Slice simulation modeling - a new approach for building very large, very detailed models

Peter Davidson

10 July 2016

In most modeling there is a trade-off between model detail and model scope. Because demand matrices grow with the square of the number of zones, there are natural limits to how large or how detailed models can become. This is true even for activity based models, where high levels of spatial details in household are often married to more aggregate skim and demand matrices. This paper presents a new modeling approach, slice simulation modeling, that overcomes these constraints. It does this through the use of Monte Carlo analysis, with all components of choice integrated into a single utility function with arbitrary selection of random variables. This utility function includes destination choice through an explicit inclusion of the utility of travel to a given attraction; mode choice through a full set of model costs; and congested route choice with transit crowding and junction delays.

This paper describes the motivation for the model, and its theoretical justification. It explores the variables that make up the utility function and discusses how the complex utility maximization process can be made algorithmically tractable. It also examines some of the complexities of the model, such as the treatment of time and directionality; the form of the attraction utility model; and methods for dealing with congestion. Finally, the paper presents some recent applications of the model, with detailed metropolitan models of Los Angeles and London; a statewide model for California; and national models of Australia and Great Britain.

Introduction

Background

The history of modeling has been dominated by aggregate, trip based approaches, typified by the classic four-step model. To manage the complexity of traffic congestion and intersections there has also been a well established history of modeling with more realistic treatment of traffic in mesoscopic models. In the last decade there has been a shift away from these approaches, with growing emphasis on micro-simulation and activity based approaches. All of these models have different areas of emphasis, and different strengths and weaknesses. They can also be characterized by the different ways in which they deal with variations in behavior; aggregation of demand; and convergence of travel costs and travel choices. This paper describes a new class of model - the slice simulation model.

The new approach is based on an integrated random utility structure, where all components of travel are included into a single utility function. Unlike many models which focus primarily on cost components, the utility function includes an explicit parameter for the utility of travel to each possible attractor. The inclusion of an attraction utility component makes it possible for destination choice to be assessed using path building algorithms, along with mode choice and route choice. The final model uses Monte Carlo

sampling (see MacKay 1998) to determine estimated demand, and assesses congestion through a large number of loadings, using the Method of Successive Averages (MSA).

Outline of principle findings

A new approach to modeling is presented in this paper - the Segmented Stochastic Slice Simulation (4S) model.

It is named for the following features:

- Segmented: Uses a comprehensive breakdown of different travel markets, and allows all behavioral 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 (samples) of the travel market across the whole model area and through the distributions
- Simulation: Uses a traveler/vehicle state-machine with very flexible transition rules to effectively simulate all aspects of travel choice

This model is based on an explicit formulation of a random utility model, without the simplifications inherent in the logit equation. It does not use zones or matrices, but instead allows all travel to occur from node to node. It avoids the problems of aggregation inherent in most strategic models, and allows very detailed networks to be used; the usual approach is to include all roads and full timetabled public transport. It also allows for very large models, covering entire metropolitan areas, states and countries.

It has many compelling advantages over existing approaches:

- It has an elegant, theoretically sound basis that allows for realistic modeling of a very wide range of issues. This includes active transport, mode choice, toll modeling, behavior change, induced demand and time-of-day analysis.

- Models can be prepared with much less effort and arbitrary coding - by eliminating zones, centroids, and centroid connectors the manual effort in putting networks together is vastly reduced. Also these aspects (zones, centroids and centroid connectors) are somewhat arbitrary abstractions that make the model highly dependent on manual inputs and individual assumptions.

- 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 traditional models' use of a whole range of separate iterations for convergence) complex models can be run with practical run times.

- Its simple core allows it to be easily extended - the current model includes intersection delays; a detailed fuel use model; and multi-commodity, multi-vehicle class freight.

As an example of the efficiency of the approach, the author has developed models of large cities (including London and Los Angeles) with every road, transit and active transport link that can run in hours. Larger regional and national models can also be used with reasonable run times; models of California, Great Britain and Australia have all been implemented to run overnight. With some simplification of the network (removing most local residential streets) a model of the whole of the United States should also possible as an overnight run.

To understand the detail included in these networks, it is instructive to compare it with a number of large US models.

- New York Best Practices Model: 52,794 one way road links, 3,000 transit routes, 73,000 transit stops
- California statewide travel demand model: 235,000 one way road links, 86,000 nodes
- 4S model for Australia: 3.5m one way road links, 1.5m nodes

The new Segmented Stochastic Slice Simulation (4S) model has many of the benefits of mesoscopic models, in that the simulation component allows for much more realistic treatment of travel costs, including intersection delays and transit crowding. Consistent with this is a much finer representation of the land use and transport systems- the model is usually run with all roads; timetable-based transit; full detail on active transport; and point based land use/demographics. By moving away from traffic analysis zones (TAZ) and allowing all travel to be point-to-point, the new model increases the realism of modeling multi-modal travel, and reduces the effort involved in creating new models.

The new approach also brings in a more flexible utility formulation, allowing it to consider a wide variety of changes to travel behavior and the interaction between demographic variables and travel choice; this brings in some of the benefits of activity modeling. But in its application and data input requirements it is closer to traditional four step models.

Description of the framework of the paper

This paper will describe the approach used by the 4S model; the theory behind it; and discuss the ways in which it has been applied over the last 7 years.

The motivation for a new approach

Traditional approaches to modeling (the four step model)

The four step model considers travel through a series of distinct (but linked) choices. These were originally seen as separate steps, but are now commonly viewed as a hierarchy, with decisions at the higher level based on the probability-weighted aggregate of lower level decisions. In order to make the problem tractable, demand is spatially aggregated to traffic zones, and the road and transit networks are generally simplified. The core description of travel is the matrix, which records all trips between each origin and destination zone; the

key input that determines people's travel choices is the skim matrix, which records the travel costs (or times) between each pair of zones.

Lack of detail

The difficulties with this model structure comes from these two simplifications - matrices and skims. The problem with demand matrices is that they are inevitably too coarse, and necessarily exclude many of the details of local demand decisions. These local demand issues have impacts on a significant portion of car travel; the "last mile" of transit demand; and almost all walking and cycle demand. Local roads are usually excluded from four step models, and intra-zonal demand is not well modeled and usually discarded. The obvious response is to increase the number of zones, but this quickly makes the model cumbersome, with long run times and high storage requirements. This is exacerbated as the study area increases. Most large metropolitan models or statewide models have very large zones; national and regional models have such large zones that almost all travel occurs intrazonally.

The other response to the problem of large zones has been to reduce the extent of the model, so that high levels of spatial detail can be used. The desire to improve the realism in treatment of roads and intersections led to the development of meso-scopic models, and then micro-simulation models. These models generally include demand matrices with much more detail in their origins and destinations, often down to individual car parks and offices. They also include a comprehensive road network, with all roads, and detailed information on intersections and lane allocation. However these models are focused on improvements to the assignment stage of the model, and often rely on inputs from a strategic, aggregate model. They also have scalability issues, and generally can only be applied to smaller areas within a city.

Lack of variability

Skim matrices have their own problems. It has long been recognized that travelers consider more than just travel time when making their decisions; they consider other direct costs (tolls, parking, transit fares, vehicle operating costs) and they value time differently for different activities. For example, most people would rather spend 10 minutes driving a car than 10 minutes walking; furthermore they would rather spend 10 minutes driving through uncongested streets than driving in traffic. The usual approach to this is to produce generalized cost skims, where all cost elements are combined to give an overall, aggregate cost. But different people value things differently - some people enjoy walking and dislike cars; some people value their time very highly and will pay high prices to save it, whereas others will be much more sensitive to costs. The differences can be in more than just perceived values - different vehicles can have different speed profiles, different operating costs, and different sensitivities to grade and congestion; transit fares are usually lower for students; and many workers have their parking paid for by their employer.

It is possible to encompass some of these variations through careful segmentation - producing different skims for different vehicle types, and perhaps for workers as opposed to students. However skim segmentation increases model run times, and furthermore it

ignores the variations that exist within each segment. The variability is inserted back into the model through techniques such as the logit mode choice or toll choice models, but these can only operate on the aggregate results from the skims. The fact that different people will choose to navigate through the network differently is ignored - the skim and assignment only reflects the decision and cost of the average traveler.

Activity based models

More recently there has been a movement towards an activity-based approach. This approach aims to improve the treatment of travel decisions within households by explicitly looking at time-allocation constraints, and interaction between household members. Activity based models (ABM) generally treat households and individuals at a disaggregate level, and produce lists of trips and tours made by synthetic travelers. However they rely on network skims done at the aggregate TAZ level, and the final estimates of traffic volumes and transit loadings are still generally converted into matrices that are then assigned to the network using traditional approaches (although dynamic assignment techniques are starting to be used).

ABM's also have difficulty in ensuring convergence across the choice structures - when run in forecast mode they require skims and accessibility measures to be determined prior to running the household choice models, but the level of network congestion is a result of the aggregate decisions made by travelers. This means that whole process must be iterated to convergence, significantly increasing already lengthy run times.

So although they represent an improvement over the four step model, ABM can still have difficulty capturing the full detail of the network and land use, and are cumbersome if extended to large areas.

Treatment of variability

Many of the differences between different models come down to how they deal with variability of travel behavior. The traditional approach uses aggregate proportions derived from analytic equations. These proportions were originally developed empirically without a clear theoretical basis; an example of this is diversion curves. Sometimes they were done by analogy with field equations, such as the gravity model. The development of discrete choice models gave a theoretical basis to the proportions; the random utility model (RUM) identified the proportion as the probability that a given choice would have a higher utility than any other. However the preference for an analytical approach necessitated the use of very simple utility formulations, at least in their treatment of variability. The most widely used approach for managing variation is the multinomial logit, which allows for variation only in a single additive term (the error term) and does not allow for variation in taste, or any complex correlation in variability (other than that allowed through nested discrete choices).

So the traditional approach is to enumerate all alternatives; find the associated costs/utility; and then determine the proportion selecting each alternative using an equation. In most cases the spatial aspect of the alternatives are defined by a zoning system

and the full enumeration of all possible alternatives is given by the skim matrix. In four-step models the demand is then allocated using the proportions; the final demand matrix may have many cells with tiny fractions on a trip, but these are aggregated using assignment algorithms to give final demand on the network. The fractional trips can not be eliminated since they contribute to the total.

In activity based models and microsimulation models the proportions are treated as probabilities, and Monte Carlo techniques are used to turn them into discrete events. Usually a uniform probability distribution is used, and the probability that any particular decision will be "realized" in the simulation is the fractional proportion. The final model will have a list of complete choice events (which may be individual vehicles in a microsimulation model, or tour and trip lists in an ABM (Castiglione, Bradley, and Gliebe (2015) Pg85)).

Because the ABMs use logit curves and log-sums, they still require full enumeration of all alternatives, necessitating the use of aggregation of cost data by zone. Travel modes are also simplified, for example the CT-RAMP models typically use only 5 modes (SOV, HOV, Walk to Transit, Drive to Transit, Non-Motorized) (see Parsons Brinckerhoff in association with Arizona State University 2010 Pg 24).

The motivation behind the 4S model can be understood by considering the way in which a utility maximizing model can be applied in a Monte Carlo framework. In any RUM only the highest utility alternative is selected; most models are focused on working out the probability of this occurring. In most formulations, the calculation of the probability that any alternative is the highest requires the determination of the full utility of all other alternatives. However if the utilities are sampled using a Monte Carlo technique then only the highest utility must be explored in detail - lower utilities only need to be considered up until the point that they can be rejected. In practice, this can be done using path building techniques, which implicitly find only the best path. The 4S model works by converting the full travel choice to a complex path build through a multi-dimensional network, as described below.

Methodology - The 4S Model

Flexible random utility

At heart the model adopts the same Random Utility Maximization (RUM) theory 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 behavior, modelers' 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.

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 traveling.

$$U_{a,m,r,n} = U_{a,n} - \beta_n C_{a,m,r,n}$$

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 traveling to attractor a by mode m on route r, β_n is a vector of random taste coefficients for individual n. The coefficients are random variables that vary over individual decision makers and are distributed with a density function that 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.

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.

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

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

The actual variables to be included in the utility formulation can be easily changed, and can include any elements of link, mode, market segment or route. Importantly it can also include any derived variables, including aggregate demand by time period; opposing flows at intersections; remaining capacity of transit services at loading; crowding of transit services; or remaining parking supply. The variables can all be estimated throughout the process, using MSA to produce successively reliable estimates. Speed of convergence for many of these elements may be slower than, for example, a classic equilibrium approach. But as shown below, the solution strategy requires a very large number of iterations/slices for reliable Monte Carlo exploration of the random variables; these are generally sufficient to allow for MSA convergence.

A typical implementation of a slice simulation model would contain the following components in the utility formulation.

- Attraction Utility - the utility associated with traveling to a given attraction
- Value of time spent traveling - influenced by the following components
  - Hourly income - usually a log normal distribution based on person type
  - Wage rate multiplier - varies by market segment
  - Value of time multiplier - varies by mode and activity (in vehicle, waiting, interchange, wait at first stop)
  - Speed of travel - based on network and market segment characteristics (including walking speed, cycling speed, intersection delays)
  - End of trip costs - varies by mode, trip length, and network characteristics (such as end of trip facilities for cyclists, or parking search time for drivers)

- Operating costs
  - Fuel costs - based on market segment/vehicle type, trip length, network characteristics (such as grade, road surface type)
  - Other vehicle operating costs
- Direct costs
  - Tolls
  - Fares
  - Flagfall cost, per distance cost, per time cost (for taxis and shared autonomous vehicles)
  - Parking charges
  - Road pricing

In addition, the model uses distributions of preferred arrival time/departure time to determine when in the day travel occurs (see section below on Continuous Time).

Solution strategies

It would seem that producing an estimate of p 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) recognize that closed form solutions to the maximized 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 the utility formulation 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 travelers 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 recognizing that the attractor can also be considered part of the route - a travel decision can be seen as a route to fulfillment 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 maximizes net utility.

Maximum utility path building

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 minimizing 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. 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 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 maximizes net utility. This algorithm can be extended to multiple modes, allowing for arbitrarily complex multi-modal trips.

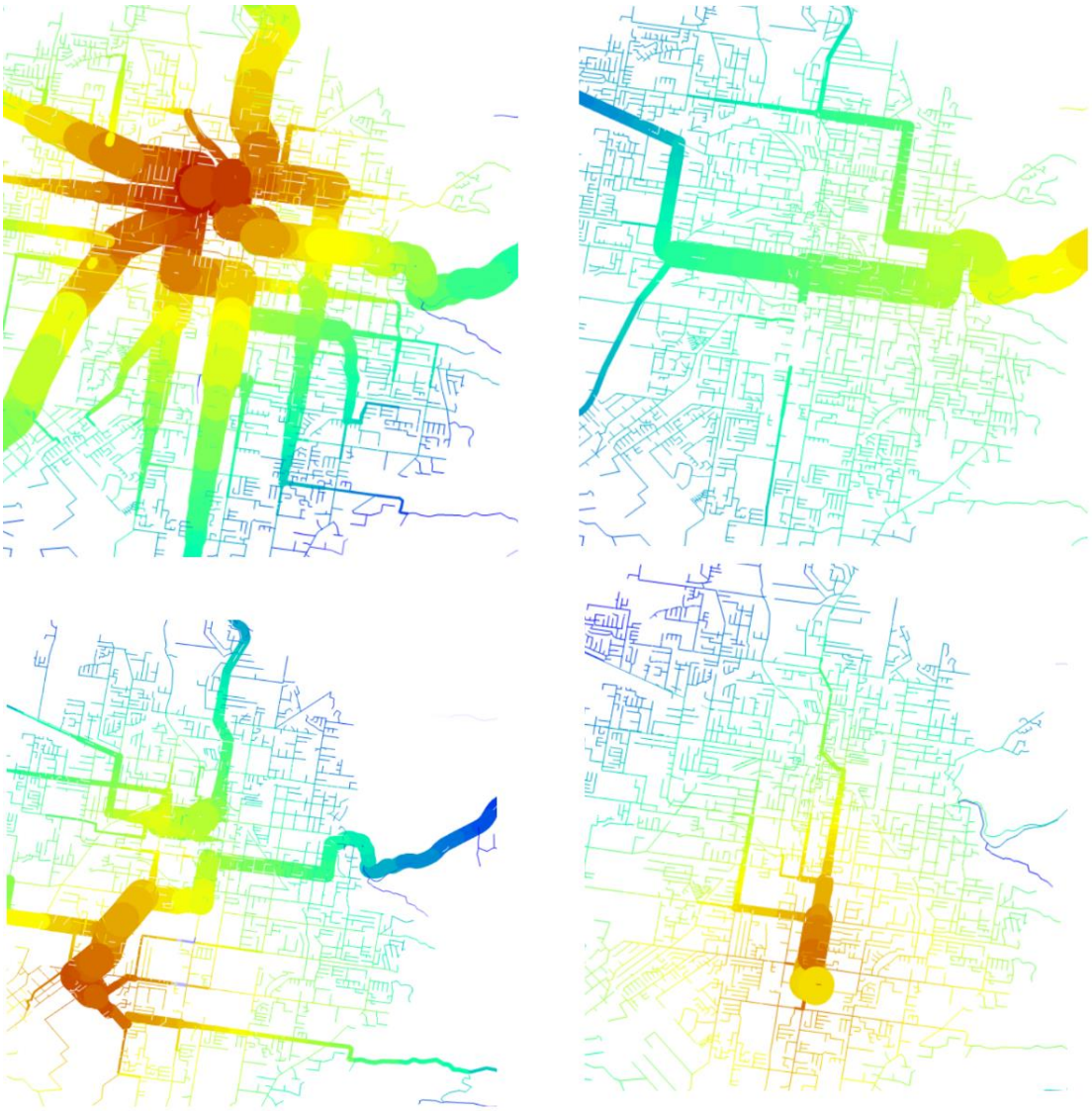

FIGURE 1: Examples showing four slices volumes

Figure 1 shows the algorithm in action. In each of the four sections of the plot a single slice result can be seen for the same section of network; the line widths show the allocated demand, and the colors show the net utility of the choice at that location. The utilities are highest at the most desirable attractor; as paths build out from this point the utilities drop, but are still higher than those that can be obtained at any other attractor. Eventually the net utility values are low enough that another different attractor becomes viable; links that lead to it will now be added to the set of viable links for consideration in the path building algorithm. At each point in the process, the next highest utility link will be considered, and all links leading to it will be explored. The algorithm can be visualized as a series of expanding catchments around attractors; at the watersheds (or dividing points between catchments) the net utilities will be approximately equal.

An important point to be noted in this algorithm - as soon as a production location is reached, the process is sure that the attractor and route to that attractor are the best

alternative for that production location. It is not possible that any better route or destination could be found, since the process always considers the highest utility option (an no path can have increasing utility). Thus as soon as the path has been found it can have demand assigned to it. This is a key efficiency improvement of this algorithm compared with a traditional skim/assignment process, where paths must be determined twice.

Continuous 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 travelers 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 traveling 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 center opening hours)

Directionality

The current implementations of the 4S model are trip based - this is a simple approach and can be solved efficiently. It is possible to structure a tour based approach using similar techniques, but the path building algorithm is more complex and is anticipated to have slower run times. This is an area of ongoing research by the author.

In the trip based formulation, both directions of trips need to be considered - from production to attraction (P2A), and from attraction to production (A2P). However, as in a

traditional destination choice model, the fundamental choice is always production to attraction, with attributes at the attractor determining its likely demand. In the maximum utility path building process, this means that paths must always be built from attractions back towards productions, with each desirable attractor forming a catchment zone of associated productions.

In order for both directions of travel (P2A and A2P) to be modeled, the algorithm must switch the directionality of the path building process. When modeling P2A trips, paths are built in reverse, from attractions to productions going back in time. In this case the time at the attraction is a preferred arrival time, and the model works back to when the trips would need to leave the production to arrive at the attraction at the nominated time. When modeling A2P trips, the paths are built in the forward direction, going forward in time. In this case the time at the attraction is the preferred departure time, and the model works out when the trip will arrive back at the production.

Attraction utility

In order for destination choice to be included in the choice set, the utility of each attraction point is explicitly included in the 4S model. Since this utility is usually implicit in destination choice models, there is no consensus on what form the attraction utility distribution should take. From the form of a classic gravity model, one would expect that the average utility should be logarithmic with respect to size. This is because in a classic model the number of trips to a particular attraction increases linearly with the size of the attractor - under a logit destination choice this equates to:

$$T_{i,j} \propto A_j e^{-C_{i,j}}$$

$$T_{i,j} \propto e^{ln(A_j)-C_{i,j}}$$

There are also some other desirable properties of an attraction utility function. It should be fat-tailed, such that the probability of a very high utility is not negligible - this ensures that long distance trips are able to sometimes occur. The requirement for a fat tailed distribution eliminates a normally distributed utility.

The utility function should also be invariant under different compositions - if a single attractor is decomposed into two attractors at the same location, each with half the size of the original, then the aggregate demand to the two new attractors should be equal to the original demand. By ensuring this property, the model can be easily aggregated as required without changing destination choice parameters.

A range of potential formulations could be found to align with these properties. The one that has been adopted is based on the idea that an attractor provides opportunities to satisfy some desire, but that desire could potentially be satisfied even at a small location. For example, if someone is shopping for clothes, they might find exactly what they want at a small, isolated boutique store. But the bigger the store, the more likely that they will find what they want. The basic utility distribution is unchanged, but the bigger store simply provides more opportunities to obtain it.

For maximum flexibility, the basic attraction utility is modeled as a gamma distribution - this is a fat tailed distribution defined by two parameters; shape and scale. These parameters vary by travel market, along with a scale factor. The scale factor is used to convert the destination size into a number of discrete opportunities. In the retail example, it might be set to 5, such that each 5 jobs at the destination provides one draw from the basic utility function. If there are 20 retail jobs at the destination, then the model will take 4 draws from the gamma distribution and then use the highest one.

An examination of the performance of this approach shows that in the limit the average utility increases with the log of the size, as desired. It also trivially satisfies the composition requirement, since a single location divided into two will still involve the same number of draws as the original, with the expectation of attracting the same number of trips.

Congestion

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 (FWA), and other uses of the Method of Successive Averages (MSA) and the Method of Successive Weighted Averages (MSWA). 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 model can assess route choice issues (including congestion) at the same time as determining travel choice. The loops needed for valid sampling of the probability distributions mean that there are many more convergence loops than is typically performed in traditional models (usually over 1000).

Furthermore, there is no need to perform additional convergence loops for other levels of equilibrium; most FSM and ABM require a converged equilibrium assignment to be performed multiple times to get convergence between the congested skims that are an input to the demand estimation, and the resulting assigned volumes. The simultaneous choice structure of the 4S models makes this unnecessary; a single loop can be used to gain convergence between all aspects the model. This includes congestion on public transport; utility-based double constraining factors; parking supply; and crowding at destinations.

The use of MSA, and the very high number of convergence loops, relaxes the monotonic constraint that usually applies to costs under the FWA. This allows the 4S model to incorporate elements of mesoscopic models, including opposing flow delay calculations at intersections, and queuing back of demand. It can also address transit congestion; double constraining factors; parking supply; and crowding at destinations.

Thorough market segmentation

Most models maintain segmentation through some processes, but aggregate at some point before assignment. Thus traffic volumes on roads are usually disaggregated only by vehicle type. In contrast, the 4S model maintains full segmentation throughout the model. This makes it possible to find the breakdown of traffic on each road by purpose, for example.

With suitable segmentation it allows transit demand on each line to be listed by income category, or age group. It also makes it easy to examine the equity distribution of any improvements to the transport system, easily identifying winners and losers from any policy.

Major results

Applications

The 4S model has been applied to a number of Australian cities over the last 5 years, and initial applications have been done in London, Los Angeles and Denver. The model has also been used to analyze wider networks, including a detailed model of the whole of California, Australia and Great Britain. These larger models are particularly useful for analyzing freight demand, where many of the movements are interurban, but these are strongly impacted by urban congestion (since most road freight must travel through cities to reach major ports).

The following figures show some broad results from the models.

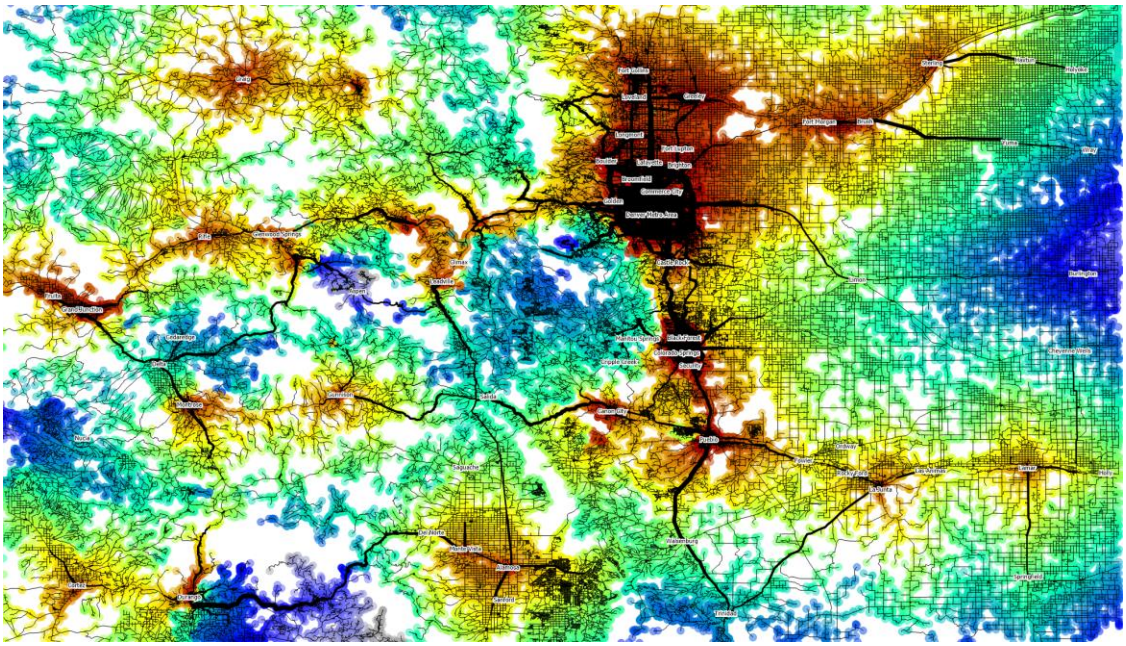

FIGURE 2: Volumes and accessibility in Denver

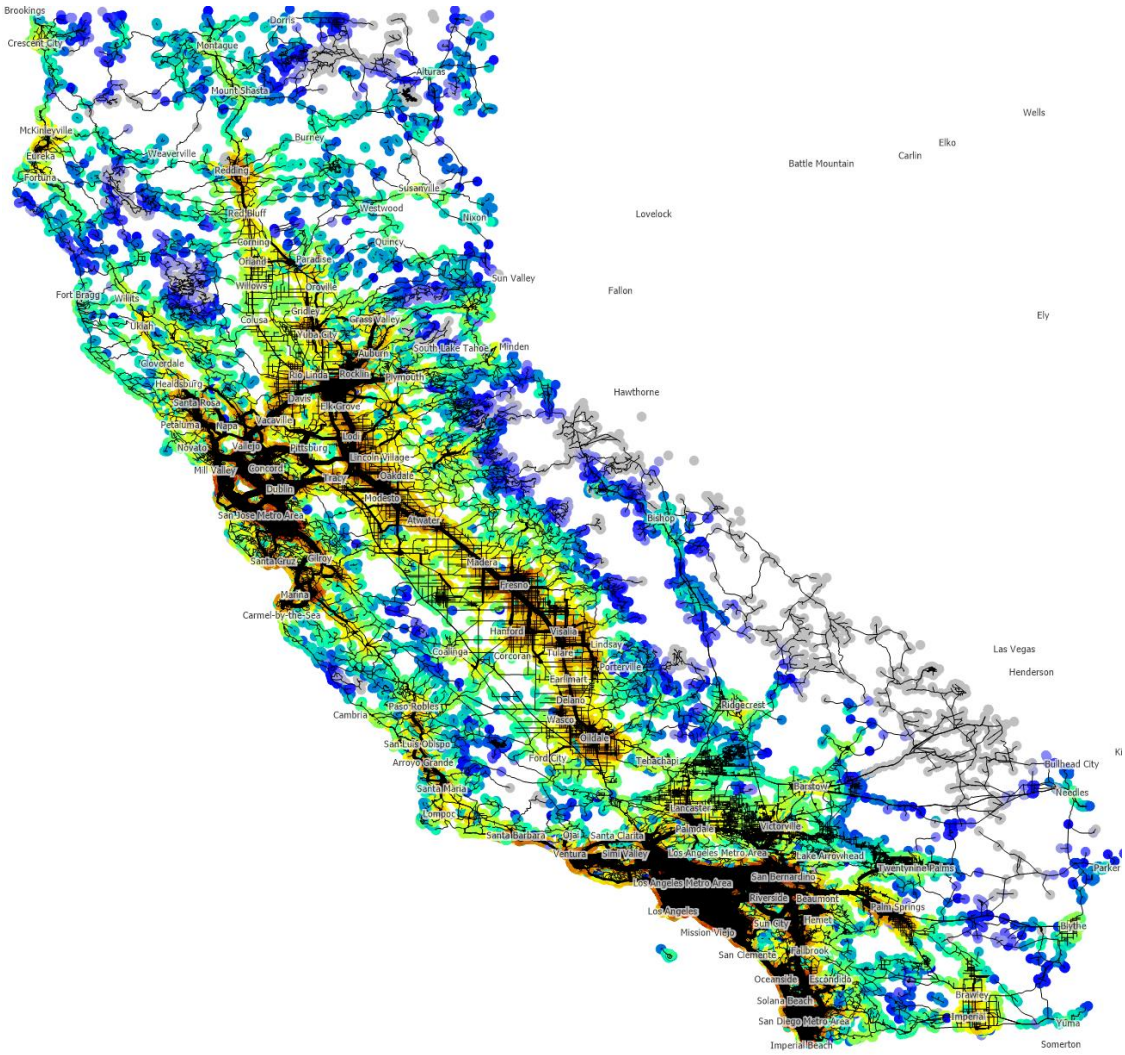

FIGURE 3: Volumes and accessibility in California

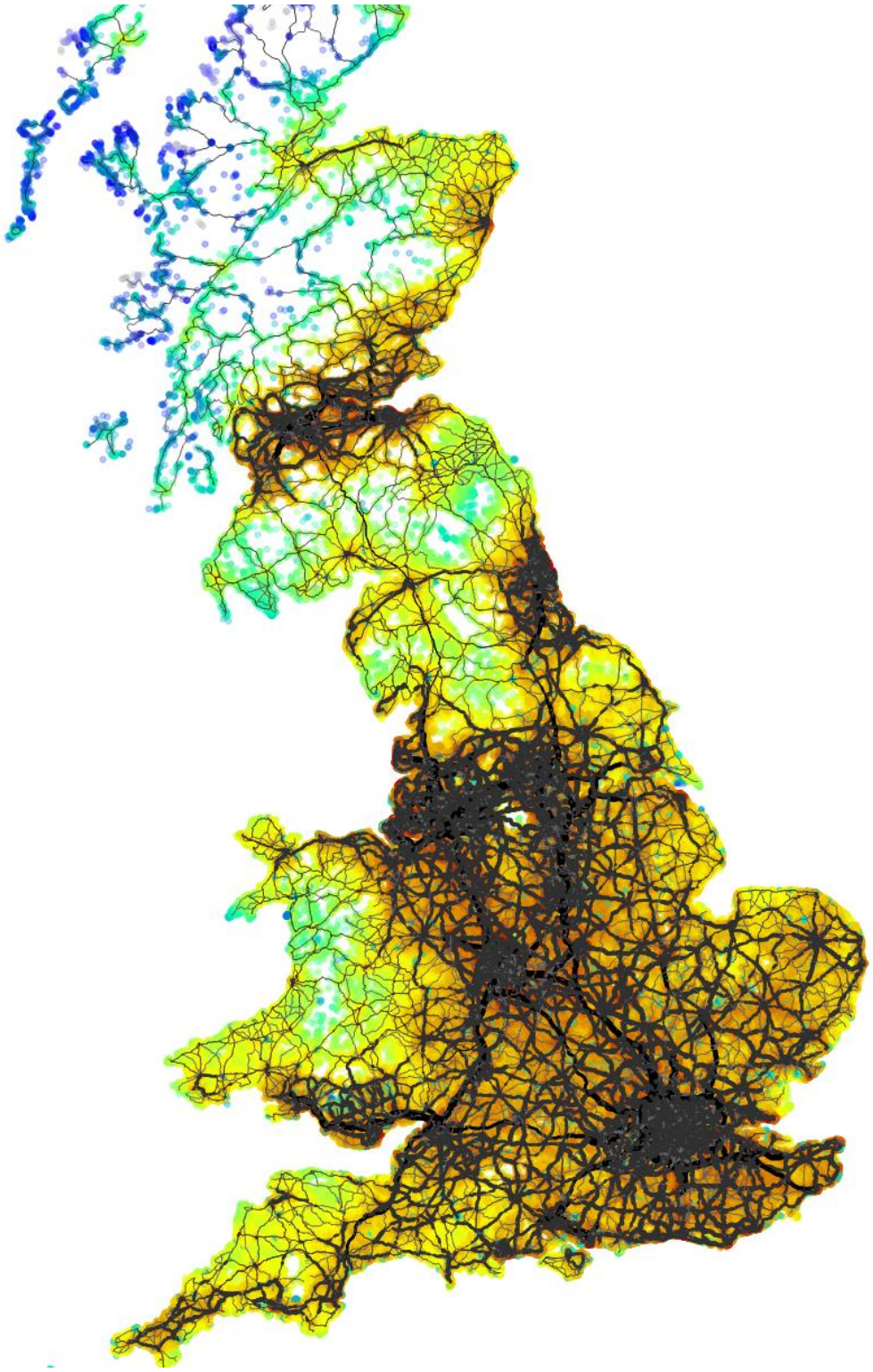

FIGURE 4: Volumes and accessibility in Great Britain

It has been applied to a range of planning issues including:

- Demand analysis for toll roads

- Development of integrated regional strategy
- Planning public transport network improvements
- Regional freight analysis
- Catchment analysis of parks, libraries, swimming pools and sports fields
- Development of cycling strategy
- Accessibility analysis for integrated land use/transport model

The models have been calibrated using traditional maximum likelihood techniques where possible, with some parameter distributions taken from other studies.

The structure of the model has been found to be very adaptable, and the models have calibrated well with relatively few parameters.

Implications for the science and/or practice of travel modeling

The 4S model is a complete redesign of transport modeling, moving away from many traditional elements such as traffic zones, centroid connectors, skim matrices and demand matrices. This significantly reduces the effort required to develop a new model, as most of the network creation can be done automatically. By distributing population and employment down to individual network nodes, the model can easily accept data from varying levels of detail, further simplifying model development and demographic forecasting.

It uses Monte Carlo sampling to very efficiently model all aspects of demand simultaneously; route choice, mode choice, destination choice and trip choice are all done using a single utility maximizing formulation. This allows for taste variation, and deep behavioral variability that affects all aspects of choice. The efficiency of the model allows it to include fully detailed networks, with every road, pathway, and timetabled transit route. The model also allows key land use activities to be specified at their exact location, rather than at some amorphous point within a traffic zone. This can be done even for very large models, covering multiple cities, entire states or countries.

This has the potential to extend the range of issues that can be modeled, and improve the realism of the analysis. The first-principles behavioral basis makes it particularly suitable to look at issues such as behavior change, autonomous vehicles, transit planning, and freight demand. The ability to model large areas but still maintain high levels of detail can remove some of the need for maintaining a hierarchy of models (statewide; strategic; mesoscopic). And finally the new approach opens up the possibility of very large models, such as the national model of Australia; and potentially detailed models of the entire US or Europe. With planned extensions of the software to run on a large network of cloud compute servers, it is even feasible to consider, for perhaps the first time, a single model of the whole world.

References

Castiglione, Joe, Mark Bradley, and John Gliebe. 2015. *Activity-Based Travel Demand Models: A Primer*. SHRP 2 Report S2-C46-RR-1.

MacKay, David JC. 1998. "Introduction to Monte Carlo Methods." In *Learning in Graphical Models*, 175–204. Springer.

Parsons Brinckerhoff in association with Arizona State University. 2010. "MAG Ct-Ramp Activity-Based Model Phase I." Maricopa Association of Governments. https://www.azmag.gov/Documents/TRANS_2010-12-06_MAG-CT-RAMP-Activity-Based-Model-PhaseI-Model-Estimation-Results.pdf.