Deploying a dynamic traffic assignment model for the Sydney region

Melissa Duell, Nima Amini, Sai Chand, Neeraj Saxena, Hanna Grzybowska, Travis Waller
University of New South Wales, Sydney, Australia
Email for correspondence: m.duell@unsw.edu.au

Abstract

To study traffic dynamics a network level, traffic flow should be evaluated from a fine resolution. Considering the ever changing traffic conditions in a city, Dynamic Traffic Assignment (DTA) models are useful to access the impact of policy measures on the travel behaviour and overall network performance at a regional level. A majority of established planning models in Australia (and around the world) are macroscopic in nature and provide aggregated performance measures for the entire network. However, these traditional models cannot represent important phenomena such as queue spillback or temporal congestion propagation due to their time-invariant framework. One solution to this issue that is receiving greater attention in both research and industry is DTA modelling.

This work presents an overview of developing and deploying a metropolitan area dynamic assignment model (MADAM) for the Sydney region. The workflow for the project involved obtaining the necessary input data, including network geometry, travel demand, signals and transit, converting the data into the form needed for the DTA platform, making model adjustments for computational efficiency, implementing and finally calibrating the model. In particular, this work focuses on the role of dynamic traffic assignment models, lessons learned during the current deployment, and presenting an overview of the calibration process and the model outputs. Future research directions will also be discussed.

This study makes a significant contribution toward developing a regional dynamic model for a metropolitan city in Australia. In the future, the MADAM model could aid in evaluating important policy decisions and infrastructural development in the context of the overall network operation. This project serves as a proof of concept and may provide valuable insight to other practitioners interested in the areas of transport planning and traffic modelling.

1. Introduction

Dynamic traffic assignment (DTA) is one of the next steps in the continuing evolution of practical traffic models. While static transport planning models have a rich history in research and in practice, it is well acknowledged that they do not meet all needs and offer numerous opportunities for improvement. Additionally, traffic micro-simulation, which is another useful prediction tool, is impractical for most large-scale applications due to the intensive data and computational requirements.

DTA is an active field of research that aims to capture time-dependent phenomena such as queue spillback, bottlenecks, and temporal congestion. These reasons make DTA an appealing alternative to traditional static traffic assignment models. However, DTA applications remain relatively scarce in practical settings, possibly due to model complexity and general confusion regarding the practicalities of large-scale implementation and calibration. Thus, this work intends to share insight and offer practical knowledge about building and implementing a large-scale DTA model, particularly for Australia.

This paper provides an overview of the development of the metropolitan area dynamic assignment model (MADAM) for Sydney, Australia. This application consists of a two-hour AM-peak network consisting of 58,583 links, 20,730 nodes, 2,282 zones, 1,262,930 vehicle
demand, 490 signalized intersections, and 1,159 bus routes. A number of the project challenges involved the data processing, which is discussed in detail in Chand et al (2015). This paper discusses the learning curve regarding the handling of relatively large datasets that was ultimately overcome by employing more visualization techniques in GIS software. While model deployment and calibration is currently an ongoing process, the initial efforts show promise.

First, this work discusses the background of DTA and other deployed DTA models in Section 2. Next, a number of aspects regarding model development are summarized in Section 3. Section 4 presents a sample of the modeling results, experiences with model calibration, and the limitations of the modeling approach. This paper concludes with a brief discussion of future model extensions.

2. Background

2.1 Dynamic traffic assignment

The static traffic assignment models that serve as the route choice component of a traditional four step planning model are generally based on the Wardropian principle of user equilibrium (UE) (Wardrop 1952). In UE, users choose the shortest path to minimize their own travel time, which collectively results in a state of network equilibrium. These models have been enduring due to numerous favorable properties, such as uniqueness, stability, and computational tractability. Static models are generally based on link performance functions and output average measure of network performance. They are not well-suited to capturing important time-dependent network phenomena such as bottlenecks, reliability, or the “peak” effect of travel demand (Peeta and Ziliaskopolous 2001).

It is well established that time dynamics play a vital role in traffic conditions. This application seeks to capture the impact of different traveler departure times on network conditions. The time dependent origin-destination (OD) demands are an input and remain constant between iterations (thus it does not account for departure time choice). Figure 1 illustrates trips over a twenty-four hour period on an average weekday in Sydney as reported in the Household Travel Survey (BTS 2013). While the peak variation shown in Figure 1 is not unique to Sydney, it does make a convincing case as to the importance of accounting for time dynamics in traffic models.

Figure 1. Illustration of time-variance in weekday trips from the New South Wales Household Travel Survey (BTS, 2013)

Dynamic traffic assignment is an established method that is well described by Chiu et al (2011). In this work, we refer to simulation-based DTA that follows a general procedure such
as in Figure 2. Based on an input demand for a chosen time period, DTA seeks the network conditions of dynamic user equilibrium, in which the travel time on all paths between an origin-destination at a departure time (ODT) is equal. DTA is generally solved using an iterative procedure in which time-dependent minimum cost paths for each ODT are found, then the updated network conditions are simulated, and the distribution of vehicles on each path between an ODT are adjusted in order to minimize a relative gap.

**Figure 2. Overview of the DTA Procedure**

The DTA platform utilised in this application is known as VISTA (Ziliaskopoulos & Waller 2000) and has also been used in a number of other deployed DTA approaches. A variety of other DTA platforms based on different methodologies are also commercially available.

### 2.2 Previously deployed DTA models

There are four primary DTA models that served as valuable knowledge bases for the development of the MADAM:

- Chicago, Illinois, USA (Chang & Ziliaskopoulos 2003)
- Dallas/Fort Worth, Texas, USA (Boyles et al 2006)
- Austin, Texas, USA (Duthie et al 2013)
- San Francisco, California, USA (PB 2012; Xyntarakis et al 2009)

A comprehensive review of each of these projects is beyond the scope of this work. However it is important to note that this project builds upon a foundation established by the previously deployed models, which experienced many of the same issues we encountered during the MADAM development. Nevertheless, the Sydney application is a large-scale effort and the first in Australia, and of course, the city has its own unique network structure that make a DTA model different from models in other cities. We believe our experiences offer a valuable addition to the existing community.

### 2.3 Study area

This project recounts the development of a DTA model for the area of Sydney, Australia. Sydney has a population of approximately 4.8 million. An average weekday features 16,670,000 trips as of 2013, of which 69% are by vehicle (BTS 2013). Traffic congestion is considered a significant problem, especially during peak hours. Sydney ranked 21st in the world in the TomTom 2014 Congestion Index, which was a drop from previous years (TomTom 2014). While traffic microsimulation models have been developed for various project-based areas around the region, this is the first attempt at a strategic level DTA model.
3. An Overview of Practical Considerations

This work describes the experiences of building a large-scale DTA model in Australia. Figure 3 illustrates the four major steps of the workflow in this project, as well as a brief description of the primary concerns for each step. The four steps included acquiring the data, processing the data, implementing the model, and finally, calibrating the model. Figure 3 represents an overview of the steps for the model development. The skills the team required to develop MADAM include: a thorough understanding of the DTA methodology, SQL scripting and Excel for large dataset processing, familiarity with Linux, ArcGIS, and Python for scripting with ArcGIS.

The remainder of this section describes the important aspects of “lessons learned” in terms of data, implementation of the model, and visualization techniques.

Figure 3. Summary of the project workflow

- Data acquisition:
  - Including: links (number of lanes, capacity, speed, location), nodes, zones, dynamic origin-destination trip matrix, signal timing plans, bus routes, bus stops, bus frequencies
- Data processing:
  - Generate dynamic data where necessary
  - Convert data from acquired formats to the format for DTA software
  - Quality control and sanity checks
- Model implementation:
  - Run model
  - Test possible measures to reduce computation time (if necessary)
  - Check relative gap
- Model calibration:
  - Acquire traffic counts
  - Compare results from model with network data using RMSE and GEH
  - Adjust model and repeat

3.1 Data

A significant portion of building a DTA model involves developing the most appropriate set of input data possible. This is a particularly challenging task for a large-scale DTA implementation, although there are still fewer requirements than for alternate traffic modeling approaches such as microsimulation. In the context of this project, some datasets were available, while other datasets had to be generated based on static planning data or other readily available data sources.

Knowledge of the area of interest is vital for quality control. A representative example in our experience was the impact of on-street parking on the number of available lanes on a link, which affects the capacity and ultimately route choice. This was an important consideration for the Sydney region. A number of major routes in the Sydney network are comprised of arterial roads that allow parking on the kerb-side lanes. However, some of these routes prohibit parking during the AM peak hours. After consulting with experienced traffic modelers in Sydney, an assumption was made that during the AM peak period, inbound corridors have 3 lanes available, while the outbound corridors generally have 2 lanes available.

For this application, the primary source of data was the Sydney Strategic Travel Model (STM3) [4] that was developed by the Bureau of Transport Statistics (BTS) and the Roads Network Model that was developed by the Roads and Maritime Services (RMS). Both of these models were built in the EMME2 platform and included approximately 80,000 links, covering an area from Newcastle about 160 km north of Sydney to Illawarra 100 km south of...
Deploying a dynamic traffic assignment model for the Sydney region

Sydney, and to the Blue Mountains region 75 km to the west. The main difference between the two models was reflected in their intended purposes. The STM is primarily a travel demand model and included more data regarding transit, disaggregated travel zones and data that impacted mode choice, while the Roads model is focused on the road network and had slightly more detailed data regarding network characteristics like link speed and capacity.

Considering the size of the dataset, comparing the two models proved challenging until the appropriate routines were developed using SQL databases and ArcGIS. The comparison showed that both models were nearly identical on the basis of locations of links and nodes. In effect, the network structure including nodes, connectors, links, and the capacities and free flow speeds of all links was obtained from the STM3 and in some locations the Roads model.

The static OD table was extracted from the STM3 for the AM peak. There are 2,722 travel zones representing both origins and destinations of the potential trips. Also, there are 7,409,284 OD pairs with nonnegative demand. The total demand of 1,665,440 equivalent passenger car units, represents the AM peak period adjusted to account for cars and trucks.

In order to capture the impact of a time-varying travel demand, the time-dependent departure profiles characterizing the car trips in selected time intervals was developed. Based on an extensive literature review, the team developed a methodology to utilise travel survey data to create a departure time profile specific to statistical sub-divisions (i.e., vehicles from outer regions may depart earlier than vehicles from inner regions). The chosen data set is a result of 5 waves (08/2009 to 12/2013) of the Household Travel Survey (HTS) performed by BTS. The length of each time interval was 15 minutes, which was considered to provide a good balance between having a sufficient sample size per interval and an appropriate length of time during which departure time choices could occur. Due to the limited number of samples some of the OD pairs were aggregated to the Statistical Subdivision (SSD). The dynamic demand matrix was obtained by applying the OD profile to the static demand obtained from the STM3 model. A representative example for two outer suburbs in the Sydney area is shown in Figure 4.

**Figure 4. Example of Departure Profiles From Parramatta and Hurstville to Sydney CBD**

To a greater degree than previously deployed DTA models, buses have a significant presence in the Sydney network, especially during AM peak and in the Sydney CBD. While...
the present study does not include transit assignment, it can include the presence of buses in the simulation procedure that determines network conditions. Therefore, 1,239 bus routes, approximately 28,000 bus stops (with an assumed dwell time of 25 seconds) were implemented in MADAM.

The data required for modeling transit for DTA included: i) the links comprising the bus route, ii) bus frequencies or bus schedules, iii) the location of stops, and iv) the dwell time at each stop. Two sources of transit data were explored: the General Transit Feed Specification (GTFS) data from New South Wales Transport Data Exchange Program, and the transit data that was included in the STM3 model.

The use of explicit signals data in DTA may allow for more realistic flow representation. The STM3 model stored green to cycle time ratio and the cycle times for each leg of a signalised intersection, which was originally based on historical traffic signal data. However, the signal data was intended to supplement link performance functions and was incomplete for the purposes of DTA. Therefore, the team developed a methodology to deduce the required signal timing based on the available data that is detailed in Chand et al (2015).

3.2 Implementation

After the data is acquired, processed, and transformed into the DTA format, the model needs to be implemented. Model implementation presents another set of challenges. One of the biggest questions for implementation is the computation time for very large scale models such as Sydney. Ultimately, the team decided to explore possibilities to decrease the size of the model and the corresponding run time.

Figure 5. The Sydney network in the MADAM project

Firstly, the external parts of the model outside of the Sydney Statistical Division were simplified, by aggregating the model areas representing Newcastle, Wollongong, and the Blue Mountains. The demand for these areas was aggregated and three new travel zones were created. Ultimately, this decreased the vehicle demand from 1.8M to 1.2M. However, it
should not impact the route choice in the city model because the demand that travelled in the city area was still included and there is essentially only a single route to access these locations. Figure 5 shows the model (not including centroid connectors) that was ultimately selected to represent the Sydney area.

Another significant source of computation time is the number of origin-destination pairs. The time-bottleneck in the DTA software is the time-dependent all-to-one shortest path algorithm for each destination. Therefore, a decrease in the number of destinations may have a significant impact of computational performance. However, too much aggregation will result in a loss of traffic interaction and internal trips that may result in a model that is too difficult to calibrate.

The demand data used for MADAM was acquired from the STM3, which as a travel demand model, required disaggregate zones in order to capture realistic walking distances to bus stops for a mode split, particularly in the CBD area. Additionally, the demand was based on a destination choice that doesn’t consider parking, which may not be realistic in a downtown area where parking is not readily available at a disaggregate destination. Finally, many OD pairs had a fractional demand, while DTA requires a discrete vehicle demand.

Travel zones were aggregated to improve computational performance and eliminate disaggregation error. While an automated aggregation process was devised, ultimately it needed to be adjusted manually through visual inspection to ensure that the aggregation was appropriate. Figure 6 illustrates the travel zones in the MADAM model (including aggregation). Ultimately, the MADAM employed 1,131 travel zones which resulted in 217,197 unique origin-destination pairs. With the reduced number of travel zones, the run time to evaluate a change to the model varied but was generally in the neighborhood of forty-eight hours.

Figure 6. Final zone representation in MADAM
3.3 Visualization

Ultimately, visualization was one of the most challenging but also valuable aspects of the MADAM application. Visualization allowed the team to identify discrepancies in the datasets. The visualization aspect of this project was achieved using ArcGIS.

Of the many lessons learned during the MADAM project, the value of visualization was one of the most important. Visual inspection was vital for comparing the sets of network data, processing the signals and the transit data. It was helpful for ensuring that zones were aggregated appropriately. It was essential for evaluating MADAM results and for the calibration process, where it allowed us to identify which routes may be overloaded with demand and to find nearby routes that may be under-loaded. Additionally, spatial analysis tools in ArcGIS allowed us to make connections between unrelated datasets and significantly reduced the manual workload.

However, links were represented distinctly for each direction but having the same nodes at the source and destination of each link. Therefore, in examining the map spatially, links were projected on top of each other, which could be deceptive when presenting results visually. The same applied to the calibration stations where we had two counts at pair of latitude/longitude coordinates. Therefore, the team developed a scheme to classify each link as “inbound” or “outbound”. The method consisted of choosing a point in the middle of the CBD and then using the “Near” tool in ArcGIS to calculate the distance between each node and that point. For each link, if the “from” node was further from the point, it was categorised as inbound. However, the Sydney network is neither perfectly radial nor grid-based. Therefore, this method was not perfect, but it does allow the visual results to distinguish between the links in each direction. In the remainder of this paper, “inbound” and “outbound” as terms imply the afore-described method.

4. Results and Calibration

This section describes the initial results and calibration process of the MADAM project.

4.1 Results

Ultimately, the DTA model included 42,628 links, 18,454 nodes, 1,131 travel zones, 14,919 centroid connectors, two hour AM peak demand of 1,262,930 vehicles, and 490 signals. The team first ran MADAM using a uniform departure profile, then added the dynamic departure time profile, signals, and transit data, in order to evaluate the changes of including different aspects of the model data. For the results at each stage of implementation, the team examined measures such as the relative gap, the total system travel time, the cost gap per vehicle, the average travel time on links for specified periods, and the volume of links for specified time periods. The relative gap remained between 5 – 15%.

Figure 7 illustrates two hour peak (7-9 AM) volumes estimated from the MADAM model, at a point that is midway into the calibration process. This means that various adjustments to the original link data have been made. Figure 7 shows a two hour estimated traffic volume, where the darker, thicker lines indicate a higher volume. The model displays a significant amount of congestion, and almost five hours of simulation time after the network loading period ends is necessary to exit all vehicles from the network.

Ultimately, examining the results from the MADAM application is an ongoing process that is closely tied to calibration, which we discuss in the next section.

4.2 Calibration

The DTA model can ultimately be used to predict future scenarios that cannot be evaluated on the basis of historical data. However, to ensure the model's predictive capabilities are as accurate as possible, the model output needs to be compared with real data such as link traffic counts or route travel time estimations. When discrepancies are identified, the model
Deploying a dynamic traffic assignment model for the Sydney region

needs to be adjusted so that it represents the reality as accurately as possible. This process is known as calibration. It is iterative and lasts until the similarity between the model and the reality is considered acceptable.

**Figure 7. Demonstration of model output (2 hour volumes)**

Thus, the calibration process requires real-life data to compare the model to. The primary calibration data is traffic counts, which can be more or less available depending on the location of interest. Other data such as turning movements, bus travel times, link speeds, or corridor travel times can also be used. The MADAM application was able to obtain data that provided a thorough coverage of the entire Sydney network. RMS has provided 165 traffic count locations for the average AM peak period as shown in Figure 8 (resulting in 322 points of comparison). The calibration consisted of comparing these counts with the output from MADAM.

The calibration process is arguably one of the most important parts of modeling and it requires skill and knowledge to execute. Ideally a model requires little calibration or a methodological calibration that can be repeated. However, the calibration for a traffic assignment model is not straightforward, even in the static case. The process utilized for the MADAM primarily involves making adjustments in the network to link capacity and link speed in order to impact the route choice mechanism and reduce or increase vehicle flows in the appropriate location, as well as adjusting data discrepancies such as the number of lanes on a link. Obviously, determining the appropriate locations is the challenge. In some cases the travel demand matrix can also be adjusted.

For the MADAM application, the calibration metrics included visual inspection, the absolute and relative difference between the count and model output, root mean square error (RMSE), and the GEH statistic. RMSE measures the differences between the values observed in the real-world and the values predicted by a model. The GEH statistic is commonly used in traffic modeling and it captures the relative nature of a count, i.e., if the difference between a predicted volume and the real count is 300 vehicles, then the implications change if the real
count is 400 or 4000. Calibration is an ongoing and challenging process, but the current outlook is detailed in the remainder of this section.

**Figure 8. Locations of calibration data counts**

As one aspect of model calibration, the team examined the locations in the network with the greatest amount of hourly traffic flow. Figure 10 illustrates the locations in the Sydney network where the measured two hour traffic volume from 7-9 AM was greater than 7,000 vehicles (highlighted by the orange points) and the places in the network where MADAM predicted a flow greater than 7,000 vehicles over the 2 hour period. The road network is represented by the light grey lines, while the small black points show all of the calibration points. The red lines represent “inbound” links with a volume greater than 7,000 vehicles, while the blue lines represent “outbound” links with a volume greater than 7,000 vehicles. The orange dot in the northeast is a misrepresentation of the Anzac Bridge.

The RMS survey locations with heavy traffic volume during the AM peak period include the M5 near the Hume Highway, General Holmes Drive near the airport, Southern Cross Drive, the Sydney Harbour Bridge, the Anzac Bridge, the M4 Motorway in Homebush, and James Ruse Drive in North Parramatta. The areas in the MADAM network with the greatest volume is along the M4 Motorway.
Next we present an overview of the current GEH statistics. In traffic modelling, a GEH less than 10 is generally considered acceptable, although that applies to a much smaller scale microsimulation model, not a strategic dynamic model. Therefore we present the GEH statistic in three categories, shown in Figure 10. The yellow points represent a GEH less than 10, while the orange points represent a GEH between 10 – 25, and the red points represent points with a GEH greater than 25 (the maximum GEH at the current stage is 72). These calibration locations are presented on top of the model's predicted volume-to-capacity ratio for each link, where grey represents a v-c ratio less than 0.3, green represents a v-c ratio between 0.3-0.6, and the red represents a v-c ratio greater than 0.6. In dynamic models, the v-c ratio cannot be greater than 1 (unlike static models). Of the 161 counts, 64 have a GEH less than 10, 59 have a GEH between 10 – 25, and 38 have a GEH greater than 25. Note that these are not considered the final calibration results.

Finally, Figure 11 highlights the same information as the previous figure but for the “outbound” links. Again, the yellow points represent a GEH less than 10, orange represents a GEH between 10 and 25, and red shows a GEH greater than 25, and the links are represented on the basis of their volume-to-capacity ratio. Currently, there are 57 points less than 10, 65 points between 10 – 25, and 39 greater than 25, out of the 161 total. Figure 11 shows a greater number of red links, indicating a higher v-c ratio, than Figure 10. This implies that traffic flow in the AM peak in Sydney is not necessarily destined for the city centre.
Figure 10. GEH statistic on “inbound” links

Figure 11. GEH statistic on “outbound” links
Calibration is an ongoing process that is constrained by the computational capabilities of the team’s hardware. In many locations, the team has found that model output can be improved by examining the consistency between real world data and model data in the areas of interest. Sydney is a corridor based network, and therefore in many of the locations where the model output is much higher than the real world traffic count, there is a location on a nearby corridor where the count is lower than the real traffic count. Thus, the calibration process is a matter of examining the locations of the calibration stations both in isolation and the network context.

5. Conclusion and recommendations

This paper describes the experiences of building the first large scale DTA model in Australia, applied to the Sydney metropolitan area. The project gathered and synthesized numerous data sources including the Sydney Strategic Travel Model (STM3), the Roads Network Model, the household travel survey, the Sydney GTFS data, Sydney SCATS signals data, and traffic count data from permanent stations acquired from the RMS journey information division. The team implemented the model and devised various techniques to address computation time. Currently, the run time is about 48 hours to evaluate updates in the model.

A calibrated DTA model presents numerous opportunities for future extension, particularly in regard to applications such as environmental impact evaluations. Of course, traffic assignment serves as an important component of a four-step transport planning model, and it would be interesting to incorporate the travel demand aspects to test if predictions change compared to the static case. More detailed transit data or even transit assignment could be included. Measures to address the computational challenges will also be useful. Finally, in order to evaluate the effects of reliability, the team intends to extend the deterministic DTA model to account for volatility in day-to-day traffic flows.

Acknowledgments

The authors gratefully acknowledge project funding from Transport for New South Wales. In addition, significant assistance and support in the form of time and data came from the Bureau of Transport Statistics (BTS) and the Roads and Maritime Services (RMS). In particular, the authors would like to thank Chris Zito from RMS, Malcolm Bradley and Matthew Jones from BTS, and James Sloan from UNSW.
REFERENCES


Parsons Brinckerhoff (PB), 2012, San Francisco Dynamic Traffic Assignment Project “DTA Anyway”: Final Methodology Report, San Francisco County Transportation Authority


Xyntarakis, M, Sall, E, Hicks, J, Charlton, B, 2009, Regional Dynamic Traffic Assignment for Real World Travel Demand Models – Trying it out in San Francisco, San Francisco County Transportation Authority
