TRANSPORT MODELLING IN THE CAPE TOWN METROPOLITAN AREA

JUSTIN BAZIMAZIKI MUNYAKAZI

A thesis submitted in partial fulfilment of the requirements for
the degree of Magister Scientiae in the Department of Mathematics
and Applied Mathematics, University of the Western Cape.

Supervisor: Dr NIZAR MARCUS

November 2005

Transport Modelling in the Cape Town Metropolitan Area

Justin Bazimaziki Munyakazi

KEYWORDS

Cape Town Metropolitan Area

Transport Modelling

Trip generation

O-D matrix

Modal split

MEPLAN

EMME/2

Random utility theory

Entropy

Gravity models

Abstract

TRANSPORT MODELLING IN THE CAPE TOWN METROPOLITAN AREA

J.B.Munyakazi

MSc Thesis, Department of Mathematics and Applied Mathematics, University of the Western Cape.

The use of MEPLAN by the Metropolitan Transport Planning Branch of the Cape Town City Council since 1984 was not successful due to apartheid anomalies. EMME/2 was then introduced in 1991 in replacement of MEPLAN. In this thesis we first introduce some aspects of transport modelling. Secondly we summarize the above-mentioned models before we undertake their comparative study in a post-apartheid situation. A mathematical proof of why MEPLAN was discarded is provided. The strengths and weaknesses of both MEPLAN and EMME/2 are recorded.

November 2005.

Declaration

I declare that Transport Modelling in the Cape Town Metropolitan Area is my own work, that it has not been submitted before for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged as complete references.

Justin Bazimaziki Munyakazi   November 2005

Signed:..

Acknowledgement

I profited immensely from the assistance of Dr Nizar Marcus as my supervisor. May he find the expression of my sense of gratitude for his time and patience.

I would like to record my sincere thanks to the following institutions and organisations which provided the necessary financial support to this work:

- The African Institute for Mathematical Sciences

- The Department of Mathematics and Applied Mathematics, University of the Western Cape

- The Faculty of Natural Sciences, University of the Western Cape

I owe a special debt of gratitude to Prof P. Witbooi, Prof R. Fray, Prof D. Kotze, Dr E. Mwambene, Mrs N. Myburgh and Mr J. Spotten who assisted me in various ways.

I would also like to thank my wife Adeline and daughters Julia and Joëlle for their patience, understanding and encouragement over many months of preoccupation.

Contents

Introduction

Any planning exercise of a city or a wider area is strongly related to land-use and transportation systems which play a major role in the socio-economic decision-making processes. The reliability and sustainability of these decisions depend on how well the models in use fit with the structure of the region under study.

The Cape Metropolitan Transport Branch of the City Council of Cape Town used MEPLAN (the Echenique Model) since 1984. MEPLAN is a land-use/transport model, strongly economic based, which was developed at the University of Cambridge by the firm MARCIAL ECHENIQUE & PARTNERS Ltd (ME& P). Due to the apartheid situation, the implementation of the model did not come up with successful results. The model assumes that low income households live closer to their workplace unlike high income households. This assumption did not meet the reality of the Cape Town Metropolitan Area. Modellers had therefore to artificially calibrate the parameters of the model. In other words, they had to 'create' attraction effects in lower income areas. This is one reason why MEPLAN was discarded and replaced by EMME/2 in 1991. The transport model EMME/2 was designed by INRO Consultants at the University of Montréal and is currently being used by the Cape Metropolitan Transport Branch. The other reason is linked to the fact that, in some areas, it was difficult to collect survey data because of a high number of informal settlements.

Since the abolition of the apartheid, people have the opportunity of choosing the place to live. This issue coupled with the redistribution of jobs closer to the low income areas and/or major employment creation in these areas (previously called the disadvantaged areas) 'might lead to better results in the prediction of future travel demand' [13] because of the change in the land-use characteristics. Therefore, the current structure of the Cape Town Metropolitan Area may partially or totally meet the requirements for the implementation of MEPLAN.

Our aim, in this thesis, is to analyse whether this is realistic. We discuss the structure, the implementation and the efficiency of MEPLAN and EMME/2. Their strengths and weaknesses are also recorded.

We hope that modellers and planners will find this work helpful in that it shows aspects in which either model is strong.

The three chapters of this work are organised as follows. The general aspects of transport modelling are overviewed in Chapter1. Chapter 2 presents a summary of MEPLAN and EMME/2 with an emphasis on their mathematical aspects. Chapter 3 deals with the comparison of these models. Their structure and implementation are discussed and their mathematics is analysed in depth allowing us to build conclusion upon the efficiency of the models. We mathematically show why the Echenique model was cancelled.

Chapter 1

Some Aspects of Transport Modelling

1.1 Introduction

The need for transportation is strongly related to the nature of human activities. Interactions between individuals, whenever it comes to business or social related issues, imply almost invariably trips of different nature, from an origin (home or workplace for example), to a destination (shop or office) where a transaction will take place. The collection of trips made by individuals and their distribution over the available network will determine a series of relevant factors to the urban planner:

- In which areas congestion is more likely to occur?

- How important are the levels of congestion?

- How are the different kinds of flows (individual vehicles, transit lines, pedestrians, etc.) distributed?

- How are emissions of pollutants distributed?

Transportation theory has given rise to numerous studies under various approaches, but in general, a common framework has emerged, which is often referred to as the 'four step' (five with the inclusion of peak hour models) or 'classical' method of travel prediction

This method is based on the hypothesis that any user, when it comes to travel, will make a series of successive independent choices:

1. to travel or not;

2. choice of destination;

3. choice of time period of travel;

4. choice of means (mode) of transportation;

5. choice of path followed.

This has led to develop five classes of models:

1. trip generation and attraction models;

2. trip distribution models;

3. teak hour estimation models;

4. modal split models;

5. route assignment models.

It should be noted that early work on travel prediction did not take peak hour estimation models into account. The average daily flows of traffic were rather estimated.

1.2 Trip Generation Modelling

The aim of the trip generation process in transport modelling is to establish formulae relating the number of trips likely to be made in the study area to its land-use characteristics as well as to the socioeconomic characteristics of the users.

1.2.1 Classification of Trips

Trips may be classified by purpose, by time of making them or by person type depending on the advantage perceived by the modeller on using any of these factors.

Compulsory (or mandatory) trips are mainly trips to work and education premises. Shopping trips, social and recreational trips are known to be discretionary (or optional) since they do not take place regularly and one can prevent oneself from making them without any major loss.

Trips are not uniformly distributed throughout the day and they are sometimes classified into peak and off-peak period trips.

The decision of making a trip and therefore the number of trips is strongly related to socioeconomic factors of individuals (income, car ownership, household size or structure).

Thus, modellers may set up a number of criteria which enable them to stratify individuals making trips according to their socioeconomic groups.

1.2.2 Factors Affecting Trip Generation

Trip production may take into consideration the following factors:

- income,

- car ownership,

- household structure,

- household size,

- value of land,

- residential density,

- accessibility.

Value of land and residential density are usually used in zonal studies. The accessibility factor is rarely taken into account in most practical studies although many authors do insist on its use (see [14]).

It should be noted that we are not considering freight trips. The reason for this is that, although their contribution to congestion and pollution is not negligible, they amount to very small proportion of trips.

1.2.3 Classification of Models

Growth Factor Method

The problem here is to find a way of determining the number of future trips T_i originating in zone i, given the current number of trips t_i. The Growth Factor Method consists of estimating the growth factor F_i so that

$$T_i = F_i t_i. \tag{1.1}$$

F_i is obviously connected to the population (P), income (I) and car ownership (C):

$$F_i = \frac{f(P_i^d, I_i^d, C_i^d)}{f(P_i^c, I_i^c, C_i^c)}, \tag{1.2}$$

d and c denote the design and current years. This method is mostly inadequate (see [12, p.127]).

Linear Regression Analysis

The linear regression model can operate at different levels of aggregation. It can be zone based to explain inter-zonal variations or household based to show the independance of zonal boundaries. At the zonal level (such as trips per zone) the regression model would be:

$$Y_i = \theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i} + \cdots + \theta_k X_{ki} + \varepsilon_i. \tag{1.3}$$

When rates or zonal means are considered (trips per household per zone), the model would be:

$$y_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \cdots + \theta_k x_{ki} + e_i \tag{1.4}$$

with $y_i = Y_i/H_i; x_i = X_i/H_i; e_i = \varepsilon_i/H_i$ and H_i the number of households in zone i. $X_{1i}, X_{2i}, \ldots, X_{ki}$ are the attributes (variables) of the model, $\theta_0, \theta_1, \ldots, \theta_k$ are parameters of the model to be determined by calibration. Equations (1.3) and (1.4) represent the same reality. However they are different in the sense that the error-terms ε_i and e_i do not bear the same distribution. In practice, (1.4) is preferred. The aggregate variables $X_{1i}, X_{2i}, \cdots, X_{ki}$ directly reflect the size of the zone, their use should imply that the magnitude of the error actually depends on zone size. Dividing these variables by H_i has the effect of reducing the variability of the variance since the model no longer depend on zone size.

The problem of intra-zonal variation may be dealt with by the reduction of zone size with the logic side effect of increasing the number of zones and consequently:

- the model may become more expensive in terms of data collection, calibration and operation;

- sampling errors may be magnified.

To avoid these inconveniences, households have been considered as the most appropriate analysis unit. In the linear regression model, it is assumed that each independent variable has a linear effect on the dependent variable. However, some variables (especially the qualitative ones such as sex, age, type of dwelling) may demonstrate non-linearity effects. To avoid this difficulty, two methods can be used ([12, p.139]):

- either transform the variables in order to linearise their effect (e.g. take logarithms, raise to a power);

- or use dummy variables.

Cross-Classification or Category Analysis

A method known as Category Analysis in the UK (Wootton and Pick 1967, cited in [12]) and Cross-classification in the USA was established and became popular in the late 1960s. The method was preferred to linear regression because it is household based and simple to handle. Furthermore, no assumptions are needed about the relationship between the independent variable and the dependent variables.

The method is based on a stratification of the population in the study area. For example, the $H(h)$ households could be categorised in the following way: m households sizes and n car ownership classes which lead to mn types h. The average number $t^p(h)$ of trips with purpose p is the total number of trips in cell h, by purpose, divided by the number of households $H(h)$ in it:

$$t^p(h) = \frac{T^p(h)}{H(h)}. \tag{1.5}$$

where $T^p(h)$ is the total number of (observed) trips in cell h. The aim is to choose the categories such that the standard deviations of the frequency distributions are minimised.

Some of the disadvantages of the method include:

- it is complex to predict the number of household, the $a_i(h)$'s (see equation (1.6) below) in categories in future;

- the method requires large samples;

- it is difficult to choose more appropriate stratifying variable.

The rate $t^p(h)$ is now used to compute trip productions with purpose p by person type n in zone i:

$$O_i^{np} = \sum_{h \in H^n(h)} a_i(h) t^p(h) \tag{1.6}$$

where $a_i(h)$ is the number of households of type h in zone i and $H^n(h)$ is the set of households of type h containing person type n.

1.3   Trip Distribution Models

1.3.1   A General Consideration

The trip distribution process which aims to estimate the future zone-to-zone trips may be achieved either by *growth-factor methods* or by the use of a *distribution model.*

The first approach supposes that a complete origin-destination survey has to be carried out. This survey results in an estimated number (base year) of trips t_{ij} being made per unit time from each origin to each destination. The estimated future number of trips T_{ij} could be written:

$$T_{ij} = A_i B_j t_{ij}. \tag{1.7}$$

Several iterative methods are used to determine the constants A_i and B_j. The most popular is the Furness method; also known as the 'balancing method'.

In the other approach, it is assumed that the number of trips per unit time between each pair of zones is proportional to a decreasing function, $f(C_{ij})$, of the cost of travelling between them. The most known forms of this function, sometimes referred to as the deterrence function, are: $\exp(-\beta C_{ij})$ (exponential function), C_{ij}^{-n} (power function) and $C_{ij}^n \exp(-\beta C_{ij})$ (combined function). It is also assumed that the number of zone-to-zone trips is proportional to the number of trips beginning at the origin zone i and ending at the destination zone j. We can therefore write:

$$t_{ij} = O_i D_j f(C_{ij}) \quad \text{and} \quad T_{ij} = A_i B_j t_{ij}, \tag{1.8}$$

where O_i and D_j are respectively the number of trips originating in zone i and attracted in zone j obtained from the trip generation models.

Both approaches lead to the same problem: finding suitable values of the balancing factors A_i and B_j. As it was said above, t_{ij} in a distribution model requires only a little survey work (since the deterrence function $f(C_{ij})$ usually contains a parameter to be calibrated) unlike in the growth-factor methods where a extensive survey is demanded. Evans (1970) proved that a sufficient condition for the convergence of the Furness method (in either approach) is $t_{ij} > 0$ for all i and j.

In this section, we briefly explore the growth-factor, the general gravity and the entropy maximising methods after a look at what the cost of travelling from a zone i to a zone j is meant to be.

1.3.2 Generalised Cost of Travel

The way zone-to-zone trips are distributed is heavily dependent upon the generalised cost of making these trips. This is a linear function of the attributes of the journey. All the aspects of the disutility towards making a trip are considered in terms of time, distance and money units. A typical form of the generalised cost for travel from zone i to zone j would be:

$$C_{ij} = a_1 t_{ij}^v + a_2 t_{ij}^w + a_3 t_{ij}^t + a_4 t_{nij} + a_5 F_{ij} + a_6 \phi_j + \delta. \tag{1.9}$$

where:

t_{ij}^v is the in-vehicle travel time between i and j;

t_{ij}^w is the walking time to and from stops or stations ;

t_{ij}^t is the waiting time at stops or stations;

t_{nij} is the interchange time, if any;

F_{ij} is the fare charged to travel between i and j;

ϕ_j is a terminal (typically parking) cost associated with the journey form i to j;

δ is a modal penalty (representing for example safety, comfort and convenience).

The parameters $a_1, \ldots, a_6$ have dimensions appropriate for conversion of all attributes to common units. It should be noted that the number of parameters in (1.9) depend on the mode on use. For instance, if users are not charged for parking, then $a_6 = 0$; $a_3 = 0$ if waiting time is not involved; $a_5 = 0$ if fare is not to be paid. Pubic transport users can be charged for parking when they leave their vehicles at the station.

1.3.3 Growth Factors Methods

Several growth-factor methods can be used according to the situation at hand. If the only information available is the growth rate τ of the number of trips for the whole study area, then $T_{ij} = \tau t_{ij}$. τ is said to be a *uniform growth-factor*. It could happen that the growth-factor is only known by origins (τ_i's) or by destinations (τ_j's). The future zone-to-zone trips could be determined by taking:

$T_{ij} = \tau_i t_{ij}$ for rates by origins,

$T_{ij} = \tau_j t_{ij}$ for rates by destinations.

This method is said to be *singly constrained*. The *doubly constrained growth-factors* method pointed out above (see equation (1.7)) supposes two sets of rates, one by origins and another by destinations. The rates A_i and B_j (the balancing factors) must be calculated so that the constraints

$$\sum_j T_{ij} = O_i, \tag{1.10}$$

$$\sum_i T_{ij} = D_j \tag{1.11}$$

are satisfied.

These methods are simple to apply. However, they are not responsive to changes in the transport network and can only be used for short period forecasting purposes. As it was said above, the base year trip matrix results from a complete survey study. This survey study must be as accurate as possible since the application of the method may amplify errors. Another serious disadvantage of the method is the zero elements in the base year trip matrix which are reproduced in the forecasting matrix.

1.3.4   General Gravity Models

Numerous trip distribution models have been developed and among them, the gravity model has been the most widely used. It adapts the Newtonian gravitational concept, introduced in 1686, to the problem of distributing traffic throughout an urban area. In this regard, the gravity model has the form:

$$T_{ij} = \frac{kP_iP_j}{d_{ij}^2} \quad \text{or more generally} \quad T_{ij} = \frac{kP_iP_j}{C_{ij}^n}$$

where P_i, P_j, d_{ij}, C_{ij} represent respectively the population at origin i, the population at destination j, the distance separating i and j and the generalised cost between i and j. The value of the square power led to several discussions and a compromise was made that it could be generalised to n. This may take a value between 0.6 and 3.5. The popularity of the model results from its simplicity in concept and its being well documented. It has been realised that there is a whole family of such models (see Wilson, 1970-A cited in [19, p.13]). Three of these are presented in the section devoted to the Entropy-maximising approach.

Several studies have been conducted to develop a gravity model which fits better with the traffic distribution problem. The last finding assumes that the effect of distance

could be modelled better by a 'separation' decreasing function of the generalised cost of travelling between the zones. The gravity model could be written as:

$$T_{ij} = \alpha O_i D_j f(C_{ij}). \qquad (1.12)$$

The proportionality factor α could be replaced by the two sets of factors A_i and B_j to yield (1.8) or alternatively,

$$T_{ij} = A_i O_i B_j D_j f(C_{ij})$$

which is the classical doubly constrained gravity model. We obtain the singly constrained versions in replacing one set of balancing factors A_i or B_j by one.

1.3.5 The Entropy-Maximising Approach.

The Idea of Entropy-Maximisation

The use of the entropy-maximising method has led to many of the advances observed today in transport modelling. A large number of models can be constructed from this approach including shopping models, location models and gravity models with different deterrence functions.

Consider a system made up of a large number of distinct elements. We need to identify three different levels of description; namely: the *micro-*, *meso-* and *macro-*level of a system. A full description of the system requires a complete specification of its micro-states. The name of each person making a trip is recorded in the appropriate origin-destination cell. This is done at the most detailed level. At a medium level of detail, the meso-level of description, we can sum up the number of names in each cell, to get T_{ij} (for the (i, j)-th cell to give the trip matrix (T_{ij})). At the higher level of description, the macro-level, we can focus only on the row and column totals, O_i and D_j, and also the total expenditure on transport, C.

It is clear that there are many possible micro-states associated with each meso-state, and many meso-states with each macro-state.

Suppose we are interested in the meso-level of description. We assume that each micro-state is equally probable and therefore the most probable meso-state is that with the greatest number of micro-states associated with it. Thus we must now calculate the number of micro-states associated with some meso-state (T_{ij}); let this number be $w(T_{ij})$. It is the number of distinct arrangements of individuals which give rise to the distribution

11

T_{ij} and is the number of ways in which T_{11} can be selected from T (the total number of trips; it is assumed that $T = \sum_i O_i = \sum_j D_j$), T_{12} from $T - T_{11},...$ and so on, and so:

$$w(T_{ij}) = \frac{T!}{T_{11}!(T - T_{11})!} \frac{T - T_{11}!}{T_{12}!(T - T_{11} - T_{12})!} \cdots = \frac{T!}{\prod_{ij} T_{ij}!}. \tag{1.13}$$

We need to find the (T_{ij}) which maximise (1.13) subject to the macro-level constraints. But this problem is equivalent to finding the (T_{ij}) which maximise

$$S = \ln w(T_{ij}) \tag{1.14}$$

since both problems have the same maximum, S being a monotone function of w. S is often referred to, in statistical mechanics, as the entropy function.

Entropy-maximisation and Gravity Models.

It is worth emphasising that the Entropy function was derived on the assumptions (1.10) and (1.11). The type of gravity model generated from the Entropy-maximisation depend on the type of additional constraints. Let us assume that the total amount spent on these trips in the region, and at the given point in time is a fixed amount C, that is:

$$\sum_i \sum_j T_{ij} C_{ij} = C. \tag{1.15}$$

Now we form the Lagrangian L:

$$L = \ln w + \sum_i \lambda_i^{(1)} \left(O_i - \sum_j T_{ij} \right) + \sum_j \lambda_j^{(2)} \left(D_j - \sum_i T_{ij} \right) + \beta \left(C - \sum_i \sum_j T_{ij} C_{ij} \right) \tag{1.16}$$

where $\lambda_i^{(1)}$, $\lambda_j^{(2)}$ and β are Lagrangian multipliers. The T_{ij}'s which maximise L, and which therefore constitute the most probable distribution of trips, are solution of

$$\frac{\partial L}{\partial T_{ij}} = 0 \tag{1.17}$$

and the constraints (1.10)-(1.11) and (1.15). The Sterling's approximation

$$\ln \mathcal{N}! = \mathcal{N} \ln \mathcal{N} - \mathcal{N} \tag{1.18}$$

leads to

$$\frac{\partial \ln \mathcal{N}!}{\partial \mathcal{N}} = \ln \mathcal{N} \tag{1.19}$$

12

and so:

$$
\begin{aligned}
\frac{\partial L}{\partial T_{ij}} &= \frac{\partial}{\partial T_{ij}}\Big(\ln w + \sum_i \lambda_i^{(1)}(O_i - \sum_j T_{ij}) + \sum_j \lambda_j^{(2)}(D_j - \sum_i T_{ij}) + \beta(C - \sum_i \sum_j T_{ij}C_{ij})\Big) \\
&= \frac{\partial}{\partial T_{ij}}\Big(\ln \frac{T!}{\prod_{ij} T_{ij}!}\Big) - \lambda_i^{(1)} - \lambda_j^{(2)} - \beta C_{ij} \\
&= \frac{\partial}{\partial T_{ij}}(\ln T!) - \frac{\partial}{\partial T_{ij}}\Big(\ln \prod_{ij} T_{ij}!\Big) - \lambda_i^{(1)} - \lambda_j^{(2)} - \beta C_{ij} \\
&= -\frac{\partial}{\partial T_{ij}}\Big(\sum_{ij} \ln T_{ij}!\Big) - \lambda_i^{(1)} - \lambda_j^{(2)} - \beta C_{ij}.
\end{aligned}
$$

Therefore:

$$
\frac{\partial L}{\partial T_{ij}} = -\ln T_{ij} - \lambda_i^{(1)} - \lambda_j^{(2)} - \beta C_{ij}. \tag{1.20}
$$

From (1.17) and (1.20), $\ln T_{ij} = -\lambda_i^{(1)} - \lambda_j^{(2)} - \beta C_{ij}$. Hence,

$$
T_{ij} = e^{-\lambda_i^{(1)} - \lambda_j^{(2)} - \beta C_{ij}}. \tag{1.21}
$$

(1.21) respectively in (1.10) and (1.11) give:

$$
e^{-\lambda_i^{(1)}} = \frac{O_i}{\sum_j e^{-\lambda_j^{(2)} - \beta C_{ij}}}, \tag{1.22}
$$

$$
e^{-\lambda_j^{(2)}} = \frac{D_j}{\sum_i e^{-\lambda_i^{(1)} - \beta C_{ij}}}. \tag{1.23}
$$

Write

$$
A_i = e^{-\lambda_i^{(1)}}/O_i, \tag{1.24}
$$

$$
B_j = e^{-\lambda_j^{(2)}}/D_j. \tag{1.25}
$$

and then

$$
T_{ij} = A_i B_j O_i D_j e^{-\beta C_{ij}}. \tag{1.26}
$$

This is the gravity model with an exponential deterrence function. From (1.22)-(1.25):

$$
A_i = \frac{1}{\sum_j B_j D_j e^{-\beta C_{ij}}}, \tag{1.27}
$$

$$
B_j = \frac{1}{\sum_i A_i O_i e^{-\beta C_{ij}}}. \tag{1.28}
$$

This model meets Evans's suggestion which states that 'the general form of $f(C_{ij})$ is unknown. The particular form $f(C_{ij}) = e^{-\beta C_{ij}}$ is satisfactory for describing urban travel

both for its theoretical properties and the fact that the resulting model seems to fit travel data well' [5].

If we replace the cost constraint (1.15) by

$$\sum_{ij} T_{ij} \ln C_{ij} = C' \tag{1.29}$$

the Lagrangian L becomes:

$$L = \ln w + \sum_i \lambda_i^{(1)} \left(O_i - \sum_j T_{ij} \right) + \sum_j \lambda_j^{(2)} \left(D_j - \sum_i T_{ij} \right) + \beta' \left(C' - \sum_i \sum_j T_{ij} \ln C_{ij} \right) \tag{1.30}$$

and therefore

$$\frac{\partial L}{\partial T_{ij}} = -\ln T_{ij} - \lambda_i^{(1)} - \lambda_j^{(2)} - \beta' \ln C_{ij} = 0. \tag{1.31}$$

Using (1.24) and (1.25), we get

$$T_{ij} = A_i O_i B_j D_j C_{ij}^{\beta'} \tag{1.32}$$

which is a gravity model with a deterrence power function. The joint use of (1.15) and (1.29) will give the following expression:

$$
\begin{aligned}
L = & \ln w + \sum_i \lambda_i^{(1)} \left(O_i - \sum_j T_{ij} \right) + \sum_j \lambda_j^{(2)} \left(D_j - \sum_i T_{ij} \right) + \beta \left(C - \sum_i \sum_j T_{ij} C_{ij} \right) \\
& + \beta' (C' - \sum_i \sum_j T_{ij} \ln C_{ij}).
\end{aligned} \tag{1.33}
$$

Using the same technique as above, we get

$$T_{ij} = A_i O_i B_j D_j C_{ij}^{-\beta'} e^{\beta C_{ij}} \tag{1.34}$$

which is the gravity model with a combined deterrence function.

1.4   A Note on Modal Split

A lot of attention has been given to modal choice models by transport researchers. The following are some non-detailed criteria of classification of those models:

a. *descriptive* models, establishing empirical relationships, *predictive* models, explaining relations between the elements introduced in the model, and *planification* models, which evaluate the consequences of different alternatives;

14

b. *deterministic* models, giving only one possible state, and *probabilistic* models, producing a probability distribution of possible states,

c. *analytical* or *statistical* models, in some mathematical form, and *simulation* models;

d. *dynamic* models and *static* models, depending whether time is introduced or not;

e. *aggregated* models, where groups are considered (category of population, resident population of a given area, etc) and average values used, and *disaggregated* models, where individual behaviour is considered.

An example of a descriptive, deterministic and aggregated model is:

$$Y = a - b_1 \ln(t^{TC}/t^A) - b_2 \ln R + b_3 \ln(P/S) + b_4 \ln(E/S) + b_5 TS$$

where

Y: proportion of trips made with public transport,

t^{TC} and t^A: times of trips with public transport and car,

R: income,

P/S: residential density,

E/S: employment density,

TS: price of parking.

In this model, some of the variables are correlated and this can dramatically affect the regression analysis.

Another example is that of descriptive, probabilistic and disaggregated model: probabilistic elements are introduced, thus modelling the inherent uncertainty of choices made. We have:

$$z = a_0 + a_1 \ln(t_2/t_1) + a_2 \ln(C_2/C_1) + a_3 \ln R + a_4 \ln(R/m)$$

where

t_1 and t_2: duration of trip for mode 1 and 2,

C_1 and C_2: cost of trip for mode 1 and 2,

R: income of household,

m: motorisation rate (number of vehicles per adult),

The probability function is given by

$$P(z) = \frac{e^{\alpha z + \beta}}{1 + e^{\alpha z + \beta}}.$$

The coefficients $a_0, a_1, \ldots, a_4, \alpha$ and β are the parameters of the model.

The entropy-maximising approach can be used to estimate the number of trips, T_{ij}^k by mode k from origin i to destination j. To do so, we need to solve the following problem:

$$\text{Maximise } S = -\sum_{ijk} \left(T_{ij}^k \ln T_{ij}^k - T_{ij}^k \right)$$

subject to

$$\sum_{jk} T_{ij}^k = O_i,$$

$$\sum_{ik} T_{ij}^k = D_j,$$

$$\sum_{ijk} T_{ij}^k C_{ij}^k = C.$$

The use of the Lagrangian multipliers method leads to the solution:

$$T_{ij}^k = A_i O_i B_j D_j e^{-\beta C_{ij}^k}; \tag{1.35}$$

$$P_{ij}^k = \frac{T_{ij}^k}{T_{ij}} = \frac{e^{-\beta C_{ij}^k}}{\sum_{k'} e^{-\beta C_{ij}^{k'}}} = \frac{1}{1 + \sum_{k' \neq k} e^{-\beta(C_{ij}^{k'} - C_{ij}^k)}}. \tag{1.36}$$

P_{ij}^k is the proportion of trips travelling from i to j by mode k. The generalised cost, C_{ij}^k, from i to j by mode k has the form:

$$C_{ij}^k = \sum_r a_r X_r(i, j, k)$$

where the X_r's are variables like fares, travel time and excess time. In the case of two modes, the above formula becomes:

$$P_{ij}^1 = \frac{T_{ij}^1}{T_{ij}} = \frac{e^{-\beta C_{ij}^1}}{e^{-\beta C_{ij}^1} + e^{-\beta C_{ij}^2}}$$

from which the following observations are obtained:

- If $C_{ij}^1 = C_{ij}^2$, then $P_{ij}^1 = P_{ij}^2 = 0.5$ meaning that if the cost of travelling from i to j by mode 1 is equal to the cost of doing so by mode 2, then there is no preference on a mode over the other.

- If C_{ij}^2 is very large compared to C_{ij}^1, then P_{ij}^1 is close to 1 meaning that all the travellers will tend to use mode 1.

A modal split model for each person type (see [18]) can also be released by a similar procedure to that outlined above:

$$\frac{T_{ij}^{kn}}{T_{ij}^n} = \frac{e^{-\beta^n c_{ij}^k}}{\sum_k e^{-\beta^n c_{ij}^k}} \tag{1.37}$$

where T_{ij}^{kn} and T_{ij}^n are the total number of trips from i to j by persons of type n using mode k and the total number of trips from i to j by persons of type n, respectively. Notice that now the mode summation is over the subset of modes available to persons of type n.

1.5 Discrete Choice Models

The Discrete Choice Models or Disaggregate Demand Models are based on observed choices by individual travellers. They deal with choice probabilities: they suppose that individuals have to select an option out of a number of alternatives. The *Random Utility Theory* which presents the general framework for the Discrete Choice Models takes into consideration the relative attractiveness of an option chosen by a traveller as it can be seen in the following postulate:

> *'The probability of an individual choosing a given option is a function of his/her socioeconomic characteristics and the relative attractiveness of the option'* (see Williams 1981, cited in [12, p.219])

The relative attractiveness of an option is associated to what is termed *utility* which is what individuals seek to maximise.

1.5.1 Utility and Consumer's Surplus

The utility of an option or an alternative is known to the user, but not to the modeller. Moreover, users do not usually have perfect information about the system. This may lead the user to make wrong choices. The modeller observes some attributes of the alternatives as faced by the user q, labelled x_{jq} for all j and some attributes of the user, labelled s_q and can specify a function that relates these observed factors to the user's (indirect) utility:

$$V_{jq} = V(x_{jq}, s_q)$$

for all j, where V_{jq} is the observed or representative utility of alternative j to user q. For these reasons, it is assumed that utility functions have deterministic (observed) utility component and a random (white noise) component:

$$U_{jq} = V_{jq} + \varepsilon_{jq} \tag{1.38}$$

where U_{jq} is the utility that user q obtains from alternative j. The ε_{jq}'s capture the factors that affect utility, but are not observable by the modeller. The user q chooses the alternative that provides the greatest utility:

$$U_q = \max_j U_{jq} \tag{1.39}$$

for all j. U_q takes account of the disutility of travel time and costs and can be referred to in terms of money units. For this purpose, the term *surplus* is used. By definition, a *person's consumer surplus* is the utility, in monetary terms, that a person receives in the choice situation. It can be calculated as follows:

$$\text{CS}_q = \frac{1}{\lambda_q} U_{jq} \tag{1.40}$$

where

U_{jq} is the utility that user q obtains from alternative j,

λ_q is the marginal utility of income and equals $\frac{\partial U_{jq}}{\partial Y_q}$ if j is chosen,

Y_q is the income of person q, and

U_q the overall utility for the person q.

Note that the division by λ_q converts utility into money units since

$$\frac{1}{\lambda_q} = \frac{\partial Y_q}{\partial U_{jq}}.$$

Taking into account (1.38), the modeller is able to calculate the expected consumer surplus by:

$$
\begin{aligned}
\mathbb{E}(\text{CS}_q) &= \frac{1}{\lambda_q} \mathbb{E}[\max_j U_{jq}], \quad \text{for all } j \\
&= \frac{1}{\lambda_q} \mathbb{E}[\max_j (V_{jq} + \varepsilon_{jq})], \quad \text{for all } j
\end{aligned}
$$

where the expectation is over all possible values of the ε_{jq}'s.

1.5.2 Logsum, Utility and Accessibility

If each ε_{jq} is independent and identically distributed as type I extreme value (the so called IID-Gumbel), then the expectation becomes:

$$\mathbb{E}(\mathrm{CS}_q) = \frac{1}{\lambda_q} \ln \left(\sum_j e^{V_{jq}} \right). \tag{1.41}$$

The term $\ln(\sum_j e^{V_{jq}})$ is called the 'logsum term'.

Recall that the IID-Gumbel function is given by $G(x) = \exp(-e^{-x})$, $\forall x \in \mathbb{R}$.

Under the usual interpretation of distribution errors, $\mathbb{E}(\mathrm{CS}_q)$ is the average consumer surplus in the subpopulation of people who have the same representative utilities as person q. The logsum is also interpreted as a measure of accessibility (see [14]): If C_q is a choice set, for multinomial logit (see subsection 1.5.4):

$$V_q' = \frac{1}{\mu} \ln \left(\sum_{j \in C_q} e^{\mu V_{jq}} \right) \tag{1.42}$$

where V_q' is the systematic component of the maximum utility i.e the measure of accessibility and μ is the scale parameter of the disturbance term ε_{jq}.

Note that the logsum is a suitable alternative of measuring 'composite cost' which can be used to obtain hierarchical logit models (see [10]).
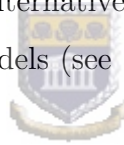
1.5.3 Basis of Discrete Choice Modelling

Equation (1.38) is not valid unless a certain homogeneity in the population under study is guaranteed. This requires that the same set of alternatives and constraints should be presented to all individuals.

On the other hand, the individual q selects the choice which presents the maximum utility to him: Individual q chooses j if and only if:

$$U_{jq} \geq U_{iq}, \quad \text{for all alternatives } i;$$

$$\text{i.e } \varepsilon_{jq} + V_{jq} - V_{iq} \geq \varepsilon_{iq}.$$

Therefore, the probability of individual q choosing alternative j is:

$$P_{jq} = \mathrm{Prob}\{\varepsilon_{iq} \leq \varepsilon_{jq} + V_{jq} - V_{iq}, \ \forall i\}. \tag{1.43}$$

Equation (1.43) is equivalent to

$$P_{jq} = \int_{R_N} f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N) d(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N) \tag{1.44}$$

19

where

$$R_N = \begin{cases} \varepsilon_{iq} \leq \varepsilon_{jq} + V_{jq} - V_{iq}, & \text{for each option } i, \\ V_{jq} + \varepsilon_{jq} \geq 0 \end{cases}$$

$f(\varepsilon_{iq})$ is the density function of the random variable ε_{iq}. Analytical models can be derived from (1.44) if the distribution of the ε_{iq} is known.

If these residual ε_{iq} are independent and identically distributed (IID), then the density function f can be decomposed in a product of utility functions $g(\varepsilon_n)$ associated with option n, for all $n \in \{1, \ldots, N\}$:

$$f(\varepsilon_1, \ldots, \varepsilon_N) = \prod_n g(\varepsilon_n).$$

Consequently, (1.44) becomes:

$$P_j = \int_{-\infty}^{\infty} g(\varepsilon_j) d(\varepsilon_j) \prod_{i \neq j} \int_{-\infty}^{\varepsilon_j + V_j - V_i} g(\varepsilon_i) d\varepsilon_i \tag{1.45}$$

where we have extended the range of both integrals to $-\infty$ in order to solve them. This inconsistency is not fatal since it introduces a very small error.

1.5.4 The Multinomial Logit Model (MNL)

If the ε_{iq} are IID Gumbel, then:

$$P_{iq} = \frac{e^{\beta V_{iq}}}{\sum_j e^{\beta V_{jq}}}. \tag{1.46}$$

The parameter β is related to the common standard deviation σ of the Gumbel variate by $\beta^2 = \pi^2/(6\sigma^2)$ (see [12, p.225]). In practise, β is taken to be equal to 1 as it cannot be estimated separately from the other parameters involved in the $V_{jq}f$'s.
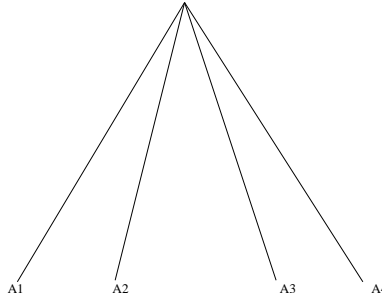

Figure 1.1: Multinomial logit model with 4 alternatives.

The MNL is the simplest and most popular practical discrete choice model. It satisfies the axiom of *independance of irrelevant alternatives* (IIA) which can be stated as:

Where any two alternatives have a non-zero probability of being chosen, the ratio of one probability over the other is unaffected by the presence or absence of any additional alternative in the choice set (Luce and Suppes cited in [12])

1.5.5   The Hierarchical Logit Model (HL)

In this type of models, also called the nested logit models, alternatives are grouped in subsets (nests) if they are expected to be more correlated. For instance, in a transport system where car, taxi, bus, rail and underground are used as means of transport, bus, rail and underground may be grouped into the nest 'Public Transport'(see Figure 1.2) since they are likely to display similar correlation patterns on the unobserved influences. For instance, they all involve waiting time, walking distance (to station or bus-stop, the
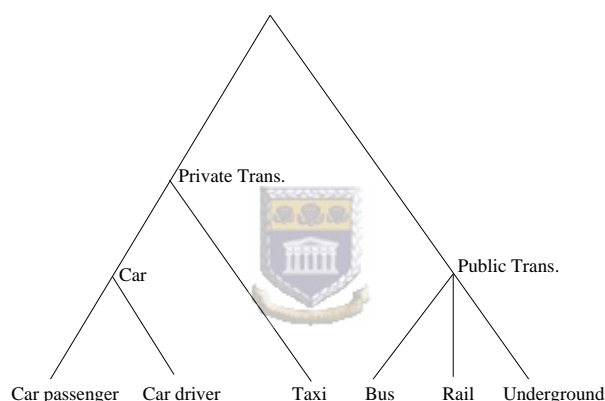
Figure 1.2: Hierarchical logit model

availability or possibility of weekly or monthly tickets (rather than the only daily tickets), etc. Car driver and car passenger are also correlated but less strongly. Indeed, Car Passenger has elements of private and public transport in that it relies on the provider for service while having door-to-door convenience.

Hierarchical structures can be estimated sequentially by determining an MNL for each nest and considering this as an alternative at the higher level of the hierarchy (In Figure 1.2, at the higher level, alternatives Bus, Rail and Underground is replaced by the composite alternative Public Transport) or simultaneously by analysing at once the full information about the system (see [8] and [12, pp.228-235]). The difference lies mainly in the calibration techniques.

Let us consider an example of bi-dimensional choices:

$$U(d, m) = U_d + U_{dm}.$$

U_d could stand for the portion of the utility specifically associated to the destination and U_{dm} the disutility associated to the cost of travelling. Using our previous notation:

$$U(d, m) = V(d, m) + \varepsilon(d, m),$$

where $\quad V(d, m) = V_d + V_{dm} \quad$ and $\quad \varepsilon(d, m) = \varepsilon_d + \varepsilon_{dm}.$

The hierarchical logit model is formed (assuming that the unobserved components of utilities ε are IID-Gumbel):

$$P(d, m) = \frac{e^{\beta(V_d + V_d^*)} e^{\lambda V_{dm}}}{\sum_{d'} e^{\beta(V_{d'} + V_{d'}^*)} \sum_{m'} e^{\lambda V_{dm'}}}. \tag{1.47}$$

This is equivalent to $P(d, m) = P_d.P_{dm}$ where P_d is the probability of choosing destination d (high level of the hierarchy) and P_{dm} is the probability of achieving destination d by mode m (low level of the hierarchy). $V_d^* = \frac{1}{\lambda} \ln \sum_{m'} e^{\lambda V_{dm'}}$ (see (1.42)) is the composite utility or expected maximum utility associated with all modes for destination d. If $\beta = \lambda$ in (1.47), then:

$$
\begin{aligned}
P(d, m) &= \frac{e^{\lambda(V_{d'} + \frac{1}{\lambda} \ln \sum_{m'} e^{\lambda V_{dm'}})} . e^{\lambda V_{dm}}}{\sum_{d'} e^{\lambda(V_{d'} + \frac{1}{\lambda} \ln \sum_{m'} e^{\lambda V_{d'm'}})} . \sum_{m'} e^{\lambda V_{dm'}}} \\
&= \frac{e^{\lambda V_d}(\sum_{m'} e^{\lambda V_{dm'}}) e^{\lambda V_{dm}}}{\sum_{d'} e^{\lambda V_{d'}}(\sum_{m'} e^{\lambda V_{d'm'}}) . \sum_{m'} e^{\lambda V_{dm'}}} \\
&= \frac{e^{\lambda(V_d + V_{dm})}}{\sum_{d'} e^{\lambda V_{d'}} \sum_{m'} e^{\lambda V_{d'm'}}} \\
&= \frac{e^{\lambda(V_d + V_{dm})}}{\sum_{d',m'} e^{\lambda(V_d' + V_{d'm'})}}.
\end{aligned}
$$

This shows that if $\beta = \lambda$, then the HL collapses to the MNL. This occurs when the source of correlation $\varepsilon_d = 0$. To make this clear, let us consider a fixed destination d and write down in full the utility expressions for a simple binary mode case:

$$U(d, 1) = V_d + V_{d1} + \varepsilon_d + \varepsilon_{d1},$$
$$U(d, 2) = V_d + V_{d2} + \varepsilon_d + \varepsilon_{d2}.$$

The term ε_d is found in both $U(d, 1)$ and $U(d, 2)$. Therefore, when ε_d becomes 0, there is no correlation and the model becomes a MNL.

In general, if i is the index representing an alternative or nest of the higher level, and j an option at the lower level inside the nest i, the utility function can be written as:

$$\begin{aligned}
U(i,j) &= U_i + U_{j/i} \\
U(i,j) &= V(i,j) + \varepsilon(i,j) \\
V(i,j) &= V_i + V_{j/i} \\
\varepsilon(i,j) &= \varepsilon_i + \varepsilon_{j/i}.
\end{aligned}$$

Then, the probability of choosing nest i, and inside it, option j is given by:

$$P_{ij} = P_i.P_{j/i} \tag{1.48}$$

$$\text{with} \quad P_{j/i} = \frac{e^{\lambda_i V_{j/i}}}{\sum_{k \in A^l(q)} e^{\lambda_i V_{k/i}}} \quad \text{and} \quad P_i = \frac{e^{\beta V_i}}{\sum_{j \in A^h(q)} e^{\beta V_j}}$$

where $V_{j/i}$ in the representative utility of option j inside nest i in which only those alternatives that vary inside the nest are considered. $A^h(q)$ is the set of alternatives at the higher level to user q. $A^l(q)$ is the set of options at the lower level to user q. The parameters β and λ_i correspond to the scale factors at the high level and nest i respectively ($\beta \geq 0$ and $\lambda_i \geq 0$). We have:

$$V_i = X_i + \frac{1}{\lambda_i} \ln \sum_{k \in A^l(q)} e^{\lambda_i V_{k/i}} \tag{1.49}$$

which means that, the utility of the nest is the expected maximum utility of all alternatives in the nest (excluding the component of utility associated to the common attributes of the nest alternatives) plus the component of utility to these common attributes. Multiplying both sides of (1.49) by β, we get:

$$\beta V_i = \beta X_i + \phi_i \ln \sum_{k \in A^l(q)} e^{\lambda_i V_{k/i}}$$

where $\phi_i = \frac{\beta}{\lambda_i}$. It is clear that $0 < \phi_i \leq 1$.

1.5.6   Other Choice Models

The MNL and HL are generated under certain sets of conditions. Different sets of conditions lead to other choice modes such as the Multinomial Probit Model (where the stochastic residuals ε are distributed multivariate Normal with mean zero and the errors have different variances and may be correlated in any fashions), the Mixed Logit, the Choice by Elimination and Satisfaction, etc. We will not elaborate on these models because they are beyond the scope of this thesis.

1.6 Assignment

1.6.1 Conceptual Framework

The demand side of Transport Modelling is made up of an indication of the number of trips by origin-destination pair and mode that would be made for a given level of service. This is achieved through trip generation, distribution and modal split. The network assignment process constitutes the supply side. A road network is represented by its links (and their associated nodes) and their costs. The costs are (strongly or less strongly) related to a number of attributes associated to the links: length, direction (from node A to node B, for instance), capacity, free-flow speed and speed-flow relationship.

The equilibrium between demand and supply, in this context, can be viewed in the following way: Suppose a fixed trip matrix of which travellers seek routes to minimise their travel costs (times). By trial and error, they will find a stable pattern. At this stage, if a given traveller changes his route, he does no longer improve his travel time, the pattern is said to be an *equilibrium*. This means that the travellers are already using the best routes available. In public-transport networks, generalised travel costs that travellers seek to minimise are affected by overcrowding, waiting and walking times, and in-vehicle times.

Network assignment is made in order to obtain good aggregate network measures such as total revenue by bus service, to estimate zone-to-zone travel costs (times) for a given level of demand and to obtain reasonable link flows and to identify heavily congested links. The estimation of routes used between each O-D pair, the analysis of which O-D pairs use a particular link or route are the secondary objectives.

Modellers need a number of elements to accomplish a network assignment:

- a demand matrix (Trip Table) for a relevant period: AM peak hour, 24 hours, 16 hours, off peak,

- a relevant network (nodes and links) for the time period,

- route choice criteria,

- assumptions about user behaviour.

As a starting point in an assignment process, travellers are assumed to select routes which offer the least perceived individual costs. Factors that influence this selection

include travel time, distance, monetary cost, congestion and queues, type of roads (e.g. highway, secondary way), road works. It is not however realistic to think of a generalised cost function which incorporates all theses factors.

Different drivers (users) often choose different routes when travelling between the same two points because they have different valuations of time (or monetary costs). Therefore, market segmentation is critical. Another reason is due to the fact that congestion effects affect shorter routes first and make their generalised costs comparable to initially less attractive routes. Drivers would experiment with all possible routes until they find a more or less stable arrangement where none can improve their travel time by switching to an other route. This is a case of *Wardrop's equilibrium*, which is discussed below. Diversion across routes in this case is due to *capacity restraint.*

Moreover, different users in the same segment have different perceptions or include different aspects in their generalised costs (or have different level of information). These differences in objectives and perceptions generate some *stochasticity* in route choice.

Particular types of models are built on the basis of one or more of the influences described above. These could be classified as follows:

- All-or nothing models: do not include either capacity restraint or stochastic effects.

- Wardrop's equilibrium models: include capacity restraint but not stochastic effects.

- Pure stochastic models: include stochastic effects but not capacity restraint.

- Stochastic equilibrium model: include both capacity restraint and stochastic effects.

Each assignment method comprises several steps and performs the following steps:

1. Identification of a set of routes which might be considered attractive to drivers: tree-building stage (build minimum cost tree);

2. Assignment of suitable proportions of the trip matrix to these routes (generation of flows on the links in the network);

3. Search for convergence: iterative techniques are usually used.

1.6.2   All-or-Nothing Methods

These methods consider the following assumptions:

- no congestion effects,

- no variations in perception of attributes for route choice.

Travel times are not adjusted (i.e link costs are fixed) by congestion since this has no effects. All users from origin i to destination j use the same route as a consequence of absence of variations in perception.

All-or-nothing assignment is a basic building block for other types of assignment techniques. It represents what drivers would like to do in absence of congestion and is not, in itself, of considerable interest to planners.

1.6.3 Stochastic Methods

They are based on the variability in drivers' perception of costs and the composite measure they seek to minimise (distance, travel time, generalised costs). To incorporate this variability, two methods are used: *Simulation-based* method which uses Monte Carlo simulations and the *Proportion* method which allocates flows to alternative routes from proportions calculated using logit-like expressions. Let us examine the later method.
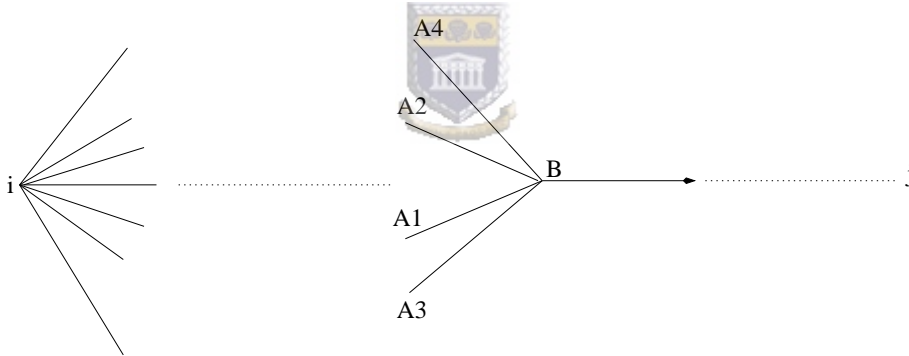
Figure 1.3: Stochastic proportional assignment.

B is a node between origin i and destination j (see Figure 1.3). Nodes A_1, A_2, A_3 and A_4 are possible entry points. Let us denote d_{A_i} the minimum cost of 0 from origin i to node A_i. The *splitting factors* f_i are defined as follows:

$$f_i = 0 \qquad \text{if } d_{A_i} \geq d_B,$$

$$0 < f_i \leq 1 \qquad \text{if } d_{A_i} < d_B.$$

The trips T_B that pass through B are divided according to the equation:

$$F(A_i, B) = T_B \frac{f_i}{\sum_i f_i}.$$

This equation shows that if an entry point A_i is further away from i than B, then link (A_i, B) is not loaded (i.e no trips are assigned to that link). Dial's method (see [12, p.338]) requires that:

$$f_i = e^{-\Omega \delta d_i}$$

where δd_i is the extra cost incurred in travelling from i to node B via mode A_i rather than via the minimum cost route. It is clear that,

if $\delta d_i = 0$, $f_i = 1$ and A_i lies in the minimum-cost route;

if $\delta d_i > 0$, $0 < f_i < 1$ showing that expensive route are less sollicited.

The split of trips from i to j among alternative routes r is:

$$T_{ijr} = T_{ij} \frac{e^{-\Omega C_{ijr}}}{\sum_r e^{-\Omega C_{ijr}}}. \tag{1.50}$$

The parameter Ω can be used to control the spread of trips among routes. C_{ijr} is the travel cost from i to j via route r. One of the weaknesses of this method is its ignorance of correlation between similar routes.

1.6.4 Congested Assignment

Wardrop's equilibrium methods (see [9, p.6-7] and [12, p.337]) emphasise capacity restraint rather than variability in drivers' perception and is stated as:

> *'Under equilibrium conditions traffic arranges itself in congested networks in such a way that no individual trip maker can reduce his path costs by switching routes'.*

Two alternative ways of assigning traffic onto a network were proposed by Wardrop in 1952. The first, known as Wardrop's first principle states that if all trip makers perceive costs in the same way:

> *'Under equilibrium conditions traffic arranges itself in congested networks such that all used routes between an O-D pair have equal and minimum costs while all unused routes have greater or equal costs'.*

Wardrop's second principle on the other hand is enunciated as follows:

> *'Under social equilibrium conditions traffic should be arranged in congested networks in such a way that the average (or total) travel cost is minimised'.*

The first principle is oriented to modelling individual drivers' behaviour trying to minimise their own trip costs. The principle is therefore sometimes referred to as *selfish* or *users' equilibrium* in opposition to the second principle known as the *social equilibrium* which is oriented towards planners and engineers trying to manage traffic to minimise travel costs and therefore achieve an optimum social equilibrium.

Considering the two relationships about costs and the make up of flows on links:

$$V_a = \sum_{ijr} \delta_{ijr}^a T_{ijr}, \tag{1.51}$$

$$C_{ijr} = \sum_a \delta_{ijr}^a C_a(V_a), \tag{1.52}$$

where T_{ijr} is the number of trips from i to j via route r, V_a, the volume on link a, C_{ijr}, the cost of travel from i to j via route r and $C_a(V_a)$, the cost on link a.

$$\delta_{ijr}^a \begin{cases} = 1, & \text{if link } a \text{ is part of route } r \\ = 0, & \text{otherwise,} \end{cases}$$

the selfish equilibrium can be translated as follows:

$$C_{ijr} \begin{cases} = C_{ij}^* & \text{if } T_{ijr}^* > 0 \\ \geq C_{ij}^* & \text{if } T_{ijr}^* = 0. \end{cases} \tag{1.53}$$

This can be set up as a mathematical programming problem:

$$\text{Minimize} \quad Z(T_{ijk}) = \sum_a \int_0^{V_a} C_a(v) dv \tag{1.54}$$

$$\text{subject to} \quad \sum_r T_{ijr} = T_{ij} \tag{1.55}$$

$$\text{and} \quad T_{ijr} \geq 0. \tag{1.56}$$

Z is convex. Indeed:

$$\begin{aligned}
\frac{\partial Z(T_{ijr})}{\partial T_{ijr}} &= \frac{\partial}{\partial T_{ijr}} \sum_a \int_0^{V_a} C_a(v) dv \\
&= \sum_a \frac{\partial}{\partial T_{ijr}} \int_0^{V_a} C_a(v) dv \\
&= \sum_a \frac{d}{dV_a} \left(\int_0^{V_a} C_a(v) dv \right) \frac{\partial V_a}{\partial T_{ijr}} \\
&= \sum_a C_a(V_a) \frac{\partial V_a}{\partial T_{ijr}} \\
&= \sum_a C_a(V_a) \delta_{ijr}^a.
\end{aligned}$$

28

$$\Rightarrow \frac{\partial Z(T_{ijr})}{\partial T_{ijr}} = C_{ijr} \qquad (1.57)$$

$$
\begin{aligned}
\frac{\partial^2 Z(T_{ijr})}{\partial T_{ijr}^2} &= \frac{\partial}{\partial T_{ijr}} \sum_a C_a(V_a)\delta_{ijr}^a \\
&= \sum_a \frac{\partial C_a(V_a)}{\partial T_{ijr}}\delta_{ijr}^a \\
&= \sum_a \frac{dC_a(V_a)}{dV_a}\frac{\partial V_a}{\partial T_{ijr}}\delta_{ijr}^a \\
&= \sum_a \frac{dC_a(V_a)}{dV_a}\delta_{ijr}^a\delta_{ijr}^a.
\end{aligned}
$$

$$(1.58)$$

$\frac{dC_a(V_a)}{dV_a}$ is positive since the cost-flow on link a is known to be an increasing function of the flow on that link and this is a general requirement for convergence of Wardrop's equilibrium. Thus:

$$\frac{\partial^2 Z(T_{ijr})}{\partial T_{ijr}^2} \geq 0.$$

The mathematical programming problem can now be solved by the 0 multipliers method:

$$L(T_{ijr}, \phi_{ij}) = Z(T_{ijr}) + \sum_{ij} \phi_{ij}\left(T_{ij} - \sum_r T_{ijr}\right).$$

We then get:

$$\frac{\partial L}{\partial T_{ijr}} = \frac{\partial Z(T_{ijr})}{\partial T_{ijr}} - \phi_{ij} = C_{ijr} - \phi_{ij}.$$

We have two possibilities with respect to value of T_{ijr} at the optimum:

If $T_{ijr}^* = 0$, then

$$\frac{\partial L}{\partial T_{ijr}} \geq 0 \quad \text{because the function is convex.}$$

If $T_{ijr}^* \geq 0$, then

$$\frac{\partial L}{\partial T_{ijr}} = 0.$$

In other words,

if $T_{ijr}^* = 0$, then $C_{ijr} \geq \phi_{ij}^*$ for all ijr,

if $T_{ijr}^* > 0$, then $C_{ijr} = \phi_{ij}^*$ for all ijr.

ϕ_{ij}^* must be equal to the minimum cost of travelling from i to j: $\phi_{ij}^* = C_{ijr}^*$. Thus, the set of T_{ijr} that minimises the objective function Z satisfies (1.53). It is worth emphasising that the problem solved here assumes that the delay on a link depends on flows on the link itself.

1.6.5   Public Transport (Transit) Assignment

Route choice and assignment for passengers using public transport are heavier than in private transport in many respects. For instance, in private transport the concern is the movement of vehicles whereas public transport considers the movement of people. Moreover monetary costs in public transport are of various kinds: variable fares with distance, flat fares (independent of distance), time limit fares (e.g valid for any number of boardings in an hour), season tickets for a fixed service (daily, weekly, monthly, etc). This variability in fares makes route choice and assignment more challenging compared to what is observed in private transport where monetary costs are proportional to travelled distance.

The Concept of Strategy

Transit assignment models are based on the hypothesis that users select a strategy, instead of a single path between origin and destination. To illustrate the concept of strategy, let us use the following example (see Figure 1.4): A small transit network consists of four bus lines and four bus stops. For each line the frequency (minutes between 2 buses, in parentheses) and the time between two bus stops are known.
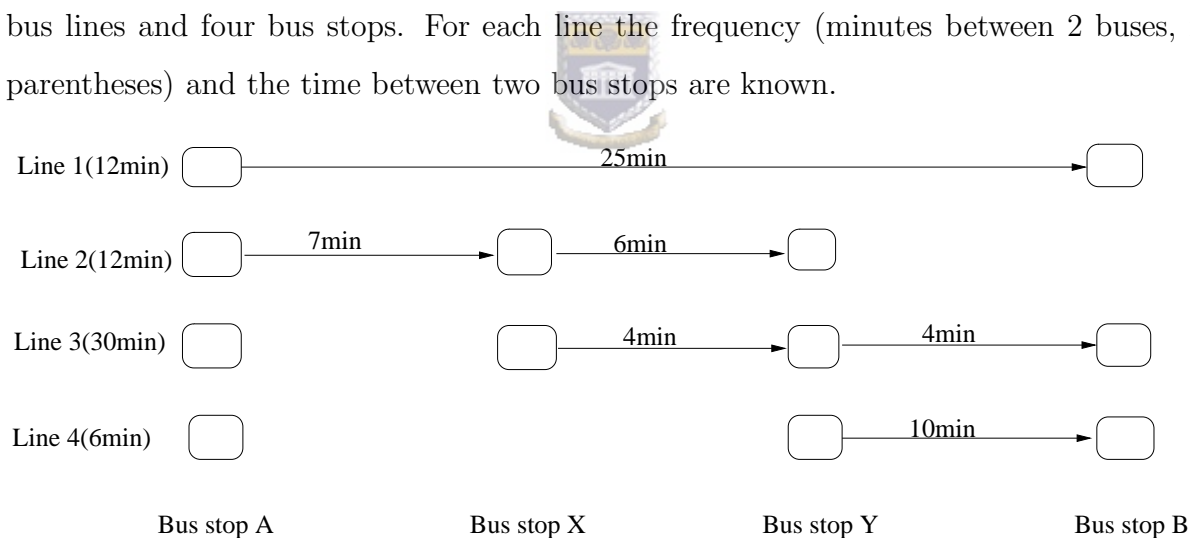
Figure 1.4: The concept of strategy - a small transit network

If we consider the case of a user that has to travel from A to B, several different paths are available, involving or not changing line at a given bus stop. This user could for example choose line 1 until B, or taking line 2 until stop X, and change for line 3 until stop B, or even take line 2 up to stop Y then change for line 4.

One would be tempted to formulate the problem as 'find the path between A and

30

B that minimises total expected travel cost'. This formulation is misleading, since any user, instead of selecting a single path, would rather board the bus that arrives next. The choice is therefore more complex than selecting a single path out of all possible paths.

We now define a strategy as a set of rules that, when applied, allows the user to reach his destination. Such a strategy could be, in the case of our example, be formulated as: 'at stop A take whatever bus arrives first. If line 1 was taken, exit at stop B. If line 2 was taken, exit at stop Y and board line 3 or line 4 depending on which arrives next.

We denote as attractive lines, the lines that allow the user, at each stop, to reach his destination. We assume the waiting time for any given line to be half of the interarrival time.

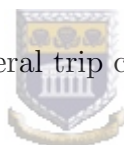The line probability is the probability that a line will be boarded and is equal to its frequency divided by the total frequency i.e at node X, where attractive lines 2→Y, 3→Y are selected, the probability of boarding line 2 is given by $\frac{5}{2+5}$, that is: [# buses/hour on line 2]/[# buses/hour line 2 + # buses/hour line 3].

The Transit Assignment Model

A transit trip consists in general of several trip components that may include some or all of the following:

- access from origin to transit stop,

- waiting for a vehicle,

- riding in a vehicle,

- alighting a vehicle,

- walking between two transit stops,

- egress from transit stop to destination.

These trip components, with the exception of 'waiting for a vehicle', are usually quantified by a nonnegative time (or cost). The component 'waiting for a vehicle' is quantified by using the statistical distribution of waiting times for the arrival of the first vehicle of a given transit line at a given stop. The trip components are represented by links $a \in A$ of a network $G = (I, A)$ where nodes $i \in I$. Each link $a \in A$ is characterised by the pair

(c_a, G_a) where c_a is a nonnegative link travel time and G_a the distribution function for the waiting time

$$G_a(x) = \text{prob}\{\text{waiting time on link } a \leq x\}.$$

The functions $G_a(x)$ can be obtained from the distribution of interarrival times (headways) and the distribution of passenger arrival times.

$$G_a(x) = \int_a^x g_a(t)dt$$

where

$$g_a(x) = \frac{1 - H_a(x)}{\int_a^\infty (1 - H_a(t))dt} : \text{ waiting time distribution for service on link } a.$$

$$H_a(x) = \int_a^x h_a(t)dt \text{ where } h_a(x) \text{ is the density function of the distribution of}$$
interarrival times (headways) of the vehicles on link a.

Note that uniform arrival of passengers at the transit stop is assumed. For a link a that does not involve waiting, we have

$$G_a(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0. \end{cases}$$

For the transit route choice problem in this generalised form, a strategy to reach destination node r is defined by a partial network $G_r = (I, \bar{A})$ that contains only those links that will be used as a consequence of this strategy. Let us write: $\bar{A}_i^+ = A_i^+ \bigcap \bar{A}$, $i \in I$. Among the links that are included in the strategy $\bar{A}$, at such node $i \in I$, a traveller boards the first vehicle that serves any of the links $a \in \bar{A}_i^+$. Hence, $\bar{A}_i^+$ corresponds to the set of attractive lines and, of course $\bar{A}_i^+ \neq \emptyset$ for $i \neq r$.

Let $W(\bar{A}_i^+)$ denote the expected waiting time for the arrival of the first vehicle serving any of the links $a \in \bar{A}_i^+$. $W(\bar{A}_i^+)$ is called the combined waiting time of links $a \in \bar{A}_i^+$. Let further $P_a(\bar{A}_i^+)$ be the probability that link a is served first among the links $\bar{A}_i^+$ ($\forall a \notin \bar{A}_i^+$, $P_a(\bar{A}_i^+) = 0$).

$W(\bar{A}_i^+)$ and $P_a(\bar{A}_i^+)$ depend on the distributions of waiting time $G_a(x)$ according to the following relationships:

$$W(\bar{A}_i^+) = \int_0^\infty \prod_{a \in \bar{A}_i^+} \{1 - G_a(x)\}dx.$$

$$P_a(\bar{A}_i^+) = \int_0^\infty g_a(x) \prod_{a' \in \bar{A}_i^+, \, a' \neq a} \{1 - G_{a'}(x)\}dx.$$

The model that will be presented here is based on the assignment of the trips from all nodes towards a single node r, that we shall denote the destination node. Let $g_i, i \in I - \{r\}$, be the demand (number of trips) from node i to the destination node r. In order to simplify the notation, we define $g_r = -\sum_{i \neq r} g_i$, $\forall i \in I - \{r\}$, $g_i > 0$. We note however, that all results presented below remain valid for $g_i \geq 0$, $i \in I - \{r\}$.

The volume at a node, which we denote $V_i, i \in I$, is the sum of the volumes of all incoming links and the demand at that node:

$$V_i = \sum_{a \in A_i^-} v_a + g_i, \quad i \in I. \tag{1.59}$$

The node volume V_i is distributed on the outgoing links according to their link probabilities under strategy $\bar{A}$

$$v_a = P_a(\bar{A}_i^+)V_i, \quad a \in \bar{A}_i^+, i \in I. \tag{1.60}$$

Since $\sum_{a \in A_i^+} P_a(\bar{A}_i^+) = 1$, from (1.59) and (1.60) we have the conservation flow equation:

$$\sum_{a \in A_i^+} v_a - \sum_{a \in A_i^-} v_a = g_i, \quad i \in I. \tag{1.61}$$

The optimal strategy $\bar{A}^*$ is the strategy that minimises the expected total travel time including waiting time.

In the following, we consider the special case in which the waiting time distribution of each link a (or transit line, in the original form of problem) is quantified by a positive parameter f_a, which will be called the frequency of a link. The expected combined waiting time and the link probabilities are derived from the frequencies in the following way:

$$W(\bar{A}_i^+) = \frac{\alpha}{\sum_{a \in \bar{A}_i^+} f_a}, \quad \alpha > 0. \tag{1.62}$$

$$P_a(\bar{A}_i^+) = \frac{f_a}{\sum_{a' \in \bar{A}_i^+} f_{a'}}, \quad a \in \bar{A}_i^+. \tag{1.63}$$

The case $\alpha = 1$ corresponds to an exponential distribution of interarrival times of the vehicles with mean $1/f_a$ and a uniform passenger arrival rate at the nodes. The case $\alpha = \frac{1}{2}$ is an approximation of a constant interarrival time $1/f_a$ for the vehicles on link a. This measure of waiting time is the most widely used approach in practise, in spite of the fact that it is based on a rough approximation.

Let us assume $\alpha = 1$. In order to take advantage of the special case (1.62) and (1.63) for the statement of the model, we need to express the strategy $\bar{A}$ in terms of the 0 - 1

variables x_a.

$$x_a = \begin{cases} 0, & \text{if } a \notin \bar{A} \\ 1, & \text{if } a \in \bar{A} \end{cases} \quad a \in A. \tag{1.64}$$

The problem of finding the optimal strategy $\bar{A}^*$ is now stated as follows:

$$\text{Min} \sum_{a \in A} c_a v_a + \sum_{i \in I} \frac{V_i}{\sum_{a \in A_i^+} f_a x_a} \tag{1.65}$$

subject to

$$v_a = \frac{x_a f_a}{\sum_{a' \in A_i^+} f_{a'} x_{a'}} V_i, \quad a \in A_i^+, \quad i \in I \tag{1.66}$$

$$V_i = \sum_{a \in A_i^-} v_a + g_i, \quad i \in I \tag{1.67}$$

$$V_i \geq 0, \quad i \in I \tag{1.68}$$

$$x_a = 0 \text{ or } 1, \quad a \in A. \tag{1.69}$$

At first sight, the problem defined by (1.65) to (1.69) has a nonlinear objective function that is subject to nonlinear constraints, and the variables are partly continuous, partly integer. However, this problem may be reduced to a much simpler linear programming problem by observing the following:

1. The nonnegativity constraints for the node volume (1.68) may be replaced by non-negativity constraints for the link volume because of (1.69).

$$v_a \geq 0, \quad a \in A. \tag{1.70}$$

2. By summing the constraints (1.66) for $a \in A_a^+$, we obtain:

$$\sum_{a \in A_i^+} v_a = \sum_{a \in A_i^+} \frac{x_a f_a}{\sum_{a' \in A_i^+} f_{a'} x_{a'}} V_i, \quad i \in A \text{ i.e}$$

$$\sum_{a \in A_i^+} v_a = V_i, \quad i \in A. \tag{1.71}$$

Equation (1.71) in (1.67):

$$\sum_{a \in A_i^+} v_a - \sum_{a \in A_i^-} v_a = g_i, \quad i \in A \tag{1.72}$$

which is the conservation of flow constraint.

34

3. By introducing new variables ω_i that denote the total waiting time for all trips at node i

$$\omega_i = \frac{V_i}{\sum_{a \in A_i^+} f_a x_a}, \quad i \in I \tag{1.73}$$

the variables $V_i, i \in I$, may be substituted.

These arguments prove that the problem defined by (1.65) to (1.69) is equivalent to the following problem:

$$\text{Min} \sum_{i \in A} c_a v_a + \sum_{i \in I} \omega_i \tag{1.74}$$

subject to (1.61), (1.69), (1.70) and

$$v_a = x_a f_a \omega_i, \quad a \in A_i^+, i \in I. \tag{1.75}$$

The objective function is now linear and the node volumes $V_i, i \in I$, no longer appear explicitly in the formulation. The 0-1 variables x_a are only used in constraints (1.75), which are the only nonlinear constraints in the above mentioned formulation of the problem. These constraints may be relaxed by replacing (1.75) with

$$v_a \leq f_a \omega_i, \quad a \in A_i^+, i \in I \tag{1.76}$$

to yield the linear programming problem. This becomes:

$$\text{Min} \sum_{i \in A} c_a v_a + \sum_{i \in I} \omega_i \tag{1.77}$$

subject to (1.61), (1.70) and (1.76). This model was developed by H. Spiess and M. Florian in [17]

1.6.6 Modal Split - Route Split

Wilson in [18] emphasised the relevance of mode choice and route choice in the assignment part of a transport model.

The composite impedance C_{ij} is constructed from the modal costs C_{ij}^k. Now, within a mode k between i and j which may consist of several routes, costs are observed on those routes.

Let γ_{ij}^r be the (observed) cost of travelling on the r-th route between i and j. A mode can be defined as a set of routes. Let $R_{ij}(k)$ be the set of routes between i and j which we define to be mode k. Let $M_{ij}(n)$ be the set of modes available to type n people from

i to j. $M_{ij}(n)$ is also the set of routes available to type n people from i to j. Let S_{ij}^{rn} be the number of trips between i and j by persons of type n on the r-th route between i and j. Then:

$$S_{ij}^n = \sum_{r \in R_{ij}(k)} S_{ij}^{rn} = \sum_{k \in M_{ij}(n)} T_{ij}^{kn} = T_{ij}^n. \tag{1.78}$$

Two possible mechanisms might determine route split within the maximum entropy methodology:

1. That people perceive route costs directly, and that a route split formula can be developed by analogy with (1.37), but using a parameter μ^n to allow for the possibility of its being different from β^n. Then:

$$\frac{S_{ij}^{rn}}{S_{ij}^n} = \frac{e^{-\mu^n \gamma_{ij}^r}}{\sum_{r \in M_{ij}(n)} e^{-\mu^n \gamma_{ij}^r}} \tag{1.79}$$

is the appropriate equation.

2. That people perceive mode costs directly, and that mode split is determined by (1.37). Route split is then determined within modes thus:

$$\frac{S_{ij}^{rn}}{T_{ij}^{kn}} = \frac{e^{-\mu^n \gamma_{ij}^r}}{\sum_{r \in R_{ij}(k)} e^{-\mu^n \gamma_{ij}^r}}. \tag{1.80}$$

The composite impedances C_{ij} are constructed out of the C_{ij}^k's according to (1.42). In a similar way, the C_{ij}^k's are constructed out of the γ_{ij}^r's.

Chapter 2

MEPLAN and EMME/2: A Summary

2.1 MEPLAN-The Echenique Model

MEPLAN is a mathematical framework and software package for modelling the spatial economies of cities or regions. It is called a 'model of land-use/transport interaction' [1].

2.1.1 Structure and Logic of the Model

Structure of the Model

The model consists of three sub-models: the metropolitan model, the land-use model and the transport model. The metropolitan model is concerned with the totals of population and employment in the study region as a whole. The land-use model deals with the location of activities and relationships between them. Activities are subdivided in two classes: *employment activities* (measured in jobs) and *residential activities* (measured in households). In the employment activities category, we have:

1. Agriculture, fishing and mining;

2. Manufacturing industry;

3. Shopping and commerce;

4. Education;

5. Other non-commercial services.

whereas in the category of residential activities

1. Low-income white households;

2. High-income white households;

3. Low-income coloured and Asians households;

4. High-income coloured and Asians households;

5. Black households

are the options.

The land-use model calculates a matrix of the spatial-functional relationships between zones. These spatial-functional relationships are of four kinds:

- flows of labour from households in each zone to employment in each zone,

- flows of goods and services from employment in each zone to households in each zone,

- flows of labour from households directly to other households (i.e. domestic service),

- flows of goods and services from industry (i.e. employment) in each zone to industry in each zone. (These are not included in the land-use model.)

The transport model takes the functional flows calculated by the land-use model and works out the personal travel needed to deliver these flows. It calculates the demand for travel given the established pattern and linkages of activities; it calculates the supply of travel in terms of the times and costs between zones by each mode and sub-mode. It also finds a partial equilibrium between supply and demand, given that the demand will cause congestion which increases times, these increments will influence the demand, and so on.

The most complex input to the transport side of the model is the description of the transport network. An important feature of the transport model is the fact that all links of the road network are allowed for private cars, buses, kombi-taxi and goods vehicles.

The different modes used in the Cape Town situation are:

1. Private car/motor cycle,

2. Conventional bus,

3. Kombi-taxi and conventional taxi,

4. Train,

5. Walk.

A separate mode for goods movements by truck was also considered. These modes are organised into the following hierarchy, to represent the decision-making pattern of each group of travellers (see Figure 2.1). At the upper level of the hierarchy, the options are car, public transport and walk, the options being, at the lower level for public transport, train, kombi and bus.
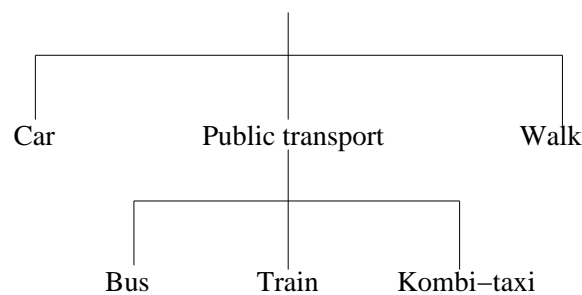
Figure 2.1: The hierarchy of modes in MEPLAN

Because MEPLAN (the multi-purpose software package developed by Marcial Echenique & Partners Ltd in 1984) is an economic model, employments, households and floorspace are treated as economic factors to be consumed. The different characteristic of these -for example, that service employment in a zone may serve consumers in many zones, but floorspace must be used where it has been constructed- have to be defined in the input to the model. All factors are both produced (or supplied) and consumed. The distinction must be made between factors that may be supplied from one zone and consumed in another, called *transportable* factors, and those that must be used where they are supplied, called *non-transportable* factors. Service employment and labour are the transportable factors in the model implementation, while floorspace is the main non-transportable factor.

Logic of MEPLAN

The building blocks of the model fit together in the theoretical basis of the model according to the diagram of Figure 2.2.

```
┌─────────────────────────────────────┐
│        Long–Run Land–use Model       │
└─────────────────────────────────────┘
                   │
                   ▼
┌─────────────────────────────────────┐
│   Short–Run Equilibrium Land–Use Model │
└─────────────────────────────────────┘
                   │
                   ▼
        ┌──────────────────────┐
        │    Flow generation    │
        └──────────────────────┘
                   │
                   ▼
          ┌──────────────────┐
          │    Modal Split    │
          └──────────────────┘
                   │
                   ▼
          ┌──────────────────┐
          │    Assignment     │
          └──────────────────┘
```
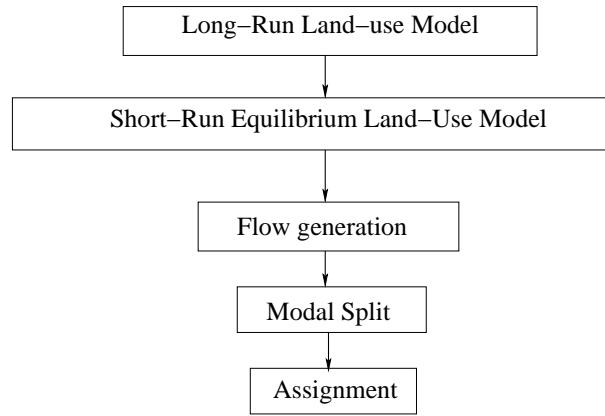
Figure 2.2: The logic of MEPLAN

The long-run, the short-run and the flow generation take place within the land use model. Modal split and Assignment are part of the transport model. The long-run model predicts changes over time in those aspects of the urban system that are least sensitive to local factors, such as employment in heavy industry, or that change only gradually, such as the stock of buildings in each zone. The short-run model predicts for a specific date the connections between activities and the location of the more mobile activities such as residence and service employment. It also estimates values for floorspace or land due to the competition for them. The connections between activities generate flows (trips), which are allocated to modes and within mode assigned to the transport network. The transport side of the model includes feedback effects, so that congestion of the system will affect travel times which will affect modal choice; changes in modal choice will alter both the time and cost of travel which may influence the number of trips made. The implementation of MEPLAN starts at the bottom of this sequence i.e the network assignment (most detailed process) and ends with the long-run land-use model.

2.1.2   Land-Use Model

The main module for the land-use model is cross-sectional: it predicts some parts of the system from information about the stage of other parts of the system at the same time. This module deals simultaneously with the full set of activities and flows. It is however preceded by two more specific modules which handle the location of 'basic' activities and the development of floorspace. These are incremental: they explicitly predict changes through time.

General Model Process

The starting point for the Main Activity Allocation Module is the set of figures for basic activity of each type in each zone. The model calculates, for each activity in turn, the amounts of further activity generated -for example, the households (labour supply) generated by employment (labour demand), or the service jobs (supply) generated by households' demand. These generated activities then have to be located in the zones from which they will send flows of labour or goods and services to the activities that generate them. In other words, the model must predict where units of activities will locate in order to supply whatever is demanded of them. The trade of labour from household to work forms the basis for calculating daily flows of passengers to and from work. The households in turn generate demand for services, for domestic labour (from other households) and for floorspace. The pattern for these trades is predicted, and in turn forms the basis for calculating flows of shopping and educational flows, and the remaining work flows. The chain of demand and supply calculation is continued until all the factors have been located and all the trades in labour and services have been generated.

There are two ways of locating activities. If the demanded activity is a household, the theory of consumer' behaviour is applied: consumers are assumed to maximise their utility subject to the constraint represented by their income. If the demanded activity is a unit of employment, then the theory of firms' behaviour is applied: firms tend to maximise their profit (minimise production costs).

It is worth noting that a change in the rent directly affects the demand for floorspace by each consumer in the zone; this then indirectly affects the utility of locating in that zone and hence the number of consumers who will choose to live there. Consumers will tend to locate in zones with greater 'total utility', but in a distribution reflecting the facts that they will also consider factors other than those modelled, and they will make varying individual assessments of the factors that are modelled.

As it is said above, because the concept of utility does not apply for the location of employment generated by demand within the region, a process related to profit-maximising is used instead. This assumes that employment will occur where industries can best supply the demand for the goods and services they produce, subject to the same variations of unmodelled factors and differing perceptions as in residential location.

Many influences and many constraints are introduced into this process. The most

important constraints are that at any point in time the supply of floorspace in each zone is fixed, and controlled in its possible uses by zoning laws of various kinds. A zoning law permits one or more activities to use the space referred to; an activity may be able to use the space under more than one zoning law. The model has to ensure that these constraints are respected in its results.

The model proceeds in this way to generate and locate this increments of each activity demanded by each activity in each zone. Having started with the generated activities demanded by basic activity, it will go on to further rounds of activities demanded by generated activity, and so on; the increment added will get smaller and smaller until the correct totals of every activity have been included and located. The process used to reach this solution is such that the following conditions will be satisfied:

- the expenditure of a household (or the expenditure per employee by a firm) will equal the amount budgeted, if it is satisfied;

- the rent paid per unit of space affected by each zoning law in each zone will be the same for all activities occupying it;

- all constraints will be satisfied i.e. the land or floorspace used will be less than or equal to the land or floorspace available; the amount of activity or density of activity within the area affected by a zoning law will be less than or equal to the appropriate maximum, minimum standards of space per household or per employee will be observed.

The Mathematics of the Land-Use Model

Basic Activity Location   The increments of basic activity are located by an incremental model of the form:

$$\Delta X_i{}^n = \Delta X_*{}^n . W_i{}^n . A^n \tag{2.1}$$

where

$\Delta X_i{}^n$ = increment of activity n allocated to zone i,

$\Delta X_*{}^n$ = total increment of activity n to be allocated,

$A^n = (\sum_i W_i{}^n)^{-1}$: balancing term,

$W_i{}^n$ = Attraction of zone i for activity n.

This W_i^n factor may be built up differently for each activity n, but will in general be a multiplicative function of a contemporary supply measure (such as the unused land available to activity i at this date) and a measure of the previous choices of zone i for location of activity n.

Floorspace Location   The model for locating additional floorspace is similar in form to the basic activity location model. The typical form involves a product of:

- the amount of floorspace permissible on relevant zones land in i at maximum allowable plot ratios,

- the profitability of floorspace in i, measured in terms of the difference between gross rent and supply cost, in the form

$$\Delta F_i = \Delta F_* \frac{(r_i^f)^{\sigma^r}[F_i^{max} - F_i^{previous}]^{\sigma^p}(F_i^{previous})^{\sigma^f}}{\sum_i (r_i^f)^{\sigma^r}[F_i^{max} - F_i^{previous}]^{\sigma^p}(F_i^{previous})^{\sigma^f}} \qquad (2.2)$$

where

r_i^f = economic rent or scarcity value of floorspace in zone i,

F_i^{max} = maximum permitted floorspace in i at the end of the period,

$F_i^{previous}$ = actual floorspace in i at the beginning of the period,

$\sigma^r, \sigma^p, \sigma^f$ = parameter for the relative importance of these effects,

ΔF_* = total floorspace to be located.

Main Activity Allocation-Equilibrium Model   One of the assumptions about the model's operation is that ratios between different activities are fixed and are not elastic with respect to prices. Accordingly the model can start from the employment in sector n and the consequently required labour supply in the form of households m

$$Y_j^m = \sum_n a^{mn}(Z_j^n + X_j^n) \qquad (2.3)$$

where

Y_j^m = households type m whose labour is demanded at work place j,

a^{mn} = technical coefficient,

Z_j^n = basic employment in sector n at j,

X_j^n = generated employment in sector n at j from previous iterations of the model initially zero.

Another assumption of the specifically urban model is that salaries at the workplace are known and fixed, although they may vary between workplaces. The model of residential location is based on how people spend these salaries or budgets to obtain a range of goods and services. There are three components to the choice of residential location: the utility of zone i (people tend to maximise this utility) for someone working at j, which is affected by the budget constraint, the cost of travel, and by the prices at i; the non-monetary disbenefit of travel between i and j; and the externally or unpriced advantages of i.

A logit model of residential choice is (omitting m)

$$\text{prob}(i|j) = \frac{e^{\lambda(uU_{ij}-d_{ij}+w_i)}}{\sum_i e^{\lambda(uU_{ij}-d_{ij}+w_i)}} \tag{2.4}$$

where

$\text{prob}(i|j) = $ probability of choosing residential location i given workplace j,

$\lambda = $ spread parameter,

$u = $ a parameter controlling the influence of locational utility,

$d_{ij} = $ non-monetary travel disbenefit (i.e travel disutility minus travel cost),

$w_i = $ zonal attractor representing advantages of locating in i that are not directly priced or charged for and are therefore outside the budget constraint; it also reflects the size of i,

$U_{ij} = $ utility of locating at i given workplace j.

The function (2.4) is applied to the total demand of labour at zone j in order to find the flow of labour from i to j:

$$F_{ij} = Y_j \frac{e^{\lambda(uU_{ij}-d_{ij}+w_i)}}{\sum_i e^{\lambda(uU_{ij}-d_{ij}+w_i)}} \tag{2.5}$$

where F_{ij}^m is the flow of labour (in households units) type m from residential zone i to workplace j.

The utility function used is of the form:

$$U_{ij}^m = \prod_l (q_{ij}^{lm} - b^{lm})^{\alpha^{lm}} \tag{2.6}$$

where:

$q_{ij}^{lm} = $ quantity of factor l consumed by a household type m with employment at j living at i,

$b^{lm} = $ minimum consumption of l by m,

44

α^{lm} = parameter describing the importance of l to the consumer m,

$q_{ij}^{lm} \geq b^{lm}$, for all factors l.

The consumers' aim is to choose quantities of consumption q_{ij}^{lm} so as to maximise U_{ij}^m subject to income constraint I. We will now show that U_{ij}^m is maximised by choosing the quantities

$$q_{ij}^{lm} = b^{lm} + \frac{(I - \sum_l b^{lm} C_i^l)\alpha^{lm}}{C_i^l \sum_l \alpha^{lm}} \tag{2.7}$$

where C_i^l is the cost per unit of l at i.

First of all, let us bear in mind that the sub- and the superscripts in equations (2.6) and (2.7) are just labels. They can thus be ignored without loss of generality. However, it is important to keep the index l which turns out to be relevant for the sake of the proof. The statement to be proven can therefore be stated as follows:

The locational utility $U = \prod_i (q_i - b_i)^{\alpha_i}$ is maximised by choosing q_i such that

$$q_i = b_i + \frac{(I - \sum_i b_i C_i)\alpha_i}{C_i \sum_i \alpha_i}$$

subject to the income constraint $I = \sum_i q_i C_i$.

U is not linear in q_i. We can therefore use the method of Lagrange multipliers. The Lagrangian can be written:

$$L = (q_1 - b_1)^{\alpha_1}(q_2 - b_2)^{\alpha_2}(q_3 - b_3)^{\alpha_3} \cdots (q_l - b_l)^{\alpha_l} + \lambda(I - \sum_i q_i C_i)$$

where λ is a Lagrange multiplier. Differentiating L with respect to q_i and λ and equating the derivatives to zero yield:

$$\frac{\partial L}{\partial q_i} = \alpha_i (q_1 - b_1)^{\alpha_1}(q_2 - b_2)^{\alpha_2}(q_3 - b_3)^{\alpha_3} \cdots (q_i - b_i)^{\alpha_i - 1} \cdots (q_l - b_l)^{\alpha_l} - \lambda C_i = 0; \tag{2.8}$$

$$\frac{\partial L}{\partial \lambda} = I - \sum_i q_i C_i = 0. \tag{2.9}$$

Multiplying (2.8) by $q_i - b_i$ leads to:

$$\alpha_i (q_1 - b_1)^{\alpha_1}(q_2 - b_2)^{\alpha_2}(q_3 - b_3)^{\alpha_3} \cdots (q_i - b_i)^{\alpha_i} \cdots (q_l - b_l)^{\alpha_l} = \lambda(q_i - b_i)C_i. \tag{2.10}$$

For any other variable q_j (or more precisely $q_j - b_j$):

$$\alpha_j (q_1 - b_1)^{\alpha_1}(q_2 - b_2)^{\alpha_2}(q_3 - b_3)^{\alpha_3} \cdots (q_j - b_j)^{\alpha_j} \cdots (q_l - b_l)^{\alpha_l} = \lambda(q_j - b_j)C_j. \tag{2.11}$$

Dividing (2.10) by (2.11), we get:

$$\frac{\alpha_i}{\alpha_j} = \frac{(q_i - b_i)C_i}{(q_j - b_j)C_j} \Rightarrow q_j - b_j = \frac{\alpha_j}{\alpha_i}\frac{C_i}{C_j}(q_i - b_i). \tag{2.12}$$

But we can transform (2.9) in the following way:

$$
\begin{aligned}
I &= \sum_{i=1}^{l} b_i C_i + \sum_{i=1}^{l}(q_i - b_i)C_i \\
&= \sum_{i=1}^{l} b_i C_i + (q_i - b_i)C_i + \sum_{j=1, j \neq i}^{l}(q_j - b_j)C_j.
\end{aligned}
$$

Substituting (2.12) in this expression yield:

$$
\begin{aligned}
I &= \sum_{i=1}^{l} b_i C_i + (q_i - b_i)C_i + \sum_{j \neq i}\frac{\alpha_j}{\alpha_i}C_i(q_i - b_i). \\
I - \sum_{i=1}^{l} b_i C_i &= (q_i - b_i)C_i + \frac{(q_i - b_i)C_i}{\alpha_i}\sum_{j \neq i}\alpha_j \\
&= (q_i - b_i)C_i \frac{\alpha_i + \sum_{j \neq i}\alpha_j}{\alpha_i}.
\end{aligned}
$$

Hence:

$$q_i = b_i + \frac{(I - \sum_i b_i C_i)\alpha_i}{C_i \sum_i \alpha_i}.$$

Equation (2.7) says that each household has an optional budget given by the term $I - \sum_l b^{lm}C_i^l$ that is their income I less the total expenditure on minimum consumption of all the factors l. Of this discretionary budget each household will spend the proportion $\alpha^{lm}/\sum_l \alpha^{lm}$ (MEPLAN assumes that $\sum_l \alpha^{lm} = 1$) on additional consumption of factor l. This sum of money is then divided by the cost of factor l, C_i^l, to get the physical quantity of l purchased out of the discretionary budget. This is added to the minimum consumption b^{lm} to get the total quantity q_{ij}^{lm}.

I is the base cost of locating at i:

$$I = C_j^m - t_{ij}^m - g_i^m$$

where

C_j^m = the income of m, i.e. the cost of a unit of m (the labour produced by the household) at the workplace j,

t_{ij}^m = the cost of travel between home i and work j for all the members of the household,

g_i^m = any direct tax on factor m locating at i.

The total households type m located at i is therefore given by:

$$X_i^m = \sum_j F_{ij}^m. \tag{2.13}$$

To ensure that the land-use model produces results consistent with the metropolitan model, the demand for service employment type n resulting from the location of residents at i is calculated as

$$Y_i^n = \sum_m a^{mn} X_i^m. \tag{2.14}$$

This demand will be met from zones j according to a different model, based on minimising costs rather than maximising utility. This is:

$$F_{ji}^n = Y_i^n \frac{L_j^n e^{\lambda^n S_{ji}^n}}{\sum_j L_j^n e^{\lambda^n S_{ji}^n}} \tag{2.15}$$

where

L_j^n : space or capacity for n at j,

λ^n : distribution parameter,

$S_{ji}^n = p_i^n - (C_j^n + \tilde{C}_j^n + r_j^n)$: a 'surplus' associated with obtaining n from j, where in turn

C_j^n : cost of producing a unit of n at j,

$\tilde{C}_j^n$: cost of transport per unit of n,

p_i^n : price that must be paid at i to obtain Y_i^n,

r_j^n : economic rent of securing a unit of n from j.

Equation (2.15) will generate flows from services to residents, F_{ji}^n. The flows from each service zone will be summed to find the generated employment there:

$$X_j^n = \sum_i F_{ji}^n. \tag{2.16}$$

Equation (2.16) is input to equation (2.3) for the next iteration of the model.

2.1.3 The Land-Use/Transport Interface

The Interface Program FREDA

The program FREDA is one of the modules integrated in MEPLAN package. It is used in two modes:

(a) From land-use model to transport model, FREDA is used to generate flows/trips represented in the transport model from the interzonal trades represented in the land-use model.

(b) From transport mode to land-use model, FREDA is used to calculate disutilities per unit of trade from disutilities per unit of flow. Through this mode, changes in transport and accessibilities are made to affect land-uses.

FREDA uses the same data for both modes **(a)** and **(b)**. FREDA has to be run with the same inputs both before and after each run of the transport model. On each occasion it will produce the following output:

(i) a file of flows which the transport model will read, split between modes and assign to routes,

(ii) a file of disutilities and costs to be used by the land-use model when distributing trades in the next time period,

(iii) a file of information which is used by the evaluation process to extract information about the travel generated by each trade.

Flow Generation

The Cape Town land-use model was run on time periods of a month (i.e. floorspace rents are rents per month), and the transport model on time periods of a 12-hour day. By adjusting the proportions of flow explicitly, peak period or peak-hour could be modelled instead of a 12-hour day.

The generation of flows from trades is the main aspect of the process of the land-use/transport interface. The inclusion of exogenous trades, the conversion from months to days and from land-use model zones to transport model zones are all necessary preliminaries to the calculation of flows.

'Trip generation' within the conventional four-stage transport models and 'flow generation' within MEPLAN are critically different. Within the conventional four-stage transport models, trip ends (origin and destination) are 'generated' separately then, they are linked together by a trip distribution model. In MEPLAN, the equivalent for the above mentioned processes is the generation of flows from trades. This is summarised as follows (see [16, p.70]):

> '*The flow volume (the number of one-directional movements of flow units from i to j) is determined by the volume of trade from i to j and/or from j to i (given that the physical flow may be a movement of the producer or the consumer of*

the trade, and may or may not include return movements) multiplied by the
flow generation rate which may be constant or may be a declining function
of the disutility of flow from i to j (implying that as the disutility of travel
increase there will be less flow for the same volume of trade)'.

The model of travel demand is based on a set of equations predicting how many one-way trips a household of each type will make per day for each purpose.

2.1.4 Transport Model

The transport component of MEPLAN is concerned with the modal split and the network assignment since the generation and destination of flow have been carried out in the land-use model.

Modal Split

The modal split model is based on the theory of utility. The starting point of the model is the hypothesis that each individual within the process studied will choose that option out of the set available to him which gives him the greatest utility, or the least disutility.

It is generally assumed that travellers prefer to make journeys that cause them least loss of money or time, or least discomfort. Therefore, it is easier to talk in terms of disutility than of utility.

Disutility being a concept which cannot be measured empirically, components of a disutility function have to be identified and their relative importance must be estimated. These components typically include time and cost of the journey.

The first step of the modal split model is to calculate the disutility of mode m for one unit of flow f from zone i to zone j. For clarity in the following formula which gives the disutility function, the subscripts i and j are omitted as the modal split is carried out separately for each pair (i, j):

$$Y^{mf} = d^{mf}D^{mf} + c^{mf}C^{mf} + t^{mf}T^{mf} + k^{mf} \qquad (2.17)$$

where

Y^{mf} : disutility of the journey made by mode m for one unit of flow f (typically one traveller),

D^m : distance of minimum generalised cost path by mode m,

C^m : cost of minimum generalised cost path by mode m,

T^m : time of minimum generalised cost path by mode m,

d^{mf}, c^{mf}, t^{mf} : parameters on distance, cost and time, specific to mode m and flow f,

k^{mf} : modal constant for mode m and flow f (or discomfort factor).

These disutilities are applied in logit models working upwards through the hierarchy. In the present case (see Figure 2.1) the model for choice between the public transport modes at the lower level will be:

$$\text{prob}^f(m|n) = \frac{e^{-\lambda_L^f Y^m f}}{\sum_m e^{\lambda_L^f Y^m f}} \tag{2.18}$$

where

the summation is over all the modes m at the lower level of the hierarchy; $\text{prob}^f(m|n)$ is the probability that a user will choose mode m out of the modes in the lower level of the hierarchy having at the upper level chosen the 'super-mode' n that leads to this lower level; and λ_L^f parameter appropriate to this flow at the lower level of the hierarchy.

The disutility for the supermode n, for example the disutility for public transport to use in the upper level choice, is calculated using the logsum type of averaging:

$$Y^{nf} = -\frac{1}{\lambda_L^f} \ln\left(\sum_m e^{-\lambda_L^f Y^m f}\right) \tag{2.19}$$

in accordance with the random utility theory. This is then used in the logit model equivalent to (2.18) for the upper level.

Assignment

The assignment model allocates flows to set of 'reasonable' paths, the choice between paths being based upon a function of distance, cost and time known as 'generalised cost'. A 'reasonable path' is one in which every successive node through which the traveller passes is nearer, in terms of generalised cost, to the destination.

MEPLAN carries out a separate assignment for each mode and for each flow group. A flow group is defined as a group of flows that have the same assignment characteristics.

The operation of the assignment process occurs in two phases. The first builds up information about paths by each mode and for each group, from every origin to each destination in turn. This information is needed to carry out the modal split calculations for each flow within that flow group, and must therefore be done before the modal split is predicted. The second part of the assignment process comes after the modal split,

taking the predicted flows by mode for all the flows in the flow group, from every origin to each destination in turn, and predicting which paths they will take through the network. The flow assigned to each link are accumulated over all the destinations, modes and flow groups so as eventually to find the total load on each link.

One run of the path-building module therefore occurs for each combination of flow group, mode and destination. This module works backwards from the destination node so as to order all the accessible modes in the network in order of ascending generalised cost to the destination. One such a ranking exists with the destination at the top and the 'distant' node at the bottom, a 'reasonable' path - on which every node visited is progressively nearer to the destination - is easily identified as one in which every node visited is successively higher in the ranking. Where more than one 'reasonable' path is available from a node to the destination, the average generalised cost is calculated using the logsum type of averaging.

The route of the minimum path is identified by recording for every relevant node the identity of what is conventionally called the 'backnode', i.e. the next node along the minimum path towards the destination. By noting the backnode of node N, and then the backnode of that backnode, and so on, one can find the whole sequence of nodes making up the minimum path from node N to the current destination. Information about the distance, cost and time of the minimum paths from each origin zone to the current destination zone forms the basis of the modal split, which will calculate new functions of the distance, cost and time for each mode and flow.

Having done the full path calculation process for one flow group, one destination and every mode, the modal split is calculated for each flow. The predicted flow by each mode is converted from flow units (e.g. passengers) to modal units (e.g. cars). Then the assignment proper is done for each mode in turn, still for the same one destination and one flow group as processed by the path calculation module.

First for all the flows in the flow group, the modal load to be assigned is attributed to the origin node in the list of nodes ranked by distance for the destination.

The assignment module starts at the bottom of the ranked list of nodes, that is, at the node more remote from the destination. If this is not an origin node, it works up the list until it finds the most remote origin not from which there is a flow to assign. It will then refer to the sorted network description so as to find all the links leaving this current node and the next nodes they connect to. These nodes will then be checked against the

ranked list of nodes. Those further from the destination will be eliminated. This should leave one or more nodes closer to the destination to which the flow can be assigned.

If no link closer to the destination is available, there is an error in the coding of the network. If there is only one linked node in the direction of the destination, the assignment is trivial: the volume of flow is added to the recorded load on the link, and is added to the volume to be assigned from the next node. If there is more than one link which offer travel towards the destination, the module will divide the volume of flow between them. To do this it will treat each available link from the current node as an alternative route. It will work out the generalised cost of the route by adding the generalised cost of the link to the previously recorded average generalised cost from the link so reached through to the destination. These alternative route costs from the current node to the destination are used together with the parameters and the capacities of the links to predict how much of the flows at the current node will take each route. The quantities calculated will then be added to the loads on the links from the current node, and added to the volumes to be assigned from the next nodes thus reached.

The way the path calculation and assignment modules operate can be summarised as follows:

- they work for one destination at a time

- the path calculation module ranks the nodes in order from the destination by calculating the generalised cost to the destination from each node.

- each link can be used only in one direction by traffic with destination; any sequence of links so used is a 'reasonable' path always getting nearer to destination.

- the assignment module works through the network towards destination j splitting the flow between the choices available at each node. The proportion depend on the length of the route which is, from a given node, the generalised cost of the link from this node to its backnode plus the composite generalised cost from this backnode to the destination. The assignment from a node may also be affected by the capacities of links adjacent to it.

It is clear that route choice is negatively related to the generalised cost of each route and positively related to the capacity of each of the 'reasonable' links.

The composite generalised cost from a node h to destination j is given by:

$$g_{hj} = -\frac{1}{\lambda} \ln \left(\sum_k e^{-\lambda(g_{hk}+g_{kj})} \right) \qquad (2.20)$$

where

g_{hj}, g_{kj} : composite generalised costs from h to j and from k to j

g_{hk} : generalised cost of link (h, k).

$g_{hk} > 0$ and k is every node such that $g_{kj} < g_{hk} + g_{hj}$ (i.e. the set of reasonable paths requiring choice at h

λ : assignment parameter.

The choice between alternative k is calculated in the following way:

$$\text{prob}(k|(h,j)) = \frac{e^{-\lambda(g_{hk}+g_{kj})}.b_{hk}^{\alpha}}{\sum_k e^{-\lambda(g_{hk}+g_{kj})}.b_{hk}^{\alpha}} \qquad (2.21)$$

where

$\text{prob}(k|(h,j))$: probability of the flow being assigned to k given that it is at h with destination j.

λ : assignment parameter (the same as in (2.20))

b_{hk} : capacity of link (h, k)

α : parameter for influence of capacity on assignment.

2.2 EMME/2

2.2.1 Trip Generation Models

In most cities, demand for trips usually grows with population and the rise of living standards. It is then of vital importance to quantify this global mobility, since it will drive the needed global capacity of the network, hence the necessary investments to be made.

Trip generation models rely on both census studies and transportation surveys. In such surveys, users are asked a series of questions regarding their income, their home and workplace, the number and type of trips they make per day, their mode of transportation, their alternatives if any. This allows the calculation of global mobility, which is usually expressed by a number of trips made per household and per day for all purposes (trips to work, shopping, etc).

Global mobility is usually divided into two elements:

- A nearly fixed component, denoted 'alternate migrations', close to two, which is the number of trips made per day to and from workplace.

- A very variable component, denoted 'other trips', which is related to the city structure, behaviour of individuals, and of course living standards (income, availability of motor vehicles, spare time) that drive shopping, leisure, or even school trips.

Once all these elements are made available, one can divide the area under study into zones, and identify the number of trips of different kinds made from one zone to another. This leads to build a base origin-destination matrix.

The prediction of future demand can be achieved in several ways:

- either actual mobility is adjusted by a factor corresponding to population growth on one hand, and by a factor related to predicted evolution of living standards (this is the actual situation of the Cape Town Metropolitan Area).

- or a comparison is made with other cities where mobility is higher to determine an average mobility on the time horizon considered.

With a base O-D matrix, predicted trip generation and attractions per zone, one can now obtain the predicted O-D matrix. This is done through trip distribution models.

2.2.2 Trip Distribution Models and Matrix Balancing

Two-Dimensional Balancing

Trip distribution models, that use two-dimensional matrix balancing, take as inputs a matrix to be balanced c_{pq}, an origin matrix O-D (production of trips at origins) and a destination matrix D_q (attractions of trips at destinations), in order to compute an O-D matrix g_{pq} (the balanced matrix) by finding origin balancing coefficients α_p and destination balancing coefficients β_q which satisfy:

$$g_{pq} = \alpha_p \beta_q c_{pq} \text{ for each O-D pair } (p, q) \tag{2.22}$$

$$\sum_q g_{pq} = O_p \text{ for each origin } p \tag{2.23}$$

$$\sum_p g_{pq} = D_q \text{ for each destination } q \tag{2.24}$$

$$g_{pq} \geq 0 \text{ for each O-D pair } (p, q). \tag{2.25}$$

It is assumed that $\sum_p O_p = \sum_q D_q$. The process by which a solution of the form (2.22) may be obtained that satisfies (2.23) and (2.24) is an iterative one, and is called the *Furness method* or the *balancing method*. This method is identical to the *biproportional process* of Bacharach [6]. The solution algorithm as it is implemented in EMME/2 is as follows:

0. *Initialization*

$$
\begin{aligned}
l &= 0 \text{ (iteration counter)} \\
\alpha_p^0 &= 1 \text{ for each origin } p \\
\beta_q^0 &= 1 \text{ for each destination } q
\end{aligned}
$$

1. *Balancing rows*

$$\alpha_p^{l+1} = \frac{O_p}{\sum_q \beta_q^l c_{pq}} \text{ for each } p$$

2. *Balancing columns*

$$\beta_q^{l+1} = \frac{D_q}{\sum_p \alpha_q^{l+1} c_{pq}} \text{ for each } q$$

3. *Stopping test*

If $\max\left(\max_p \frac{\alpha_p^{l+1}-\alpha_p^l}{\alpha_p^{l+1}}, \max_q \frac{\beta_q^{l+1}-\beta_q^l}{\beta_q^{l+1}} \right) \le \varepsilon$ or if $l+1 = lmax$ then **stop**

Otherwise $l = l + 1$ and return to **step 1**.

The balanced matrix is then given by

$$g_{pq} = \alpha_p^{l+1}\beta_q^{l+1}c_{pq}. \tag{2.26}$$

It is worth noting that the production and attraction vectors correspond to predicted production and attractions; the resulting balanced matrix (2.26) is the predicted O-D matrix.

Three-Dimensional Balancing

Three-dimensional trip distribution models, which use an additional stratification of trips other than by origins and destinations, use the three-dimensional balancing procedure of Evans and Kirby [6]

In order to illustrate the 'third dimension', consider a screen line that divides an urban area into two separate parts, say A and B. The traffic counts obtained from the screen

line give the number of trips from A to B and from B to A. If the predicted matrix g_{pq} is to be compatible with this information, one could assign the O-D pairs (p, q) into four classes:

- Class 1: O-D pairs with trips from A to B

- Class 2: O-D pairs with trips from B to A

- Class 3: O-D pairs with trips from A to A

- Class 4: O-D pairs with trips from B to B

The total number of trips associated with Class 1 is the screen line count of trips from A to B and with class 2, the screen line count of trips from B to A. The totals associated with classes 3 and 4 are the appropriate number of trips from A to A and from B to B. The aim is to obtain the matrix g_{pq} that satisfies these additional conditions, and also respects the usual production and attraction totals.

As another example of the 'third dimension', consider the following subdivision of the O-D pairs (p, q) into k classes which are based on the impedance (=travel time) u_{pq} of making the trip from p to q. O-D pair (p, q) belongs to class k if the travel impedance u_{pq} is such that $\underline{u}_k \leq u_{pq} < \overline{u}_k$ where $\underline{u}_k$ and $\overline{u}_k$ are the lower and upper bounds of impedance interval k. A matrix with elements k_{pq} is used to identify that O-D pair (p, q) belongs to interval k.

Trip distribution models that use three-dimensional balancing take as inputs a matrix c_{pq}, an origin matrix O_p (the trip productions), a destination matrix D_q (the trip attractions), the third dimension totals F_k, for each interval k, and the third dimension matrix k_{pq}. They compute an origin-destination matrix g_{pq} (the balanced matrix), by finding origin balancing coefficients α_p, destination balancing coefficients β_q and third-dimension balancing coefficients $\gamma_{k_{pq}}$ which satisfy:

$$g_{pq} = \alpha_p \beta_q \gamma_{k_{pq}} c_{pq} \text{ for each O-D pair } (p, q) \tag{2.27}$$

$$\sum_q g_{pq} = O_p \text{ for each origin } p \tag{2.28}$$

$$\sum_p g_{pq} = D_q \text{ for each destination } q \tag{2.29}$$

$$\sum_{(p,q)|k_{pq}=k} g_{pq} = F_k \text{ for each interval } k \tag{2.30}$$

$$g_{pq} \geq 0 \text{ for each O-D pair } (p, q).$$

Note that, it is assumed that $\sum_p O_p = \sum_q D_q = \sum_k F_k$. The solution g_{pq} here is said to be triproportional to c_{pq}. It seems natural that a triproportional procedure, similar to the biproportional one, is used to obtain the solution g_{pq}. This process is regarded as being an extension to the three dimensions of the Furness method of iterations:

0. *Initialization*

$$
\begin{aligned}
l &= 0 \text{ (iteration count)} \\
\alpha_p^0 &= 1 \text{ for each origin } p \\
\beta_q^0 &= 1 \text{ for each destination } q \\
\gamma_{k_{pq}}^0 &= 1 \text{ for each class } k.
\end{aligned}
$$

1. *Balancing rows*

$$\alpha_p^{l+1} = \frac{O_p}{\sum_q \beta_q^l \gamma_{k_{pq}}^l c_{pq}} \text{ for each } p.$$

2. *Balancing columns*

$$\beta_q^{l+1} = \frac{D_q}{\sum_p \alpha_q^{l+1} \gamma_{k_{pq}}^l c_{pq}} \text{ for each } q.$$

3. *Balancing third dimension totals*

$$\gamma_{k_{pq}}^{l+1} = \frac{F_k}{\sum_{(p,q)|k_{pq}=k} \alpha_q^{l+1} \beta_q^{l+1} c_{pq}} \text{ for each interval } k.$$

4. *Stopping test*
   If $\max\left(\max_p \frac{\alpha_p^{l+1} - \alpha_p^l}{\alpha_p^{l+1}}, \max_q \frac{\beta_q^{l+1} - \beta_q^l}{\beta_q^{l+1}}, \max_k \frac{\gamma_k^{l+1} - \gamma_k^l}{\gamma_k^{l+1}} \right) \leq \varepsilon$ or if $l + 1 = lmax$ then **stop**.
   Otherwise $l = l + 1$ and return to **step 1**.

When the algorithm terminates, the balanced matrix is given by

$$g_{pq} = \alpha_p^{l+1} \beta_q^{l+1} \gamma_{k_{pq}}^{l+1} c_{pq}.$$

2.2.3   Route Assignment Models

The following assignments are implemented within the EMME/2 system ([9, p12 of Chap1]):

- equilibrium assignment on the auto network with one or more classes of users, with fixed demand,

- equilibrium assignment on the auto network with one or more classes of users, with variable demand for one class,

- multipath transit assignment with fixed demand,

- disaggregate transit assignment for individual trips,

- timetable-based transit assignment.

The fixed and variable demand auto assignment implemented in EMME/2 are based on Wardrop's user optimal principle (see below) and hence yield flows such that all paths used are of equal time (or impedance).

The transit assignment of aggregate or individual trips are based on the concept of 'strategy' (see below) which is as generalisation of the concept of a path. It is assumed that the transit rider wants to minimise his expected travel time (including waiting, in-vehicle, walking time).

An important feature of the EMME/2 assignment module, is that the auto assignment may use data related to the transit network and the transit may use data that results from the auto assignment. For instance, the congestion effect due to buses can be included in the auto volume-delay functions. Conversely, transit time functions may depend on the auto times resulting from an auto assignment.

Below, we point out the fixed and variable auto assignment models as well as the standard transit assignment model as they are given in [9].

Auto Assignment

General Principle of the Equilibrium Auto Assignment Auto assignment models are based on the assumption that each user will choose the path which he perceives as the best; if a shorter route than the one he is currently using exists, then he will select it. This will produce flows satisfying *Wardrop's user optimal principle*, that is, at equilibrium, no user can improve his travel time by changing routes. Thus, all used paths between origin and destination are of equal time.

The solution to the equilibrium traffic assignment problem is equivalent to solve the problem illustrated in Figure 2.3. The problem is to assign to each arc a flow v_1 and v_2, with 'costs' $s_1(v_1)$ and $s_2(v_2)$, in such a way to minimise total cost and that cost is the same on each arc, that is : $s_1(v_1) = s_2(v_2)$.
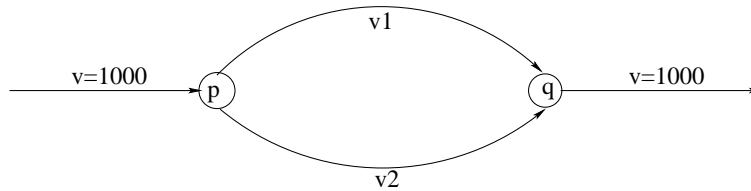
Figure 2.3: Simple equilibrium assignment

The solution to the equilibrium traffic assignment problem is equivalent to solving a problem where the area under the volume-delay curves is minimised.

Several methods (including the linear approximation method, incremental assignment, capacity restraint and the successive average method) can be used to perform an equilibrium auto assignment. However, they are not equally efficient.

Before we examine the results produced for a simple example by the linear approximation method, we present the four above mentioned methods.

a. The Linear Approximation Method   The linear approximation method (Frank and Wolfe, 1956 cited in [7]) has the advantage that, at each iteration, the total area under the volume-delay curves decreases and a measure of the difference between the current flows and the equilibrium flow can easily be estimated. It has the following general steps:
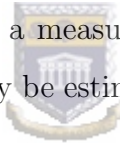
0. *Initialisation*

   Initial solution v^0 is obtained by an all-or-nothing assignment of demand g on shortest paths computed with arc costs $s^0 = s(0)$;

   $k = 0$ (iteration count).

1. *Update link costs*

$$k = k + 1$$
$$s^k = s(v^{k-1}).$$

2. *Descent direction*

   y^k is obtained by an all-or-nothing assignment of demand g on shortest paths computed with arc costs s^k.

3. *Compute optimal step size*

   λ^k is the value of λ of which the area under the volume-delay curves is minimised, for the flow $v^{k-1} + \lambda(y^k - v^{k-1})$, $0 \leq \lambda \leq 1$.

4. *Update link flows*

$$v^k = v^{k-1} + \lambda^k(y^k - v^{k-1}).$$

5. *Stopping Criterion*

If $|s^k v^{k-1} - s^k y^k| > \varepsilon$ return to step 1 (total travel time still significantly different from total travel time on shortest paths).

Otherwise $v^* = v^k$, $s^* = s(v^*)$ and stop.

b. The Incremental Method The incremental method proceeds through the following general steps:

0. *Initialisation*

Define number of increments N;

$v^0 = 0$

$k = 0$ (iteration count).

1. *Update link costs*

$$\begin{aligned} k &= k + 1 \\ s^k &= s(v^{k-1}). \end{aligned}$$

2. *Load Increment of Demand*

y^k is obtained by an all-or-nothing assignment of demand g/N on shortest paths computed with arc costs s^k.

3. *Update Link Flow*

$$v^k = v^{k-1} + y^k.$$

4. *Stopping Criterion*

If $k < N$ return to step 1.

Otherwise $v^* = v^k$, $s^* = s(v^k)$ and stop.

c. The Capacity Restraint Method The capacity restraint method is probably one of the first heuristic methods used for the emulation of equilibrium flows. It proceeds through the following general steps:

0. *Initialisation*

Define number of iterations N; initial solution y^0 is obtained by an all-or-nothing assignment of demand g on shortest paths computed with arc costs $s^0 = s(0)$; $k = 0$ (iteration count).

1. *Update link costs*

$$
\begin{aligned}
k &= k + 1 \\
s^k &= 0.75 s^{k-1} + 0.25 s(y^{k-1}).
\end{aligned}
$$

2. *Load Demand*

y^k is obtained by an all-or-nothing assignment on shortest paths computed with arc costs s^k.

3. *Stopping Criterion*

If $k < N$ return to step 1;

Otherwise $v^* = \frac{1}{4} \sum_{k=0}^{3} y^{N-k}$, $s^* = s(v^*)$ and stop.

d. The Successive Average Method This method is known to be a convergent method but its convergence is very slow and there is no reasonable stopping criterion, other than an arbitrary number of iterations. The method resembles the linear approximation method, except that the step size, λ, is arbitrarily fixed to yield a solution in which each of the all-or-nothing flows y^k have the same weight. The general steps of the method are:

0. *Initialisation*

Define number of iterations N; Initial solution v^0 is obtained by an all-or-nothing assignment of demand g on shortest paths computed with arc costs $s^0 = s(0)$; $k = 0$ (iteration count).

1. *Update link costs*

$$
\begin{aligned}
k &= k + 1 \\
s^k &= s(v^{k-1}).
\end{aligned}
$$

2. *All-or-nothing Assignment*

y^k is obtained by loading demand g on shortest paths computed with arc costs s^k.

3. *Compute Step Size*

$$\lambda^k = \frac{1}{k+1}.$$

4. *Update Link Flows*

    If $k < N$ return to step 1.

    Otherwise $v^* = v^k$, $s^* = s(v^k)$ and stop.

Example: Total demand is 1000 trips from p to q. The problem is to assign trips to links in order to minimise total travel time and that the travel time on each link is the same (see Figure 2.4).
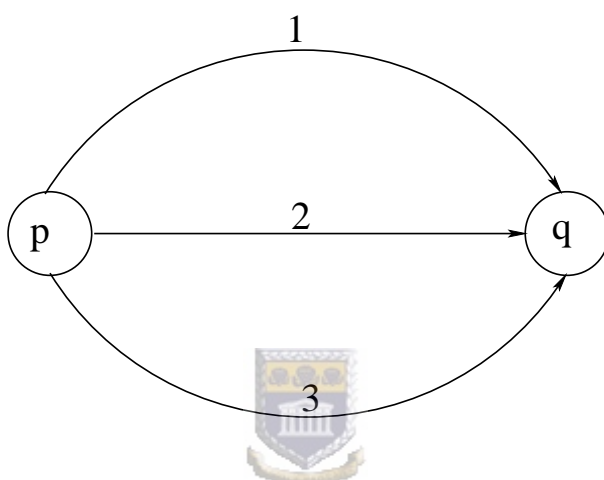

Figure 2.4: An example of an equilibrium assignment

The travel time (volume-delay) functions for each link are given by:

$$s_1(v_1) = 10\left[1 + 0.15\left(\frac{v_1}{200}\right)^4\right],$$
$$s_2(v_2) = 20\left[1 + 0.15\left(\frac{v_2}{400}\right)^4\right],$$
$$s_3(v_3) = 25\left[1 + 0.15\left(\frac{v_3}{300}\right)^4\right].$$

The problem can be formulated as:

$$\text{Min} \sum_i s_i(v_i)$$

subject to :

$$\sum_i v_i = 1000$$
$$s_i(v_i) = s_j(v_j) \text{ for each pair } (i,j)$$
$$v_i \geq 0.$$

62

The results obtained after the first nine iterations of the linear approximation method, as well as the optimal solution, where all paths used are of equal length, are given in Table 2.1. For an illustrative purpose, let us manually compute a few entries of the table. We must first calculate the travel time on each link and then get the load on those links according to the algorithm.

$$\text{For } k = 0 : s_1^0 = s_1(0) = 10, s_2^0(0) = 20 \text{ and } s_3^0(0) = 25. \tag{2.31}$$

The all-or-nothing assignment suggests that all the demand is assigned to the path with less travel time, that is to path 1 (10), meaning that $v_1^0 = 1000$, $v_2^0 = 0$ and $v_3^0 = 0$. Thus $F(v^0) = \int_0^{1000} 10\left[1 + 0.15\left(\frac{v_1}{200}\right)^4\right] dv_1 = 197500$

$$\text{For } k = 1 : \quad s_1^1(v_1^0) = 10\left[1 + 0.15\left(\frac{1000}{200}\right)^4\right] = 947.5$$

$$s_2^1(v_2^0) = 20\left[1 + 0.15\left(\frac{0}{400}\right)^4\right] = 20$$

$$s_3^1(v_3^0) = 25\left[1 + 0.15\left(\frac{0}{300}\right)^4\right] = 25.$$

Now we need to compute λ^1. From equation (2.33), we have:

$$s_1(v_1^0 + \lambda^1(y_1^1 - v_1^0))(y_1^1 - v_1^0) + s_2(v_2^0 + \lambda^1(y_2^1 - v_2^0))(y_2^1 - v_2^0)$$

$$+ s_3(v_3^0 + \lambda^1(y_3^1 - v_3^0))(y_3^1 - v_3^0) = 0$$

$$\Rightarrow s_1(1000 - 1000\lambda^1)(-1000) + s_2(1000\lambda^1).1000 + s_3(0 - 0) = 0$$

$$\Rightarrow -1000 \times 10\left[1 + 0.15\left(\frac{(1000(1-\lambda))^4}{200^4}\right)\right] + 1000 \times 20\left[1 + 0.15\left(\frac{(1000\lambda)^4}{400^4}\right)\right] = 0$$

$$\Rightarrow 11.71875\lambda^4 - 93.75(1-\lambda)^4 + 1 = 0.$$

Solving this equation using MUPAD, we get the following set of solution: 0.5965430164, 2.468798867, 0.7530433442 - 0.4479433565i, 0.7530433442 + 0.4479433565i. Other mathematical softwares for symbolic computation such as MAPLE, MATHEMATICA or XMAX-IMA could be used. λ^1 is required to be a real number between 0 and 1. Therefore, $\lambda^1 = 0.59654$. This value of λ is then used to get the volumes on the different links at iteration 1:

$$v_1^1 = v_1^0 + \lambda^1(y_1^1 - v_1^0) = 1000 + 0.59654(0 - 1000) = 403$$

$$v_2^1 = v_2^0 + \lambda^1(y_2^1 - v_2^0) = 0 + 0.59654(1000 - 0) = 597$$

$$v_3^1 = v_3^0 + \lambda^1(y_3^1 - v_3^0) = 0 + 0.59654(0 - 0) = 0.$$

$$F(v^1) = \int_0^{403} 10\left[1 + 0.15\left(\frac{v_1}{200}\right)^4\right] dv_1 + \int_0^{597} 20\left[1 + 0.15\left(\frac{v_2}{200}\right)^4\right] dv_2 = 19740.$$

Table 2.1: Results of the linear approximation on an equilibrium assignment problem, see **Example**

Iteration k	s_1^k	s_2^k	s_3^k	v_1^k	v_2^k	v_3^k	$F(v^k)$	λ^k
0	10.00	20.00	25.00	1000	0	0	197500	1.00000
1	947.50	20.00	25.00	403	597	0	19740	0.59654
2	34.73	34.88	25.00	338	500	161	18999	0.16113
3	22.30	27.35	25.31	362	483	155	18945	0.03555
4	26.09	26.36	25.27	355	473	173	18936	0.02040
5	24.82	25.86	25.41	359	469	171	18934	0.00719
6	25.61	25.69	25.40	357	467	176	18933	0.00536
7	25.28	25.57	25.44	359	466	175	18933	0.02000
8	25.50	25.52	25.44	358	465	177	18933	0.00156
9	25.40	25.49	25.45	358	465	177	18933	0.00059
Opt. sol	s_1^*	s_2^*	s_3^*	v_1^*	v_2^*	v_3^*	$F(v*)$	
	25.46	25.46	25.46	358	465	177	18933	

For $k = 2$:
$$s_1^2(v_1^1) = 10\Big[1 + 0.15\Big(\frac{403}{200}\Big)^4\Big] = 34.73$$
$$s_2^2(v_2^1) = 20\Big[1 + 0.15\Big(\frac{597}{400}\Big)^4\Big] = 34.88$$
$$s_3^2(v_3^1) = 25\Big[1 + 0.15\Big(\frac{0}{300}\Big)^4\Big] = 25.$$

From equation (2.33), $\lambda^2 = 0.16113$. The volumes on the links at iteration 2 are:

$$v_1^2 = v_1^1 + \lambda^2(y_1^2 - v_1^1) = 338$$
$$v_2^2 = v_2^1 + \lambda^2(y_2^2 - v_2^1) = 500$$
$$v_3^2 = v_3^1 + \lambda^2(y_3^2 - v_3^1) = 161.$$

The figures under the column $F(v^k)$ give the area under the volume-delay curves which is minimised where the flows are so-called 'equilibrium' flows.

The fixed demand auto assignment model.

The auto assignment model implemented in EMME/2 computes the equilibrium flows

and travel times by solving the fixed demand problem:

$$\text{Min } f(v) = \sum_{a \in A} \int_0^{v_a} s_a(v + x_a)dv + \sum_{i \in \bar{I}} \sum_{a_1 \in A_i^-} \sum_{a_2 \in A_i^+} \int_0^{v_{a_1 a_2}} p_{a_1 a_2}(v + x_{a_1 a_2})dv$$

subject to:

$$v_a = \sum_{k \in K} \delta_{ak} h_k \qquad a \in A$$

$$v_{a_1 a_2} = \sum_{k \in K} \delta_{a_1 k} \delta_{a_2 k} h_k \qquad a_1 \in A_i^-, a_2 \in A_i^+, i \in I$$

$$\sum_{k \in K_{pq}} h_k = \frac{g_{pq}}{\eta_{pq}} + \gamma_{pq} \qquad p \in P, q \in Q$$

$$h_k \geq 0 \qquad k \in K_{pq}, p \in P, q \in Q.$$

The notation used is described below:

Indices and sets:

$p \in P$: origin zones,

$q \in Q$: destination zones,

$i \in I$: modes of the auto network,

$i \in \bar{I}$: modes corresponding to intersections with turn penalties,

$a \in A$: links of the auto network,

$a \in A_i^-$: links 'ending'at node i,

$a \in A_i^+$: links 'starting' at node at i,

$k \in K_{pq}$: directed paths linking p to q,

$k \in K$: all directed paths.

Constants

δ_{ak}: 1 if link a belongs to path k,

g_{pq}: auto demand from p to q (persons),

η_{pq}: car occupancy for O-D pair (p, q) (persons/car),

γ_{pq}: additional demand (vehicles),

x_a: additional volume on link a (vehicles),

$x_{a_1 a_2}$: additional volume on turn $a_1 a_2$ (vehicles),

Functions

$s_a(v_a)$: volume-delay or cost function on link a,

$p_{a_1 a_2}(v_{a_1 a_2})$: penalty function on the turn (a_1, a_2).

(The volume-delay and penalty functions are non-decreasing functions of the auto volumes)

<u>Variables</u>

v_a: auto volume on link a,

$v_{a_1a_2}$: auto volume on turn (a_1a_2),

h_k: flow on path k.

This model is solved by the linear approximation method: starting from an initial feasible solution v, the linear approximation method obtains a feasible (descent) direction $y - v$ by linearising the objective function, solving a linear programming subproblem and then finding an improved solution on the line segment between the current solution and the solution of the subproblem.

For the fixed demand network equilibrium, the linearised subproblem is:

$$\text{Min} \sum_{a \in A} y_a s_a(v_a + x_a) + \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} y_{a_1a_2} p_{a_1a_2}(v_{a_1a_2} + x_{a_1a_2}).$$

which may be solved by assigning all the demand to the shortest paths (all-or-nothing assignment) that consider explicitly the turn penalties at penalised intersections.

The manual [9] provides a shortest path algorithm with turn penalties (from origin p to all destinations $q \in Q$). This algorithm is applied to the auto network with costs

$$
\begin{aligned}
c_a &= s_a(v_a), \quad a \in A, \\
c_{a_1a_2} &= p_{a_1a_2}(v_{a_1a_2}), \quad a_1 \in A_i^-, a_2 \in A_i^+, i \in \bar{I}.
\end{aligned}
$$

and then the demand g_{pq} is assigned to the found paths in order to obtain the volumes y.

An iteration of the linear approximation algorithm is completed by finding the solution of

$$\text{Min}_{0 \leq \lambda \leq 1} \sum_{a \in A} \int_0^{(1-\lambda)v_a^{k-1} + \lambda y_a^k} s_a(v)dv + \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} \int_0^{(1-\lambda)v_{a_1a_2}^{k-1} + \lambda y_{a_1a_2}^k} p_{a_1a_2}(v)dv \quad (2.32)$$

or equivalently annulling its derivative, that is find λ for which

$$\sum_{a \in A} s_a((1-\lambda)v_a^{k-1} + \lambda y_a^k)(y_a^k - v_a^{k-1})$$
$$+ \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} p_{a_1a_2}((1-\lambda)v_{a_1a_2}^{k-1} + \lambda y_{a_1a_2}^k)(y_{a_1a_2}^k - v_{a_1a_2}^{k-1}) = 0. \qquad (2.33)$$

For numerical reasons (stability), it is preferable to find λ by solving (2.33). The linearisation is repeated for $k = 1, 2, \ldots$ until a satisfactory solution is obtained: It corresponds

to the optimal step length, λ^*.

$$f'(\lambda) = \frac{df(\lambda)}{d\lambda} = \sum_{a \in A} s_a((1-\lambda)v_a + \lambda y_a)(y_a - v_a)$$

$$+ \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} p_{a_1 a_2}((1-\lambda)v_{a_1 a_2} + \lambda y_{a_1 a_2})(y_{a_1 a_2} - v_{a_1 a_2}) = 0. \qquad (2.34)$$

- If $f'(0) \leq \varepsilon$, then $\lambda^* = 0$ and the algorithm terminates with the solution $v^* = v$.

- If $f'(1) < 0$, then $\lambda^* = 1$; that is v is replaced by y.

- Otherwise, the optimal value of λ is the one that annuls the gradient $\frac{df(\lambda^*)}{d\lambda} = 0$, $0 \leq \lambda^* \leq 1$.

At each iteration of the linear approximation method, the solution of the subproblem provides a lower bound, LB, for the optimal value of the objective function $f(v^*)$, which is

$$LB = f(v) + \sum_{a \in A} s_a(v_a + x_a)(y_a - v_a) + \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} p_{a_1 a_2}(v_{a_1 a_2} + x_{a_1 a_2})(y_{a_1 a_2} - v_{a_1 a_2})$$

due to the fact that the objective function is convex. The function $f(v)$ is the current value of the objective function. The numerical value of the optimal solution of the subproblem, $f(v) - LB$ is referred to as the current gap, or GAP:

$$f(v) - LB = -\sum_{a \in A} s_a(v_a + x_a)(y_a - v_a) - \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} p_{a_1 a_2}(v_{a_1 a_2} + x_{a_1 a_2})(y_{a_1 a_2} - v_{a_1 a_2}).$$

The best current lower bound, BLB, is the largest value of the LB obtained up to the current iteration. The relative gap, which is a measure of the closeness of the current assignment to a perfect equilibrium assignment, is computed as

$$\text{Relative Gap} = \frac{f(v) - BLB}{f(v)} \times 100.$$

Empirically, assignments that are characterised by a relative gap of 1% or less, are considered sufficiently close to a perfect equilibrium assignment. The solution of the subproblem provides another criterion for characterising the closeness of an assignment to a perfect equilibrium assignment. If one rewrites GAP as $T - S$ where

$$T = \sum_{a \in A} s_a(v_a + x_a)v_a + \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} p_{a_1 a_2}(v_{a_1 a_2} + x_{a_1 a_2})v_{a_1 a_2},$$

$$S = \sum_{a \in A} s_a(v_a + x_a)y_a + \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} p_{a_1 a_2}(v_{a_1 a_2} + x_{a_1 a_2})y_{a_1 a_2},$$

it is easily recognised that S represents the total travel times on the current shortest paths and T is the total travel time on the currently used paths. For a perfect equilibrium assignment, $T = S$.

Clearly, when S is sufficiently close to T, one may terminate computations. In fact, it is intuitive to define the travel time difference to be the mean trip time on network less the mean minimal trip time (or mean trip times on shortest paths):

$$\frac{T - S}{\sum_{p \in P} \sum_{q \in Q} \left(\frac{g_{pq}}{\eta_{pq}} + \gamma_{pq} \right)}$$

and to select as a stopping criterion a suitable small value for this difference. As noted above, $T - S$ is the value of the current GAP. The mean trip time less the mean minimal trip time is the GAP divided by the total number of trips, which is also referred to as the normalised gap.

The variable demand auto assignment model

The variable demand auto assignment model implemented in EMME/2 computes the equilibrium auto demand, link flows and travel times by solving the problem:

$$\text{Min } f(v, g) = \sum_{a \in A} \int_0^{v_a} s_a(v + x_a) dv + \sum_{i \in \bar{I}} \sum_{a_1 \in A_i^-} \sum_{a_2 \in A_i^+} \int_0^{v_{a_1 a_2}} p_{a_1 a_2}(v + x_{a_1 a_2}) dv$$
$$- \sum_{p \in P} \sum_{q \in Q} \frac{1}{\eta_{pq}} \int_0^{g_{pq}} W_{pq}(y) dy$$

subject to:

$$v_a = \sum_{k \in K} \delta_{ak} h_k \qquad a \in A,$$

$$v_{a_1 a_2} = \sum_{k \in K} \delta_{a_1 k} \delta_{a_2 k} h_k \qquad a_1 \in A_i^-, a_2 \in A_i^+, i \in \bar{I},$$

$$\sum_{k \in K_{pq}} h_k - (g_{pq}/\eta_{pq}) - \gamma_{pq} = 0 \qquad p \in P, q \in Q,$$

$$h_k \geq 0 \qquad k \in K_{pq}, p \in P, q \in Q,$$

$$g_{pq} \geq 0 \qquad p \in P, q \in Q.$$

The additional notation used is described below.

<u>Functions</u>

G_{pq}: auto demand for O-D pair (p, q),

W_{pq}: inverse of auto demand function for O-D pair (p, q).

(The auto demand functions are non-decreasing functions of the travel time).

<u>Variables</u>

g_{pq}: auto demand for O-D pair (p, q),

w_{pq}: inverse of auto demand for O-D pair (p, q),

u_{pq}: current value of shortest path for O-D pair (p, q),

f_{pq}: value of auto demand function evaluated with u_{pq} for O-D pair (p, q).

The partial linear approximation method finds, given a current solution (v, g), a descent direction $(y - v, f - g)$, by solving a subproblem where the link and turn delay functions are linearised, but the inverse demand functions are not. The resulting subproblem is:

$$\text{Min} \sum_{a \in A} y_a s_a(v_a + x_a) + \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} y_{a_1 a_2} p_{a_1 a_2}(v_{a_1 a_2} + x_{a_1 a_2})$$
$$- \sum_{p \in P} \sum_{q \in Q} \frac{1}{\eta_{pq}} W_{pq}(f_{pq})$$

This subproblem may be solved by computing u_{pq}, the shortest paths based on the current link and transfer times, and then determining the corresponding demand $f_{pq} = G_{pq}(u_{pq})$. Then the demand f_{pq} is assigned to the shortest paths found in computing u_{pq}, in order to find the link and turn volumes $(y_a, y_{a_1 a_2})$.

The optimal step length, λ^*, for the direction of descent $(y - v, f - g)$ is the one that minimises the objective function, that is

$$\text{Min } f(\lambda) = \sum_{a \in A} \int_0^{(1-\lambda)v_a + \lambda y_a} s_a(v + x_a) dv$$
$$+ \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} \int_0^{(1-\lambda)v_{a_1 a_2} + \lambda y_{a_1 a_2}} p_{a_1 a_2}(v + x_{a_1 a_2}) dv$$
$$+ \sum_{p \in P} \sum_{q \in Q} \frac{1}{\eta_{pq}} \int_0^{(1-\lambda)g_{pq} + \lambda f_{pq}} W_{pq}(y) dy.$$

In order to ensure numerical stability and also to avoid the evaluation of integrals of the inverse demand function, which is not available in an analytical form, it is preferable to annul the gradient of the function $f'(\lambda)$:

$$f'(\lambda) = \frac{df(\lambda)}{d\lambda} = \sum_{a \in A} s_a((1-\lambda)v_a + \lambda y_a)(y_a - v_a)$$
$$+ \sum_{i \in \bar{I}} \sum_{a_i \in A_i^-} \sum_{a_i \in A_i^+} p_{a_1 a_2}((1-\lambda)v_{a_1 a_2} + \lambda y_{a_1 a_2})(y_{a_1 a_2} - v_{a_1 a_2})$$
$$+ \sum_{p \in P} \sum_{q \in Q} (\frac{1}{\eta_{pq}}) W_{pq}((1-\lambda)g_{pq} + \lambda f_{pq})(f_{pq} - g_{pq}).$$

- If $f'(0) \leq \varepsilon$, then $\lambda^* = 0$ and the algorithm terminates with the solution $v^* = v$, $g^* = g$.

- If $f'(1) < 0$, then $\lambda^* = 1$; that is v is replaced by y and g is replaced by f.

- Otherwise, the optimal value of λ is the one that annuls the gradient, $\frac{df(\lambda^*)}{d\lambda} = 0$, $0 \leq \lambda^* \leq 1$.

Chapter 3

Comparative Study of MEPLAN and EMME/2

MEPLAN and EMME/2 have different structures since the former has two components, land-use and transport and the latter deals only with transport. Therefore, an efficient comparison can only be undertaken on the transport aspect of the two models.

3.1   A Point on the Population Groups

MEPLAN used five groups of population in order to represent the South African racial situation as enumerated in Chapter 2. The segregational basis of this stratification does not justify its use in an economic model. More precisely, a stratification of the population according to their income would be more suitable for MEPLAN than the one which is based on a racial criterion. Furthermore, 'the number (of groups) seems to be larger than desirable' [2].

The three income groups (high, middle and low income) used currently in EMME/2 seem to suit MEPLAN in that 'the choice of location is based upon the hypothesis that each household locates with respect to a fixed place of work where it receives a fixed income out of which it must obtain housing, pay travel costs and other expenses. This hypothesis (the utility maximising hypothesis) specifies that households of a given socio-economic group will tend to maximise their utility that they get from spending their income '[16].

If follows that one would expect far better results if MEPLAN were implemented in a non-apartheid situation of the Cape Town Metropolitan Area with the three above-

mentioned income groups compared to EMME/2.

3.2 A Point on the Zoning System

A zoning system is used to aggregate the individual households and premises into manageable entities for modelling purposes. The main issues of a zoning system are the number of zones and their size. Obviously, the two are related: the greater the number of zones, the smaller they can be to cover the same study area. The number of zones in the study area depends on a compromise between a series of criteria discussed below. For example, the analysis of traffic management schemes will generally call for smaller zones, whereas strategic studies, on the other hand, will often be carried out on the basis of much larger zones.

Zones are represented in the computer models as if all their attributes and properties were concentrated in a single point, the zone centroid. Centroids are attached to the network through centroid connectors representing the average costs (time, distance) of joining the transport system for trips with origin or destination in that zone. In modelling, centroids and centroid connectors are important in defining zone boundaries.

Some zoning criteria drawn from modellers experience in several practical studies can be outlined:

- Zoning size must be such that the aggregation error caused by the assumption that all activities are concentrated at the centroid is not too large. It is convenient to operate a system with many small zones as this may be aggregated in various ways later depending on the nature of the projects to be evaluated.

- The zoning system must be compatible with other administrative divisions, particularly with census zones.

- Zones should be as homogeneous as possible in their land-use and population composition; census zones with clear differences in this respect (i.e residential sectors with vastly different income levels) should not be aggregated, even if they are very small.

- Zones do not have to be of equal size; if anything, they could be of similar dimensions in travel time units, therefore generating smaller zones in congested than in uncongested areas.

For the Cape Town Metropolitan Area, many factors could lead EMME/2 and ME-PLAN modellers to consider different zoning system:

- These models were utilised at different periods.

- The objectives could be identical but the administration system might interfere (MEPLAN was in use during apartheid, EMME/2 is a post-apartheid model). We have the feeling that, it was possible to get larger homogeneous zones during the apartheid than it is nowadays since people had to reside in zones according to their race.

- The number of population groups (3 in EMME/2 and 5 in MEPLAN).

- The nature of the model (MEPLAN is a logit model, EMME/2 is a synthetic model).

- The population size: It is a fact that the population size is increasing. The increase is even enhanced by immigration effects.

All the abovementioned factors justify the small number of zones in MEPLAN as compared to EMME/2. In fact, MEPLAN used 60 transport zones aggregated into 28 macro or strategic zones (land-use zones) (see [16]). A layer of 470 transport planning zones were defined in terms of the 1991 census enumerator subdistricts in Cape Town. EMME/2 aggregated these zones into 39 macro zones. Another 7 macro external zones were identified (see [2]).

The low number of zones in MEPLAN is a major problem for Traffic Engineers who need traffic volumes on specific roads and not just corridor flows.

3.3 Trip Distribution in MEPLAN and EMME/2.

3.3.1 Trip Distribution and Modal Split: What First?

Another point of discussion would be focussed on the modal split in the models. Wilson (1969) in [18] provides information in this regard and shows that modal split process can be performed before or after the completion of trip distribution. Unlike in EMME/2, the number of trips between any O-D pair (i, j) is first calculated in MEPLAN. Then, for each O-D pair (i, j) the proportion of the trips (and therefore the number of trips) using each mode k is determined. In EMME/2, the proportion of trips by each mode is first

computed, then the trips are distributed amongst different O-D pairs (i, j). This occurs in the following way. For each income group (each person type), the base year matrix (to be balanced) by each mode, σ_{ij}, is determined. These matrices are balanced to give the predicted modal distribution of trips for each income group. The balanced matrices are then aggregated over all modes and income groups to provide the distribution of trips. We have:

$$T_{ij} = A_i B_j \sigma_{ij} \tag{3.1}$$

where

- T_{ij}: predicted trips from origin i to destination j for each mode and each income group.

- σ_{ij}: 'base year' trips from origin i to destination j for each mode and each income group.

- A_i and B_j are origin and destination balancing coefficients respectively.

3.3.2   Trip Distribution: MEPLAN versus EMME/2

In order to estimate σ_{ij}, a complete or partial survey study is carried out on a population sample, then a function is evaluated over the whole population (in the case of a partial survey) in the study area. Evans and Kirby [6] suggest that σ_{ij} has the form

$$\sigma_{ij} = O_i D_j f(C_{ij}) \tag{3.2}$$

where:

- C_{ij} is the cost of travelling between i and j,

- $f(C_{ij})$ is a decreasing function of C_{ij} (sometimes called the cost function or the separation function),

- O_i is the number of trips beginning at the origin zone i,

- D_j is the number of trips ending at the destination zone j.

Therefore,

$$T_{ij} = A_i B_j O_i D_j f(C_{ij}) \tag{3.3}$$

and, according to the entropy maximising method, $f(C_{ij})$ could be a negative exponential function of C_{ij} and hence:

$$T_{ij} = A_i B_j O_i D_j e^{-\beta C_{ij}}. \tag{3.4}$$

The parameters A_i and B_j can be determined iteratively by the balancing method introduced in Chapter 1.

Although EMME/2, through the model (3.4), considers income earned by households (from survey studies), the number of trips generated at i and attracted at j and the cost of travelling from i to j, other factors describing socio-economic behaviour of households -for instance land cost, rent of locating a particular zone or a tax on exercising a particular business- are just ignored. In this regard, MEPLAN is better.

However, MEPLAN fails at the point of view of the calibration. Indeed, equations (3.7) and (3.4) both are expected to predict the number of trips between an origin and a destination. But in (3.7) the number of inputs (and consequently the number of parameters) is far large than in (3.4). This argument is supported by Kirby HR in [11]: 'it is clear that the problem of estimating the parameters that satisfy calibration requirements becomes more complex as the number of parameters increases'.

To prove this, let us suppose that the study area is subdivided into N transport zones. Then the number of parameters to calibrate in (3.4) is $2N + 1$ (N A_i's, N B_j's and β). On the other hand, the utility U_{ij} in (3.7) has the form

$$U_{ij} = \sum_i \theta_{ij} X_{ij}.$$

X_{ij} are the attributes of the locational utility and θ_{ij} are N^2 parameters to calibrate. $N^2 \geq 2N + 1$ as N is supposed to be large (as it does not make sense to develop a transport model for a study area with, for example, 2 transport zones). Equation (3.7) contains much more parameters than the only N^2 involved in U_{ij}.

This makes MEPLAN expensive to calibrate and increases the probability of making errors.

3.3.3 MEPLAN and the Multinomial Logit Model

MEPLAN uses a multinomial logit model (MNL) to distribute trips according to the utility that each origin (locational zones) offers to travellers with given destination (workplace). Being a disaggregate demand model, it is based on observed choices made by individual travellers rather than on averages. This makes it transferable in time and space.

Furthermore, the utility function (used in logit models) allows any number of variables (attributes), as opposed to the case of the generalised cost function (used in conventional methods) which is generally limited and has several fixed parameters. Therefore, the policy variables considered relevant in the modelling exercise are more flexibly represented in disaggregate models and the relative importance of variables are reflected by their coefficients. This method is expected to produce more realistic results than the Furness method.

Simmonds [15] suggests that the flow of labour F_{ij}^{mn} of equation (2.5) needs to be converted into trips in the following way:

$$T_{ij}^{mn} = t_{ij}^{mn} F_{ij}^{mn} \tag{3.5}$$

for trips by residents type m to work in employment type n.

$$T_{ij}^{mn} = t_{ij}^{mn} F_{ji}^{nm} \tag{3.6}$$

for trips by residents of type m to obtain services type n, where

T_{ij}^{mn} = trips by residents of type m living at i to j for purpose n, per month,

t_{ij}^{mn} = trip rate per month per unit of F_{ij}^{mn}.

The above follows from the fact that 'the Cape Town land-use model was run on time periods of a month (i.e floorspace rents are rents per month), and the transport model on time period of a 12-hour day' [16, p.67]. As a consequence, F_{ij}^{mn} is calculated on the basis of a time period of a month and T_{ij}^{mn} are one-way trips per month between home end i and non-home end j. T_{ij}^{mn} must be turned into one way trips per day or per peak period.

Equation (2.5) in (3.5) produces:

$$T_{ij}^{mn} = t_{ij}^{mn} . Y_j^{\ m} \frac{e^{\lambda^m(uU_{ij}^m - d_{ij}^m + w_i^m)}}{\sum_i e^{\lambda^m(uU_{ij}^m - d_{ij}^m + w_i^m)}}. \tag{3.7}$$

Taking into consideration the fact that U_{ij}^m is given by (2.6) and is maximum provided that (2.7) is satisfied, we conclude that the distribution of trips is a function of travel time and cost, land costs and elasticity of demand for land, and income. Equation (3.7) gives the distribution of trips between each O-D pair (i, j) by person type regardless of modes used to achieve these trips.

Let $T_i^m \geq 0$ be the number of trips generated at origin i and let $\lambda^m w_i^m = \ln T_i^n$. Then the distribution model (3.7) becomes

$$T_{ij}^{mn} = T_i^m Y_j^{\ m} t_{ij}^{mn} \frac{e^{\lambda^m(uU_{ij}^m - d_{ij}^m)}}{\sum_i T_i^m e^{\lambda^m(uU_{ij}^m - d_{ij}^m)}}. \tag{3.8}$$

Thus, the logit model (3.7) is defined by a gravity model with an exponential cost function.

3.3.4 EMME/2 and the Gravity Model

One of the main advantages of EMME/2 is that the modeller can specify his own model. For instance, one may decide to use the Multinomial Logit Model for trip distribution, but then the cost of calibration and the number of zones become a problem.

In Cape Town, the Gravity Model is used and calibrated with the trip length distribution according to the equation

$$T_{ij} = A_i B_j O_i D_j e^{-\beta c_{ij}}$$

where A_i and B_j are the balancing factors and c_{ij} is the travel cost between each O-D pair (i, j). The parameter β determines the average trip length and the sensitivity of people to trip length. The estimation of the parameter β requires some survey work unlike in growth factor methods where a complete O-D survey is needed to determine the base year matrix. Moreover, the problem of zero elements highlighted in Chapter 1 cannot arise in Gravity Models as it does in growth factor methods.

The parameters A_i and B_j are estimated as part of the Furness (biproportional) balancing factor operations. The parameters β must be calibrated to make sure that the trip length distribution is reproduced as closely as possible. This is not an easy task for a single parameter. The A_i's and the B_j's are functions of β through the two sets of equations (1.27) and (1.28) and yet a practical technique to estimate the best value for β is needed.

A naive approach to this task is to 'guess' a value of β, run the Gravity Model and then extract the modelled trip length distribution (MTLD). This should be compared with the observed trip length distribution (OTLD). If they are not sufficiently close, a new guess for β can be used and the process repeated until a satisfactory fit between MTLD and OTLD is achieved.

3.3.5 Trip Distribution in MEPLAN with the Apartheid Policy.

MEPLAN envisages a locational framework mainly constrained by people's income. In other words, people tend to reside in the zones which offer them a maximum utility according to their income (see (2.4)). This means that a person working in a zone j might have several alternatives of residential zones i and chooses the one he finds to be more convenient but affordable in terms of rents and other related expenses (how far from the workplace j, from the nearest hospital, from schools or university, etc).

However, the apartheid system in South Africa in general and in Cape Town in particular did not allow such flexibility. People were forced to reside in zones according to their race rather than their income. Therefore, the use of the concept of locational utility was inappropriate. Furthermore, it was seen that people with low income had to travel long distances to work unlike their high income counterpart and this contrasts with MEPLAN features. In addition to this, several informal settlements observed in most townships of the Cape Town Metropolitan Area had made survey studies difficult even impossible, therefore complicating the calibration process.

Modellers had to artificially 'create' attractors to justify, in terms of the model, the fact that some people 'chose' (they were actually forced) to live in less attractive zones. As a logical consequence, the application of MEPLAN in such a situation could not come up with successful results since it is a model derived from the Random Utility Theory.

The abolition of Apartheid has shown a clear evolution in the choice of residential places which is now income dependent. The institutionalization of democracy in South Africa, the people's empowerment and job creation policies are factors among others which would match and encourage transport modelling with a software such as MEPLAN.

Indeed, in the Cape Town metropolitan area, the calibration of a distribution model on observed trip patterns would lead to a perpetuation of certain apartheid anomalies (for instance, long trip lengths). On the other hand, it is unrealistic to assume that the post-apartheid Cape Town will become a city where most inhabitants will find suitable residential accommodation within walking distance from their places of employment.

The above arguments justify the high rate of relocation or change of employment, or both by households in the 'new' Cape Town. This should provide reasonable opportunities for achieving some degree of optimisation between places of residence and employment. As a consequence, average commuter trip lengths tend to gradually reduce during the course of years to come.

3.4   Modal Split in MEPLAN and EMME/2

Modes are structured in MEPLAN as shown in Figure 1.1 whereas, in EMME/2 the modal structure is presented in Figure 3.1. The seemingly hierarchical structure of modes in EMME/2 is not observed in applying the model since the modal split precedes the distribution of trips.
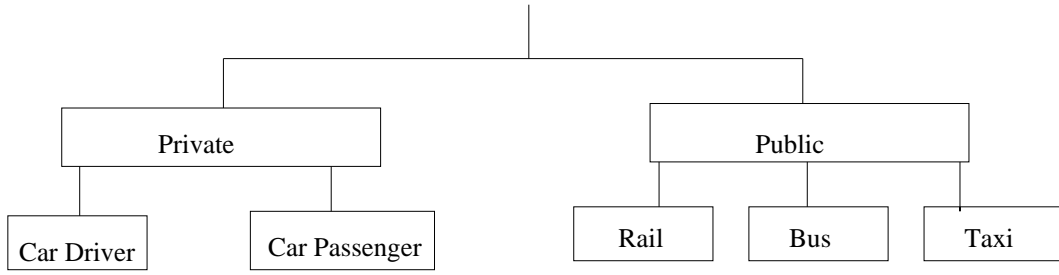
Figure 3.1: The hierarchy of modes in EMME/2

The choice of transport mode in MEPLAN depends on the disutility (or generalised cost) of the journey made by that mode according to equation (2.17) which is a linear function of distance, cost and time. The weights d^{mf}, c^{mf}, and t^{mf} as well as the mode parameter k^{mf} are to be estimated by calibration.

Equation (2.18) computes the proportion of people travelling by a particular mode m out of the modes in the lower level of the hierarchy. The composite cost function for the supermode n that leads to this lower level is given by:

$$Y^{nf} = -\frac{1}{\lambda_L^f} \ln \sum_m e^{-\lambda_L^f Y^{mf}}. \tag{3.9}$$

λ_L^f measures the sensitivity of travellers to the costs of different modes (see [18]). If λ_L^f is small, there is little price discrimination between modes; but if it is large, the majority of people travel by the minimum cost mode. To show this, let us assume that $Y^1 = \text{Min}_i\{Y^i\}$, $i = 1, 2, \ldots, m$ where Y^i is the disutility by mode i (f is omitted for clarity). By L'Hôpital's rule, we have:

$$
\begin{aligned}
\lim_{\lambda_L \to \infty} Y^n &= \lim_{\lambda_L \to \infty} \left[-\frac{e^{-\lambda_L Y^1}(-Y^1 - Y^2 e^{-\lambda_L(Y^2 - Y^1)} \ldots - Y^m e^{-\lambda_L(Y^m - Y^1)})}{e^{-\lambda_L Y^1}(1 + e^{-\lambda_L(Y^2 - Y^1)} + \ldots + e^{-\lambda_L(Y^m - Y^1)})} \right] \\
&= Y^1 \text{ since } Y^i > Y^1, \quad \forall i \in \{2, \ldots, m\} \\
&= \text{Min}_i Y^i.
\end{aligned}
$$

The composite cost function for the nest n is less than or equal to the minimum of the modal costs inside the nest. Indeed, if $Y^1 = \text{Min}_i\{Y^i\}$, $i = 1, 2, \ldots, m$, then:

$$
\begin{aligned}
Y^n &= -\frac{1}{\lambda_L} \ln \left[e^{-\lambda_L Y^1} \left(1 + e^{-\lambda_L(Y^2 - Y^1)} + \ldots + e^{-\lambda_L(Y^m - Y^1)} \right) \right] \\
&= -\frac{1}{\lambda_L} \ln e^{-\lambda_L Y^1} - \frac{1}{\lambda_L} \ln \left(1 + e^{-\lambda_L(Y^2 - Y^1)} + \ldots + e^{-\lambda_L(Y^m - Y^1)} \right) \\
&= Y^1 - \frac{1}{\lambda_L} \ln \left(1 + e^{-\lambda_L(Y^2 - Y^1)} + \ldots + e^{-\lambda_L(Y^m - Y^1)} \right) < Y^1.
\end{aligned}
$$

79

Moreover, the proportion of travellers using mode m in the nest n is given by $\frac{\partial Y^n}{\partial Y^m}$:

$$
\begin{aligned}
\frac{\partial Y^n}{\partial Y^m} &= -\frac{1}{\lambda_L}\frac{\partial}{\partial Y^m}\Big(\sum_m \ln e^{-\lambda_L Y^m}\Big) \\
&= \frac{1}{\lambda_L}\frac{\lambda_L e^{-\lambda_L Y^m}}{\sum_{m'} e^{-\lambda_L Y^{m'}}} \\
&= \frac{e^{-\lambda_L Y^m}}{\sum_{m'} e^{-\lambda_L Y^{m'}}}.
\end{aligned}
$$

3.5 Network Assignment in MEPLAN and EMME/2

3.5.1 Assignment in MEPLAN

Another feature of MEPLAN supposes that low income households live closer to their workplace unlike high income households. This subscribes to Random Utility Theory: long trips imply high travel costs (in terms of money, time and discomfort). As a consequence for individuals with limited income, they imply little utility. The spatial distribution of jobs and residential locations of households did not match with the above-mentioned assumption, thus complicating the implementation of MEPLAN in the Cape Town metropolitan area.

It was observed that people decided first of the transport mode to adopt before they actually made any choices about the route to follow since the difference in costs between different routes within a mode are not much pronounced as they are between different modes. Therefore people perceived mode costs directly rather than route costs and route split is determined within modes according to (1.80). This is the base for the assignment in MEPLAN.

The concept of reasonable path applies and in the case of several reasonable paths, the composite generalised cost is calculated:

$$
C_{ij}^k = -\frac{1}{\lambda}\ln\sum_{r\in M_{ij}(k)} e^{-\lambda\gamma_{ij}^r}. \tag{3.10}
$$

It is worth recalling that γ_{ij}^r is the observed cost on the r-th route between i and j and C_{ij}^k is the perceived cost for mode k between i and j. Equations (3.10) and (2.20) represent the same situation. Within a mode, loads on routes are positively related to their capacity. Hence, (1.80) might be updated, if route capacity were to be considered to become

$$
\frac{S_{ij}^{rn}}{T_{ij}^{kn}} = \frac{e^{-\mu^n\gamma_{ij}^r}.b_{ij}^r}{\sum_{r\in R_{ij}(k)} e^{-\mu^n\gamma_{ij}^r}.b_{ij}^r}. \tag{3.11}
$$

which corresponds to equation (2.21).

3.5.2 Assignment in EMME/2

The assignment models in EMME/2 seek to attain the Wardrop's user optimal principle or the optimal strategy principle depending on whether it is an auto assignment or a transit assignment. Optimal strategies consider various factors that include transit headways, transit speeds, transit accessibility, and ultimately choose the shortest transit path with respect to time. The assignment process is performed separately over the available network. Public transport is first loaded because they use (fixed) regular lines, then private cars are assigned to all remaining unsaturated links. This is to avoid the situation in which some links are overloaded whereas others are not fully used, in accordance to link capacity.

The assignment must be performed in such a way to alleviate congestion of the network. The objective function in the actual problem to solve is therefore non-linear because 'whenever congestion phenomena are present, the cost functions that are employed to reflect such situations are nonlinear' [7]. Some mathematical methods are then used to linearise the problem.

EMME/2 uses Volume Delay Functions (VDFs) of the form

$$s(v) = t_0[1 + \beta(v/c)^\alpha] \tag{3.12}$$

for each link of the auto assignment where,

$s(v)$: impedance (travel time) perceived by an individual user.

t_0: free flow travel time on a link

v/c: volume to capacity ratio

α, β: constants that vary based on the link type.

Like in most transport modelling systems, one of the primary goals of the Cape Town Metropolitan Transport Modelling is to alleviate congestion on the network. Through EMME/2, this would be achieved by minimising the overall network travel time and therefore maximising the system wide benefits.

The Cape Town Metropolitan Area network is still experiencing strong congestion, mostly during morning and afternoon peak periods causing considerable loss to the economy of the area due to delays (at work, for instance).

An appropriate toll pricing strategy for all the facilities in the system could be set up in order to perform the maximisation of the system wide benefits. The use of augmented VDFs concept which was developed based on the work of Dr. Randall Podzena cited in [3] would be suggested for the Cape Town Metropolitan situation.

The augmented VDF is defined in consideration of marginal social cost in the system as opposed to the marginal individual cost that is usually perceived by the individual user. Individual users perceive their own delay; they do not perceive the incremental delay that their vehicles impose on other users. The key principle in the augmented VDF approach is that the toll paid by an individual user of the system should equal the incremental delay (cost) he/she is causing to the system. Hence, the total impedance to individual user is the sum of individual delay and the incremental delay. The regular VDF given by (3.12) is modified to an augmented VDF and has two components, the individual delay component and the toll component. Hence, the toll in this case is applied in terms of time.

$$s(v)_{aug} = s(v) + s(v)_{toll} = t_0[1 + (\alpha + 1)\beta(v/c)^\alpha] \tag{3.13}$$

where,

$s(v)_{aug}$: the augmented VDF

$s(v)_{toll} = t_0\alpha\beta(v/c)^\alpha$: impedance due to the pricing (incremental delay).

The use of the augmented VDF may have as advantage the reduction of travel on the most congested routes which will improve their flow and operations. In fact, with the augmented VDF, users perceive greater impedance on the links and this causes changes in travel behaviour, including alternative destinations with shorter travel distance, diversion to alternative routes, and shift to alternative modes. The possible application of the augmented VDF in the Cape Town Metropolitan Area is not expected to modify the overall person trip productions and attractions in response to tolls. However, travel demand will be redistributed among alternative routes and modes, which will result in some trips being shortened. This Congestion Pricing Model was set up for the Pouget Sound Region in the State of Washington (see [3]) on the basis of the following assumptions:

- Fixed level of travel demand i.e total productions and attractions of person trips remains constant

- Tolling is applied in terms of time

- Tolling is applied to only General Purpose vehicles. Transit Vehicles and HOV's(High Occupancy Vehicles) are not being tolled

- A modified/augmented volume delay function by facility type, as described previously, are being used to simulated tolls

- The augmented VDFs are not applied to HOV links. HOV links have regular VDFs.

- Tolls are applied on all auto mode links in the model. However, different classification of facilities (links) may have different toll values corresponding to v/c ratios and the VDF.

- Tolls are applied all day.

These assumptions may be revisited (relaxed or strengthened) for the Cape Town Metropolitan situation.

Conclusion

Modelling transport intends to solve planning problems by first identifying the policy variables (and objectives), then constructing balanced behaviourally rich and computationally tractable models. Some of the elements usually considered are the zoning system, the population groups, the level of aggregation (individuals, households or zones) and the transportation systems' performance.

This thesis has highlighted some important sources of demarcation between MEPLAN and EMME/2. After we have shown the inappropriateness of MEPLAN in the apartheid situation of the Cape Town metropolitan area, we released a number of aspects in which either model is stronger than the other in the post-apartheid context.

While presenting some aspects of transport modelling in Chapter1, MEPLAN and EMME/2 were overviewed in Chapter 2. Chapter 3 discussed the population groups, the zoning systems and the mathematical expressions of the two models.

Grounded in Random Utility Theory, MEPLAN appears to be behaviourally richer than EMME/2. Moreover, trip distribution is given by a logit model in opposition to EMME/2 synthetic (entropy) distribution model. We have showed that the calibration exercise of MEPLAN is difficult as compared to that of EMME/2. As a summary, the computational tractability of EMME/2 makes it popular. However MEPLAN is flexible and behaviourally rich therefore more efficient.

The above arguments could be supported by the implementation of the same data with both MEPLAN and EMME/2 to make this work more complete. We failed to achieve this goal because of the lack of MEPLAN version which could run on the computers available to us.

Bibliography

[1] Abraham J.E (1998), 'A review of the MEPLAN modelling framework from a perspective of urban economies', University of Calgary, Department of Civil Engineering, Research Report No CE98-2

[2] Cape Town Metropolitan Transport Model, User Notes, Moving Ahead 3/31/2004

[3] Davuluri S, Longo N, Adury MK and Scott C, 'Application of EMME/2 and Enif for a Congestion Relief Analysis Study in the Puget Sound Region', paper submitted for presentation at the 18th Annual International EMME/2 Users Conference, Mexico City, Mexico, October 20-22, 2004.

[4] Evans AW(1970), Some Properties of Trip Distribution Methods, *Transportation Research* **4**, 19-36.

[5] Evans A.W(1971), The Calibration of Trip Distribution Models with Exponential or Similar Cost Functions, *Transportation Research* **5**, 15-38.

[6] Evans S.P. and Kirby H.R (1974), A three-dimensional Furness procedure for calibrating gravity models, *Transportation Research* **8**, 105-123.

[7] Florian M.(1986), Nonlinear Cost Network Models in Transportation Analysis, *Mathematical Programming Study* **26**, 167-196.

[8] Hensher D.A. 1986), Sequential and full information maximum likelihood estimation of a nested logit model, *The Review of Economics and Statistics* **68**, 657-667.

[9] INRO Consultants (1998), EMME/2 User's Manual Release 9, Montréal, Canada.

[10] Jong G, Daly A, Pieters M and Hoorn T., 'The logsum as an evaluation measure: review of the literature and new results', paper presented at the 45th congress of the

European Regional Science Association, Vrije Universiteit Amsterdam, 25-27 August 2005.

[11] Kirby H.R(1973), Theoretical requirements for calibrating gravity models, *Transport Research* **8**, 97-103.

[12] Ortuzar J. de D. and Willumsen L.G.(2005), *Modelling Transport*, 3rd Edition, John Wiley & Sons, West Sussex, England.

[13] Rontiris K. and Crous W. 'Integrating land-use and Transport Planning Procedures with EMME/2 in Cape Town, South Africa', from 12th International EMME/2 Users Conference, San Francisco, California, 23-24 Oct, 1997

[14] Safwat K.N.A and Magnanti T.L(1988), A Combined Trip Generation, Trip Distribution, Modal Split, and Trip Assignment Model, *Transportation Science* **18**, 14-30.

[15] Simmonds D. C., 'Cape Town Metropolitan Transport Area Land-Use/Transport Study, Report 1: Model Design', Marcial Echenique & Partners Ltd, Cambridge, for Cape Town City Council, 1984

[16] Simmonds D. C., 'Cape Town Metropolitan Transport Area Land-Use/Transport Study, Report 4: Calibration', Marcial Echenique & Partners Ltd, Cambridge, for Cape Town City Council, 1986

[17] Spiess H. and Florian M. (1989), Optimal strategies: A new assignment model for transit networks, *Transportation Research* **23B**, 83-102

[18] Wilson AG(1969), The use of entropy maximising models in the theory of trip distribution, mode split and route split, *Journal of Transp. Econ. & Policy* **3**, 108-126

[19] Wilson AG & Kirkby MJ, *Mathematics for Geographers and Planners*, Clarendon Press, Oxford, 1975.