

A new approach to transport modelling - the Stochastic Segmented Slice Simulation (4S) model and its recent applications

Peter Davidson¹

¹TransPosition, Level 5 199 George St, Brisbane Q4000

Email for correspondence: peter@transposition.com.au

Abstract

Recently there has been renewed interest in alternative modelling approaches, driven by recognised deficiencies in the traditional four step model. One stream of development has come through improved modelling of individual and household travel scheduling and joint decision making, under the umbrella of tour-based and activity-based modelling. Another stream of development has been improved approaches to discrete choice modelling, and the incorporation of more complex Random Utility Model (RUM) formulations, such as Mixed Logit.

In this paper we present a new multi-modal transport model structure which incorporates many of the elements of these recent developments, but within a formulation that overcomes many of the obstacles associated with moving away from traditional models. In particular, we show that the new 4S model is very computationally efficient, and is extremely flexible and extensible, while having similar data requirements to traditional models.

The model uses Monte-Carlo simulation with a unified trip generation, destination, mode, route and time-of-day choice. It allows for detailed behavioural modelling including; cycling end of trip facilities, multi-modal freight, road pricing and public transport congestion.

We describe the structure of the model and discuss the theory behind it. We then give examples of the application of the model to recent projects in South East Queensland and demonstrate the new types of analysis that can be done with the model. We also demonstrate the efficiency of the model by describing its practical application to the whole of Queensland at the lot-level, and the whole of Australia at the collector-road level.

1. Introduction

Current transport modelling practice is dominated by the traditional Four-Step-Model (FSM), first developed in the 1950's. Recently a number of alternative approaches have been adopted to improve on the FSM, incorporating tour-based and activity-based insights. This paper will review the problems with the traditional models, and the strengths and weaknesses of the newer alternatives. It will then go on to develop a novel approach to modelling that incorporates many of the best features of what has come before. The new Stochastic Slice Simulation, is based on Random Utility Theory and can be formulated in trip-based, tour-based and activity-based approaches, but allows more flexibility in the specifications of utility, particularly transport costs. It is also very efficient, allowing large networks and complex behaviours to be modelled without making running times too onerous. The paper describes the fundamentals of the model, and presents some initial applications to a range of issues including toll analysis, active transport demand, accessibility, and equity analysis. The current work on extending the model is also described.

2. The traditional approach – four step modelling

2.1 Background

The four-step modelling approach was first developed in the 1950's, notably in the Detroit Transport Study in 1953 and the Chicago Area Transport Study in 1955 (Weiner 1992). The traditional four step model (FSM) has the following sequence.

- Trip Generation
- Trip Distribution
- Model Choice
- Route Assignment

Sometimes the model is extended with a Time-Period Choice step, and the model is almost always preceded by land use projections and followed with evaluation processes, typically giving a 6 or 7 step model. Many papers and books described the FSM, including (McNally 2000b; Ortuzar & Willumsen 2011).

This approach was initially developed as a fairly practical way of planning for highways, without much of a theoretical basis. Two alternative theoretical foundations to the model were retrospectively applied in the 1970's – Entropy Maximisation and Random Utility Theory. The second of these approaches has been more widely used, and is the foundation to almost all mode-choice modelling, usually in the form of the logit curve. With the addition of equilibrium assignment in the 1970's, the basic approach has remained almost unchanged, although vastly improved computing resources have allowed later models to have more detailed zoning systems and more highly segmented behaviour.

There have long been criticisms of the four-step-modelling approach the nature of which has changed somewhat over the history of the profession, as planners have changed their emphasis.

An interesting perspective on the problems associated with integrated transport land use models can be found in (Lee 1973), who identifies the seven sins of large scale models as:

- Hyper-comprehensiveness (too broad a scope and too many variables)
- Hungriness (need for too much data)
- Grossness (too much aggregation)
- Wrong-headedness (ill-conceived ideas, invalid theory)
- Complicatedness (black box, with too many hidden constraints needed to get sensible outputs)
- Mechanicalness (affected by inaccurate calculations and indeterminate iterative solutions)
- Expensiveness (cost too much)

These, and other, criticisms may be grouped into three categories of problems – theoretical, operational and functional.

2.2 Theoretical problems

2.2.1 Trip based

One of the most common recent criticisms of the four step model is the use of trips as the primary unit of analysis. (McNally 2000a) summarises a number of weaknesses of trip based models which:

- Ignore the fact that travel is a derived demand
- Focus on individual trips and ignore the interaction of trips in time and space
- Do not consider the complex constraints on travel and activity choice

- Neglect the consideration of interactions between household members in their choice and scheduling of activities
- Overemphasise utility maximisation, without considering alternative decision strategies, relationship dynamics, habitual behaviour, information and choice complexity

He argues that the key problem is that the model does not link travel to activities, and the constraints, dependencies and motivations for these activities.

These limitations form part of the motivation for the move to tour based and activity based models.

2.2.2 Aggregate (Grossness)

The four step modelling approach is characterised by aggregate behaviour (even though some components of the model are calibrated using disaggregate approaches) and largely sequential operation. These characteristics ensure that the models generally remain fairly tractable, with moderate data requirements and computing power. Each step of the FSM aggregates all of the trips coming to and from an area into a zone, and looks at travel between zone centroids through skims and matrices. These inevitably lead to loss of spatial detail and mean that, at the local level, the model is very strongly influenced by the choice of centroid location, and the specification of centroid connectors. Furthermore, the aggregate nature of traffic zones typically leads to the use of aggregate networks – local roads are abstracted away into centroid connectors and only higher order roads are included.

Increasingly there is a desire by modellers to increase the number of zones, particularly when looking at active transport or public transport (which usually includes an active transport component). But since the FSM works so heavily with matrices, the computational cost of increasing zone numbers is high – run time and storage requirements often increase with the square of the number of zones.

2.2.3 Not based on realism in process

Ideas expressed by (Cowing & McFadden 1984) and restated by (Hensher et al. 1992) look at how we should assess models. When models are being used to prepare forecasts that go a long way into the future or that have substantial changes to current conditions, their assessment should focus on “realism in process”. This means that the behavioural sensitivities in the model should have the right form, and the models should be based on a realistic representation of the underlying motivations for travel. The more usual assessment of models effectiveness is “realism in performance”, which focuses on comparing the models outputs with empirically observed data.

The FSM emerged as an operational solution, without an underlying theoretical framework. Later work reinterpreted some of the steps of the FSM within a theoretical framework, such as recasting the gravity model into a utility based discrete choice model. But the idea of separate sequential steps for trip generation, trip distribution, mode choice and assignment does not reflect the reality of how people make their travel decisions (Ben-Akiva et al. 1998).

2.3 Operational problems

2.3.1 Data needs (hungriness)

The traditional FSM generally requires substantial data, both in model inputs and in data required for calibration. A large part of the requirement for large-scale household travel surveys comes from the needs of modellers to calibrate their models. Furthermore, the aggregate and indirect nature of many of the model parameters means that they cannot be easily estimated in the absence of data, and they are often not directly transferrable from one region to another.

The FSM traditionally requires another set of reasonably expensive data inputs – the transport networks themselves. Prior to the widespread use of Geographical Information Systems (GIS), transport networks were manually coded, with hand drawn maps and individually applied a-nodes and b-nodes. Now that good quality spatial representations of the road network are widespread, many model networks are initially populated with data directly from the GIS. However the model still requires a number of modeller-created elements, including zoning systems and zone centroid connectors, as well as individually determined free-flow speeds and capacities. Often the model calibration process is focused on modifying individual speeds and capacities to give a better fit to observed traffic counts. Very quickly the transport network becomes a separate entity from the source GIS data from which it was created, and is not able to make use of the frequent updates that are usually available for the source GIS data.

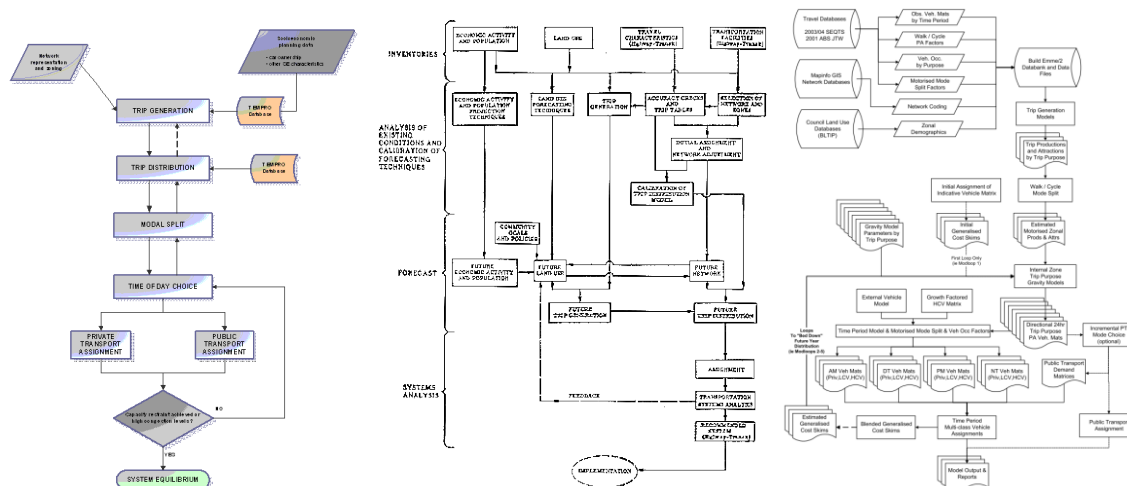
It should be noted that the data needs of the FSM are significantly less than those required for newer activity models. As well as requiring much more information on people's activities throughout the day rather than just their trips, activity models usually contain a significant synthetic component. The synthesis of populations and activity schedules require more comprehensive and more detailed data.

2.3.2 Difficult to extend

The FSM has a relatively simple structure – the eponymous four steps. However it quickly becomes more complex. Should distribution happen before mode choice? Where does the time of day come in? Where should the toll model be applied? How do we incorporate induced demand? Since the four step model is a series of discrete steps, the obvious way of extending the model is to add more steps. But since the feedback between different elements is important, each new step is often associated with a new feedback loop. The model quickly becomes byzantine in structure.

As an example, here are a couple of selected model structure diagrams for four step models that show how the addition of new steps can lead to quite complex structures.

Figure 1: Four step model structures



Sources: (WebTAG 2007, p.3.11.1; Weiner 1992; Brisbane City Council et al. 2006)

In a sense it is not that the model is difficult to extend – any additional step can be bolted onto the side. But it is difficult to extend coherently, in a way that leads to understandable, predictable result. The bolt-on and iterate approach used to extend the FSM leads to models that are difficult to understand, difficult to calibrate and difficult to be sure of the outputs.

2.3.3 Inefficient and long running times (Expensiveness)

The key computational element in the FSM is the matrix. Costs are calculated from the network and stored in zone-zone skim matrices, probabilities are calculated using matrix operations and travel demand is stored in trip matrices by purpose and mode. Almost without exception, the matrices are not sparse, but dense – that is every cell has a value. In order for the distribution model or the mode choice model to work, each cell of each skim matrix must have a value. Given the tendency to increase the behavioural resolution of models by adding more purposes, more time periods and more modes this means that the model needs to calculate values for a huge number of cells. The inefficiency comes from the fact that most of these cells contribute very little to the final result.

Consider, for example, a regional model for South East Queensland. In order to calculate the number of trips that travel from a residence in Noosa to the local primary school, the model must enumerate the skim matrix for all possible schools that could be reached from that residence. That means that the model will find the shortest path to all other zones, including those three hours away at the Coolangatta. These other destinations will contribute almost nothing to the destination choice – the trips will be very close to zero. But since it takes more time to build paths for a long distance than a short one, it is clear that the model spends most of its time finding paths that will not be chosen. This situation is compounded when multiple modes and time periods are added, and even further exacerbated when the number of zones increases and when more iterations are added to the model (see the next section).

This inefficiency is helped somewhat by the exponential increases in computer speeds, but the desire to better model public transport, behaviour change, pricing and freight has led to more complex models and generally increased run times. Some Australian capital city models take over 24 hrs to run, and some models for bigger cities are reported to take a number of days (according to (Chiao 2004), the New York Best Practice Model, a tour-based model, took over 78 hours to run, with 50 hours just for the transit and highway assignment).

2.3.4 Nested iterations

In the FSM travel is modelled as a series of discrete steps. But because there are many feed-back processes within the travel system many of the steps cannot be processed just once. The usual approach to this is to perform iterative feedback loops that will (hopefully) lead to convergence. The classic FSM has two such loops – an inner one in the assignment step and an outer one used to feed congested road speeds into the distribution model. Other new steps usually add new convergence loops – for example, iterating the toll choice model to get convergence between free routes and tolled routes, or iterating a public transport congestion step through the mode choice model to reflect the impact of PT capacity on PT mode share.

2.4 Functional problems

2.4.1 Lack of Policy Sensitivity

One of the most damning criticisms of the FSM, and one of the key motivations for developing new models is the lack of policy sensitivity in traditional models.

The FSM was developed primarily to assist in the planning and expansion of the road network, particularly in the development of new highways. Many have criticised traditional models for encouraging the “predict and provide” approach to planning, where models are used to identify future traffic congestion and to propose new roads to (purportedly) mitigate the congestion (Evans et al. 2007).

The current issues facing transport planners are very different from those facing planners in the ‘60s and ‘70s when the FSM was developed. Solutions are much more likely to include

public and active transport, demand management and behaviour change, and pricing, and none of these are dealt with very well in the FSM.

This paper cannot address all of these in detail, and there may be some disagreement about the degree to which these issues can be addressed, but the following list shows policy questions that are problematic for traditional models.

- Active transport planning
- Air quality and noise
- Behaviour change and travel demand management
- Greenhouse gases, climate change and carbon pricing
- Induced demand
- Land use/transport interaction
- Peak oil and energy security
- Peak spreading, flexible working hours
- Public transport congestion
- Risk analysis and network security
- Socio-economic linkages, and equity impacts
- Toll analysis and road pricing

3. A new alternative – Segmented Stochastic Slice Simulation

Over the last few years, the author has conceived a new alternative approach to transport modelling, and his company, TransPosition, has developed software to implement the model, applying it to a range of real world problems. The Segmented Stochastic Slice Simulation (4S) model is named for the following features:

- **Segmented:** Uses a comprehensive breakdown of different travel markets, and allows all behavioural parameters to vary by market segment (value of time, tolls, destination utilities etc)
- **Stochastic:** Uses Monte Carlo methods to draw values from probability distributions. Every parameter can be a random variable
- **Slice:** Takes very efficient slices of the travel market across the whole model area and through the distributions
- **Simulation:** Uses state-machine with very flexible transition rules to effectively simulate all aspects of travel choice

It differs in many ways from the traditional FSM, and has many compelling advantages over many of the newer models as well.

- It has an elegant, theoretically sound basis that allows for realistic modelling of a very wide range of issues. This includes active transport, toll modelling, behaviour change, induced demand and time-of-day analysis.
- Models can be prepared with much less effort – by eliminating zones, centroids, and centroid connectors the manual effort in putting networks together is vastly reduced
- It is very computationally efficient – by focusing all of the computational effort on tasks that are likely to contribute to the final outcome, and by having a single iterative structure (rather than a whole range separate iterations for convergence in traditional models) complex models can be run with practical run times. As an example of this, TransPosition have applied models of the whole of Queensland at the lot/local street level, and the whole of Australia at the Collection District (CD) and collector road level.
- Its simple core allows it to be extended to include time choice models, tour-based models, activity models, links to micro-simulation, latent class models and land-use/transport interaction.

This paper can only provide a brief overview of the model and leaves many details untouched, in particular the processes for parameter estimation. It also has little space to present the applications and outputs of the model. These will be addressed in future papers.

3.1 Theoretical basis

At heart the model adopts the same Random Utility Maximization (RUM) theory (McFadden 1974) that underlies most discrete choice models. It assumes that when choosing between alternatives, each individual makes an assessment of the utility of each option and chooses the one that will yield the greatest utility. All of the utility values are random variables (due to variation in people's behaviour, modellers' ignorance and changing circumstance) so preferred choice will vary, leading to choice probabilities.

The standard formulation of RUM divides the utility into a systematic part and a random part. The 4S Model does not make this distinction, but instead builds the utility measures out of random variables.

So for the 4S model,

$$U_{in} = \beta_n x_{in} \quad (1)$$

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn}, \forall j \in S_n, j \neq i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Where U_{in} gives the (random) utility of alternative i as perceived by individual n ; x_{in} is a vector of variables that give the characteristics of alternative i as evaluated by individual n ; β_n is a vector of random taste coefficients for individual n and S_n is the set of alternatives available to individual n . The coefficients are random variables that vary over individual decisions makers and are distributed with a density function $f(\beta)$. The density is described by a distribution (such as normal, uniform, triangular, gamma, log-normal) and parameters (such as mean and standard deviation) of β 's across the market segment under consideration.

Train (2003) shows that the random coefficients model (as described here) is equivalent to the error components model which is the basis for the mixed logit model which in turn is a generalised form of the standard logit model. In the standard logit model the taste coefficients are fixed and an iid extreme valued error term is added. In the 4S approach the extreme valued error term is not necessary, but could be added as an extra β applied to a fixed characteristic variable x .

The model works with market segments, which defines the random variables from which the utilities for each individual are drawn. In order to work as a travel choice model, the assumption is made that the utility of a particular travel choice is composed of two parts – the utility of the attractor and the disutilities of travelling.

$$U_{a,m,r,n} = U_{a,n} - \beta_{a,n} C_{a,m,r,n} \quad (3)$$

where $U_{a,n}$ is the intrinsic utility of the attractor a to the individual n , and $C_{a,m,r,n}$ is the vector of cost components of travelling to attractor a by mode m on route r .

Thus the trips between a particular production location p and an attractor i , by a particular mode m and route r for a market segment s is dependent on the size of the market segment at the production $S_{p,s}$ and the probability that the attractor, mode, route combination is optimal

$$T_{p,a,m,r,s} = S_{p,s} p(U_{i,m,r,s} > U_{j,m,r,s}, \forall j \in A_s, j \neq i) \quad (4)$$

Where A_s is the set of all possible attractors for market segment s .

Now it would seem that producing an estimate of p in Equation (4) would require the enumeration of every attraction, mode and route combination for each production area. In fact the traditional combined model attempts to do this, but by reducing the problem complexity using the following simplifications

- Aggregate production areas and attraction areas into zones
- Aggregate modes into simple sets (e.g. combine bus, rail and their access modes into a single mode PT)
- Consider only the shortest route in each mode
- Simplify the utility formulation (through assumptions such as independent and identically distributed IID error terms), to give an algebraic solution, such as multinomial logit

If all of these simplifications are adopted, the resulting model would look very similar to the traditional four step model with nested logit for route choice, mode choice and destination choice. Use of a mixed logit model would allow more complex specification of the utility structure, but would typically still require the full enumeration of costs for all modes, routes and attractors.

The full enumeration of all options is the sticking point – for the traditional discrete choice model, probabilities can only be established by calculating the non-random costs for all alternatives, leading to traditional cost skim matrices by mode. The full enumeration is infeasible for routes, so traditional models generally work with deterministic route costs and calculate a single shortest path (for each mode and origin-destination pair).

Many of the newer discrete choice models (such as mixed logit) recognise that closed form solutions to the maximised utility probability are impossible without severe constraints on the utility specification, so incorporate a stochastic sampling procedure to estimate the probabilities. The 4S Model embraces this stochastic sampling, and does away with full enumeration of alternatives.

Nonetheless, the problem space inherent in Equation (4) seems too large to effectively sample. Fortunately it can be conceptually simplified without loss of generality. First, the distinction between mode and route is artificial. In reality mode is an abstraction, and travellers simply choose a route. There are still likely to be mode-related preferences that influence the correlation between the disutilities of different routes, but since the model makes no assumptions about the form of the random variables making up the disutilities this can be easily handled.

The second simplification comes by recognising that the attractor can also be considered part of the route – a travel decision can be seen as a route to fulfilment of a particular desire, and that route can pass through a range of possible destinations, each one giving a different boost to utility.

So the choice problem becomes to find the route through the full choice space (which incorporates destination and mode) that maximises net utility.

3.2 Maximum utility path

Fortunately this problem can be solved efficiently, given a Monte Carlo draw for each of the random variables. The solution is a modified version of the standard shortest path algorithm, but rather than minimising costs the algorithm must find the maximum utility path.

The problem may be formulated as follows. Consider a standard transport network with links and nodes, with disutilities (negative valued costs) on each link – at this stage include only roads; the generalisation to include other modes is examined below. Some (or all) of the nodes represent productions – these are the locations where the individuals making the

travel choice are located. There are also a number of locations in the network that connect to attractors for the market segment that is being examined. Connect each of these attractors to a single meta-attractor, using a shadow link with a positive utility. A modified Dijkstra algorithm (Dijkstra 1959) can then be used to build paths from the single meta-attractor back to all production nodes. The algorithm must always start at the attractor, and build back towards the production nodes, and the greedy algorithm must always choose the next option that maximises net utility.

The process of the algorithm as it finds solutions may be visualised as follows. The algorithm will find the attractor with the highest utility and start building paths out from that location. At each point the net utility of the nodes as it builds out will be the highest possible net utility for that location, and the attractor link that the path builds from will be the best choice for that location. At some point the net utility around the highest valued attractor will be reduced by the network links costs, and the next most valued attractor will emerge, and paths will be built out from both attractors. As the algorithm proceeds, some attractors will be chosen, but the locations to which they connect have already been visited, indicating that those locations have a better attractor elsewhere. Finally all production nodes will have been visited and the algorithm can stop. Each node in the network will have a net utility value, and its path history will lead back along its route to the preferred attractor for that location.

A consideration of the process of the maximum utility algorithm shows the efficiency gains over the traditional shortest path algorithm used for full enumeration of skim matrices in traditional models. A traditional implementation of the shortest path process will build paths from single origin to all destinations. To find the optimal destination the algorithm must be repeated for every possible destination, and a full skim matrix produced. But much of this work is wasted, since time is spent building paths to destinations that are not optimal. In fact, because so many trips are quite short, the optimal destination is quite close, and the majority of the time in the algorithm is spent finding routes to unnecessary, not-optimal destinations.

In contrast, the modified algorithm used in the 4S model builds paths from all production nodes to their best attractor in a single pass.

This process can also be contrasted to the algorithms that build from a single origin to a single destination (such as the A* algorithm) that might be used in an individual simulation model, such as TRANSIMS. These single path algorithms use heuristics to prune the search tree to minimise the exploration of the network that is unlikely to contain the shortest path. Nonetheless, they still must examine a significant portion of the full network to find the path for a single traveller. The 4S model examines the whole network, but efficiently finds paths for every producer to their optimal attractor in a single step. In this way it can be considered to be a slice across the whole network – each pass of the algorithm finds the optimal route and attractor for a slice of the market at that producer. As discussed below, each pass can also be seen as taking a slice from the probability space given by the probability distributions of the random variables that describe utility and cost.

3.3 Productions/attractions, origins/destinations and trips/tours

Models traditionally make a distinction between the origin and destination of a trip, and the production and attraction of that trip. That distinction is particularly important in the 4S model. As shown in Equation (3), the model has the decision maker (or market segment) at the production end and the location choice utility at the attraction end. The maximum utility path build works requires that at every step the top of the priority queue of elements to visit must have the highest net utility. Since the positive utility comes from the attractor itself, the path building process must always start with the attractor and build back to the producer.

This paper focuses on the use of the 4S framework within a trip-based model. But the same basic approach can be used within a tour-based 4S model, or within an activity-based 4S model. In the trip based model, the forward leg of the trip must be solved separately from the reverse leg of the trip to allow for asymmetries in the transport network. When this is done,

the production to attraction trip must be built from the attractor back to the producer using a back-in-time, reverse link path build. The attractor to production trip can be built on the forward network.

3.4 Description of the central model process

The core process of the model is the utility maximising slice, where in a single step the optimal attractor is found for each production location, and in the process the route to that attractor is found and the net utility calculated. The overall sequence of the model is given in the following pseudo-code.

For each market segment s

Determine the number of slices (n) to take for s

For each slice

Randomly draw values for each behavioural parameter from the random variables for s

Randomly draw values for each attractor in A_s

Seed the maximum utility path build algorithm with the utility values for each attractor

While there are still more production nodes to visit

Find the next highest net utility production node p

Store the net utility (and other path variables) for p

Find the size of market segment s at node p ($S_{p,s}$)

Calculate the number of trips in this slice = $r \frac{S_{p,s}}{n}$

Allocate these trips to the path to the optimal attractor

The final result of this process is aggregate path information at each production node (that can be used to give accessibility, catchment areas etc) as well as volumes on links.

TransPosition's implementation of the 4S model actually allows any path variable to be computed and stored for production nodes, attraction nodes, links and turns. This allows for generalised accessibility analysis, select path analysis, analysis of externalities etc.

3.5 Treatment of time

Each slice of the 4S model is based on finding the optimal route and attractor for each production node, and it may be seen as a simulation of travel through the network for that slice. Thus it is possible to incorporate other aspects of individual simulation, such as the accurate treatment of time.

This can be done quite easily by including a random distribution of preferred arrival/departure time and then keeping track of time throughout the path build. For the trip based model, separate paths are built for production to attraction trips and for the reverse, but all paths are built out from the attractor. Thus the seeding of the simulation time must always be given from the perspective of the attractor. The consequence of this is that the forward P-A trips are specified in terms of preferred arrival time, and the reverse A-P trips are specified in terms of departure time. Thus the model assumes that the time constraints are always specified at the attraction end; work travellers are assumed to have to arrive at work at a particular time and then leave work at a particular time. The times at the production end (leaving for work in the morning, or returning from work in the afternoon) emerge from the model and depend on the time spent travelling through the network.

The treatment of time in this way allows the model to do a number of things much more effectively than traditional models. In particular, it can easily incorporate time-based network elements such as

- Scheduled public transport
- Timetabled vehicular ferries
- Time dependent tolls and parking costs
- Peak hour clearways
- Off peak on-road parking
- Different availability times for attractors (e.g. shopping centre opening hours)

It also allows the output of results by arbitrary time periods. Rather than having to separately apply a peak hour model and an off-peak model, the 4S model can seamlessly investigate travel across the whole day. Depending on the number of slices taken, the model can give outputs broken down by any time period, giving, for example, hourly traffic volumes throughout the day. Furthermore, by including congestion (as discussed below) the model can incorporate time-based congestion variation.

Note that the model can keep track of the different types of time, and apply different weightings to each. Thus it is possible to apply different weights for the scheduled delay for the first public transport stop, the unscheduled variation in public transport (due to non-timetable compliance), the waiting time involved in transferring between public transport services, as well as different weights for different travel modes.

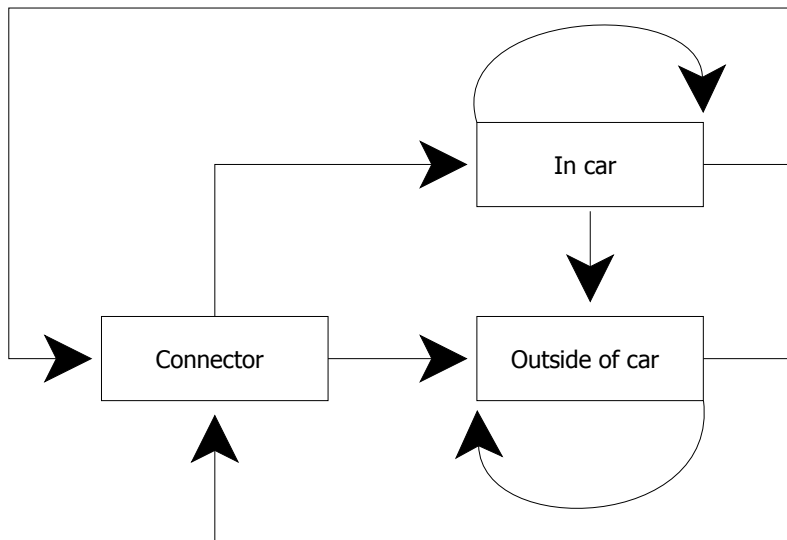
3.6 Incorporation of multiple modes

The previous discussion has assumed that the maximum utility path building process was working with a single mode. To generalise it to multiple modes the network just needs to be made more complete, and keep track of the state of travellers throughout the path build. The process is similar to the Label Constrained Shortest Path described by (Barrett et al. 2007). The model defines a series of states (labels) and permissible state transitions (grammar). It also defines a series of modes, and each mode has a list of permissible states. Finally, the network contains links, and each link defines the allowable modes (and indirectly the states) that are permissible on that link.

So, for example, to allow for the simple case of public transport than can be walked to or driven to, but not driven from, we could define the following states. The initial state for the algorithm is **Connector**, which is the state that applies on an automatically created virtual link that connects all production or attraction nodes to then network. The path is not allowed to stay in this state as there is no transition that points back to the **Connector** state. This is to prevent fully internal trips that start and end at the same location. From the connector state, the path can go into the **In Car** state and use car links, or go into the **Outside of Car** state and use active transport or public transport. It is possible to go from the **In Car** state to the **Outside of Car** state, since drivers can park and move onto public transport, or park at a car park and walk. By correct specification of the network these transitions can be limited to park and ride facilities or car parks. But there is no possibility to move from **Outside of Car** to **In Car** so people cannot hop off public transport and drive away. Note that this set of transitions is for the Production to Attraction direction; it is possible to have a different set of transitions for the Attraction to Production that does not allow the use of the car before being in public transport but does allow travellers to drive away from public transport back to their homes.

This set of states is quite simple but can easily be extended to incorporate almost any rules for mode use, including allowing bicycles on some public transport services, ride sharing, vehicular ferries and end of trip facilities for cyclists (the state system can be used to track whether travellers are sweaty or whether they have showered).

Figure 2: Example model states



The power of this system is that it does not require artificial distinctions between modes – paths can use any combination of modes that make sense, and all link volumes associated with travel will automatically be included. This means that the driving component of public transport trips is automatically included in traffic volumes (and impacted by traffic congestion) as are walk trips between different public transport services, and walking trips from car parks to destinations.

The other aspect of public transport – scheduled services, is addressed in Section 3.5. By separating out the concept of mode from the concept of scheduled services it is possible to include on-demand public transport, as well as scheduled road links, such as vehicular ferries.

3.7 Congestion

Traditional models treat traffic assignment as a separate process from the preparation of demand matrices – assignment is typically done with fixed matrices. Furthermore, most assignment is done with deterministic costs. There are algorithms for stochastic user equilibrium assignment (Burrell 1968; Dial 1971; Daganzo & Sheffi 1977) but these are not widely used in practice. One of the problems with traditional stochastic assignment techniques is that they require a closed form solution to the discrete choice problem, in part because the set of alternatives for route choice is so large. This means that traditional stochastic assignment assumes that the random error term for discrete route choice is Gumbel distributed, and that the variations in costs on each route are iid (independently, identically distributed). In reality this is unrealistic, as many routes will be almost identical, differing by only one or two links, with strongly correlated costs. It should be noted that there are alternative formulations that make a more realistic assumption, such as random-coefficient mixed probit models, but again these are rarely used in practice.

The key element that all equilibrium processes have in common is some iterative loading of demands onto links, and some updating of cost based on the loaded volumes. These range from simple incremental loading, through to the very popular Frank-Wolfe algorithm, and other uses of the Method of Successive Averages (MSA) and the Method of Successive Weighted Averages (MSWA) (Liu et al. 2009). The key difference between these approaches is their efficiency at converging on the equilibrium solution.

Within the 4S model, the efficiency of convergence does not matter as much as it does in traditional models. The reason that convergence efficiency matters so much for traditional models is that assignment is a totally separate step from demand estimation, and extra

iterations for convergence just add computational cost to the model. In contrast, due to its integrated choice framework, the 4S model can assess route choice issues (including congestion) at the same time as decisions about whether to travel, where to go and which mode to use. This means that extra iterations to converge congestion are also useful in providing additional sampling of the probability distributions and improve the robustness of the model output.

The incorporation of congestion into the algorithm given above is quite simple – simply assign trips to paths as each slice is simulated, and calculate updated costs after each slice based on the aggregate volumes of all previous slices. There does need to be a randomising of the ordering of the slices (the pseudo-code given above had each market segment analysed in turn) but the overall process is very similar.

The simplest implementation, and the one currently used by TransPosition, is to use an incremental assignment – volumes are accumulated with each slice, and congested costs calculated before the path building of the next slice. It would be possible to use more complex aggregation processes (such as one of the MSA or MSWA equilibrium rules), but with the large number of slices needed to reduce sampling error, it is unlikely to significantly improve the equilibrium outputs (although (Nielsen 2005) has a different perspective).

Because the 4S model has an integrated choice framework this congestion approach can be used to model not only road congestion, but also public transport capacity, parking supply, destination capacity/level of service, and attraction constraints (that are usually managed through doubly-constrained distribution models).

4. Key benefits of stochastic slice simulation

4.1 Comparison with the Four Step Model

Compared with the traditional FSM, the 4S model offers a variety of benefits.

4.1.1 Appropriateness

The most significant is that it has the policy sensitivity to address the issues that currently face transport planners. Every issue in 2.4.1 that the FSM struggles to address can be modelled effectively in the 4S model. This flexibility comes from the integrated nature of the choice framework within the 4S, the unconstrained specification of cost and utility functions, and the structural simplicity of the model. The trip-based 4S model discussed in this paper still struggles to address issues related to inter-person scheduling, but many of the issues of time and space are resolved through the proper treatment of time in the 4S approach. Furthermore, as discussed below, work is being done to extend the 4S model to tour-based and then activity-based formulations.

4.1.2 Efficiency

Unlike most models that increase policy sensitivity, the 4S model does not involve increasing run times. In fact the model is much more efficient than the FSM since it eliminates most of the wasted computation associated with full enumeration of choice sets. It also breaks away from the curse of large models – it is possible to increase the physical extent, spatial detail and behavioural sophistication without massively increasing run times. This makes it possible to run regional or national models at a high level of spatial detail, and makes it possible to have highly segmented purposes and travel markets. For example, a recent model developed by TransPosition has 74 purposes/market segments (including detailed freight markets), full mode breakdown including public transport, walking and cycling, 64,000 links, and lot-level land use inputs and yet can run in only a few hours. Another model covers the whole of Australia, with just under 300,000 road links, 90,000 PT links and Collector District demographics – this model runs overnight.

4.1.3 Extensibility

The simple core at the heart of the 4S model is an integrated random utility equation. There is no need for separate mode choice models, or toll models, or induced demand models; all of these are dealt with directly by the core model. Any other factor that has an impact on costs or utilities can be easily included by adding new random variables to the utility equation, or for more complex cases, new modes or states to the maximum utility path building system. This has already been done to incorporate issues such as multi-modal freight; the impact of grade on vehicle speeds and operating costs and walking and cycling; and time-of-day variations in parking charges.

It is also very easy (and computationally inexpensive) to add new market segments and to have varying behaviour for each market segment (including varying route choice behaviour). This is difficult to do in the FSM since each new purpose means a new skim, a new gravity model, a new mode choice model and a new assignment (although purposes are often aggregated for some of these steps). Each new purpose adds a fixed amount of extra computation, and the model spends just as much time modelling a significant purpose, like Home-Based-Work, as it does a much more minor purpose. In the sampling approach used by the 4S model each market segment is generally sampled according to its relative importance. In many cases the computational effort required to perform a 4S model is constrained by the need to get a statistically significant number of slices – in this case the addition of new purposes or market segments adds no extra burden to the model.

4.1.4 Expressiveness

The FSM is applied as a series of steps, and at each step information is lost. Skim matrices summarise costs, but lose information about routes. Distribution models summarise trips, but lose information about the trade-off between destination utilities and costs. At each step outputs are summarised and aggregated. To simplify computational effort purposes are often combined – we might have 10 purposes at trip generation and distribution, but only 5 segments in the mode choice model, and only 3 vehicle types at assignment.

In the 4S model, information about purposes is retained right through the model, including assignment. Also full cost information is available for analysis, and this can be linked to purposes. This makes it possible to, for example, report traffic volumes by purpose; break down travel time savings by socio-economic variables; or report total fare revenues collected by person type. In fact obtaining outputs from the model is similar to querying the results of a household travel survey – any sort of cross tabulation can be done. Except that unlike a household travel survey the model also has information about routes and costs.

It should be noted that the model also shares another feature with travel surveys – the more detailed the output the more samples are required. For the 4S model this means that a highly segmented set of outputs will require a large number of slices and a corresponding increase in run times.

4.1.5 Practicality

The FSM remains the preferred choice for most modellers around the world because it is seen as a useful, practical model. There has been resistance to moving towards newer models because they seem to add difficulty and complexity without offering clear benefits. The new models are seen to need large amounts of data that is not currently collected, significant increases in complexity, and even more lengthy run times (Vanasse Hangen Brustlin, Inc. 2007, p.99).

The 4S model is a practical solution – it can be calibrated using similar data to the FSM; it has a simple conceptual core; it can work with networks directly from GIS, without extensive coding of zone centroids or centroid connectors; it has short run times; and it uses behavioural parameters that are intuitive and comprehensible.

Initial experience with the model has been very positive. New models have been constructed very quickly, using automatically generated networks and transferred parameters. The elimination of zoning systems means that the model can directly make use of data from a wide range of sources – one application used lot level projections in some areas, traffic zone projections from a traditional transport model in other areas and Local Government Area projections in other areas.

5. Applications

Although the 4S model is fairly new it has been used in a range of practical applications. The work has been focused in South East Queensland, with models developed for Brisbane, the Sunshine Coast, Logan City and Toowoomba. Also a prototype model has been developed for the whole of Queensland, including all roads (including residential streets) and a significant portion of the public transport network (all of TransLink services plus some regional services). All of the models that have been developed are multi-modal, with timetable based public transport and detailed active transport (including the assignment of walking and cycling demand).

A multi-modal model has also been developed for the whole of Australia, using collector district demographics and all roads classes from collector up. There is ongoing work to extend this model to incorporate multi-modal freight and air travel.

The problem domains that have been examined using the model have included the following

- Demand estimation
- Public transport evaluation (including detailed access mode analysis)
- Toll choice analysis
- Accessibility
- Detailed/regional modelling
- Active transport
- Equity analysis
- Freight analysis
- Land use forecasting

Details on the use of the outputs of 4S models for land use forecasting are presented in a companion paper (Kidd & Davidson 2011), and some discussion of the toll choice impacts are presented in another companion piece (Davidson 2011). The appendix to this paper presents some typical results from the model, along with discussion on those results. The result included in the appendix should be considered examples only, and they are presented without much context or analysis. They simply show the range and character of outputs that can be produced from 4S models.

Details on some of the other applications will be presented in forthcoming papers.

6. Conclusion

The Stochastic Segmented Slice Simulation model is an improvement to the current state of practice. It is a theoretically sound model that overcomes many of the deficiencies of traditional modelling. It is also a practical model that offers many compelling advantages over current approaches. It is much more computationally efficient than other models, and this broadens the scope of problems that can be modelled and reduced modelling turn-around. It is also very efficient for model development – new models can be developed in a matter of days and weeks instead of months and years. It has progressed beyond prototype stage and has been used for a wide range of applications.

7. References

- Barrett, C. et al. (2007) *Label constrained shortest path algorithms: An experimental evaluation using transportation networks*, Technical report, Virginia Tech (USA), Arizona State University (USA), and Karlsruhe University (Germany).
- Ben-Akiva, M. et al. (1998) Behavioural realism in urban transportation planning models. *Transportation Models in the Policy-Making Process: Uses, Misuses and Lessons for the Future*, pp.4-6.
- Brisbane City Council, SKM & Connell Wagner (2006) Hale Street Link Feasibility Study - Appendix C: Business Case Traffic and Transport.
- Burrell, J.E. (1968) Multipath route assignment and its application to capacity restraint. In *Proceedings of the Fourth International Symposium on Transportation and Traffic Theory*.
- Chiao, K.-A. (2004) The New Trend of Travel Demand Model - Lessons learned from the New York Best Practice Model. Available at: <http://www.nymtc.org/project/bpm/model/NYBPM.ppt>.
- Cowing, T.G. & McFadden, D. (1984) *Microeconomic modeling and policy analysis: Studies in residential energy demand*, Academic Press.
- Daganzo, C.F. & Sheffi, Y. (1977) On stochastic models of traffic assignment. *Transportation Science*, 11(3), p.253.
- Davidson, P. (2011) Modelling toll roads - Where have we gone wrong? *Australian Transport Research Forum*, 2011.
- Dial, R.B. (1971) A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research/UK*, 5.
- Dijkstra, E.W. (1959) A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1), pp.269-271.
- Evans, R.J., Burke, M.I. & Dodson, J.R. (2007) Clothing the Emperor?: Transport modelling and decision-making in Australian cities. In *Proceedings of State of Australian Cities National Conference 2007*. URI <http://www.unisa.edu.au/>.
- Hensher, D.A., Smith, N.C. & Group, M.U.T.R. (1992) *Dimensions of Automobile Demand: A Longitudinal Study of Household Automobile Ownership and Use*, North-Holland.
- Kidd, P. & Davidson, P. (2011) Detailed land use projections - a new land-use/transport model based on supply-side constraints, accessibility and developer profitability. *Australian Transport Research Forum*, 2011.
- Lee, D.B. (1973) Requiem for large-scale models. *Journal of the American Planning Association*, 39(3), pp.163-178.
- Liu, H.X., He, X. & He, B. (2009) Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. *Networks and Spatial Economics*, 9(4), pp.485-503.
- McFadden, D. (1974) The measurement of urban travel demand. *Journal of public economics*, 3(4), pp.303-328.
- McNally, M.G. (2000a) The activity-based approach. *Handbook of transport modelling*, pp.53-69.
- McNally, M.G. (2000b) The four step model. *Handbook of transport modelling*, 1, pp.35-52.
- Nielsen, O.A. (2005) Does the theoretical development in assignment procedures matter in practice?
- Ortuzar, J.D. & Willumsen, L.G. (2011) *Modelling transport*, Wiley.

Vanasse Hangen Brustlin, Inc. (2007) Determination of the State of the Practice in Metropolitan Area Travel Forecasting - Findings of the Surveys of Metropolitan Planning Organizations.

WebTAG (2007) Transport Analysis Guidance - UK Department for Transport.

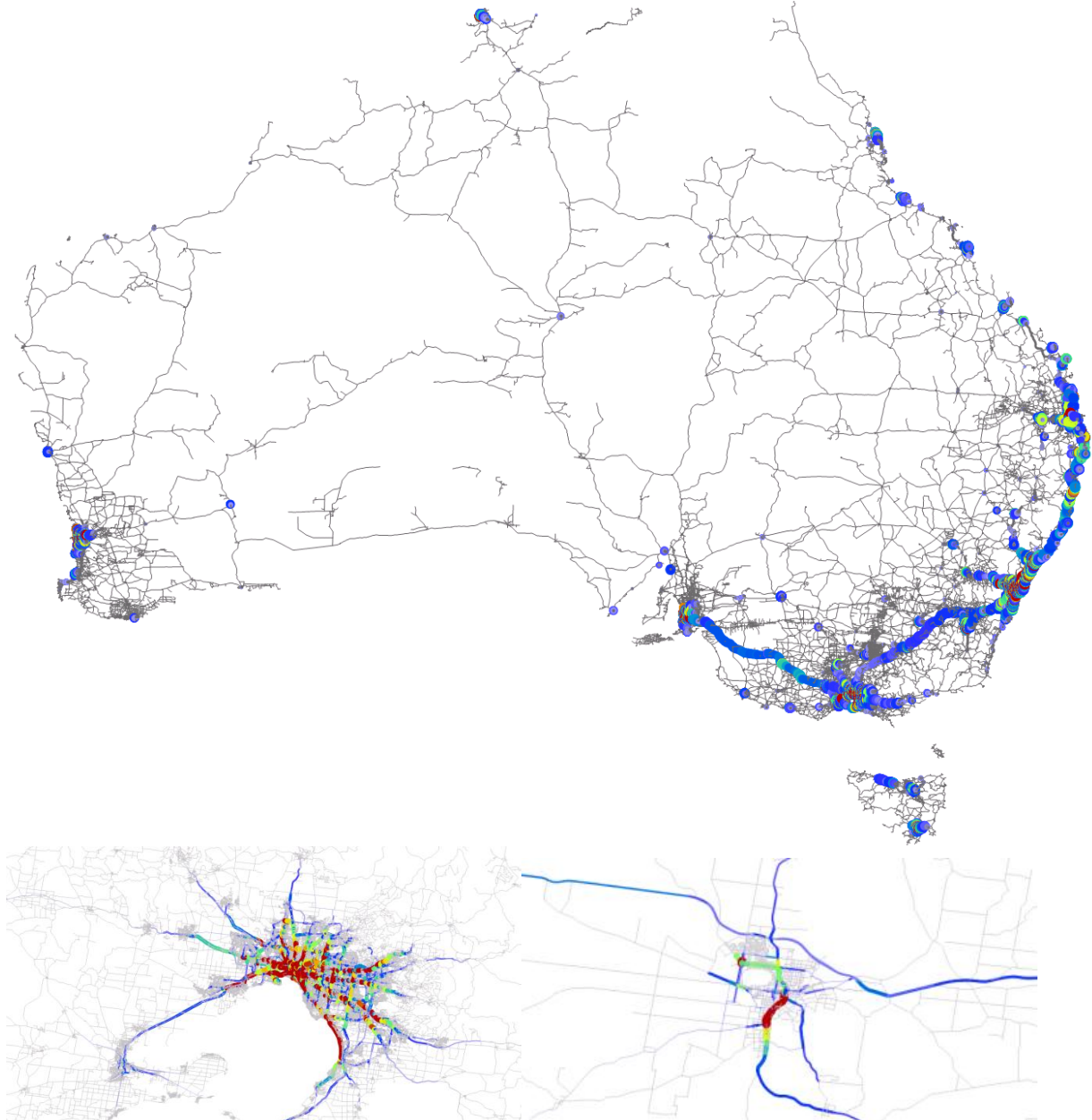
Weiner, E. (1992) *Urban transportation planning in the United States: An historical overview*, US Dept. of Transportation.

Appendix: Some example outputs from various 4S models

Note: The result included in this appendix should be considered examples only, and they are presented without much context or analysis. They simply show the range and character of outputs that can be produced from 4S models.

National / Regional / Local Modelling

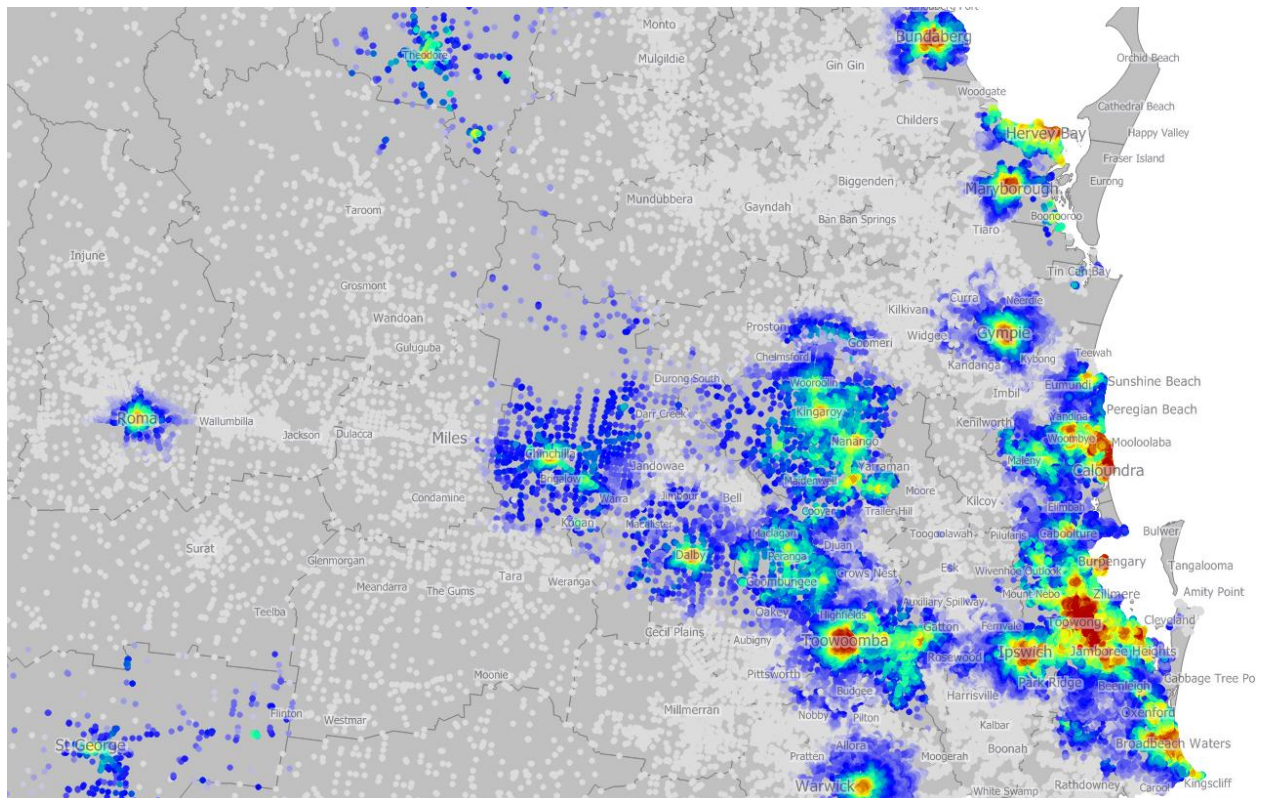
Figure 3: 4S Model of Australia, with detail in Melbourne and Ballarat



The efficiency of the model, and the fact that it does not require zones or zone centroid connectors, makes it possible to quickly construct models that cover a very wide area. These plots show traffic volumes from a 4S model of the whole of Australia, still with a very high level of spatial detail (all roads excluding local connector roads), and with full multi-modal modelling (including active transport, public transport, and air travel). We can quickly go between national issues, down to regional-level issues (the Melbourne network is shown above left), right down to local issues even in smaller cities and towns (the local network around Ballarat is shown above right). Note that in these plots, the colour bands change at different levels to reflect the very different traffic volume ranges.

Accessibility and Economic Analysis

Figure 4: Accessibility to health services in South East Queensland

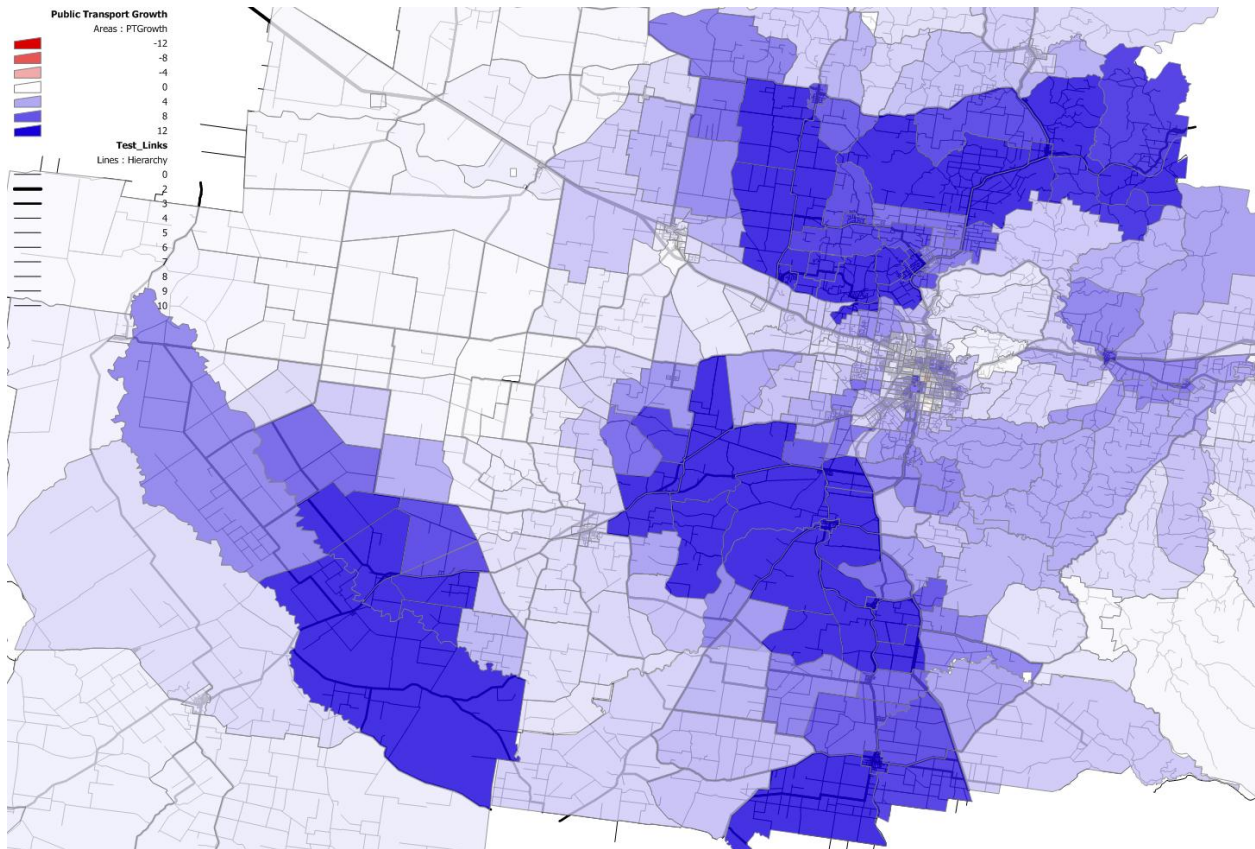


Since the model uses utility directly in the choice framework, it is easy to extract utility-based accessibility measures, broken down by any market segment detail. Here is a heat plot of accessibility to Health Services in South East Queensland. Those areas coloured red have good connections to health services, fading down through orange, yellow and blue into grey for those areas with poor access. This approach to accessibility combines information on the size and quality of services available, with multi-modal transport costs. It is also possible to reflect the different transport choices available to different market segments and find, for example, the accessibility to health services for those without a car.

The net utility outputs, along with the detailed cost outputs, can be used to undertake detailed and realistic economic assessment. And because the model explicitly allows for suppressed/induced demand, the economic assessment will correctly deal with changing demand, without requiring simplifications such as the rule of half. It is also possible to produce estimates of fuel use and emissions based on detailed instantaneous power models that vary by market segment and vehicle type with Monte Carlo selections of engine parameters.

Multi-modal analysis and links to traditional models

Figure 5: Increase in PT mode share in Toowoomba

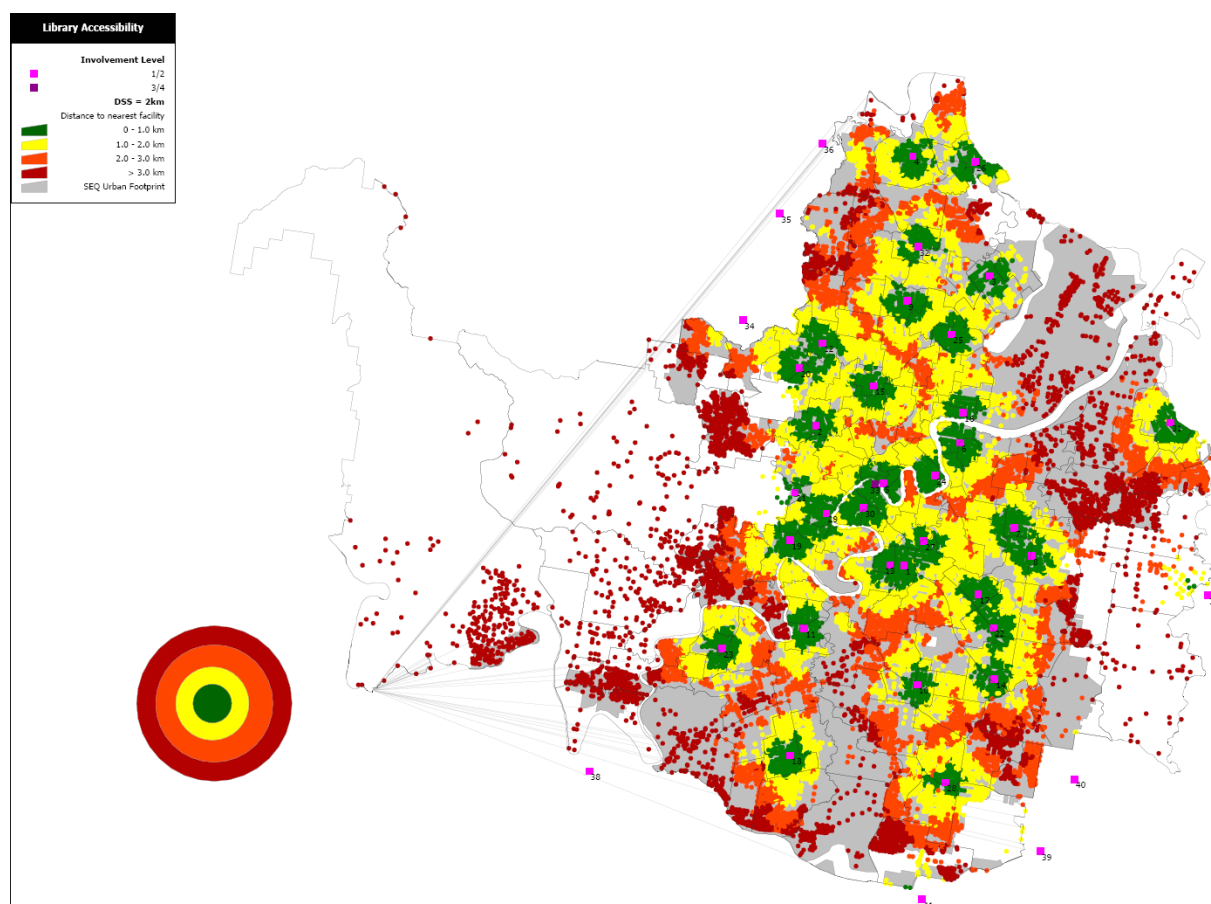


Although the model does not use zones, it is possible to aggregate the model outputs to any spatial system, including traditional traffic zones. This plot shows the increase in PT ridership in the Toowoomba Region as a result of adding new regional bus services. The very large park and ride catchments emerge organically from the model due to a trade-off between parking costs in the city and the free parking at the park and ride facilities, as well as travel times and costs. The model can also estimate the parking requirements, and take into account the effect of capacity constraints on park and ride.

(Produced for Toowoomba Regional Council in 2010 as part of the report on Transport Strategy Proposals to 2031, http://www.toowoombarc.qld.gov.au/docs/doc_download/3290-transport-study.html Page 147)

Local Issues – Active transport and community facilities

Figure 6 : Compliance with desired standards of service for libraries in Brisbane



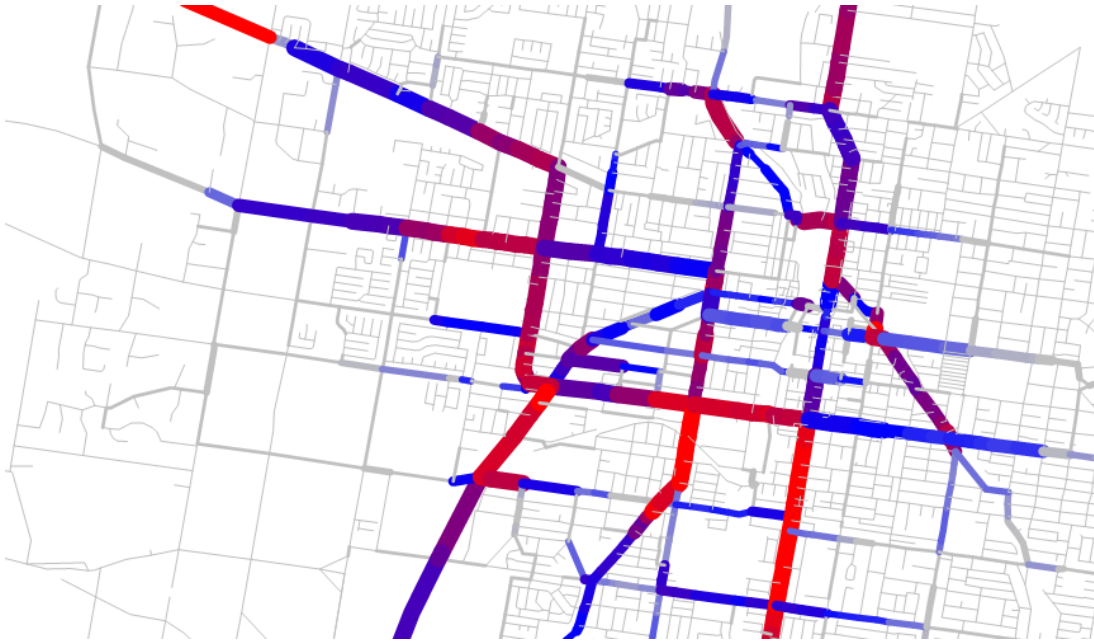
The level of spatial and behavioural detail allows modelling to be done for quite local issues. This plot shows local accessibility to libraries in Brisbane and how well it complies with the Desired Standards of Service (DSS). All local streets, and most off-road walking paths are included in the analysis, and facilities are located at exact locations (rather than somewhere within a zoning system).

This work has also been extended to analysing parks, open space, sporting facilities, swimming pools and community halls. It has also been used to investigate capacity issues in community facilities, and to compare alternative schemes for infrastructure provision. Note that the model allows new market segments to be specified very easily, so this analysis can be done in conjunction with analysis of the wider travel market to find the effects of congestion, public transport, and community severance.

(Produced for Brisbane City Council, 2009)

Network analysis

Figure 7 : Volume/capacity ratio in Toowoomba (all vehicle types)



The model can be used to produce all of the outputs prepared by traditional models, such as traffic volumes, volume/capacity ratios, sector matrices, select link outputs etc. This plot shows AM peak hour volume/capacity ratios in Toowoomba (colour) and traffic volumes (width).

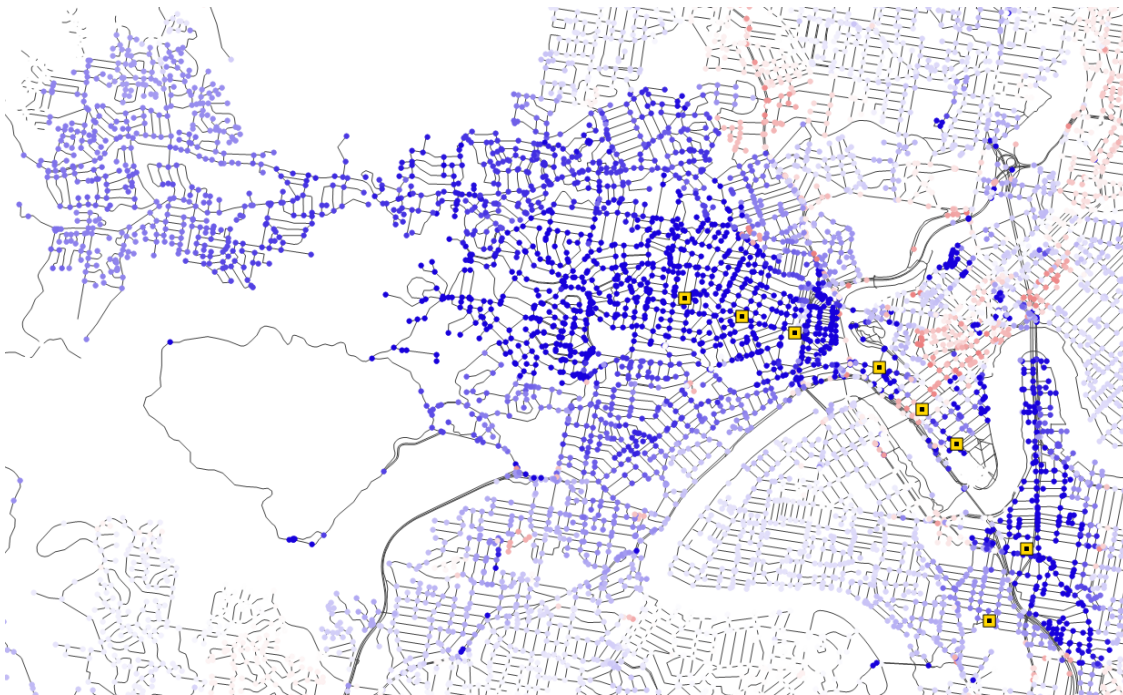
Figure 8: Cycling demand in Townsville



As well as traditional car volumes, the model can easily produce volumes by multiple vehicle classes, active transport volumes and public transport patronage. This plot shows estimated cycling demand in Townsville, taken from a Queensland-wide model.

Public transport – Benefit footprints and access mode analysis

Figure 9: Benefit footprint for hypothetical rapid transit in Brisbane



Because the model has, at its core, multi-modal utilities, it can be used to perform utility-based project evaluation. This plot shows the user benefit of a (hypothetical) rapid transport project in Brisbane. The total economic benefit can be calculated by summing over all populations and markets. By linking benefits and costs to locations and market segments, it is also possible to perform horizontal and vertical equity analysis.

Figure 10: Walking access catchments for hypothetical rapid transit in Brisbane



Detailed multi-modal networks allow PT access to be investigated. This plot shows regions of significant walk-to-rapid-transit mode share for the hypothetical project mentioned above. Walking catchments are determined organically and can take into account topography, severance and competing modes. Similar analysis can be done for bicycle, park-and-ride or kiss-and-ride, giving information on parking requirements, end-of-trip facilities and interchange capacities.