ability to address traffic congestion, buildup, spillback, and oversaturated conditions through the explicit consideration of time-dependent flows and the representation of the traffic network at a high spatial resolution (1)

Studies that have compared STA and DTA for the same problem have found large differences in model performance measures. Boyles et. al. concluded: “The results indicate that traditional static models have the potential to significantly underestimate network congestion levels in traffic networks, and the ability of DTA models to account for variable demand and traffic dynamics under a policy of congestion pricing can be critical” (3). In a study of choice between managed lanes (ML) and general purpose lanes (GPL) by the Florida Department of Transportation, it was concluded that: “the difference in the travel time of using the GPL or the alternative ML, and the resulting number of travelers that decide to choose the ML, is considerably underestimated by static assignment” (4).

METHODOLOGY
In this paper, STA and DTA are applied to a series of simple congested networks to investigate the differences in congestion-related performance metrics. These case studies include:

1) Comparison of 2-link and 3-link networks with fixed demand
2) Extension of the 3-link network to managed lanes choice
3) Subarea network with route choice.

STA modeling was done in a spreadsheet for the simplest cases with fixed demand and no route choice. STA modeling for the more complicated networks was done using TransCAD.

The DTA software used in these tests is DTALite (5). The DTALite developers state: “DTALite, an open-source mesoscopic DTA simulation package, in conjunction with the Network eXplorer for Traffic Analysis (NeXTA) graphic user interface, has been developed to provide transportation planners, engineers, and researchers with a theoretically rigorous and
computationally efficient traffic network modeling tool” (6). STA models are “macroscopic” with no representation of individual vehicles. Microsimulation models are “microscopic” with full representation of individual vehicles. DTA models including DTALite are “mesoscopic” and represent vehicle behavior using aggregates. DTALite uses a queue-based approach (6). The tests described below were done using the DTALite default Newell’s kinematic wave model.

This tests all use a general protocol varying demand by time period. The simulations include:

a) 30-minute initialization period with moderate demand
b) 60-minute period with high demand
c) 60-minute period returning to moderate demand

This protocol allows portrayal of temporary traffic spillback, which ultimately clears as traffic levels diminish. The traffic metrics reported are for the average conditions over the final 2 hours of the simulation, i.e. without the initialization period.

DYNAMIC ILLUSTRATION OF BRAESS PARADOX
The first network problem was chosen to illustrate how results between STA and DTA can be very different, even in an extremely simple network. Braess established that there are cases where adding roadway capacity in STA increases total travel time. This is the well known “Braess paradox” (7). Knoop, Hoogendoorn and van Arem (8) illustrate that the Braess paradox also is applicable with DTA and illustrate this with a very simple network with 3 one-way links (reproduced as Figure 1)

FIGURE 1: Simple Illustrative Network Reproduced from Knoop (I)

In the left hand diagram, there are only 2 travel links. The capacity of both links 1 and 2 is greater than the flow, and there is no congestion. In the right-hand diagram, a 3rd travel link is added parallel to link 2. This new link 3 offers a faster travel time than link 2. Therefore, in the dynamic case, all vehicles will choose link 3 and no vehicles will choose link 2. However, in this illustration, the capacity of link 3 is lower than the flow. Therefore, vehicles will queue on link 1 waiting to get to link 3. The additional roadway capacity has introduced significant congestion where there was no congestion before. This result is somewhat counterintuitive, and may be especially surprising for those long experience in STA modeling, which as discussed below, performs very differently. However, Knoop’s interpretation is clearly correct. Once, a driver has
Marshall

1 reached the downstream end of link 1, they will choose the faster route which is link 3.
2 Assuming perfect information, no drivers will choose link 2.
3 This example was implemented in DTALite using the parameters shown in Figure 2. This
4 includes 3 lanes at 55 mph for links 1 and 2, and 2 lanes at 65 mph for link 3. Each of the links is
5 1.0 miles in length.

![Diagram of 2-link and 3-link systems](image)

(a) Without extra link  
(b) With extra link

**FIGURE 2: Implementation of Knoop Braess Paradox in DTALite**

In this first test, the test volumes are 3000 vehicles per hour for the moderate periods
(initialization and final hour) and 4000 vehicles per hour for the high-demand hour. All DTALite
parameters and default methods were kept except the capacity per freeway lane per hour was
increased to 2000 from 1900. This was done to simplify calculations, and also because a capacity
of 2000 is common in STA models. With the input numbers described above, the volume-to-capacity
ratio (V/C) in the 2-link system for the moderate traffic period is 3000 vehicles / (3
lanes * 2000 capacity/lane) = 0.5. It increases to 0.67 in the high-volume hour. The STA Bureau
of Public Roads (BPR) volume coefficients were set to average values for large MPOs in
NCHRP 716 (1): i.e. alpha = 0.48 and beta = 6.95.

Figure 3 shows speed and delay metrics for the 2-link system using both STA and DTA.
The results are very similar, with average speed of 54 mph in both models and little delay.

When the third link is added, the models perform very differently. With DTA, link 3 is
faster than link 2 and is chosen by all vehicles. In the high traffic period, the volume-to-capacity
ratio on link 3 increases to 1.0 and traffic spills back onto link 1. However, each vehicle getting
to the downstream end of link 1 will choose link 3, and link 2 receives no traffic. As shown in
Figure 4, the average travel speed over the 2-hour period drops by more than half to 24 mph and
modeled delay skyrocket from 4 hours to 320 hours. In the STA, traffic splits along links 2 and
3 so that both links operate at the same speed. This increases the average speed slightly over the
2-link case. There is more delay than in the base case, but only 1/16 as much delay as with DTA.
The DTA model accurately portrays this case; the STA model does not.
FIGURE 3: Static and Dynamic Model Congestion Metrics – Without Link 3

FIGURE 4: Static and Dynamic Model Congestion Metrics – With Link 3
The Braess paradox simulation described above may seem like a contrived special case that is not applicable to the real world. However, the Braess paradox network actually is very similar to sections of managed lanes networks being planned and implemented today. This is illustrated with a network with 4 freeway lanes dividing into 2 barrels, each with 2 lanes (Figure 5).

![Managed Lanes Network Section in DTALite](image)

**FIGURE 5: Managed Lanes Network Section in DTALite**

The traffic volumes tested are 4800 vehicles per hour in the moderate traffic period and 6400 vehicles per hour in the high traffic hour. With equal lane utilization, these volumes correspond to $V/C = 0.6$ and $0.8$ in the two periods. Two cases are tested – one with a 50-50 traffic split, and one with 60 percent assignment to the general purpose lanes and 40 percent to the managed lanes. The 60-40 split is representative of real-world projects and often is achieved through a combination of free access for high-occupancy vehicles and tolled single-occupant vehicles (HOT lanes).

The links are only 1.0 miles in length; therefore it is assumed that vehicles have mostly sorted out to the proper lanes prior to entering this network. The results of the managed lane simulation is shown in Figure 6. These results are very similar to the Braess paradox simulation results. Traffic flows well for the 50-50 split, but there is so much spillback behind the fork in the 60-40 split scenario that the average speed for all vehicles for the 2-hour period is only 26 m.p.h. Furthermore, the travel time for the managed lanes vehicles is only slightly lower than for the general purpose lanes because almost all of the delay is behind the split.
FIGURE 6: Static and Dynamic Model Congestion Metrics for Managed Lanes Simulation

Sometimes it is suggested that there is an “optimal” traffic split for managed lanes. What is optimal depends on what is meant by optimal. In both STA and DTA models, the highest speeds and the least congestion are with a 50-50 split, i.e. without managed lanes. In STA models, the two barrels are treated as independent of each other and also independent of upstream and downstream networks. Therefore, congestion can easily be isolated to the general purpose lanes. The simulation described above indicates that in the real world, it can be hard to isolate congestion to general purpose lanes without having them spillover onto the managed lanes. These spillover effects can be avoided by relying only on direct-connect ramps. However, such systems are extremely expensive to construct. Therefore, most managed lane designs combine a combination of direct connect ramps along with merge and diverge sections interacting with the general purpose lanes. The assumed traffic volumes in the simulation described above are moderate compared to forecast general purpose traffic volume in some HOT lane studies. Here 60% of 6400 vehicles per hour is 3840 vehicles per hour or 1920 vehicles per lane per hour (V/C = 0.96) for a single peak hour before traffic volumes drop. This is high enough to cause significant spillback so that all vehicles are delayed behind the split. In fact, the travel time savings in the simulation for the managed lane users is only a few seconds – much less than would be estimated with STA.
ROUTE CHOICE

There are two major limitations in all of the modeling discussed above: 1) routes are fixed, and 2) demand is fixed. In this section, routing changes are explored using a subarea model. The DTALite distribution includes a sample network for a section Fort Worth Texas south of the downtown along I-35W between I-30 and I-20 (Figure 7).

The subarea network includes I-35W plus frontage roads and the more important streets. Streets in the network are as close as ¼ mile apart in the northern part of the subarea but average about a mile apart in the southern part. The subarea network is fairly detailed with signal control information entered for 92 nodes. The urban form represented in the subarea network is typical of many urban areas in the United States. I-35W express lanes in the model (but not yet constructed) were removed for the sake of simplicity, leaving four freeway lanes in each direction. As in the previous case studies, the freeway capacity per lane per hour was increased to 2000 and the same BPR volume-delay coefficients for freeways were used in STA. Similarly, arterial volume-delay coefficients for arterials were taken from NCHRP 716 (1): i.e. alpha = 0.53 and beta = 4.40.

The sample network distributed was used only for modeling construction impacts and does not include traffic volumes for the entire network. Therefore, illustrative traffic volumes were developed. In the real world, I-35W and other urban freeways carry a mix of traffic that can be divided among:

a) through traffic (XX),

b) traffic with both origin and destination within the study area (II), and
c) traffic with either an origin or destination in the study area (IX/XI).

In order to keep the model as simple and understandable as possible, the modeled traffic volumes include only XX and II traffic. This simplification focuses the exercise on the impacts of local traffic jumping on and off an urban freeway.

For the through XX traffic, the modeled volumes in each direction are 5000 for the moderate periods and 6000 per hour in the heavier traffic period. With 4 lanes in each direction and a capacity of 2000 per lane per hour, the through traffic alone accounts for a volume-to-capacity ratio on the freeway of 0.625 in the moderate period and 0.75 in the high traffic period.

The distributed network includes 11 internal transportation analysis zones (TAZs), most of which load traffic at multiple nodes. A trip table was developed with traffic volumes from each internal TAZ to every other internal TAZ. A total of 110 II cells are filled (11 origin TAZs, each with 10 destination TAZs). For the sake of simplicity, each of these trip table cells has the same value. 55% of the two-hour volume is assigned to the first of the analysis hour and the remaining 45% of traffic to the second hour. A commonly used measure of congestion is the Travel Time Index (TTI) developed by the Texas Transportation Institute for their Urban Mobility Report series (9). The TTI is the ratio of total congested travel time to total free-flow travel time. In this exercise, the II trip table was scaled up until the level of congestion in the network for the two-hour analysis period until it was comparable to the TTI reported in the 2012 Urban Mobility Report for moderately-congested regions, 1.22. This level of congestion is achieved with trip matrix cell values of 138 in the first hour and 112 in the second hour.
If all travel demand could be completed with free-flow travel speeds, most local (II) trips traveling in a generally north-south direction would use I-35W for a portion of the trip because the speed limit on I-35W is much higher than the 40 mph speed limit on the frontage roads and other arterials, and actual maximum arterials speeds are even lower due to traffic signals. This idealized behavior of everyone using the shortest free-flow path is generally consistent with a road hierarchy where each trip segment is completed with the highest class of roadway possible. This model was tested using 1-iteration DTA and results in immediate gridlock. The primary...
cause is multiple instances of the Braess paradox network embedded in the Fort Worth subarea network (and probably in every other region in the U.S.) As illustrated in Figure 8, in a 1-iteration model, vehicles choose the freeway because it is faster. However, capacity is constrained by single-lane on-ramps and off-ramps. Without rerouting, queues quickly form on the arterial street segments upstream of the on-ramps and queues also from on the freeways upstream of the off-ramps.

![Braess Paradox Network Embedded in Fort Worth Subarea Network](image)

To create more realistic traffic assignments, 100 iterations were done with the default fixed switching rate of 15%. This method converges to a stable solution faster than the alternative Method of Successive Averages (MSA). This resulted in a TTI of 1.22 over the 2-hour analysis period. It is important to note that the freeway cannot function properly with congestion to moderate the demand for access and egress.

With STA (done in TransCAD), the TTI is higher (1.09) but the total travel time is 4% less than with DTA. This combination of lower time and higher TTI results from more traffic on the freeway (Figure 9). STA is not constrained to keep link flows under capacity. In STA, two of the ramps have a volume of capacity ratio of over 1.0 for the two-hour period and others have high V/C just below 1.0.
FIGURE 9: 2-Hour Static Assignment Volume – Dynamic Assignment Volume (Red higher than 500 or more vehicles, Green lower by 500 or more vehicles)
In the base subarea example presented in the previous section, the through XX traffic alone is 11,000 in each direction for the 2-hour analysis period, representing 69% of capacity. With DTA, the highest freeway segment volume is 14,106 or 88% of capacity. With STA, the highest freeway segment volume is 15,343 or 96% of capacity. As it is assumed that 55% of the 2-hour traffic is in the peak hour, these traffic levels correspond to peak hour V/C = 97% for DTA and 105% for STA.

The standard response to this congested situation is to add one or more freeway lanes in each direction. This was tested with both models by adding a fifth through lane in each direction. With DTA, the maximum freeway link volume increases by 1,038 to 15,144. With STA, the maximum freeway link volume increases by 799 to 16,142. Total travel time drops by 12% in both models. The DTA TTI declines from 1.22 to 1.08. The STA TTI declines from 1.08 to 1.04 (but this is achieved with unrealistic ramp V/C as high as 1.28). These indicators suggest that the freeway widening is a successful enhancement to the network. However, as is demonstrated in the next section, the expanded network is very fragile.

TIME-OF-DAY CHOICE

A major limitation of all of the analyses presented above is that demand is fixed. A recent review of the research literature published on induced travel between 1997 and 2012 concluded: “Thus, the best estimate for the long-run effect of highway capacity on VMT is an elasticity close to 1.0, implying that in congested metropolitan areas, adding new capacity to the existing system of limited-access highways is unlikely to reduce congestion or associated GHG [greenhouse gas] in the long-run” (10).

Even if daily demand were fixed, adding capacity to a congested network would cause shifts in traffic into peak travel periods. This peak shifting is tested with DTA with 5 freeway lanes in each direction by keeping the same traffic volumes for the two-hour period, but changing the temporal split from the 55%/45% in the base case to 57%/43%. This fairly modest change is sufficient to overload the freeway access points and results in gridlock. Compared with the base case with 4 freeway lanes in each direction and a 55%/45% split in traffic volumes, the 5 freeway lanes case with a 57%/43% split results in an increase in VHD of 265% and an increase in TTI from 1.22 to 1.79 for the combined two-hour period.

Results from all of the Fort Worth subarea tests are summarized in Figure 10. Key differences between the metrics for the two models include:

- Freeway and ramp assignments are higher in STA than DTA and can exceed capacity.
- Free-flow travel time along assigned paths is significantly greater for DTA compared to STA due to less reliance on the freeway for short trips.
- Vehicle hours of delay (VHD) is significantly greater in DTA relative to STA because of the delays associated with queuing.
- Freeway widening is a clear benefit in STA, but in DTA, the effects are uncertain and depend on the demand response.
Figure 10: Summary of Fort Worth Subarea Dynamic and Static Model Tests Showing Free-Flow Travel Time and Vehicle Hours of Delay (Truncated Scale)

**DISCUSSION**

Regional Transportation Plans (RTP) and roadway Environmental Impact Statements (EIS) rely on static assignment models for distant horizon years. In larger U.S. regions, these horizon year models typically show many important roadway links over capacity. Here is an extreme example from a 2015 Draft EIS for a proposed tunnel in southern California, the SR 710 North Extension Project (11). For I-710 northbound at I-10 (the primary upstream source of northbound tunnel traffic), model files for 2035 for the full tunnel alternative show an average traffic volume of 2214 vehicles per lane per hour for the 13-hour period from 6 a.m. to 7 p.m. The static model calculates a delay of about a minute and a half for this roadway segment, but this much travel would really cause many thousands of vehicles to spillback and never reach the tunnel at all.

In cases like this tunnel study, future traffic demand would be much lower than modeled with STA. When future static models show volumes on important links over capacity, the first conclusion that should be drawn is that the model is wrong. The traffic volumes shown are impossible, and also outside the range of traffic volumes that were observed in model validation. The second conclusion to draw is that if traffic volumes are anyway near as great as shown, then delays are greatly underestimated because of the model’s failure to account for queues.
Furthermore, these delays will be concentrated in bottlenecks that STA is not well equipped even to identify.

The growing practice of linking STA outputs to microsimulation tools including VISSIM does not address these deficiencies. The underlying traffic volumes used in the microsimulation model are from the STA. When these STA results include over-capacity links (which cannot be modeled in microsimulation without queues of ever-increasing length), the problem has been addressed in two different ways. Most commonly, the scope of the microsimulation is constrained to the project area where sufficient capacity is assumed. In these cases, issues about unrealistic upstream and/or downstream traffic volumes generally are not addressed. The other approach is to arbitrarily scale down the traffic volumes so that the traffic can be simulated. Neither approach results in realistic network traffic volumes.

Apparent benefits shown in STA from added freeway capacity in urban areas should be treated with special skepticism. As illustrated above, STA will calculate benefits in overloading ramps and broader access routes in ways that are not physically possible. If the STA includes over-capacity links at access point, the STA performance measures are invalid. Furthermore, unintended changes in induced traffic and or time-of-day distributions could cause specific freeway capacity projects to make congestion worse rather than better. In a regression analysis of 74 regions, it was found that more arterial capacity is strongly related to less congestion, but that more freeway capacity is not (12).

As discussed above, congestion is an inherent consequence of building freeways in dense urban areas. Unless the freeway can carry all of the parallel traffic (generally impossible), the demand for freeway access at free-flow travel times will exceed capacity. For the freeway and its access points to function, this demand must be moderated by higher travel times, i.e. by congestion. The only alternative to congestion is pricing all freeway lanes – not just a subset of freeway lanes. Whether specific freeway capacity improvements would result in more or less congestion is a very complicated question that cannot be answered adequately with STA.

Given advances in DTA algorithms and increased computing power (including parallel processing across many cores as is being applied in large Activity-Based Models), it likely it is now possible to implement DTA in many regional travel demand models in the near term, and this is recommended. In addition to the computer requirements, switching from STA to DTA would mean a substantial learning curve as DTA forecasts would be very different than the STA forecasts and might include unexpected results.

Mentally grappling with DTA models could result in smarter transportation network design. For example, Knoop (8) recommends a simple modification of the Braess paradox network that is reproduced in Figure 1 above. In this revised network (Figure 11), the problem is eliminated by relegating the queue that forms behind the fast but low capacity link 3 to its initial segment labeled 3a. This prevents the queue from blocking link 2. If a substantial queue forms on link 3a, its travel time will equilibrate with the travel time on link 2 which will tend to limit the length of the link 3a queue. This simple heuristic could help improve network designs. For example, to address the freeway/ramp queuing issues, the ramps could be given considerably more storage as is sometimes done in ramp metering projects.
CONCLUSIONS AND RECOMMENDATIONS

1) Static traffic assignment (STA) cannot accurately estimate vehicle delay or other congestion metrics in congested urban networks.

2) STA is particularly bad at estimating the effects of added freeway capacity in congested urban networks.

3) STA should be replaced with dynamic traffic assignment (DTA) in regional travel demand models as soon as possible.

4) The training of transportation planners and engineers should include exercises with DTA and microsimulation so that the professionals have a better understanding of the complex interactions in urban networks.

Table 1: Properties of the links

<table>
<thead>
<tr>
<th>Link</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>3a</th>
<th>3b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity $C$</td>
<td>$&gt; Q$</td>
<td>$&gt; Q$</td>
<td>$&lt; Q$</td>
<td>$&gt; C_{3b}$</td>
<td>$= C_3$</td>
</tr>
<tr>
<td>Free flow travel time</td>
<td>$T_{1\text{free flow}}$</td>
<td>$T_{2\text{free flow}}$</td>
<td>$T_{2\text{free flow}}^{\tau}$</td>
<td>$T_{3a\text{free flow}}$</td>
<td>$= T_{3\text{free flow}} - T_{3a\text{free flow}}$</td>
</tr>
</tbody>
</table>

Figure 2: The solution avoiding extra travel time

FIGURE 10: Knoops Solution to the Braess Paradox – Reproduced from (8)
REFERENCES


4) Florida Department of Transportation, Lehman Center for Transportation Research, URS Corporation and Citilabs. Application of Dynamic Traffic Assignment to Advanced Managed Lane Modeling, November 2013


