The road of discovery – heuristic methods in traffic assignment

By Nicholas B Taylor, Research Fellow, TRL, Crowthorne House, Nine Mile Ride, RG40 3GA, UK
ntaylor@trl.co.uk

Short abstract and objective of the paper

Semi-aggregate or mesoscopic traffic and assignment modelling methods can capture the essentials of real traffic without detail which may not contribute to evaluation, as well as providing determinate results for scheme comparisons. However they depend on heuristic approaches to assignment and convergence, and need to be used with an appropriate level of data aggregation. The paper discusses these issues, in addition to the relative merits of mesoscopic, microscopic and macroscopic simulation, and shows that there is potential for improved computational efficiency, especially when dealing with large regional networks which are increasingly common. In particular, aggregating small origin-destination flows can reduce model run time substantially, without requiring software changes, and having only a small effect on estimated network costs and hence on the outcome of scheme appraisals.

Introduction – macro-, meso- and micro-scopic simulation

A primary objective of traffic assignment modelling is to enable estimation of the costs and benefits of alternative development or management strategies. On these results may rest important choices and decisions, involving large commitments of expenditure. Model results should be determinate, even if subject to quantifiable uncertainty, so they can be used to answer questions like: what is the benefit of the scheme; or how does one option differ from another? Microscopic simulation of traffic may deliver a realistic model of the situation on a single day, but many trials are required to get average results for appraisal. The attraction of more theoretical macroscopic modelling approaches is that they deliver a single repeatable result. However, they tend to depend on algorithms which have no direct real-world analogue, such as minimisation of a global objective function.

Mesoscopic modelling is not just a compromise between microscopic and macroscopic, but a whole concept of modelling in its own right. It can be summed up as seeking to represent the *effect* of real-world processes by embodying their *essential behaviour*, but without *detail* which may have no observable effect, or is impossible to verify, or does not constitute a *prediction* because it is only a random sample. Thus mesoscopic aims to achieve the predictive power of microscopic on all but the shortest time scales, with much greater computational efficiency. Its disadvantage is that it has to rely on ‘heuristic’ methods which appear to have neither the transparency of individual behaviour nor the rigour of formal theory. But because mesoscopic models try to encapsulate measurable real world processes, their *assumptions* can be considered more transparent than either macro- or micro-approaches, making it easier to see where they could fail to be realistic. This paper touches on general principles and performance, and techniques in assignment, and devotes space to describing aggregation of demand, which shows promise for reducing the computation time of large network models.

Limitations of static assignment

Another objective of modelling road networks is to identify the socially and economically important effects of congestion, so it is essential that this be modelled realistically. Accurate representation of transient effects such as congestion can be guaranteed only by time-dependent modelling. Macroscopic traffic models generally have limited provision for this. Some rely on monotonic volume/delay relationships which, due to their mathematical form, allow global optimsation of an overall objective function, resulting typically in a user-optimal traffic pattern. These functions cannot deal with volumes exceeding capacity, since they predict infinite delay, so the models just strip off any 'unassigned demand'. In reality 'excess' demand, if not suppressed or redistributed, is spread over a longer period, and an accurate delay calculation ought to reflect this. But there is no simple relationship between demand and delay, due to the fact that the network has 'memory', so static delay functions would need to be calibrated for each particular scenario.

Other models use time-dependent formulae to calculate the *average* delay across a peak, which is then fed back into a static assignment. These cannot account for details such as the relative timing of conflicting movements, to which delay can be highly sensitive. Nor can they properly deal with the lag from demand peak to delay peak. If two neighbouring factories discharge workers at the same time, the delay can be more than twice as great as if each they discharge at different times, and queues may persist to cause delay after the original traffic has moved on. Finally, static models usually build up a traffic loading by an averaging process rather than simulating individual journeys, limiting their ability to represent disaggregate route choice and responses to information and Intelligent Transportation Systems (ITS).

Time-dependent modelling

The response in the late 1970s was to develop models which accommodate transient over-capacity naturally. An early example is TRANSYT (Vincent, Mitchell and Robertson 1980; Binning, Crabtree and Burtenshaw 2003), which also provided the conceptual basis for the traffic model SATSIM within the SATURN suite (van Vliet 1982). TRANSYT does not have assignment or full time-dependence, but simulates and optimises coordinated signal systems using a repeated cyclic profile. However, it accommodates over-capacity through the use of time-dependent queuing formulae which can deal with more general situations.

Traffic in TRANSYT and many other models is represented by aggregate flows. This necessarily leaves out the random variations in flow which occur in reality. In an urban network, whose capacity is dominated by junctions, these mainly affect queues. Allowance for random variation is therefore built into the queuing model (eg with exponential headway distribution). By applying this model over a succession of time slices, mean queue development can be calculated continuously through a peak of any shape and duration. An extended version of this model is described by Kimber and Hollis (1979) and discussed by Taylor (2003). Variation in vehicle speeds, whose effect is modest in urban networks, is allowed for by a simple dispersion model.

More recently, there have been many papers in the literature describing 'dynamic' traffic assignment (DTA), that is time-dependent assignment [1]. Some are based on macroscopic mathematical formulations designed to admit a global user-equilibrium solution (DUE). Commonly, it is required to optimise the assignment in the presence of responsive demand patterns, signal settings, or congestion pricing tolls, requiring bi-level optimisation. In practice, heuristic methods may be required because of the intractability of the aggregate formulation, or its inability to represent an important feature of real traffic (eg Barceló and Casas, 2001). Traffic may be disaggregated to allow simulation of individual behaviour. 'Macroparticles', as Prof. Mahmassani calls them, have existed for many years in mesoscopic models, as a way of representing the natural disaggregation of traffic.

Mesoscopic assignment

At about the same time as TRANSYT, SATURN and other well known models like EMME, the CONTRAM mesoscopic assignment model was developed (Leonard and Gower 1982, Taylor 2003)[2]. For most of its existence, this model has been the only time-dependent assignment model in regular use for scheme appraisals, because of its ability to deal realistically with all planning scenarios except for optimising linked signal systems and describing the full geometry of motorway queues. Because of its simulation approach, however, it was never considered entirely 'respectable' by the academic

[1] Strictly, 'dynamic' ought to be reserved for methods which simulate real time behaviour.
[2] See also www.contram.com

community, although this attitude has softened as complex real world problems have declined to yield to formal methods. CONTRAM divides traffic into ‘packets’ containing one or more vehicles[3], the term being derived from an analogy with communications networks. Each packet is assigned to its own optimum route according to the time-variable network conditions relative to its start time, and any constraints such as banned movements. CONTRAM resembles an agent-based approach, in the sense that the packets are autonomous and could in principle have internal states, memory, etc, subject also to lacunae and error. For historical and practical reasons, their behaviour is usually limited to user-optimal with perfect perception, and cost functions which depend only on user class, but additional flexibility has been provided for modelling responses to ITS and travel time variability.

The methods used in practice are strongly influenced by the need for computational and data efficiency. In particular:

- Traffic modelling is event-based. It deals explicitly with significant events such as arrival at or departure from a queue, but only implicitly with travel along links or within queues. Non-queuing travel time is calculated using fixed speeds or speed/flow relationships. Events need not be modelled in time order, provided that the causal relationships between them are effected through iteration.

- Network volumes and capacities are averaged over each time slice, though in principle they could be made functions of time at the cost of additional complexity and memory use. Users can increase resolution as necessary by using shorter time slices.

- The delay each packet suffers at a junction is calculated deterministically by a time-dependent queuing model which uses aggregate volume and capacity data and incorporates random effects mathematically,. This reflects the assumption of FIFO discipline in queuing vehicles.

- Assignment is disaggregated to reflect the choice available to individuals. Packets are assigned by common rules (eg minimum generalised cost), using common information about network volumes and costs, though with provision for different user classes with different cost functions, access restrictions etc.

- The ITS modelling structure extends this by allowing individual packets to respond to some kinds of information (eg incidents, diversion routes). Without altering its basic structure this could be further extended to represent perceptions, attitudes and even memories of individuals, but for efficiency these would be represented conceptually as a ‘layer’ or ‘filter’ between the packet layer and network layer rather than using object methods.

Appropriate technology – meso-/micro- comparisons

For two decades, there was no effective competition to the mesoscopic approach to time-dependent modelling. However, it was not universally adopted because sometimes a static approach was considered sufficient, and the static models were generally better integrated with demand modelling software, emerging graphics, and other tools. It could also be difficult to assemble the large amount of data needed to construct reliable time-dependent origin-destination (OD) matrices. Finally, the calibration and computational burden of a time-dependent model is multiplied by the number of time periods represented, making such models expensive to construct and run. In the late 1980s, increasing computer power and graphics began to make microscopic simulation of networks practical, but methodology also developed to contribute to increasing problem size and complexity. In the UK, following a report by the SACTRA Committee (1997), it was recognised that the reaction of demand to changes in supply must be allowed for, and the effects of time variation ought to be dealt with

[3] But not exclusively, as, eg, pedestrian networks can also be modelled. Also, packet sizes need not be integral.

accurately, otherwise traffic growth and infrastructure provision would lead to over-prediction and under–prediction of congestion respectively. The UK Department for Transport does not *require* dynamic assignment in scheme appraisal, but it *does* now require justification of a decision *not* to use it. It has also developed a variable demand framework (software: DIADEM; advice: VaDMA) to ensure that trip generation and suppression are integrated into the modelling process[4].

In the past few years there has been renewed interest in time-dependent assignment, and several model producers have developed time-dependent tools compatible with their existing static software. DYNAMEQ, developed by INRO (2006), is of interest due to its ‘minimal’ approach. Each OD flow in a 15 minute time slice is assigned as a single block of traffic, iteratively to multiple paths. In an anecdotal example provided by INRO, networks of comparable size, around 240-300 zones, were modelled towards equilibrium by DYNAMEQ and by a microscopic simulation applied iteratively. DYNAMEQ simulated 200,000 vehicles covering a 3 hour peak, achieving a relative gap[5] under 5% after 40 iterations and 1.7 hours computation time. The microsimulation program is reported to have needed 720 hours to simulate 100,000 vehicles covering a 1 hour period, performing 50 iterations to achieve “reasonable” converged results[6]. On a slightly larger network of 315 zones and around 260,000 vehicles covering a 3 hour peak period, using ‘minimal’ disaggregation of one packet per OD per time slice, CONTRAM achieved a gap of under 3% after 6 iterations, taking about 1 hour on a high-specification PC[7]. These results confirm that mesoscopic models still have the edge in quantitative appraisal.

An issue from such comparisons is that, whereas microscopic simulation is normally expected to simulate every individual vehicle in the demand, the computation time of a mesoscopic model depends almost directly on the number of units into which the traffic is divided, which is largely under the control of the user. This economy of a mesoscopic model could come at the cost of reduced accuracy – but how much reduced? Another caveat is that delay can be very sensitive to the way flow is distributed during a congested period. What is the effect on accuracy of methods which either use time-averaged network values, or concentrate all flow in the middle of time slices, as in the examples above? These questions are addressed in what follows, but first we ask what level of accuracy is needed for appraisal.

What is significant error in appraisal?

There is no absolute standard of ‘significant error’ in appraisal. Equilibrium assignment models usually reckon their degree of convergence in terms of relative gap. Taking this as a measure of accuracy presupposes that the equilibrium solution they are aiming for is the true description, but gap can also be seen as a measure of model reliability. Gordon *et al* (2005) have concluded that a relative gap around 1% is required to deliver a reliable estimate of major scheme benefits in cases where the scheme is surrounded by a large catchment or buffer region. To appraise correctly a minor scheme embedded in a large network, error would need to be a fraction of 1%, which appears impractical. On the other hand, INRO has suggested that 5% gap is a realistic achievable criterion for time-dependent modelling. Uncertainty about overall traffic growth can be accommodated by low and high growth forecasts[8]. However, as far as the author is aware, there is no evidence on the degree of certainty *theoretically* achievable about traffic and economic values in large-scale appraisals. Uncertainty about local traffic values can be large: for example, under the same average conditions, day-to-day variation in equilibrium queue length in a congested peak can be 50%. The variance of queue lengths produced by time-dependent demand (see Taylor 2005) is only one component of uncertainty, of course, but it can have practical consequences. For example, one often hears that there was a massive jam at some

---

4 http://www.dft.gov.uk/stellent/groups/dft_econappr/documents/divisionhomepage/032181.hcsp
5 Defined as the difference between current route costs and optimal route costs under the same loading.
6 INRO User Group Meeting, London, 14 February 2006.
7 CONTRAM can require relatively few iterations because it completely replaces the loading during each iteration, rather than combining new and old loadings using a step length.
8 UK DfT recommends a low-to-high range of $2.5\sqrt{Y}\,\%$, where Y is the number of years forecast ahead.

point for no obvious reason. This could be due simply to the extended shape of the queue length probability distribution. It is likely that perception of the reliability of a network affects route choice and other travel decisions.

The effect of varying time resolution

Because most appraisals involve demand with some kind of peak or variable profile, many different examples of demand profiles could be found and a large sample would be needed to give 'typical' results of varying time resolution. Here we give only a simple and quite symmetrical example based on a small 5 zone network with a congested peak, modelled by CONTRAM. The original version of the model has a demand profile based on a central hour divided into four 15-minute periods, and two shoulder hours each divided into two 30-minute periods. This not only gives quite a smooth profile but focuses computation on the sensitive part of the peak. In two variants of the model, the profile is reduced to three 1-hour averaged periods, and to a single 3-hour overall average. The results are shown in Figure 1. Each reduction in resolution reduces the predicted total travel time by about 10%, and since delay is only part of travel time it is reduced even more, by about one third each time, so using a 'flat' profile underestimates true delay by a factor greater than two. This is far greater than the levels of uncertainty or non-convergence error cited in the previous section, and there is little compensating benefit of reduced run time.

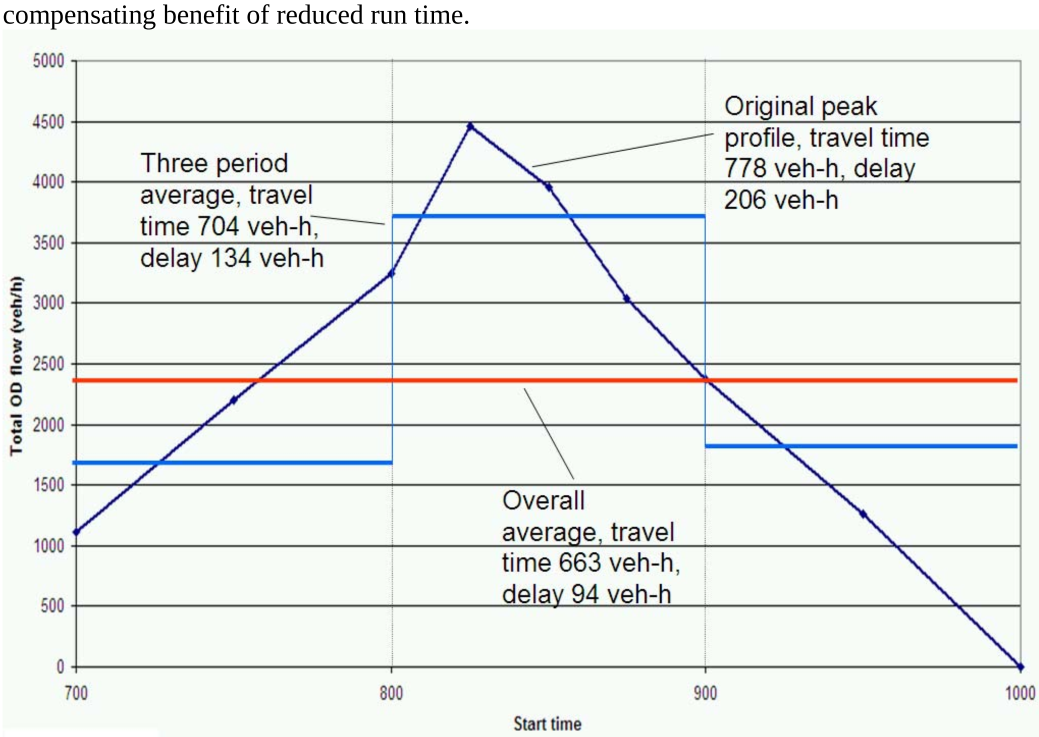


Figure 1. Effect of modelling a peak with reduced time resolution

Aggregation of OD flows in large models

Assignment modelling began originally in urban centres, but today is frequently applied to entire cities or interurban regions. As networks get larger the individual OD movements through them tend to get smaller. By illustration, the number of zones in a network model is broadly proportional to the number of links (found range 5-20 links per zone). The total traffic passing through a network is roughly proportional to the number of links since they have finite capacity, but the number of possible OD movements is proportional to the *square* of the number of zones, so the average size of each tends to be inversely proportional to network size. Large time-dependent OD matrices also tend to be highly synthesised, so small OD flows may be spread uniformly over many time periods, creating even smaller, often fractional, unit flows. Models with around 300 zones exist with *average* packet sizes of 0.2-0.3 vehicle, the smallest being *less than 0.01 vehicle*. The realism of this is questionable. In principle, such small flows represent the probability or frequency of a journey. In practice, it is doubtful whether they could be surveyed accurately enough, so their function becomes simply to distribute traffic uniformly in ignorance of the true OD pattern.

Because the computational cost of this is so high, we consider whether OD matrices could be aggregated systematically without introducing significant error. This could be done in several ways, for example:

- Combining small OD flows spread across several time slices;
- Combining zones according to their proximity and/or commonality of routes;
- Combining trips from the same origin in 'flights' (most static models do this).

Of these three, so far only aggregation over time has been tried – combining zones is tricky and constructing flights would require software changes, whereas aggregation over time can be achieved using quite simple external tools.

Aggregation over time – the SEURAT method

The first method has been implemented in a program called SEURAT, which performs the steps shown in Figure 2.

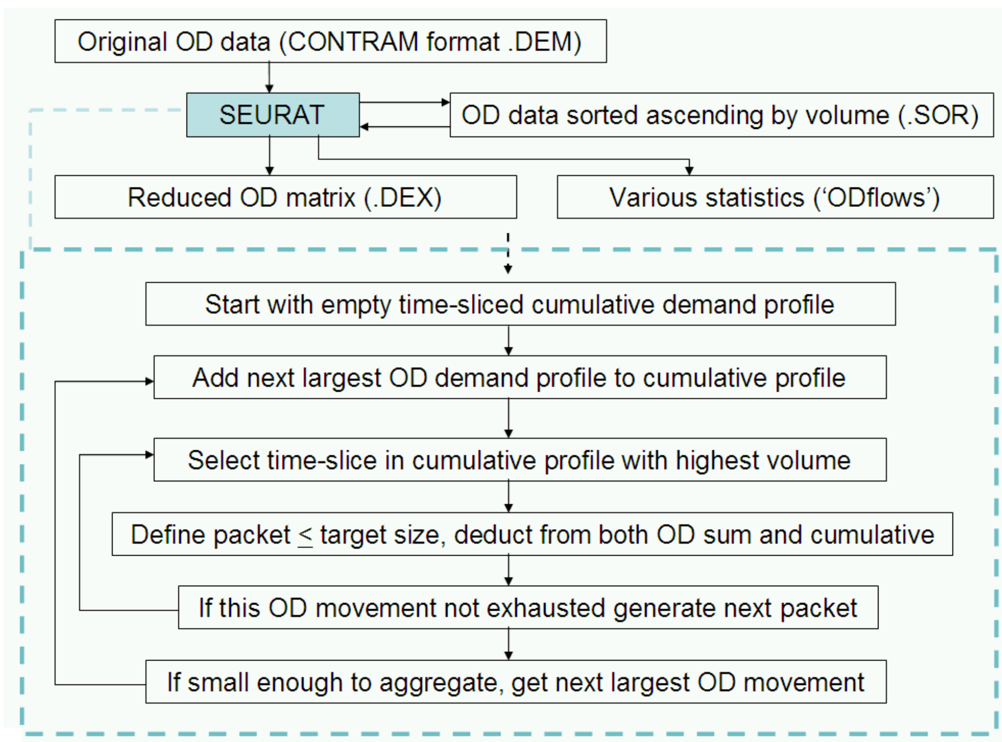
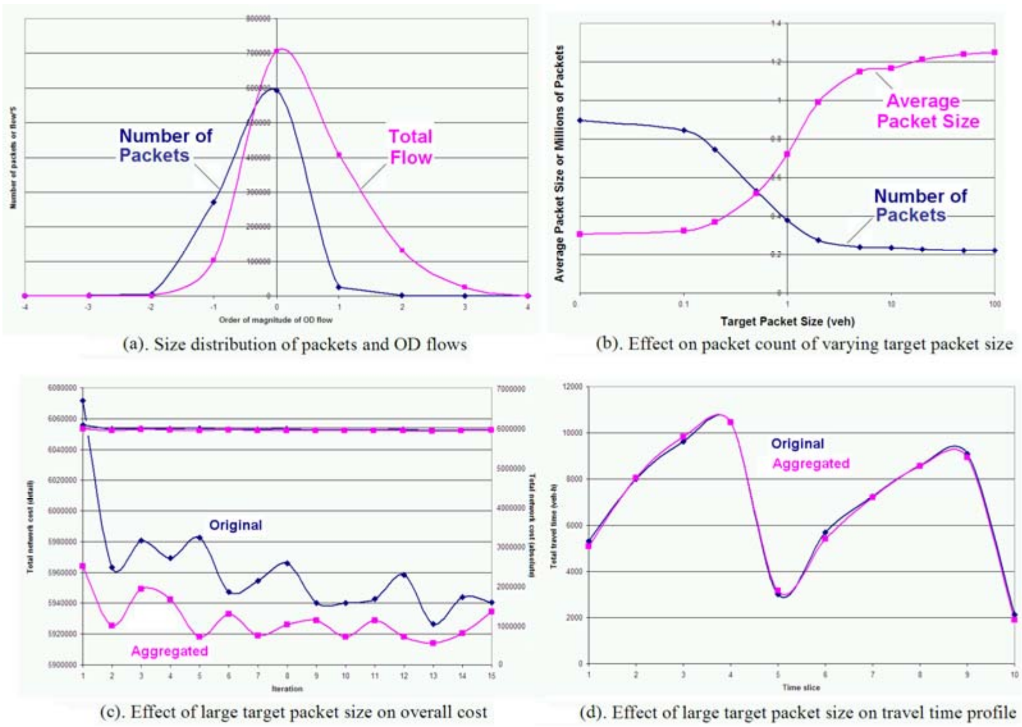


Figure 2. SEURAT flow diagram

The effect is to condense the OD flows, and distribute them to different times, while preserving the original overall demand in each time slice to high precision – the OD matrix being in effect approximated in *pointilliste* fashion. Some results are shown in Figures 3 (a-d). Figure 3(a) shows how in one city-region network a partly synthesised OD matrix includes a large number of small packets, which together contribute only a small fraction of the traffic but require the same computational effort as larger packets representing more important movements. Figure 3(b) shows that specifying a higher target packet size reduces the number of packets generated, though there is a practical limit in this case to the amount of aggregation achievable[9]. Using a large target value, Figure 3(c) shows that the effect on total network cost is small, around 0.5% (the lower pair of graphs is on a greatly magnified scale), and Figure 3(d) that the effect on the overall travel time profile is barely noticeable. On the other hand, run time is reduced by a factor of four and memory requirement more than halved. Saving of memory alone can have a major effect on run time if it avoids virtual memory being paged to disc.



(a). Size distribution of packets and OD flows

(b). Effect on packet count of varying target packet size

(c). Effect of large target packet size on overall cost

(d). Effect of large target packet size on travel time profile

Figure 3. Effects of aggregating small OD flows

SEURAT can be effective only where OD flows are initially distributed over many time slices. Combining zones is potentially more generally effective, but complex to perform. It must take account of physical barriers like rivers or railway lines, and should be guided by the degree of commonality of routes between OD movements, requiring a previous full assignment to establish typical route sets.

How much detail should be kept?

It was pointed out earlier that treating all demand as concentrated in the middle of time slices (minimal disaggregation) could affect accuracy. As an illustration, a city-region network referred to earlier has been modelled with its ODs in their original form, as reduced by SEURAT, and as reduced further by using minimal disaggregation. Statistics are given in Table 1, and graphs of convergence in Figure 4.

The original OD matrix has a large number of small movements, generating 1.15 M packets. It requires several hours to run on a high specification PC with >3 GHz processor and at least 1 GB of memory[10] (without gap calculation, to save time). Aggregation by SEURAT, which affects only *small* OD movements, produces a slight, ~0.1%, inaccuracy in the total demand while reducing the number

[9] Specifying a large target packet size is harmless since the process determines only aggregation of the matrix, not actual packet size when the model is run, only aggregation. As a result the number of packets launched by the larger OD movements tends not to be affected.

[10] Much of the memory requirement arises in the final iteration when generating tables of the OD/route skim and detailing the routes of all packets. This memory tends to escalate with increasing network size.

of packets by 77%, allowing the model to be run on a low-spec PC with 256 MB of memory. The error in total network cost compared to the 'full' assignment is an acceptable 1.37% and the gap 1.38%.

Table 1. Comparison of city-region results (315 zones and 3 hour modelled period in 15 minute time slices) with various levels of aggregation (run for 7 iterations).

Case:	**Original ODs**	**After SEURAT**	**Minimal Aggreg'n**
Packet size setting	automatic	automatic	large value
Demand (vehicles)	260608	260941	260941
Packets assigned	1151316	263238	173372
Average packet size	0.226	0.991	1.512
Routes used	497404	172506	149132
Network cost	6111902	6195417	6423958
Difference in cost %	(reference)	1.37	5.11
Stability %	1.14	1.01	3.09
Gap %	(not calculated)	1.38	2.88

The third case is where specifying a large target packet size at run time has forced the assignment to launch no more than one packet per OD movement per time slice, exactly in the middle of the time slice. This has the greatest effect on *large* OD movements, which would otherwise be split into several evenly-distributed packets. It almost quadruples the error in cost, and doubles the gap, while achieving only 8% *additional* saving of computation time.

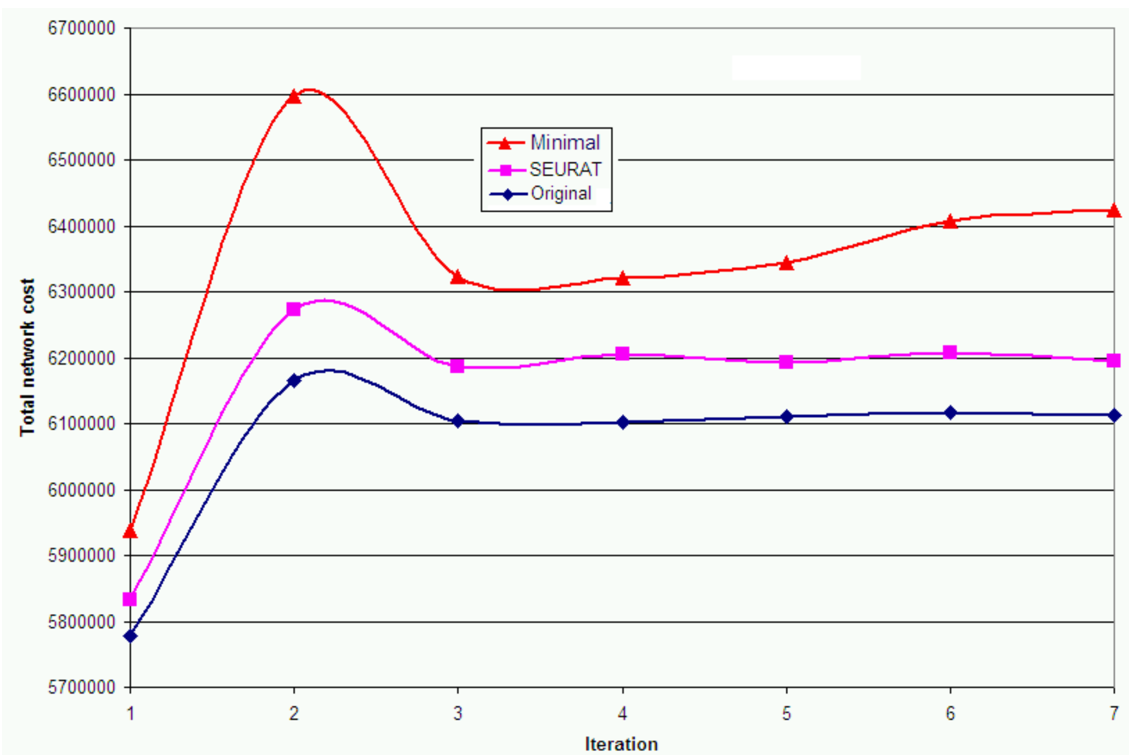


Figure 4. Convergence of total network cost for cases in Table 1

This demonstrates again the importance of accurately representing the way major demands are distributed through time. If all traffic is modelled to arrive on a network at the same moment, it tends to produce more delay than if spread out over time, as it would be in reality. However, it also demonstrates that accurate distribution of demand need not be costly in computation time provided it is targeted on the most significant demands.

Heuristics in route choice and assignment

A different approach to reducing run time is to make the process of assignment or convergence more efficient. Two facilities already exist in CONTRAM:

- Heuristic costs from points in the network to all destinations, to reduce the proportion of the network explored when searching for optimum routes;

- Packet splitting (multi-routing) combined with suitable algorithms to shift flow between the alternative routes available to each packet.

The first method is a form of what is sometimes called the 'A* algorithm' (Hart and Nilson 1968) although the present author developed it independently. It is a standard feature of the software and works automatically. The 'heuristic' costs are minimum free-flow travel costs from each network node to each destination. Since these never exceed real costs, they cannot affect routes chosen, but it is estimated they can reduce the number of link cost evaluations by a factor of 3 to10, provided that the network is not too congested. Cleverer heuristics would more nearly reflect actual travel costs, but could theoretically lead to wrong routes being chosen, so a label correcting algorithm might be required. Figure 5 shows the relationship between average number of links costed during route-finding and average number of links in the final route, for several networks ranging in size up to about 6000 links. The majority show good 'focus' around the final route (lower points), but in some cases, for reasons not yet understood, the heuristics are less effective (two upper points).

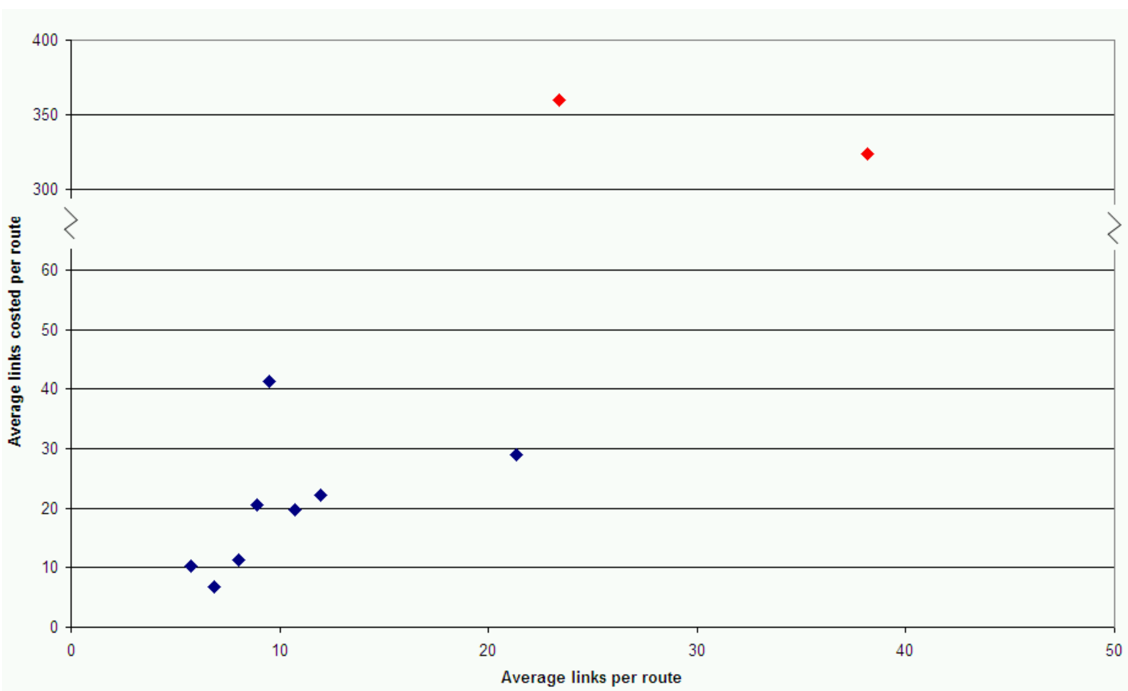


Figure 5. 'Focus' of link costing around final route, using heuristic costs to destination

Packet splitting was introduced with the aim of improving convergence of assignment when iterated with matrix estimation or variable demand modelling. Instead of changing the whole route of each packet at each iteration, packet splitting adds new routes to the 'packet set' as they arise (as the current minimum cost routes) and shifts flow between them so as to bring alternative route costs into equilibrium. All component route flows are launched at the same model time, and the total flow of each packet set remains fixed. Initial results suggest that packet splitting may be most effective in improving the performance of models with minimal disaggregation.

Several flow shifting algorithms, including a form of MSA volume-averaging [11], have been implemented. Unlike in 'standard' CONTRAM, each requires a step size parameter, which is currently a user-definable constant, but a well-known problem with MSA is its slowness of convergence. These methods are being researched as resources permit. In a simulation type model, it is difficult to estimate descent slopes so as to optimise step lengths. Further work is needed to 'shake down' the choice of flow-shifting algorithms and to develop a method of calculating efficient step lengths. Too small a step length could not only prolong run time but fail to find the optimum solution. Larger step lengths can avoid local minima, but initial results suggest that there may be little headroom before instability develops, so a mechanism is required to forestall this.

A similar problem arose some years ago when the idea of general parallel computing was popular, stimulated by the availability of the T805 Transputer and later the Intel I860. Initial experiments (Greenwood and Taylor 1993) led to the idea of developing assignment software which could run on a conventional serial computer but could be ported easily to a parallel virtual machine running on any multi-processor or distributed network. In the implementation known as CONTRAM6[12] (Taylor and Carmichael 1995), packets were replaced by 'tranches' of traffic, each occupying one time slice but with multi-routing and allowance for costs to vary over the period of the time slice. Tranches were assigned logically in parallel, so a step-size algorithm was needed to combine the results of current and previous iterations. Like the serial version, there was no overall objective function to be minimised, and the only linkage between different OD movements and time slices was through an aggregate network traffic model. The model maintained a stored 'gain' value independently for each time-sliced OD movement, which determined the amount of *split* (ie proportion of total tranche flow) moved between pairs of alternative routes according to the formulae:

> Initial gain = (a constant) / (number of routes * cost difference between two cheapest routes)
>
> IF a new minimum cost route is found OR virtually all flow uses one route THEN gain is decreased by a common factor ELSE gain is increased by a common factor, subject to not exceeding a ceiling (eg the initial gain) unless gap has also decreased
>
> Δ split (higher to lower cost) = gain * (cost difference between pair of routes)

A feature of this algorithm is that step-length (gain) is deliberately made non-convergent. By analogy, one could say that the 'temperature' is kept high to agitate the system, unlike MSA where the system is steadily 'cooled'. The result was quite rapid but somewhat temperamental convergence. Despite encouraging initial results, the work lapsed as the prospect of general parallel computing faded, while interest in non-equilibrium ITS modelling increased, for which the packet approach appeared more suitable. The cost of optimising the convergence algorithm was also a deterrent, whereas serial CONTRAM appears to converge naturally without forcing algorithms or factors which could influence the solution. However, the issue is reappearing in relation to packet splitting.

Heuristic methods in optimisation and appraisal

Any iterative assignment or bi-level model, in which equilibrium is an emergent outcome of an interplay of individual objectives and preferences, can be seen as representing learning or adaptive behaviour. However, there is no certainty that real travellers acquire either sufficiently reliable knowledge, or optimal strategies, to produce a user-optimal assignment or even a stable assignment. This scepticism is embodied in a microscopic simulation model called DRACULA, developed by the University of Leeds, which simulates learning explicitly (Liu *et al* 200?). Another non-equilibrium concept is 'bounded rationality' where the act of researching and choosing an improved route involves an unpredictable cost which the user may not choose to bear (eg Jayakrishnan *et al* 1994). Real

[11] After iteration n: AveragedFlow(n) = (1-1/n) AveragedFlow(n-1) + (1/n) NewFlow(n). This averages the results of all iterations which have been performed, so each iteration adds less information than the one before.
[12] Despite the name, the program shared no code with the serial CONTRAM program.

travellers, working with incomplete information, undoubtedly use simplified models, stored experience and heuristic short cuts to make route choices and other decisions[13]. One result could be alternative 'sub-optimal' flow patterns between which traffic can switch unpredictably. If so, at least some of the uncertainty of modelling results could be inherent in the travel-choice process rather than linked to the more regular uncertainties in data. It is important to appreciate all 'natural' sources of uncertainty and variability, to avoid devoting excessive human and computational effort to data definition and model runs in an attempt to show a theoretical scheme benefit which may not be measurable in reality.

Conclusion

Given the uncertainties in traffic, there may be an optimum level of human and computational effort devoted to data definition and model runs when analysing networks and estimating scheme benefits. In spite of increasing computer power, and even some parallel processing, many models are now so large that each run using time-dependent modelling can take a day or longer. The potential benefit of reducing this even by a factor in single figures is considerable. That mesoscopic modelling can achieve this through improved algorithms and aggregation methods demonstrates its continuing value when used appropriately. This inevitably calls into play heuristic methods which are neither formally provable nor microscopically exact. Appropriate use depends on balancing the content and detail at each level of the modelling process with the necessary or practically achievable accuracy of data and results. There is potential for research on improved heuristic methods, although not all of it will be fruitful. By the nature of the problems addressed, it is impossible to predict which approaches will turn be the most successful.

References

Barceló J and Casas J (2001). Heuristic dynamic assignment based on microscopic traffic simulation. Universitat Politècnica de Catalunya.

Binning J C, Crabtree M and Burtenshaw G (2003). *TRANSYT 12 User Guide*. TRL Application Guide AG 48 (Issue B). TRL, Crowthorne House, Wokingham, UK.

Greenwood D G and Taylor N B (1993). Parallelising the CONTRAM traffic assignment model. *Proc. ESM'93 Simulation Multiconference*, Lyon, June 1993.

Gordon A, White C, Oladeinde F and Collins J (2005). Developments in variable demand modelling. *Third PTRC Transport Practitoners' Meeting.*

Hart P E and Nilson N J (1968). A formal basis of the heuristic determination of minimum cost paths. *IEEE Transactions of Systems Science and Cybernetics* 4(2), 100-107.

INRO (2004). *Calibration and application of a simulation-based dynamic traffic assignment model.* INRO, Montréal, Canada.

INRO (2006). *DYNAMEQ White Paper*. INRO, Montréal, Canada. See http://www.inro.ca/en/products/dynameq/whitepaper_req.php

[13] At the time this was being finalised, a management magazine published a brief article on heuristics describing them as "invisible baggage". Having defined four types of heuristic, it ended with the statement "Taken together, this provides an insight into how complex a reasoning processor the human brain really is – and how fallible!" (Jay Redfern, Project Magazine, July 2006). This seems to miss the point of whether it is possible, at practical cost, to determine the 'correct' answer to a complex problem, by which to judge the 'heuristic' solution.

Jayakrishnan R, Mahmassani H and Hu T-Y (1994). An evaluation tool for advance traffic information and management systems in urban networks. *Transportation Research* **2**C, 129-147.

Kimber R M and Hollis E M (1979). *Traffic queues and delays at road junctions.* TRL Report LR 909. TRL, Crowthorne House, Wokingham, UK.

Leonard D R and Gower P (1982). *User Guide to CONTRAM Version 4.* TRL Report SR 735. TRL, Crowthorne House, Wokingham, UK.

Liu R, van Vliet D and Watling D (200?). *DRACULA: a day-to-day microscopic framework for modelling traffic network.* Institute for Transport Studies, University of Leeds, UK. http://www.its.leeds.ac.uk/software/dracula/tnote3.pdf

SACTRA (1997). *Transport and the economy*. Department for Transport, London.

Taylor N B and Carmichael C (1995). Distributable time-dependent traffic assignment. *Proc. PTRC Summer Annual Meeting 1995*, Seminar E, 129-142.

Taylor N B (2003). The CONTRAM dynamic traffic assignment model. *Networks and Spatial Economics Journal – special issue on Dynamic Traffic Assignment,* **3(2003)**, 297-322. Kluwer.

Taylor N B (2005). Variance and accuracy of the sheared queue model. *Proc. IMA Mathematics in Transport Conference, London, 7-9 September 2005*. UCL.

van Vliet D (1982). SATURN – A modern assignment model. *Traffic Engineering and Control*, **23**, 578-581, December 1982.

Vincent R A, Mitchell A I and Robertson D R (1980). *User Guide to TRANSYT Version 8.* TRL Report LR 888. TRL, Crowthorne House, Wokingham, UK.