Practical Modelling of trip rescheduling under congested conditions

Paper presented at 45th European Congress of the Regional Science Association
Aug. 23 - 27, 2005, Amsterdam, special session: "Choice analysis (N5)"

John Bates, Transport Planning Consultant, 1 Picklers Hill, Abingdon, OX14 2BA, UK
email: 113300.3565@compuserve.com

ABSTRACT

There is plenty of anecdotal evidence that drivers may make small changes in their time of travel to take advantage of lower levels of congestion. However, progress in modelling such “micro” re-scheduling within peak period traffic remains slow. While there exist research papers describing theoretical solutions, there are no techniques available for practical use. Most commonly used assignment programs are temporally aggregate, while packages which do allow some “dynamic assignment” typically assume a fixed demand profile.

The aim of the paper is to present a more heuristic method which could at least be used on an interim basis. The assumption is that the demand profile can be segmented into a number of mutually exclusive “windows” in relation to the “preferred arrival time”, while on the assignment side, independently defined sequential “timeslices” are used in order to respect some of the dynamic processes relating to the build-up of queues. The demand process, whereby some drivers shift away from their preferred window, leads to an iterative procedure with the aim of achieving reasonable convergence.

Using the well-known scheduling theory developed by Vickrey, Small, and Arnott, de Palma & Lindsey, the basic approach can be described, extending from the simple “bottleneck”, to which the theory was originally applied, to a general network. So far, insufficient research funds have been made available to test the approach properly. It is hoped that by bringing the ideas into the public domain, further research into this area may be stimulated.

Keywords: rescheduling, congestion, practical

Acknowledgement:
Part of the work reported here was developed in the course of a project for the UK Department for Transport, and their support is gratefully acknowledged. However, this paper is the product of the author’s views alone, and does not necessarily represent those of the Department.

Practical Modelling of trip rescheduling under congested conditions

John Bates, Transport Planning Consultant, Abingdon, UK

1 Introduction

The aim of this paper is to assist in the practical specification of models of transport demand and supply which can be used as planning tools. While much of practical transport modelling remains dominated by networks, often with assumptions of fixed demand at any given time, there has been a growing acceptance of the fact that the (generalised) costs of travel affect the level of demand, whether these costs derive from deliberate policy (eg road pricing) or from capacity changes.

For practical purposes, therefore, a model needs to reflect both demand and supply effects, and, while its validity may be disputed, a convenient working assumption is that the two must be in equilibrium. In other words, the generalized cost that gives rise to a given pattern of demand must be compatible with the network's ability (largely related to highway congestion) to accommodate the demand at that cost.

Prompted initially by the findings of the SACTRA (1994) report, there has been ongoing effort in the UK to develop reasonably standardized modelling approaches which allow for demand responses in a way that is accessible to practical modellers. There has been reasonable agreement about the need to reflect the responses of frequency, distribution (or destination choice) and mode choice within the demand model, as well as route choice (though this is handled routinely within the assignment module). It has proved more difficult to make progress on the choice of time of travel, but here too some general guidelines are being achieved.

One of the key issues of time of travel choice is the **level** at which it may be considered. Following a classification originally due to Bates (1996), it has become common to make a distinction between “macro” and “micro” shifts, where “macro time-shifting” allows for the possibility of transferring between defined broad periods (typically 2 or 3 hours), as between peak and interpeak, and “micro time-shifting” may be defined as relatively small changes in arrival and/or departure time. Generally, micro-shifts are motivated by changes in the temporal **profile** of journey times - ie, the variation in road journey time between a fixed origin and destination, dependent on the exact departure time. If travellers have a preferred arrival (or departure) time, they will only shift from this if they receive some benefit in the form of reduced travel times. Typically, this will have implications for the **arrival** time, which, at least in the morning peak, is likely to be critical.

Such variation in travel times is predominantly a manifestation of the build-up and dispersal of **queues** at various points in the network. Because this is essentially a dynamic process, it is not possible to represent it using “standard” assignment methods, which are “time-aggregate” - some account of the development and dispersal of queues is required. Ideally, this should be done using a fully dynamic assignment in continuous time.

There is plenty of anecdotal evidence that drivers may make small changes in their time of travel to take advantage of lower levels of congestion. However, progress in modelling such “micro” re-scheduling within peak period traffic remains slow. While there exist research papers describing theoretical solutions, there are no techniques available for practical use. As noted, most commonly used assignment programs are temporally aggregate, while packages which do allow some “dynamic assignment” typically assume a fixed demand profile.

At the same time, the main demand responses noted earlier (destination choice etc) are also typically modelled on a time-aggregate basis, even if broad distinctions (eg between peak and off-peak) are reflected.

There are thus two essential questions. Firstly, how can a better understanding of the build-up of congestion within the peak be obtained, using reasonably accessible models? Secondly, are there potential errors imported into the overall demand-supply modelling process by not taking account of the profile **within** congested periods? The standard approach is to take a two or three hour matrix for the am peak, factor it to give a single hour (with practice varying between average and "peak") and assign it, making use of standard congested network algorithms (equilibrium assignment).

The aim of the paper is to present a more heuristic method which could at least be used on an interim basis. Currently this remains untested. It is hoped that by bringing the ideas into the public domain, further research into this area may be stimulated.

The key assumption is that the demand profile can be divided into a number of mutually exclusive "windows" in relation to the "preferred arrival time", while on the assignment side, independently defined sequential "timeslices" are used in order to respect some of the dynamic processes relating to the build-up of queues. The demand process, whereby some drivers shift away from their preferred window, leads to an iterative procedure with the aim of achieving reasonable convergence. There are some difficult decisions to make as to what it means to say that demand is allocated to any particular time period – do we mean that it **starts** within that period?

Using the well-known scheduling theory developed by Vickrey (1969), Small (1982, 1992), and Arnott, de Palma & Lindsey (1994), the basic approach can be described, extending from the simple "bottleneck", to which the theory was originally applied, to a general network.

2 Preliminaries

2.1 Theoretical Background

The key starting point is the “schedule delay” formula, initially developed by Vickrey (1969), and further extended in the work of Small (1982, 1992), and a series of papers by Arnott, de Palma & Lindsey (ADL: eg ADL (1994)). A major review is given in Bates (1996), available on the UK DfT Website. However, the key aspects will be briefly presented here.

In the earliest expositions, all travellers wish to arrive at the same “preferred arrival time” (PAT), and the system is treated as having a single O-D pair. As long as the capacity of the network is sufficient, all travellers can arrive at PAT. However, once capacity problems occur, this is no longer possible, and some travellers will be early or late.

The schedule delay formula is a functional form for the utility of arriving at times other than the PAT, taking account of the possible time advantages of so doing. If we denote τ as the actual arrival time, and $\xi(\tau)$ as the travel time for those who arrive at τ, then by far the most popular proposal for this utility is that due to Small (1982), a development of Vickrey (1969), whereby

$$U(\tau) = -\alpha\,\xi(\tau) - \beta\,\text{SDE} - \gamma\,\text{SDL} - \delta\,d_L \qquad (1)$$

where all four terms α, β, γ, δ are **positive**, and the terms SDE ("Schedule Delay Early"), SDL ("Schedule Delay Late"), and d_L ("dummy (0,1) for late arrival") are defined as

$$\text{SDE} = \text{Max}(\text{PAT}-\tau, 0) \qquad (2a)$$

$$\text{SDL} = \text{Max}(\tau-\text{PAT}, 0) \qquad (2b)$$

$$d_L = 1 \text{ if } \tau > \text{PAT},\ 0 \text{ otherwise.} \qquad (2c)$$

Note that the terms in the utility function [β SDE, γ SDL, δ d_L] give the variations in utility associated with each possible arrival time *per se*: the sum of these terms constitute the **schedule utility**. Clearly this is at a maximum (of 0) when τ = PAT.

It may be noted that δ, which represents a penalty for being late *per se*, is in fact omitted from many of the studies using this general formulation: implicitly, it is set to zero or subsumed within the γ parameter. For reasons of simplicity we henceforth ignore the δ parameter.

If now we maximise the utility with respect to the arrival time τ, we obtain the well-known key demand-side first order conditions on the gradient of travel time $\xi'(\tau)$:

for early shift (τ < PAT) $\xi'(\tau) = \beta/\alpha$ (3a)

for late shift (τ > PAT) $\xi'(\tau) = -\gamma/\alpha$ (3b)

The interpretation of these conditions is that **if** shifting is to take place, the network will need to deliver the appropriate travel time gradients. It may be noted that in most of the theoretical work (by the authors previously cited), a particularly simple form of network is used - the so-called "bottleneck" model, whereby free-flow times are maintained as long as capacity is not exceeded, and thereafter a deterministic queuing process begins, in which the (additional) travel time is directly proportional to the length of the queue. This allows the departure time profile to be derived analytically. With a "real" network, this is no longer possible, and an iterative procedure is required.

As noted, the theoretical work tends to assume a homogeneous population in respect of PAT and the utility parameters, though work has been done by Small and ADL to relax these restrictions. Ideally the demand should be expressed in continuous time, but, as noted earlier, for practical purposes we will assume that it can be expressed in terms of discrete "windows" of PAT.

For convenience of exposition, we assume that total demand over the whole of the peak period is fixed, and that outside this period (in what we refer to as the pre-peak and the post-peak), the level of congestion is not affected by variations in demand, though we allow for the fact that it could be **different** between the pre- and post-peaks.

2.2 Notation

Based on the recommendations of Bates (1996), it is crucial to avoid any confusion between variables which are indexed by departure time and those which are indexed by arrival time, and this approach is therefore followed here. In essence, the key notation is set out in the following paragraphs:

As a general convention, we use "t" to indicate departure time, and "τ" to indicate arrival time. The difference between these, $\tau - t$, represents the journey time, or, perhaps better, journey duration.

Viewed from the arrival point of view, we write the journey duration, given arrival at time τ, as $\xi(\tau)$, and correspondingly, viewed from the departure point of view, we write the journey duration, given departure at time t, as $\Theta(t)$. The fundamental linking identities can then be written as:

$$\tau \equiv t + \Theta(t) \qquad (4a)$$

$$t \equiv \tau - \xi(\tau) \qquad (4b)$$

$$\xi(\tau) \equiv \Theta(\tau - \xi(\tau)) \qquad (4c)$$

$$\Theta(t) \equiv \xi(t + \Theta(t)) \qquad (4d)$$

We assume a base demand matrix for the peak period **T**, in the sense that if free-flow times prevailed, this is the level of demand which we would assign using an time-aggregate approach. It is convenient (though not entirely uncontroversial!) to act as if this represents all demand wishing to **arrive** within the peak period.

Since **all** calculations are on a matrix level, we will suppress any “ij” notation, but it is implicit throughout.

3 Outline of the Approach

3.1 Temporal disaggregation

3.1.1 Segmenting by PAT

We now assume that this base demand is segmented by PAT bands or “windows”. The notation allows for any number of such bands within the peak period, and we index the bands as “k”. There is an implication that the bands need to be relatively narrow, and there is probably value in keeping the intervals the same width. Then band k is defined on the interval $J_k = [PAT_{k1}, PAT_{k2}]$, and the demand that falls within this band will be written as A_k. It is implied that:

$$\Sigma_k A_k = T \qquad (5)$$

and that the bands cover the whole peak period.

The split into matrices $\mathbf{A}_k$ is independent of any network considerations, and is thus fixed. We are ignoring here **how** it would be done in practice, and from now on, we assume that the split has been achieved. For convenience, we may assume that the distribution **within** each PAT window is uniform.

3.1.2 Assignment

In order to introduce some “dynamics” into the procedure, a sequential assignment procedure is carried out for different timeslices within the peak period. The actual way in which this is done will depend on the assignment program. Here we try to give a general description.

It is convenient to assume that the separate time-slices relate to **departure periods**, though it is recognised that there are some questions of interpretation here. We will treat this as a technical issue, and not discuss it further at this stage.

There is no requirement, in principle, for the assignment time-slices to bear any relationship to the PAT windows, and we treat them quite independently. We index the assignment time-slices by "r". Once again, they need to be defined as abutting intervals which we write as $I_r = [t_{r1}, t_{r2}]$. For each assignment time-slice, we require a demand matrix T_r and we assume that the internal dynamics (eg queue-passing) are effectively handled, so that the assignments of successive time slices are **not** independent.

The result is then a series of cost matrices which, since we concentrate on the journey duration **only**, we write as $\boldsymbol{\Theta}_r$. In addition, we have the fixed matrices $\boldsymbol{\Theta}^1$ and $\boldsymbol{\Theta}^2$ from the pre-peak and post-peak assignments.

3.1.3 Time Period Choice

It is useful to view the departure time model as essentially carrying out the following task:

for each PAT segment k, calculate the proportion $p_{k,h}$ of total demand A_k allocated to each **Arrival time window** h

Although it will need to be checked in actual circumstances, we expect the profile of continuous journey times $\xi(\tau)$ to gradually rise to a peak value, and thereafter decline. We will denote the arrival time associated with the highest $\xi(\tau)$ as τ^*. Moreover, we may expect that the gradient on the "early" side ($\tau < \tau^*$) will typically be shallower than that on the late side – this is related to the general expectation that $\beta/\alpha < \gamma/\alpha$, reflecting that most travellers would rather be early than late.

Consider the behaviour of travellers in PAT segment k, with interpolated average travel time ξ_k. These travellers have three possible choices for their acceptable arrival times h:

within PAT window for segment	h =k
earlier window	h < k
later window	h > k

Which option they choose will depend on the gradient of ξ (and, of course, their values of the scheduling parameters – we are here assuming homogeneity in this respect).

Assuming a uniform interval size J for each PAT window, there will be no late shifting

if $\gamma\,(h–k)\,J + \alpha\,\xi_h < \alpha\,\xi_k$ for h > k,

and no early shifting if $\beta\,(k–h)\,J + \alpha\,\xi_h < \alpha\,\xi_k$ for h < k

In other words, for early shifting (h < k), we require:

$$\alpha\,(\xi_k - \xi_h) \le \beta\,(k–h)\,J \quad \Rightarrow \quad (\xi_k - \xi_h) \ge \beta/\alpha\,(k–h)\,J \qquad (6a)$$

while for late shifting (h > k), we require:

$$\alpha\,(\xi_h - \xi_k) \ge -\gamma\,(h–k)\,J \quad \Rightarrow \quad (\xi_h - \xi_k) \le -\gamma/\alpha\,(h–k)\,J \qquad (6b)$$

This, of course, rules out, on the early side, any shifting if the gradient ξ' is negative, and on the late side, any shifting if the gradient ξ' is positive.

While these conditions are straightforward to derive, their implications are less straightforward. Suppose, for example, that we are on the early side, considering the choices for PAT segment k. The gradient $(\xi_k - \xi_{k-1}) \ge \beta/\alpha\ J$, and the gradient $(\xi_k - \xi_{k-2}) \ge 2\ \beta/\alpha\ J$. However, the "incremental gradient" $(\xi_{k-1} - \xi_{k-2}) < \beta/\alpha\ J$.

This makes it clear that the conditions just stated are not in fact complete. In these circumstances, travellers will not shift all the way to window (k–2), since having got as far as (k–1), the further shift cannot be justified. Essentially, this is a consequence of dealing with discrete intervals, rather than requiring a continuous condition on the gradient. What is actually happening is that, somewhere between the midpoints of arrival segments (k–1) and (k–2), the gradient falls below the critical value. Without doing further (non-linear) interpolation, it is not obvious how much, if any, of the total demand for PAT segment k should be allocated to arrival segment (k–2). This uncertainty implies that we should limit our expectations of the level of convergence that may be attained.

The essential equilibrium conditions on the demand side can be stated as:

for each PAT segment k

$$\overline{V}_{kh} \leq \overline{V}_k * \text{ if } p_{k,h} = 0$$

$$\overline{V}_{kh} = \overline{V}_k * \text{ if } p_{k,h} \geq 0$$

where

$$\overline{V}_{kh} = \begin{cases} \beta\,(k-h)\,J + \alpha\,\xi_h & \text{if } h < k \\ \alpha\,\xi_k & \text{if } h = k \\ \gamma\,(h-k)\,J + \alpha\,\xi_h & \text{if } h > k \end{cases} \qquad (7)$$

It will be seen that these correspond with the standard Kuhn-Tucker conditions for a constrained optimisation problem.

These conditions also have to be compatible with the travel times delivered by the network, given the pattern of departure times. What we therefore need is a procedure which will a) deliver the quantities ξ_h and b) allow the quantities $p_{k,h}$ to be calculated.

By summing over PAT windows, we can then obtain the implied actual arrival demand for each arrival time window h. This can then be translated into the departure demand timeslices r, allowing for the travel time ξ_k.

3.2 General algorithmic approach

3.2.1 The equilibrium principle

An analogy can be constructed with equilibrium assignment in route choice. Suppose we enumerate all possible routes, applying some “sensible” criterion to avoid “cycles”. The equilibrium conditions tell us that all routes actually used must have the same cost. But we do not deduce from this that **all** the enumerated routes have the same cost! Rather, we have to compute how many of the enumerated routes need to be used.

Another useful way of envisaging **both** problems (ie route choice and departure time choice) is along the following lines. Assume a fixed shape matrix but allow the total demand to be variable, and consider what happens as we allow it to increase. When the total demand is low, relative to the network capacity, all vehicles will select the single minimum cost route for each O-D pair, and concomitantly, all vehicles will choose the departure time which allows them to arrive at their PAT. As total demand increases, the performance of the (free-flow) minimum cost route will gradually deteriorate, until it is equal to the 2^{nd} best route, and at this stage both routes will be brought into use. These two routes will then “deteriorate” at the same rate until they reach the cost of the third best route etc, so that the size of the set of used routes depends on the volume of demand. Precisely similar developments relate to departure time choice – the window of acceptable arrival times expands with the volume of demand.

In practice, of course, the interdependence of links in a network makes this more complex, so that, for example, what was the third best route under free-flow conditions need not be the same as the third route which is actually brought into use as the volume increases. Further, we have appealed to the network “capacity” which is difficult to define in practice. Nonetheless, none of this subtracts from the essential validity of the principle described.

We have described the "raw material" of the procedure, and what is now required is to describe the interfaces which permit the algorithm to proceed. For ease of illustration, we will assume a single homogeneous segment of demand with respect to the schedule parameters α β γ. However, the approach can in principle be extended to multiple populations with different schedule parameters.

3.2.2 Interfaces

The equilibrium conditions were set out in 3.1.2 above, in terms of the travel time ξ_h, that is, the travel time viewed from the standpoint of the arrival time window. However, we do not have direct access to this – we only have the values from the assignment, which relates to different period definitions. Since we have assumed that these yield Θ_r rather than ξ_r, these need to be converted (essentially using Identity 4d above). The details of this conversion will depend on the assignment package: for example, with CONTRAM the actual arrival times τ are known, so ξ can be obtained directly. However, we treat this as a technical issue which is in principle soluble. Here we are effectively assuming that a package such as SATURN will be run for successive "time-slices", with "queue-passing".

Assuming therefore that we have ξ_r relating to an arrival time τ_r for each assignment time-slice, we now use linear interpolation to translate these into the required ξ_k values defined at the midpoint of each PAT segment. We can proceed to calculate the utilities for each arrival timeslice, and hence the allocation of demand, separately for each PAT segment.

Because of the assignment time-slice procedure, a further calculation is required to map the contributions of each PAT segment to the assignment timeslices. We are now implicitly working in terms of **departure** times. Summing over each PAT segment k, we calculate what proportion of the total demand A falls in each assignment time-slice r, thus giving us the required assignment matrices $\mathbf{T}_r$.

We may illustrate the essential progress of the iterations as follows:

$\mathbf{T}_r \xrightarrow{\text{assign}} \mathbf{\Theta}_r \xrightarrow{\text{convert}} \boldsymbol{\xi}_r \xrightarrow{\text{interpolate}} \boldsymbol{\xi}_k \xrightarrow{\text{arrival time choice}} \mathbf{\Omega}_k \xrightarrow{\text{convert/interpolate}} \mathbf{T}_r$

The quantity $\mathbf{\Omega}_k$ represents the cumulative distribution of arrivals at time τ. The corresponding cumulative distribution of departures at time t is written as Q_t.

There are three essential elements in the procedure.

Firstly, the skimmed quantities from the assignment need to be converted to arrival time windows: this involves the use of identity (4a) and interpolation.

Secondly, separately for each of the PAT windows, a method is required to allocate the total demand among the possible arrival time windows. Given the inherent approximation in the method, a straightforward procedure such as MSA[1] would seem to be appropriate. To implement this, within each iteration n the arrival time with maximum utility must be chosen, and the demand then averaged with the estimate from the preceding iteration according to the proportion 1:n–1. After adding across all PAT windows, this leads to an estimate of the total allocation to each arrival window.

Thirdly, the implied cumulative distribution of arrival times needs to be converted back to a departure time window basis: this again involves interpolation, together with identity (4b).

3.3 Proposed Algorithm (using MSA)

In any iteration, assume that we have travel time estimates ξ_k for each PAT segment k: these have been obtained by a) conversion of the assignment cost matrices Θ_r to ξ_r by associating the costs with the appropriate **arrival** times, and b) interpolating across the entire peak period.

[1] "Method of Successive Averages"

Then, for each segment k, evaluate the “average” utility for each arrival window h:

$$
\begin{aligned}
\overline{V}_{kh} &= \beta\,(k\text{–}h)\,J + \alpha\,\xi_h && \text{if } h < k \\
&= \alpha\,\xi_k && \text{if } h = k \\
&= \gamma\,(h\text{–}k)\,J + \alpha\,\xi_h && \text{if } h > k
\end{aligned}
\qquad (8)
$$

Find h for which $\overline{V}_{kh}$ is maximised, and allocate the entire demand for segment k to the “auxiliary” (or “target”) estimate:

$$Y_{k,h} = A_k \text{ for h min, 0 otherwise} \qquad (9)$$

Combine with previous estimates $X_{k,h}$ using “MSA” weights

$$X'_{k,h} = (1\text{–}\lambda).\ X_{k,h} + \lambda.\ Y_{k,h} \qquad (10)$$

where, according to the MSA approach, λ is taken as the reciprocal of the iteration number. Note that this maintains the property that, for each PAT segment k, $\Sigma_h\ X'_{k,h} = A_k$, ie all the base demand is allocated to some arrival window h.

The “segmented arrival time matrices” $X'_{k,h}$ need to be stored for the following iteration. However, for the assignment the PAT index k is not required. Thus we proceed by calculating:

$$X'_{*,h} = \Sigma_k\ X'_{k,h} \qquad (11)$$

This gives the total arrival demand in each arrival window h. We now need to translate this back to departure time windows. This can be done by means of the interpolated values of ξ at the boundaries of the arrival time windows. For example, for arrival time window h, the boundaries are [PAT_{h1}, PAT_{h2}], so that the corresponding departure time points are [$PAT_{h1} – \xi(PAT_{h1})$, $PAT_{h2} – \xi(PAT_{h2})$]. This gives the cumulative demand in terms of the departure time.

We can now interpolate again to apportion the curve to the departure time windows. The difference between the cumulative curve at the start- and end-points of each window gives the amount of demand to be assigned in that timeslice.

In this way, all the demands $X'_{*,h}$ are allocated to a departure timeslice. Note that some of the allocations may be **outside** the span of the peak assignments: these will not be assigned, but assumed to have the appropriate journey time characteristics of the pre- or post-peak.

The temporally dependent assignments are now carried out, yielding new estimates Θ_r and the iterative sequence continues.

The stabilising properties of the MSA algorithm should ensure that the profile of $\xi(\tau)$ demonstrates **reasonable** continuity, even if true convergence is hard to achieve. While it would be feasible to adopt more powerful forms of optimisation (eg the Frank-Wolfe method), it may be doubted whether this would be worth it, given the fact that an essentially continuous process is being treated as a discrete problem. A more promising alternative might be to attempt a stochastic allocation over the possible arrival time windows, by means of a logit model, for example.

Note also that if the gradient ξ' never reaches the critical value of β/α, on the early side, or $-\gamma/\alpha$, on the late side, then no shifting will take place. In this case, the entire demand for PAT segment k will be allocated to the arrival time window for k.

Finally, as noted, that the same approach could be used if we allowed a further demand segmentation by schedule parameters.

3.4 Illustration

The travel time choice module requires similar quantities to be calculated on both departure time and arrival time windows, and the notation is potentially confusing.

At least for the purposes of illustration, it is helpful to consider the variables in tabular form, along the following lines (the values for entries in the tables are arbitrary, but intended to be indicative):

Table 1 Departure Time Intervals (for assignment) [minutes]

				from assignment			interpolate	
r	Start I_{r-1}	End I_r	midpoint $\bar{t}_r$	GC $G(\bar{t}_r)$	Duration $\Theta(\bar{t}_r)$	Arrival $[\tau r]_r$	Cumulative $Q(I_r)$	Hence T_r
(0)		**0**	< 0	ff	ff	**0**	0.0825	
1	0	15	7.5		20	27.5	0.18027	0.0978
2	15	30	22.5		21.5	44	0.30125	0.1210
3	30	45	37.5		23.5	61	0.45013	0.1489
..								
..								

Note that the arrival column is still in terms of the departure time windows, and hence is written [τr]. The columns beyond the bold vertical line are only available subsequent to the choice of arrival time windows.

The figure below plots the travel duration Θ_r against the midpoint of interval r. It then plots the same value against the implied arrival time, calculated by adding the duration to the interval midpoint. This is therefore a representation of $\xi(\tau)$.

Travel duration plotted against departure and arrival times

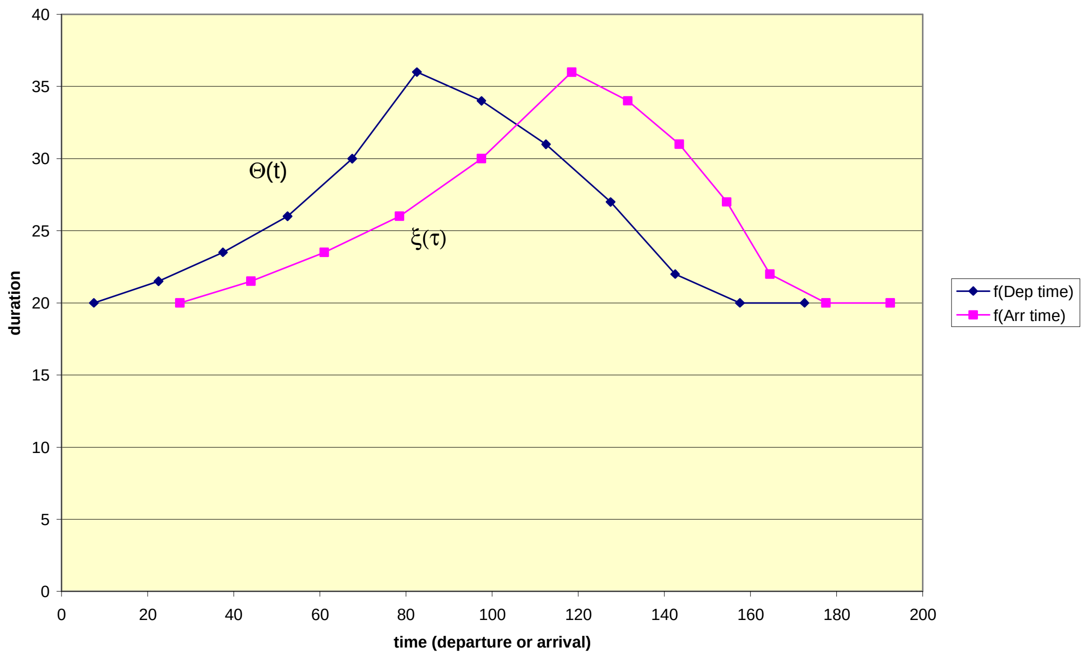

Figure 1: travel duration Θ(t) and ξ(τ)

Table 2 Arrival Time Intervals (for time period choice) [minutes]

h (k)	Start J_{h-1}	End J_h	midpoint $\bar{\tau}_h$	interpolated GC $\Gamma(\bar{\tau}_h)$	Duration $\xi(\bar{\tau}_h)$	Departure $[th]_h$	Cumulative Ω_h	midpoint $\Omega(\bar{\tau}_h)$
							0	
1	0	10	5		20	-15	0.03	0.015
2	10	20	15		20	-5	0.08	0.055
3	20	30	25		20	5	0.14	0.110
4	30	40	35		20.6818	14.3182	0.21	0.175
..								
..								

Note that the departure column is still in terms of the arrival time windows, and hence is written [th].

The next stage is to interpolate the points on the ξ(τ) curve to obtain an estimate at the midpoint of the **arrival time windows**. This can usually be done with good accuracy, as shown in the next figure. Note that we also calculate the implied departure time associated with the arrival time midpoint (by subtracting the interpolated value of $\xi(\tau)$).

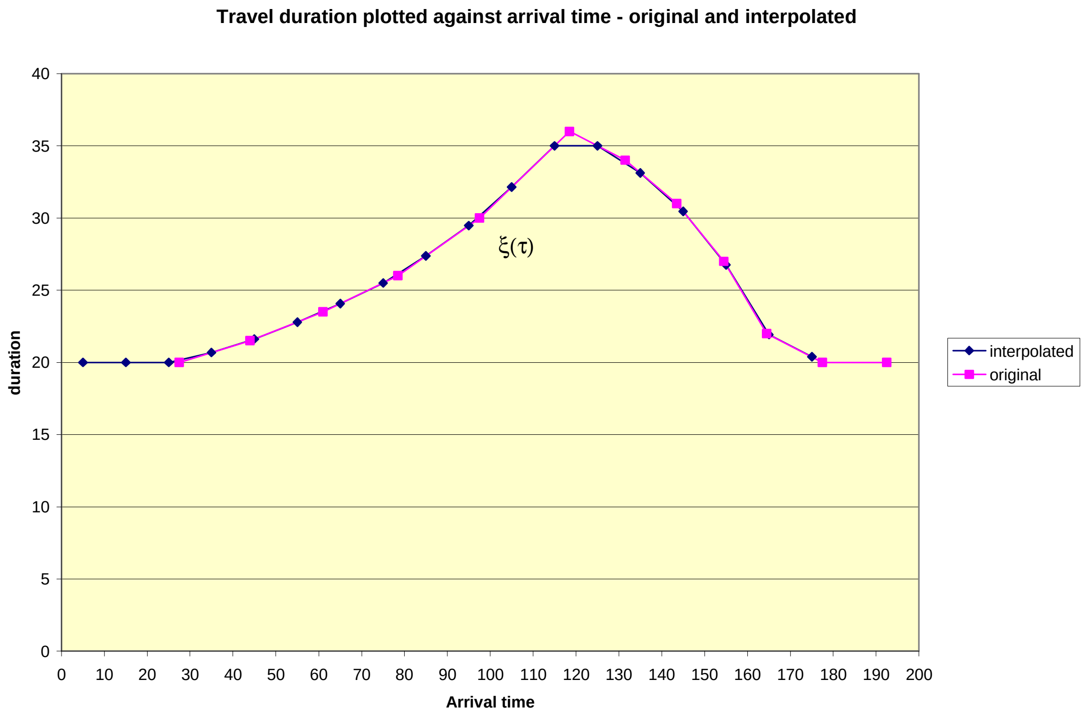

Figure 2: interpolation of $\xi(\tau)$

We can now calculate the arrival time with the maximum utility (separately for each PAT segment k), and by means of the MSA procedure, obtain the current estimate of the demand for each arrival time window. These can be summed over PAT segments, and from this we can infer a **cumulative** arrival time distribution $\Omega(\tau)$, here given as a frequency. This can be plotted both against the arrival time window midpoint **and** the implied departure time (as an estimate of $Q(t)$), as in the figure below:

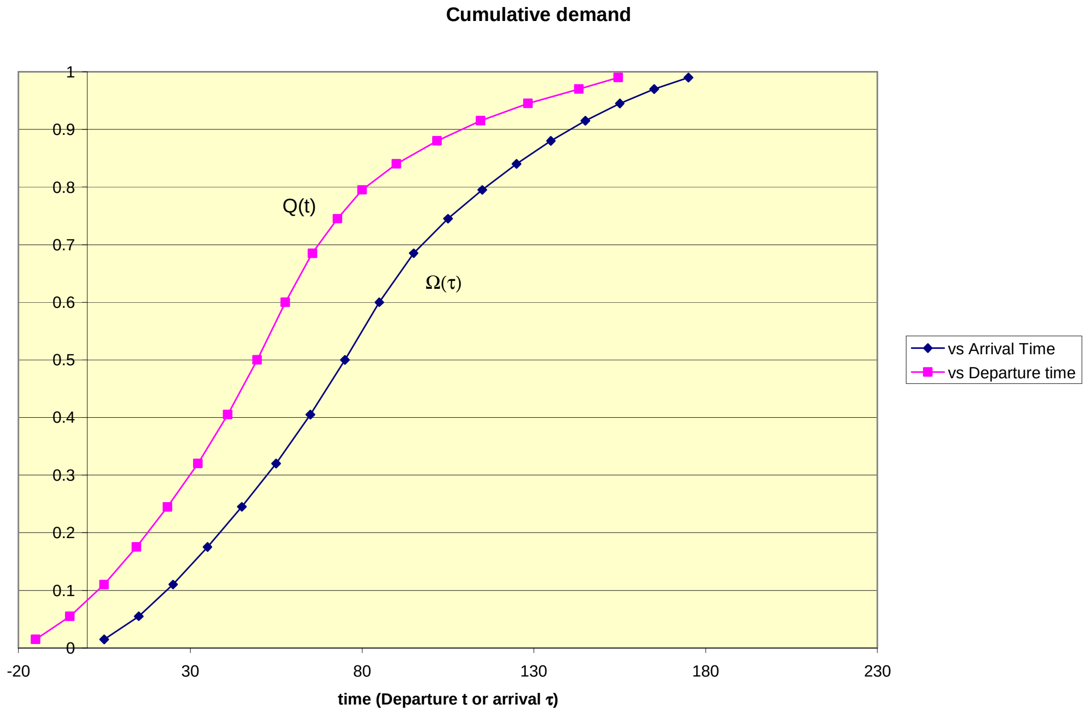

Figure 3: cumulative demand Q(t) and Ω(τ)

The final requirement is to interpolate the cumulative demand points on the Q(t) curve to obtain an estimate at the midpoint of the **departure time windows**. Once again, this can usually be done with good accuracy, as shown in the next figure.

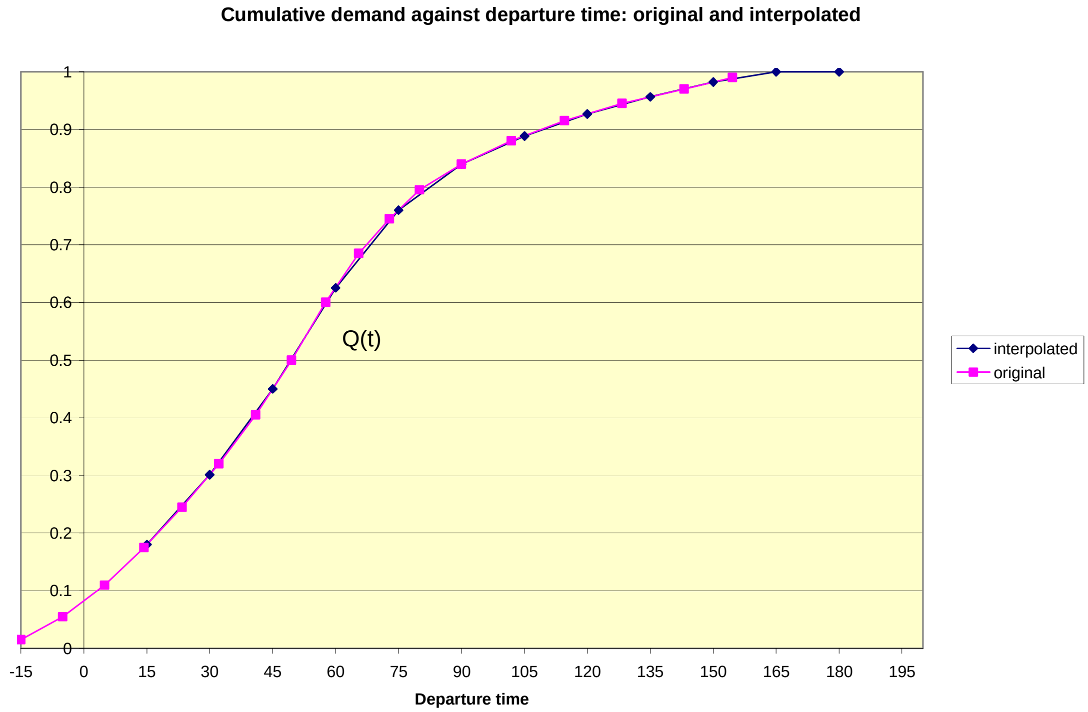

Figure 4: interpolation of Q(t)

We are now back to Table 1, above. Given the cumulative distribution, we can obtain the proportion of demand to be allocated to each assignment time slice r.

3.5 Convergence

There is a distinction to be made between true convergence measures (which determine how close to **equilibrium** we are) and stopping measures, which merely report on whether the algorithm is making progress.

The implication of the equilibrium condition is that, if for any ij pair and PAT segment k, $V_{Max} = \text{Max}_h\ \overline{V}_{kh}$, we should have

$$\Sigma_h X_{k,h} (V_{Max} - \overline{V}_{kh}) = 0 \qquad \forall \text{ ij, k} \qquad (12)$$

while the same quantity could be > 0 if the process had not converged. Hence the overall quantity:

$$\Sigma_{ij}\ \Sigma_k\ \Sigma_h\ X_{k,h}\ (V_{Max} - \overline{V}_{kh}) \qquad (13)$$

is an indicator of convergence, the smaller the better.

It is usual to scale such quantities to obtain a "δ-like" quantity such as:

$$\frac{\sum_{ij}\sum_{k}\sum_{h} X_{kh}^{ij}.(V_{Max\,kh}^{\;\;ij} - \overline{V}_{kh}^{ij})}{\sum_{ij}\sum_{k}\sum_{h} X_{kh}^{ij}.(V_{Max\,kh}^{\;\;ij})} \qquad (14)$$

Note that while this is a single overall measure, it could be broken down to inspect the convergence of individual ij pairs and, within that, individual PAT segments k. It would also be possible to substitute the denominator by the quantity $\Sigma_{ij}\ \Sigma_k\ \Sigma_h\ X_{k,h}$: this would then give the average "gap", in units of utility (here, minutes), between the optimum and the current position, and would give a more intuitive indication of the seriousness of any such gap.

It is also possible to construct comparable indicators between successive iterations. We give a number of examples, all of which can be further disaggregated by ij pair and PAT segment:

$$\frac{\sum_{ij}\sum_{k}\sum_{h}(X_{kh}^{ij\,(n)} - X_{kh}^{ij\,(n-1)})}{\sum_{ij}\sum_{k}\sum_{h} X_{kh}^{ij\,(n)}} \text{ or } \frac{\sum_{ij}\sum_{k}\sum_{h}\left| X_{kh}^{ij\,(n)} - X_{kh}^{ij\,(n-1)}\right|}{\sum_{ij}\sum_{k}\sum_{h} X_{kh}^{ij\,(n)}},$$

$$\frac{\sum_{ij}\sum_{k}\sum_{h}(\overline{V}_{kh}^{ij(n)} - \overline{V}_{kh}^{ij(n-1)})}{\sum_{ij}\sum_{k}\sum_{h} X_{kh}^{ij\,(n)}},$$

$$\frac{\sum_{ij}\sum_{k}\sum_{h}\left| X_{kh}^{ij\,(n)}.\overline{V}_{kh}^{ij(n)} - X_{kh}^{ij\,(n-1)}.\overline{V}_{kh}^{ij(n-1)}\right|}{\sum_{ij}\sum_{k}\sum_{h} X_{kh}^{ij\,(n)}.\overline{V}_{kh}^{ij(n)}}$$ etc., where n denotes the iteration number.

4 Conclusion

A method has been described, essentially heuristic in concept, for reflecting the changes in travel (departure) time which may occur within the peak period as a result of changes in the network (essentially a function of the ratio of demand to capacity). While it makes use of general principles, it is intentionally postulated in terms of tools which are reasonably available within current modelling practice.

The method has not been tested, and it may be anticipated that its convergence properties will be relatively weak. Nonetheless, it is hoped that it could provide a method for investigating a phenomenon which is well attested but poorly understood – the “flattening of the peak” as demand grows, and the “return to the peak” which occurs when additional capacity is introduced. It is hoped that the relative simplicity of the approach will encourage other researchers to try and implement it.

The method could also provide a potentially better estimate of the time-aggregate costs which could be used elsewhere in the demand model. Although it has not been specifically discussed, one reasonable candidate, having achieved adequate convergence of the allocation of peak period demand, is to calculate a flow-weighted average of the costs over the arrival time slices. In this way it is also possible to reflect the scheduling costs, along the lines discussed by Small (1992).

Finally, we should note an important feature of the general approach. While it is convenient to state the conditions in terms of the **gradient** of travel time, the treatment of discrete PAT segments means that a certain **absolute** difference is required to induce shifting. By the nature of things, a given absolute difference is more likely to occur for longer trips than for shorter, and it is therefore useful to look at certain combinations of trip length (eg, for a typical urban area, Inner to Central, Outer to Central, Through trips etc.).

REFERENCES

ADL (Arnott, R., de Palma, A., and Lindsey, R) (1994), 'Welfare Effects of Congestion Tolls with Heterogeneous Commuters', *Journal of Transport Economics and Policy* May 1994

Bates J J (1996), Time Period Choice Modelling - A Preliminary Review, Final Report to Department of Transport, HETA Division

SACTRA (1994), Trunk Roads and the Generation of Traffic, HMSO, London

Small K (1982) The scheduling of consumer activities: work trips *American Economic Review*, **72**(3), 467-479

Small, K.A. (1992) 'Urban Transportation Economics', Fundamentals of Pure and Applied Economics 51, Harwood Academic Publishers

Vickrey WS (1969) Congestion Theory and Transport Investment *American Economic Review (Papers and Proceedings)*, **59**,251-261