Railway Transport Planning Key Concepts

Edited by George Ardavanis, Ph.D.

Abstract: This paper attempts to provide an introduction to a variety of passenger railway planning problems. Specifically, the objectives of planning any new railway should be to develop and achieve the most cost-effective and appropriate system which will convey the required numbers of passengers to their desired destinations: safely, reliably, in a user friendly, reasonably comfortable, clean, attractive and environmentally acceptable manner, with competitive total journey times and fares, and at minimum total capital, operating and maintenance life cycle costs. I have attempted to cover the following four key areas: (1) Railway Planning; (2) Strategic and business planning for railways; (3) Market research for passenger railways; and (4) Public Transport Network Modelling.

1. RAILWAY PLANNING

The railway planning is very important for all those organizations, specifically those railways which are capital intensive and their project lead-times are long.

The objectives of planning any new railway should develop and achieve the most cost effective and appropriate system which will convey the required numbers of passengers to their desired destinations safely, reliably, in a user-friendly, reasonably comfortable, clean, attractive and environmentally acceptable manner, with competitive total journey times and fares, and at minimum total capital, operating and maintenance life cycle costs

Modelling is an increasingly important part of planning passenger railways, as it enables the effects of different demand levels and capacities to be tested without incurring significant expense. The results of modelling are measured in a variety of ways, and are discussed in an appraisal section. Nevertheless, the various activities which may be subsumed under the heading "planning", are not random, specifically, there are six groups of activities:

1. Policy issues, which ought to come first
2. Demand estimation
3. Engineering issues
4. Operability issues
5. Environmental issues
6. Appraisal

Key Concepts

There are a number of key concepts which underlie transport planning.

The first key concept is that "travel is a derived demand". Because of the inherent differences in activities (i.e. going to work leads to earning money, whilst going shopping involves spending it), journey purpose is a key determinant of travel behaviour.

The second key concept is of "generalised cost". Under this concept, it is assumed that passengers act so as to minimise the disutility of travelling, where disutility is measured by generalised cost in the form:

$$GC = f + b_0 + b_1A + b_2W + b_3R$$

Where:

GC = Generalised Cost

f = fare

b_0 = error term

b_1, b_2, b_3 = weighting coefficients

A = access time

W = waiting time

R = running (in-vehicle) time

Because passengers value time spent in the different ways the generalised differently, weightings are conventionally applied to the time elements. By multiplying either the generalised times or the fare spent by Values of Time (for which standard figures are available[1]), generalised cost can be expressed in either units of money or time; a common unit is required. Other attributes can be added to the generalised cost equation if data is available. The relative weightings are of considerable importance. For example, if $b_1 = 2b_3$, then passengers are twice as responsive to improvements in access time as they are for journey time improvements, and operators need to pay attention to such a fact. In fact, this is generally the case; b_1, and b_2 usually are twice b_3.

Also, note that the error term includes elements which are independent of time. For example, there is a disutility associated with changing trains, irrespective of the time penalty incurred, and interchange penalties of up to 30 minutes have been identified for some market segments.

The third key concept is that of "elasticity". An elasticity is defined as:

$$\frac{\text{proportionate change in y}}{\text{proportionate change in x}} = \frac{\frac{\Delta y}{y}}{\frac{\Delta x}{x}}$$

Where:

x = Independent variable within the control of the operator.

The most commonly used elasticity measure is the fares elasticity, values of which are generally negative, since increasing fares leads *ceteris paribus* to reduced demand. Typical values for rail elasticities are given by journey purpose in Table 1, but it should be noted that elasticities are situation specific, and great care should be taken if using these figures without independent verification, or at fare levels substantially different (say>10%) of the fare level at which they were calibrated.

Table 1. Typical British Rail Fare Elasticities

Commuting	-0.5
Business	-04
Shopping	-0.8
Visiting Friends & Relatives	-1.1

Nevertheless, readers should be familiar with the difference between "point elasticities" and "arc elasticities"

Point elasticity is the gradient of the demand curve at a particular point. This underscores the importance of quoting elasticities with an absolute value, a time (i.e. year), and the environment in which measured, since they vary with all three.

Arc elasticities are the gradient of a chord linking two points on the demand curve, and are therefore directly measurable (i.e. between two fare levels). They are, of course, equal to the equivalent point elasticity in the limiting condition, but are more appropriate for larger changes in the independent variable.

Yet, other elasticities are also used for planning purposes, specifically the service elasticity, which relates the level of generalised cost to demand; again values are generally negative, since increasing generalised cost makes a service less attractive, and therefore the demand falls. Individual elements within the umbrella of generalised cost (such as waiting) also have their own elasticities, but these do not add up to the overall Generalized Cost elasticity, even if all the Generalized Cost elements are known, unless the weighting coefficients are all equal. However, it should be noted that it is not always possible to change variables independently; for example, railway electrification may lead to increases in comfort as well as speed and frequency, and there is usually a residual increase in demand of 5-10% due to such a sparks effect.

The fourth key concept is the “trip matrix”. Data on trips between zonal origins and destinations conventionally stored in matrix form for the time-period under consideration, with traffic originating from the zones represented in the rows and having destinations in the column zones. Trips may also be found in the leading diagonal edge, and these are known as intra-zonal trips.
Forecasting methods need to take into account whether the independent variables inherently affect origins or destinations.

Key Assumptions

Transport is a relatively young area of scientific inquiry, so there are a number of issues as yet not fully explored, and assumptions are commonly made as to their effects, so as to minimise problem size. Some of the key assumptions to be considered are the following:

1. Trips are often assumed to be “single-purpose”. This is reasonable assumption for modelling the a.m. peak but, at other times of day, many people make linked trips (i.e. going shopping on the way home, rather than going home directly). This makes the allocation of trips to a trip purpose (and hence forecast of future trends) difficult. 30% of trips have been found to be multi-purpose.
2. The “attributes of passengers” (i.e. their response to interchange) are often assumed to be the same across all passengers. Although. It is possible elasticities to vary walking and waiting time weightings are usually assumed equal for all trips, even though this is patently not the case for passengers of differing mobilities with differing amounts of luggage.
3. Once travel decisions are taken behaviour becomes “habitual”. Specifically, habits, thresholds and resistance to change lead to discontinuities in the demand curve, until choices are re-evaluated. This may occur through dissatisfaction with service, unsolicited information from marketing activities, changes in home or job location, or changes in lifestyle (i.e. having children).
4. The model assumes “perfect passenger knowledge”. Although one might reasonably assume that commuters are indeed aware of alternatives, this is becoming less the case as people move house more frequently, and such an assumption is much weaker in

relation to off-peak travellers. The severity of this assumption can, nevertheless, be reduced by probabilistic, rather than deterministic, planning; however, the extent to which the other options considered are sub-optimal, and how many passengers are assumed to use them remain somewhat arbitrary.

Moreover, a knowledge of alternative requires effort (the Generalised Cost of obtaining information is not zero, as it is in the economists' perfect market), and individuals will only invest the effort if, in economic terms, the perceived gains in their utility are large enough. Marketing can, however, be used to limit the leggedness of passenger response.

Factors Not Included

A number of factors are often considered to be outside the scope of railway planning. These variables include:

- The attractiveness of particular cities as centres for commerce, tourism or any other purpose (an issue for land user planners, but usually an assumption for transport planners)
- Changes in land or property values
- Safety on the rail network (except insofar as severe overcrowding is included as a time deterrent)
- The environment (either in terms of noise or pollution)
- Road safety
- Improved passenger comfort in new rolling stock that is introduced over the forecasting period

Occasionally, these factors may be crucial in determining the demand for rail transport, even away from those situations where travel is consumed for its own sake. Furthermore, passengers may perceive some of these issues to be important (i.e. safety).

Land-use Planning

Land use planning and transport planning matters are often closely related, in what should be an iterative process. Too often, however, councils responsible for land-use planning do not take into account the local railway network (which is generally centrally funded), unlike the local highway network, which is more under their control.

Railways may therefore find themselves running through an area which is run-down or an area not suitable for development; the high fixed costs of rail construction may leave the railways in a difficult financial position. Strategic land-use planning, taking account of the railways' ability to move large numbers of passengers, is relatively rare.

Conclusion

It is assumed that overall strategic an land-use planning matters are not within the scope of most railways to affect directly (although railway operators can hope to be included as part of the consultation process)

2. STRATEGIC AND BUSINESS PLANNING FOR RAILWAYS

The strategic and business planning for railways is considered a policy issue. It is difficult to generalise about the best strategy for railways. Not only do rail systems and the environments in which they operate differ greatly, but different sets of directors and policy-makers can set varying objectives within the same scenario.

The urban travel market in which mass transit systems usually operate can be characterized as being:

- High volume, but of low individual trip value which, with the exception of the car, displays little brand loyalty
- Considered by both supplier and buyer almost as a distress purchase, from suppliers operating in a semi-monopolistic environment
- Subject to major peaks and troughs of demand: day, week and year
- Expanding, changing rapidly and very diverse in terms of modes, users and usage
- Becoming more fragmented as competition increases

Strategic planning within the context of such a mass transit system thus seeks to match the requirements and capabilities of the business with the needs of the customer over extremely long timescales. It thus:

- Defines and drives the scope of an organization and its activities
- Fits and matches the organization's activities to the future environment
- Acquires and matches resources to activities
- Allocates or reallocates resources to suit the desired and agreed goal(s) set
- Sets up internal and external values and expectations
- Determines the long term direction of the company and its response(s) to competition

- Drives the internal organization structure and workings of the business

The business planning process itself consists of the following natural chain of events: (a) Analysis, (b) Option Generation, (c) Option Evaluation, (d) Option Selection, and (e) Implementation.

Such a business planning process can, in classical planning theory, be either an incremental or a step-change process. The incrementalist continuously adjusts strategic policies and directions as the market and the environment changes. On the other hand, the proponents of step-change planning wait longer between major changes in direction, arguing that short term adjustments confuse the organization and do not reflect real long term shifts within the planning environment. Historically, a balance has been struck between occasional major strategic reviewers and continued adjustment through such mechanisms as the annual business plan.

In the early 1980's the strategies undertaken by major railway companies in Western Europe demonstrated significant shifts across a number of dimensions of the business in:

- Market volume, share and usage
- Legislative environment
- Safety performance
- Internal efficiency and effectiveness
- Funding regime and asset health

Each of these changes and imperatives influenced the shape of the railways' future planning. For example, in terms of customers the London Underground had experienced a significant and continued loss of trade since the late 1950's. The nadir of its fortunes occurred in 1982 when the business carried 600 million trips per annum. In the years following 1982 a 60% increase in usage occurred leading to problems with crowding, congestion and travel comfort. Unfortunately, the 1982 strategic plan, and its subsequent revisions in the annual business plan, failed to predict this increase in usage. This meant that a step change was now overdue. In the following six years the legislative environment changed through significant efforts in order to improve efficiency of service delivered to the customer. However, these were primarily focused on cost minimisation rather than quality and value maximisation. It was believed, that as the basis of competition and service delivery changed towards quality that London Underground would need to alter its focus. Lastly, whilst there had been significant investment over the years in new infrastructure only a limited amount of funding had been directed at maintaining an increasingly aged infrastructure and coping with the burgeoning demand.

Planning Process and Tools

The planning process should look at both the internal and external factors in order to drive a railway company forward or influence its direction. Only then could a series of strategic options be developed. An analysis of internal and external factors can run in parallel, with the modelling of internal factors to be focused on such aspects such as:

- The impact of changes in opening hours
- The quantity and quality of service delivered
- The size of the network and impact of crowding on customer usage
- New lines and extensions

Another major tool in this process is the use of the railway company's network model linked to a financial evaluation model.

Any analysis of the internal capabilities of a railway company has as an objective to establish sensible boundaries and derive fundamental relationships between cost, income, investment and demand.

External factors to be considered and analysed include:

- Market position and competing modes
- Buyers, suppliers, new entrants and substitute products
- Base demand forecasts
- Likely changes in the legislative environment

Additionally, the view of the outside environment is usually directed at developing a realistic range of demand forecasts split into three categories: (a) peak, (b) off-peak, and (c) weekends. These demand scenarios take into account:

- The relationship between supply and demand for services
- Changing demographic and employment profiles at a borough and county level
- Forecast increases in real disposable income and thus propensity to use the service
- Increased consumer spending and changes in the leisure, domestic and foreign tourist markets

On top of these demand scenarios additional potential impacts from changes in work patterns and changes in legislation can be factored. As a result of this analysis five key strategic issues are usually identified:

1. Passenger volumes exceeding capacity, which can only be enhanced slowly

2. Commercial performance may be improved, yet much is remained to be done
3. The structure and management capabilities usually need adjustment
4. Significant investment might be required under any scenario, even to keep the asset base at its present condition
5. Quality and safety standards must be improved in the customer base is to be retained

These immediate issues are in turn exacerbated by the forecast growth in demand.
To meet the five strategic issues four different strategic options are developed. These options are ranged from an expanded network carrying large volumes of customers, to a more commercial and financial driven business. No one option was clearly superior, but key elements of them were combined to form a single strategic direction. The selected strategy was then tested for its sensitivity.

Strategic Direction

The selected strategic direction usually has three major five year phases designed to address the issues in a logical sequence.

- Phase 1: To be focused on internal management reform, service quality and safety improvement and securing sound financial performance. This phase is developed to provide a firm foundation for the later phases
- Phase 2: To be consolidated on the first stage with further improvements in facilities, quality and reliability. In this phase capacity shall be brought on stream to cope with demand in the central area and potentially congested corridors
- Phase 3: To be looked towards expanding the system to tap new market opportunities

This phased approach provides the flexibility to accelerate, merge or delay the phases and hence reflect fluctuation in the market and financing.

General Implications of Strategic Alternatives

Regardless of the system or planning environment, there is a range of outcomes that will provide a variety of satisfactory solutions ranging from:

- Network requiring low capital cost, providing adequate quality and safety carrying a relatively low volume of customers at high fares

To the other extreme of

- High investment, high quality, low fares networks carrying very large numbers of customers

Whichever strategic alternative is adopted implications exist for the strategic planning, design and operation of networks across a wide range of dimensions.

Demand and Planning

The long timescales associated with the design, construction and operation of new lines require that:

- Sensitivity analysis becomes an integral part of the planning process
- The strategies adopted are sufficiently flexible to allow for change as the future unfolds
- Fundamental market research is crucial
- Quick and decisive action based on changing circumstances is vital given the long lead-time for system development

Financial

- Maximising the output from the existing asset base will in general provide a greater return than massive new capital investment in new lines
- Initial capital cost design must recognise that later expansion through anything other than by purchasing more rolling stock can be prohibitively expensive
- Operational and capital costs must be minimised if commercial borrowing or equity is to be employed
- Private contribution from developers and other beneficiaries is essential rather than optional in today's environment

Business and Operational

- Speed of response to the market is vital
- Competition will be increasingly based on quality rather than cos, thus driving the need to integrate the production and marketing elements of the business
- Information on customer needs, attitudes and movements is becoming more essential
- The human interface between server and served must be properly designed and managed
- Fares and operational systems must be designed to balance load and capacity across the network both spatially and temporarily

- For existing capacity-constrained systems market penetration allied to product extension strategies are of more value than market extension

<u>Technological and Design</u>

- Interconnectivity between systems must be maximised to harness the "network effect", where improvements in two elements can provide benefits greater than the sum of doing them individually
- Technology must be put to competitive use in terms of system performance, operational/maintenance flexibility, customer and management information and the general travelling environment
- The design of new lines must allow for:
  - Ease of upgrading of both fixed and moving infrastructure
  - Appropriate maintenance in terms of facilities and available skills
  - The increasingly rapid change in technology
  - Global sourcing must be adopted to tap the new technology available
  - Counter increasing supplier concentration
  - An explicit choice must be made between the risks and benefits of new technologies, or whether to stick with older but proven methods

3. Market Research for Passenger Railways

Market research is highly valued by passenger railways, especially in UK. Specifically, British Rail have invested quite heavily and are leading the way in innovative research techniques and approaches.

Research Usage

Research has typically been made use for the following purposes:

1. Customer profiles (to obtain information on demand, market shares, market structure, customer activity)
   - On train passenger counts/station flow counts
   - Profile surveys/monitors (e.g. origin/destination, ticket held, fare paid, age, sex, etc.)
   - Surveys of facility usage (e.g. on train catering, toilets, etc.)
   - General surveys of won/competitive mode usage
2. Attitudinal research (to measure public/customer ratings of marketplace performance)

- Customer satisfaction rating surveys (i.e. for train punctuality, train comfort, catering standards, etc.)
- Image and awareness studies (i.e. of company, products offered, etc.)

3. Pricing Research (to explore price sensitivities and tariff structure development within a competitive environment)
    - Awareness studies (i.e. rail vs. competitive modes)
    - Appraisal of tariff structures through price elasticity examination (i.e. differentials 1st/Standard class, railcard discounts, etc.)
4. Design research (to assist in design and marketing processes)
    - For infrastructure development (i.e. new stations, rolling stock)
    - Input into redevelopment/refurbishment programmes
    - Product packaging and pricing
5. Demand forecasting (to provide inputs to appraisal process)
    - For product reformulation (i.e. changed ticket restrictions, revised tariff differentials)
    - Infrastructure investment (i.e. new "park and ride" stations, new rolling stock, new service/route, electrification)
6. Advertising research (to help develop and test advertising campaign strategy development)
    - Copy development/testing
    - Pre and post advertising awareness surveys
7. New product development (to assist in optimal new product development strategy)
    - Opportunity identification
    - Concept development/testing
    - Product and positioning specification
    - Forecasting public response (i.e. awareness, attitudes, take-up)

Some areas are clearly more specialised than others with the more specialised purposes, as well as being less commonly pursued, typically also drawing upon more specialist suppliers ad approaches.

Research Approaches

Passenger railway organizations make use of the broad range of research techniques available, embracing both qualitative and quantitative approaches.

Qualitative approaches essentially comprise either individual face-to-face in-depth interviews or focus groups of up to eight or ten respondents guided by an experienced moderator. Such approaches are intended to provide a detailed, broad review of the issues involved rather than quantified responses that can be associated with statistical tests of robustness. The research location (i.e. on train, in offices or at home) needs to be appropriate to the task in hand.

There is a variety of Quantitative approaches available:

- Simple passenger counts, perhaps using technological aids like video cameras
- Self-completion surveys (i.e. on train)
- Postal surveys
- Telephone interviews
- Face-to-face interviews (i.e. at home, hall tests)

There is a broad number of issues which need to be covered when designing such research:

- Is the approach suitable, given the complexity of the research issues involved?
- Are the right questions being asked?
- Are they being asked in the right way?
- Will the sampling approach avoid unplanned biases (i.e. can all potential respondents can be reached in this way, or will a high proportion of non-responses lead to potential bias?)

Very often qualitative research is used to help satisfy the researchers' concerns in regard to the first three questions. Sampling theory can assist with the fourth. As a further safety net, it is highly recommended that any approach be piloted sufficiently to help identify potential problem areas in advance of a commitment to be a major fieldwork programme.

This less expensive techniques are typically also the more restrictive, potentially leading to higher proportions of non-response (i.e. postal surveys) and a less "customised" approach (i.e. one less tailored to a specific respondent's circumstances). Judgement is needed to weigh up the advantages/disadvantages of alternative approaches.

Data collected in such surveys can be analysed using econometric techniques, whereby statistically significant causal relationships can be derived from the results. Other modelling techniques may also be used. For example the "gravity model" is sometimes used for the estimation of trip distribution.

There is one further stream of research used by the railways and not really covered by the above which is known in Europe as "stated preference or trade-off research". Essentially this is a powerful specialist research which comprises both qualitative and quantitative research elements.

Increasing use is being made of computer-based interviewing techniques, not only with stated preference approaches. This trend has also been evident during the past few years in both telephone (computer assisted telephone interviewing) and face-to-face (computer assisted personal interviewing) quantitative approaches.

Different types of research tend to be appropriate for different market research purposes. Customer profile research tends to make predominant use of highly structured quantitative approaches whilst a mixture of qualitative and quantitative approaches will often feature in attitudinal, pricing, design and demand forecasting areas. Advertising research featuring quite heavily also in new product development work.

Any use of stated preference research will tend to be restricted to pricing, demand forecasting, new product development and, perhaps to a lesser extent, design research.

Stated Preferences

This research approach was introduced into Europe from USA in the late 1970s. In USA it had tended to be mainly used for new product development work in non-service “fast moving consumer good” sectors and marketed under a variety of different terms; for example, conjoint analysis, functional measurement. It contrasts with traditional techniques measuring actual behaviour (Revealed Preference (RP)) methods by investigating more hypothetical situations.

Essentially, it is a technique which enables the relative importance of product components to be examined in a quantified way, enabling more focused product rationalization and development. It works by asking potential customers to evaluate a series of experimentally designed product options which are presented as alternative packages.

Evaluation can be by way of ranking of packages into the respondent’s order of preference, by rating each or by making choices between two or more packages.

An important issue in the use of SP methods is the quality of the survey and the context in which the survey questions are asked. If useful results are to be obtained from SP methods, the survey needs to be of the highest possible quality and the context in which the SP questions are asked should be as realistic as possible. For this reason most practitioners have a strong preference for face-to-face interviews, conducted by experienced interviewers and also structured to ensure:

- That the background to the respondent’s evaluation process (i.e. situational constraints, demographic characteristics, planning processes) is fully understood by the researcher
- That the respondent is not “educated” about any misperceptions he might currently hold about a particular or service, or led by “enticing” stimuli (such as a new station

photographed on a very sunny day) to react more or less positively to change than he might otherwise have done

- That complete alternative specifications are provided, since individuals perceive concepts as a whole rather than as the sum of a number of separate factors

Typically, the information is provided to respondents by way of computer screens. For example, a preference based approach might well require a respondent to appraise some computer screens each comprising one option made up of differing levels of the variable being examined. Some of the information is normally preprinted on to the screens, other items may well need to be filled in by the interviewer in line with the experience of the respondent (i.e. related to the fare actually paid), providing a customised approach. In order to maximise the usefulness of the results (and therefore to minimise time spent interviewing) initial questions may also be asked to isolate those thresholds at which respondents are genuinely trading attributes off against each other. The probability of the respondent actually choosing an alternative can then be estimated by a logit model. For example:

$$P(a) = \frac{\exp(a)}{(\exp(a) + \exp(b))}$$

Where:

P(a) = the probability of choosing option a

exp (a) and exp (b) = represent the exponential of the generalised cost functions chosen for the options

Applications of Stated Preference

One of the first applications in Europe was a major study for British Rail which took over two years to complete. This research project was concerned with defining demand elasticities for primary service characteristics of long distance rail travel price, journey time, service frequency and interchange. Revealed Preference approaches involving the development of econometric models from ticket sales data had proved to be unsuccessful in separately identifying values for all of these variables and such models were certainly unable to distinguish between distinct market segment groupings (i.e. leisure of business travel). However, the study proved to be enormously successful and was well received.

Since then British Rail has invested heavily in the development of the technique using it predominantly to estimate further market segment based elasticities, for input into feasibility appraisals (i.e. station investment), for product rationalization (i.e. changed railcard discounts), and for new product development.

Development of the technique has been quite dramatic with a major breakthrough emerging in the mid-1980s, since the development of approaches capable of more direct examination of consumer choice behaviour between competitor products. Previously such techniques had been used predominantly to look at preferences between specific product variations.

These recent developments have increased the attraction of the approach substantially, paving the way for an important new research avenue – the use of stated preference in the process of forecasting demand and market shares for potential light railway transit systems. Typically, these projects have involved the calibration of hybrid revealed preference/stated preference forecasting models, where the two data sources are treated as complementary – the stated preference data providing information for areas for which no suitable revealed preference information is available.

More recent initiatives have involved exploring how such techniques could be developed to estimate the potential for generated demand rather than re-allocated demand (from one mode to another or one time period to another) which has been the focus to date. This stream of research known as “stated intentions” is still in a very early developmental stage.

Conclusion

Market research is used by railways for a large number of applications, including as a basic part of planning future developments. Both qualitative and quantitative methods are used within the latter, there is a split between Revealed Preference and the developing technique Stated Preference.

Stated Preference has a number of advantages over Reveal Preference being particularly useful for quantifying the effects of qualitative variables, and in new product situations where no Reveal Preference data exists. It can also be more data-efficient, as by asking a series of hypothetical questions, several observations per individual can be collected. It is also statistically efficient; for example, surveys can be designed to avoid correlation. Moreover, it facilitates the examination of non-linearities; for example, different levels of rail frequency might be specified separately, a technique sometimes known as piecewise estimation.

However, there are also potential problems, and customization of the approach is required to avoid undue bias. Respondents’ abilities to provide sensible answers limit how hypothetically situations can be whilst, when used for forecasting, a potential technical difficulty known as the scale factor problem needs to be overcome.

4. Public Transport Network Modelling

Network models are computer-based simulations of transport networks, in which the network configuration is an important element; therefore they are specifically important where the network exhibits considerable connectivity, where changes on one link have important effects elsewhere on the system. Although subject to simplification, the networks are comprised of zones (where traffic enters and leaves the system), nodes (which may be seen as junctions in the network), and links (joining the nodes to each other and to the zones). Associated with each link are a number of characteristics, for example directionality (that is one way or two ways), length, and time required to traverse it, and mode (i.e. walk, car, rail, etc.) whilst associated with the network is a demand matrix of trips between every zone as origin and every zone as destination.

An algorithm is used within the software to allocate the trips in the demand matrix, to the network, and this requires a number of parameters (i.e. the Value of Time) to be input. Network modeling is therefore especially suitable in major urban areas where the rail network is not merely radial and where route choice and good survey bases exist.

Applications of Network Models

The main uses of network models may be seen to be:

- The analysis of the spatial variability of demand, the identification of pinch points, and the optimum allocation of resources
- A cheap environment for testing changes to transport infrastructure
- Providing a method of disaggregation for demand forecasts

Results for the first two problems can alternatively be derived from econometric modelling work, but only at a much more global level. However, variations in demand between areas and periods may be sufficiently great so as to invalidate conclusions drawn from such global analyses.

The second main use of network models is in option analysis, right down from scenario testing to examining the effects of changes in train service levels. It is much more cost effective to simulate options on a computer in the office rather than trying to experiment with the passengers themselves.

A third use is in disaggregate demand forecasting, where known changes in network or services can be combined with matrix manipulations on the base matrix to produce demand forecasts for every link and node on the system; such as information is particularly useful as a management tool for a wide range of planning purposes. For example, in staff requirements, capacity enhancements, services changes, evacuation times etc.

The History of Network Models until 1980

Network modeling did not develop until the first computer revolution in the 1950s. As a recognized technique, it had its origins in the early 1960s, and soon came to be used as an important part of the evaluation framework for new roads, where it has long been recognized that new roads impact on the demand for existing ones. The 1970s saw developments to algorithms away from an all-or-nothing approach (in which all traffic on any origin: destination pair is allocated to the route with the smallest generalized cost) to multi-routing approaches (where some traffic is allocated to the shortest route, but other proportions are allocated to other routes with similar g.c.s.). Algorithms also began to include public transport services, where assumptions are required as to standard parameters such as weightings for walking and waiting.

Types of Network Models

It is important at this stage to distinguish between two main types of network models, which we may call zone-based and station-based. In zone-based models, the region under study is split into zones corresponding to political or administrative areas, for which data (i.e. population) is readily available. A trip generation sub-model is required to determine the number of journeys generated by such a population (or attracted by the employment of the areas) in the modelled time-period, and a modal split model is used to allocate the trips between modes, before assignment to the network occurs. Whilst this has the advantage of being soundly based on facts, this extra modelling stage introduces potential for error before the assignment stage occurs.

Station-based modelling, on the other hand, takes station entries as fixed, and assignment occurs directly. For public transport operators with good survey bases, this will provide more accurate data from which to work, but it does suffer from the disadvantage that some trip-making occurs outside the model. For example, park and ride to station A may be replaced in reality (but not in the model) by park and ride to station B. Moreover, modelling the impact of the introduction of a new station, for example, can only be done explicitly outside the mode.

Developments in the 1980s

Three key factors have provided the stimuli for the development of network models in the 1980s. The second computer revolution, an increase in environmental awareness, and a significant growth in the use of public transport in some major city centres.

The second computer revolution meant that by 1990, you could do calculations on a stand-alone personal computer which would have required a mini-computer, if not a mainframe, ten years earlier. This put network analysis within the reach of an increasing number of authorities and organizations, even if few of them realized the opportunity. For those who did, costs fell dramatically.

The increase in environmental awareness (which took place in the late 1970s in Europe) had the effect of moving transport up the political agenda. It also increased the importance of public transport relative to private transport, demanded environmental outputs from transport modelling (such as pollution emitted) and, backed up by strong NIMBYism (NIMBY = Not in my back yard) and, based on an increasingly affluent society, created much greater public interest in proposals for changes to transport infrastructure.

In some countries, public transport (but notably rail) enjoyed a renaissance in the 1980s, fuelled by increasing road congestion and the heightened environmental awareness already mentioned. Where this coincided with significant employment growth (which, in an era of continued deindustrialization, was primarily in areas dominated by office employment) railways increased not only their protection with new customers but even their mode share at the expense of private transport. All these put demand on the facilities available for network modelling, and encouraged methodological developments.

So these key factors provided the impetus for three significant developments in network modelling – better graphics, better algorithms, and improved methods of dealing with public transport congestion.

Algorithms

The first algorithms (such as those developed by Burrel and Dial) were all-or-nothing. This weakness has been removed more slowly on the public transport side of network modelling than on the highway side, and that it was not until the late 1980s that proportional assignment algorithms became available and began to be used.

The early algorithms, however, had other problems, mainly in dealing with particular public transport assignment problems, such as fares, which are rarely a linear function of distance travelled. Stage-based fares have tended to be treated by analogy to a function of the form

$$F = B + a_1 D$$

Where

F = fare

B = boarding penalty

D = distance

a_1 = constant

but most of the British metropolitan areas, as well as many centres abroad (notably the whole of Holland) use a zonal fares system, where this approach is unsatisfactory. The existence of an annual fare zone system in London where a great amount of network modelling was being carried out, stimulated development in this field, and latest thoughts seem to centre around the allocation of nodes or links to fare-zones, since the number of fare zones traversed is a basic indicator of the fare due. This problem, however, remains acute for the time being for orbital network developments – such as London Undergound's East London Line extensions – where not only is an unwanted trip into and out of the Central area avoided, but so is that part of the fare.

Fares, however, are not the only algorithmic problem facing software writers. On routes where not all services stop at all nodes, combination rules are required to determine what the real frequency available to passengers is. In some cases, the underlying market research is not always at hand. It is assumed that passengers will not board a service to a destination if a following departure is timed to arrive first, but in severe weather, on an unreliable line, some passengers with no pressing hurry are quite likely to take the first train that actually appears.

Quite complicated options may be apparent to passengers which involve decision making en-route. For example, it is rarely possible to get bus and train services from the same spot, so a choice has to be made as to where to stand in real life, a choice that needs to be modelled. Gradually, the assumptions that have hitherto underlain network modelling (i.e. rational behavior and perfect passenger information) are being tested.

Congestion

At first amongst algorithmic improvements, perhaps, is the development of software to treat overcrowding on public transport. In the 1970s and 1980s, congestion on public transport was sufficiently negligible (except on the Hong Kong metro system) to be completely ignored, but traffic and tighter resource management had made congestion endemic by the end of the decade. Software developments have had to take account of passengers' options:

1) To travel by an alternative route
2) To supress the demand for that trip (either by making a motorised trip by non-motorised means, or by moving out of the area under study altogether)
3) To travel at an alternative time

Coping with (1) has been relatively easy for software writers. The approach generally adopted has been to determine a perceived speed, flow relationship from market research, to apply the resulting delays, to the base times, and then to recalculate routes based on the revised network. This becomes an iterative process which would converge eventually, although flipping between two adjacent and heavily congested routes may occur, where traffic takes one route in the even numbered iterations and the other in the odd numbered iterations.

London Underground were amongst the first to incorporate demand suppression into this framework, with a program module which, for every origin: Destination pair, compared the times for successive iterations, and applied an elasticity to the base trip matrix to give a now trip matrix for the next iteration. This process also works in reverse, in that trips can be generated if congestion improves between iterations. In a zone-based, rather than a station-based, model, this problem is reduced, but not eliminated traffic can switch between modes, it cannot otherwise, however, be modelled to follow the industrialist moving his factory to a less-congested part of the country, or a trip being replaced by a non-motorised one.

Congestion has also become of concern, not only on public transport vehicles, but elsewhere in the public transport operators' domain. Underground railway operators in UK have faced tougher safety legislation following the fire at King's Cross in 1987, requiring them to be able to evacuate stations within 6 minutes. This necessitates an improved understanding about pedestrian flow characteristics, not only at individual stations, but also across the network as a whole.

The problem of time-switching is still under investigation. Little market research is available to show how many passengers (commuters, specifically, since it is obviously in the peak periods that the worst congestion occurs) can change their travel time and, of those, how many would do so, and what incentives they might need, and which direction they might change in (earlier or later)

Further Developments

Currently is under investigation the dynamic reassignment of travellers (i.e. modelling the ability of travellers to alter their journey en-route dependent upon congestion encountered.

This and the time switching issue are both under investigation for highway networks, but not yet for public transport networks, yet this is an important area for operators of congested systems to understand how will their passengers respond.

References

1. O'Farrell, P.N. & Markham, J. (1974) "Commuter Perceptions of Public Transport Work Journeys", Envt. & Plng. A16
2. Association of Metropolitan Authorities and Department of Transport (1984), Technical Seminar on Public Transport Demand Elasticities, Transport Studies Unit, University of Oxford
3. Coombe, D. (1990) Urban Transport Strategy Development: Top-down or Bottom-up? Transpn. Planning Systs, pp 3-10
4. Liivamagi, P. (1974) "Perception of Travel Times", Transport Studies Group, University College London (unpublished)
5. Goodwin, P.B. & Hensher, D.A. (1978) "The Transport Determinants of Travel Choices: An Overview", ch. 1 in Hensher, D.A. & Dalvi, Q. "Determinants of travel Choice"
6. Porter M.E. (1980) "Competitive Strategy: Techniques for Analysing Industries and Competitors", The Free Press, MacMillan Publishing Co. Inc., New York USA
7. Department of Transport, British Rail Network South East, London Regional Transport & London Underground Ltd. (1989) "Central London Rail Study"
8. Johnson, G. & Scholes, K. (1984) "Exploring Corporate Strategy", Prentice Hall International, London UK
9. Charles W. L. Hill, Gareth R. Jones (2004) "Strategic Management: An Integrated Approach", Houghton Mifflin Co., Boston USA
10. Davidson, P.W. (1989) "An Introduction to Stated Preference Techniques", PTRC Course" A Short Course on Stated Preference Techniques"
11. Kroes E., and Sheldon R., (1988) "Stated Preference Methods: An Introduction"
12. Ortuzar J. de D., & Willummsen L.G., (1990) "Modelling Transport", Wiley pp375
13. Ray, B., Davidson, P.W. & Chard, B. (1991) "Developing a Realistic Public Transport Strategy for the city of Truro from Public Attitudes", PTRC Summer Annual Meeting H pp 253-278
14. Goulcher, A. (1991) "Fares Elasticities for Underground Travel", PTRC Summer Annual Meeting G pp 99-110
15. Sheldon R., Bottom C., and Colob, T., (1986) "An Examination of the Priorities that Passengers Place upon Features of the London Underground System"
16. Bach M., & Benschop J. (1990) "Developments in Transportation Planning Software: MVGRAF – an Interactive Network Graphics Program", Traff. Engng. & Ctrl. pp 215-220

17. De Cea J., Bunster, J.P., Zubleta L & Florian M. (1988) “Optimal Strategies and Optimal Routes in Public Transit Assignment Models: an Empirical Comparison”, Traff. Engng. & Ctrl pp 520-526
18. Harris N.G. (1989) “Capacity Restraint Simulation in a Public Transport Environment” Traff. Engng. & Ctrl. pp 312-315
19. Harris N.G. (1991) “Modelling Walk Link Congestion and the Prioritisation of Congestion Relief”, Traff. Engng. & Ctrl pp 78-80
20. Maw J.R. & Dix M. (1990) “Appraisal of Station Congestion Relied Schemes on London Underground”, Proc. PTRC Summer Annual Mtg., University of Sussex pp H167-178
21. Miller S.G., Vincent J.F. & Wooltorton G.A. (1990) “EMME/2: A New Generation in Software”, Trans. Plng. Pp 21-32
22. Rickard J.M. & Fearon E.J. (1990) “advances in Public Transport Assignment Modelling: Application of proportional Assignment in Areas of Inner City Regeneration”, Proc. PTRC Summer Annual Mtg., University of Sussex pp H179-190
23. Wilson M. (1990) “Recent Developments in the Q-View Graphical Traffic Modelling Package”, Trans. Plng. Systems pp 49-60