Transpn Res.-B Vol. 14B, pp. 281-293
Pergamon Press Ltd., 1980. Printed in Great Britain

THE MOST LIKELY TRIP MATRIX ESTIMATED FROM TRAFFIC COUNTS

HENK J. VAN ZUYLEN
Verkeersakademie, 5022DD Tilburg, Holland

and

LUIS G. WILLUMSEN
The University of Leeds, Leeds LS2 9JT, England

(*Received* 25 *April* 1979; *in revised form* 7 *June* 1979)

Abstract—For a large number of applications conventional methods for estimating an origin destination matrix become too expensive to use. Two models, based on information minimisation and entropy maximisation principles, have been developed by the authors to estimate an O–D matrix from traffic counts. The models assume knowledge of the paths followed by the vehicles over the network. The models then use the traffic counts to estimate the most likely O–D matrix consistent with the link volumes available and any prior information about the trip matrix. Both models can be used to update and improve a previous O–D matrix. An algorithm to find a solution to the model is then described. The models have been tested with artificial data and performed reasonably well. Further research is being carried out to validate the models with real data.

1. INTRODUCTION

The normal method for obtaining an origin–destination (O–D) trip matrix developed for large scale transport modelling exercises employs a combination of home interviews and roadside surveys. However, the high cost of this approach precludes its use for most other applications. Alternative methods suggested for smaller studies, including limited roadside interviews, flagging techniques (e.g. number plate surveys), aerial photography and "car following", also tend to be expensive in terms of manpower requirements, disruptions and/or processing. In addition all these methods, with the possible exception of aerial photography, imply sampling and thus an estimation of the full O–D matrix by means of grossing up factors.

This paper proposes a method of obtaining O–D matrices directly from one of the most common pieces of traffic information, traffic counts. Since they are relatively inexpensive to obtain, are usually collected for several purposes (accident studies, maintenance planning, intersection improvements) and their automation is relatively advanced, the possible development of such methods is very attractive. A small number of such techniques have been suggested in the past few years for the following areas of application:

(a) Modelling rural or interurban transport demand, especially where other data is expensive to obtain, unreliable or out of date.

(b) Modelling transport demand in free standing towns and cities where a full scale study is too expensive.

(c) The assessment and design of traffic management schemes in inner city or local areas.

The organisation of the remainder of this paper is as follows. First, a more rigorous definition of the problem is presented and discussed in Section 2. Section 3 discusses the conditions under which a trip matrix may be estimated. Section 4 briefly describes O–D matrix estimation techniques based on some explicit travel demand model, usually of the gravity type. Section 5 describes an alternative model for the estimations of a trip matrix developed by one of the authors (HJVZ) and based on an information minimising approach. A similar model derived from maximum entropy considerations and developed by the second author (LGW) is presented in Section 6. Section 7 gives the link between the models and Section 8 describes an algorithm for the solution. Section 9 describes tests on an artificial network for these last two models. Finally Section 10 identifies some tentative conclusions regarding the application of these approaches.

2. THE NATURE OF THE ESTIMATION PROBLEM

Consider a study area which has been divided into n zones (centroids) with trips from all origins to all destinations but where we can safely disregard intrazonal trips. We denote the trip matrix by $\{T_{ij}\}$ and we force $T_{ij} = 0$ for all $i = j$. For the purpose of the analysis the road network is represented by a set of nodes and links joining node pairs. A set of observed link flows has been obtained by counting traffic at the corresponding points on the road network.

A key issue in the estimation of a trip matrix from traffic counts is the identification of the origin–destination pairs whose trips use a particular link. Throughout this paper we shall use the variable p_{ij}^a, to represent the proportion from origin i to destination j which use link a. In general:

$$0 \leq p_{ij}^a \leq 1.$$

If we denote the flow on link a by V_a, the fundamental equation in the estimation of a trip matrix from traffic counts is:

$$V_a = \sum_i \sum_j p_{ij}^a T_{ij}. \tag{1}$$

The type of assignment thought to be realistic is a crucial assumption in the determination of p_{ij}^a. Two general types of traffic assignment are of importance here:

(a) *Proportional assignment.* In this case p_{ij}^a is assumed to be independent of traffic volumes, i.e. congestion. In other words, the proportion of drivers choosing each route depends on assumed drivers and route characteristics but not on flow levels. The values of p_{ij}^a can be determined independently of the trip matrix T_{ij} and before any estimation is made of it. Most stochastic assignment models, for example Dial's, fall into this group. All-or-nothing is a particular case of proportional assignment and can be characterised by:

$$p_{ij}^a = \begin{cases} 0 \text{ if link } a \text{ is not used by trips from } i \text{ to } j \\ 1 \text{ if link } a \text{ is used.} \end{cases}$$

(b) *Capacity restrained assignment.* Wherever congestion effects are thought to be more important than the differences in perceived costs, capacity restrained assignment is usually considered more realistic. The cost of travelling on a link depends on the flow on, and the cost-flow relationship for, that link. Today, equilibrium assignment techniques, which try to satisfy Wardrop's first principle that "equilibrium" is reached when no driver can reduce his/her travel costs by switching to another route, are often used in these cases. There are several methods, based on both heuristic and mathematical programming techniques, which either satisfy this condition or provide a reasonable approximation. In this case, the value of p_{ij}^a depends on the flow levels on each link and cannot be determined independently of the trip matrix estimation process.

Throughout this paper only proportional assignment cases will be treated. Congestion effects introduce additional complications to the problem which have not yet been satisfactorily solved. Assuming that one knows which type of proportional assignment is most realistic (all-or-nothing, stochastic, etc.) the problem of estimating the trip matrix is that of finding the values for the $n(n-1)$ unknowns $T_{ij}(i \neq j)$ from as many equations (1) as these are counted flows in the network. In most study areas $n^2 - n$ will be greater than the *total* number of links so that it will not be possible to uniquely determine $\{T_{ij}\}$ and the problem is underspecified. Assumptions about trip making behaviour may help to reduce the number of unknowns. The different techniques which have been proposed differ on the nature of these assumptions.

3. INDEPENDENCE AND INCONSISTENCY OF TRAFFIC COUNTS

The estimation problem is further complicated by the fact that some counts may well fail to add any information. Not all equations (1) will be independent since it should be possible to write, for each node m, flow continuity equations of the type

$$\sum_{l} V_{lm} - \sum_{k} V_{mk} = 0. \tag{2}$$

where V_{lm} is a "flow in" and V_{mk} is a "flow out" from node m.
Thus if we know the total "flow out" from a node, specifying all but one of the "flows in" automatically determines the other "flow in".

On the other hand, traffic counts are certainly not error free. The simplified description of the network, counting errors and the fact that not all counts are made at the same time, are probably the main sources of error. This error component will, in general, prevent conditions (2) from being met, and the observed flows will be considered internally inconsistent.

There are two general approaches for dealing with the inconsistency problem, the most common being to develop trip matrix estimation models which accept inconsistent flows. The approaches based on a gravity model described in Section 4 do this. However, the approach followed by the authors is to use a maximum likelihood method put forward by Hamerslag and Huisman (1978) to remove inconsistencies and produce a better estimation of the link flows.

Assuming that observed flows are Poisson distributed with mean V_a the probability of counting $\hat{V}_a$ vehicles is

$$P(\hat{V}_a) = \frac{e^{-V_a} V_a^{\hat{V}_a}}{\hat{V}_a!}. \tag{3}$$

Then the likelihood of observing a set of $\hat{V}_a$ flows on a network is

$$L = \pi_a \, e^{-V_a} V_a^{\hat{V}_a} / \hat{V}_a!.$$

The best estimates for V_a are those which maximise L or $\log_e L$ subject to the constraints (2). The solution to this is found by forming the Lagrangean and differentiating with respect to V_a. The formal solution can be written as

$$V_a = \hat{V}_a (1 + \delta_l \eta_l - \delta_m \eta_m) \tag{4}$$

where $\delta_m = 1$ when there is a constraint on node m and $\delta_m = 0$ otherwise. The values of η are found by substitution of eqn (4) into eqn (2). It is in principle possible that the equations (2) are mutually dependent. In that case the solution for η is not determined uniquely.

This method is thought superior to a minimum squares estimate in that there is no risk of a modified flow turning out to be negative.

4. PREVIOUS WORK

Most of the alternative approaches to the estimation of a trip matrix from traffic counts assume that trip making behaviour is reasonably represented by a gravity model. In this way the number of unknowns is radically reduced and the problem usually becomes one of over-specification rather than underspecification. Robillard (1975) suggested a generalised gravity model of the type:

$$T_{ij} = R_i S_j f(C_{ij}) \tag{5}$$

where $f(C_{ij})$ is a deterrence function of the cost of travelling between i and j and R_i and S_j are zonal parameters to be calibrated from the traffic counts through eqns (1).

Low (1972), Hogberg (1976) and Holm *et al.* (1976) put forward more conventional gravity models requiring additional but easily obtainable data, e.g. population and employment by zone. The use of eqns (1) to calibrate the parameters of these models leads to a non-linear regression approach in Hogberg's case, linear regression for Low and maximum likelihood for Holm *et al.*

Symons *et al.* (1976) combine some concepts of Central Place Theory to produce a gravity type of model for intercity travel.

Nguyen (1977) has suggested an equilibrium assignment approach but, as he himself observed, his method fails to find a unique trip matrix due to the underspecification problem. Gur *et al.* (1978) have adapted Nguyen's approach for small study areas. They treat the underspecification problem by using a "target" trip matrix and search for the solution to Nguyen's problem which is closest to it. This "target" O–D matrix is determined using a distribution model specially developed for small areas. A more comprehensive review of these methods can be found in Willumsen (1978a). The main criticism that can be levied against the abovementioned models is that by forcing the trip matrix to follow a gravity type pattern they are not making full use of the information contained in the counts. This is critical in small or inner city areas where most of the trip length will be outside the area. A better approach would be to solve the underspecification problem by introducing the minimum external information. This is what the authors have done following two very similar lines: first, information minimisation and second, entropy maximisation.

5. INFORMATION MINIMISING APPROACH

One possible starting point is to calculate the information contained in a trip matrix T_{ij}. Since the information available in the traffic counts on the links is insufficient to determine a complete trip matrix, it seems reasonable to choose a trip matrix that adds as little information as possible to the knowledge contained in eqns (1). This approach has been followed by Van Zuylen (1978) using Brillouin's information measure.

The information contained in a set of N observations where the state k has been observed n_k times is defined by Brillouin (1956) as:

$$I = \log_e N!\, \pi_k [q_k^{nk}/n_k!] \tag{6}$$

where q_k is the *a priori* probability of observing state k. If the observations are counts on a particular link it is possible to define state ij as the state in which the vehicle observed has been travelling between origin i and destination j. So,

$$n_{ij}^a = T_{ij}p_{ij}^a. \tag{7}$$

We can also express the *a priori* probability of observing state ij on link a as a function of *a priori* information about the O–D matrix as

$$q_{ij}^a = \frac{t_{ij}p_{ij}^a}{\sum_{ij} t_{ij}p_{ij}^a} \tag{8}$$

where t_{ij} is the *a priori* number of trips between i and j provided, for example, by an old O–D matrix.

The information contained in V_a counts on a link is then

$$I_a = -\log_e V_a!\, \pi_{ij} \left\{\frac{t_{ij}p_{ij}^a}{S^a}\right\}^{T_{ij}p_{ij}^a} \Big/ \pi_{ij} (T_{ij}p_{ij}^a)! \tag{9}$$

where

$$S^a = \sum_{ij} t_{ij}p_{ij}^a. \tag{10}$$

Using Stirling's approximation, $\log_e X! \simeq X\log_e X - X$, it is possible to obtain

$$I_a \simeq \sum_{ij} T_{ij}p_{ij}^a \log_e \frac{T_{ij}\cdot S^a}{V_a t_{ij}}. \tag{11}$$

Summing up over all the links in the network with counts:

$$I_a = \sum_a \sum_{ij} T_{ij} p_{ij}^a \log_e \frac{T_{ij} \cdot S^a}{V_a t_{ij}} \tag{12}$$

is the total information contained in the observed flows. The problem of finding an O–D matrix consistent with the observations and adding a minimum of extra information to them is *equivalent to minimizing* I (12) subject to the flow constraints (1).

The formal solution to this problem can be obtained by differentiation of the Lagrangean.

$$\mathscr{L} = \sum_a \sum_{ij} T_{ij} p_{ij}^a \log_e \frac{T_{ij} S^a}{V_a t_{ij}} + \sum_a \lambda_a \left(\sum_{ij} T_{ij} p_{ij}^a - V_a \right) \tag{13}$$

where λ_a is the Lagrangian multiplier corresponding to the count a. Thus:

$$\frac{\delta \mathscr{L}}{\delta T_{ij}} = \sum_a p_{ij}^a \log_e \left[\frac{T_{ij} S^a}{V_a t_{ij}} \right] + \sum_a T_{ij} p_{ij}^a \cdot \frac{1}{T_{ij}} + \sum_a \lambda_a p_{ij}^a \tag{14}$$

$$\frac{\delta \mathscr{L}}{\delta T_{ij}} = \sum_a p_{ij}^a \log_e \left[\frac{T_{ij} S^a}{V_a t_{ij}} \right] + \sum_a p_{ij}^a (1 + \lambda_a) = 0 \tag{15}$$

$$\log_e \left\{ \pi_a \left[\frac{T_{ij} S^a}{V_a t_{ij}} \right]^{p_{ij}^a} \right\} = -\sum_a p_{ij}^a (1 + \lambda_a) \tag{16}$$

$$\left(\frac{T_{ij}}{t_{ij}} \right)^{\Sigma_a p_{ij}^a} = \pi_a \left(\frac{V_a}{S^a} \right)^{p_{ij}^a} \exp \left[-\sum_a p_{ij}^a (1 + \lambda_a) \right]. \tag{17}$$

By making

$$g_{ij} = \sum_a p_{ij}^a \tag{18}$$

we end up with

$$T_{ij} = t_{ij} \pi_a \left\{ \frac{V_a}{S^a} \cdot e^{-(1+\lambda_a)} \right\}^{p_{ij}^a / g_{ij}}. \tag{19}$$

By replacing

$$X_a = \frac{V_a}{S^a} \cdot e^{-(1+\lambda_a)} \tag{20}$$

we obtain

$$T_{ij} = t_{ij} \pi_a X_a^{p_{ij}^a / g_{ij}}, \tag{21}$$

where the X_a can be calculated from eqns (1). This problem has the form of a multi-proportional model.

6. AN ENTROPY MAXIMISING FORMULATION

Willumsen (1978a, 1978b) has tackled the same problem following an entropy maximising approach and generated a similar model. This derivation parallels the now conventional method of obtaining a doubly-constrained gravity model from maximum entropy considerations.

Following Wilson (1970) we define as the most likely O–D trip matrix the one having a greatest number of micro-states associated with it. The number of ways of selecting an O–D trip matrix T_{ij} with a total number of trips T is then:

$$W(\{T_{ij}\}) = \frac{T!}{\pi_{ij} T_{ij}!}. \tag{22}$$

We seek to find the trip matrix which maximises $W(\{T_{ij}\})$ or a monotonic function of it. For convenience we maximise

$$S = \log_e W(\{T_{ij}\}). \tag{23}$$

We replace here the conventional trip end and total expenditure constraints by simply the traffic counts constraints (1). Using Stirling's approximation for $\log X!$ the problem becomes

$$\text{Max } S = \log_e T! - \sum_{ij} (T_{ij} \log_e T_{ij} - T_{ij}). \tag{24}$$

We assume here the total number of trips T to be a constant which can then be excluded from the maximisation problem but we shall return to this point later. The problem is thus reduced to

$$\text{maximise } S' = -\sum_{ij} (T_{ij} \log_e T_{ij} - T_{ij})$$

$$\text{subject to } V_a - \sum_{ij} T_{ij} p_{ij}^a = 0(1) \tag{25}$$

for all counted links a.

The formal solution is obtained by forming the Lagrangean

$$£ = \sum_{ij} (T_{ij} \log_e T_{ij} - T_{ij}) + \sum \lambda_a \left(V_a - \sum_{ij} T_{ij} p_{ij}^a \right) \tag{26}$$

and differentiating it

$$\frac{\delta \mathcal{L}}{\delta T_{ij}} = -\log_e T_{ij} - \sum_a \lambda_a p_{ij}^a = 0. \tag{27}$$

Thus

$$T_{ij} = \exp\left[-\sum_a \lambda_a p_{ij}^a \right] \tag{28}$$

and by making

$$e^{-\lambda_a} = X_a \tag{29}$$

we obtain

$$T_{ij} = \pi_a X_a^{p_{ij}^a}. \tag{30}$$

This is another case of a multi-proportional problem.

We can also extend this model to make use of prior information following the lines suggested by Batty and March (1976) and Nathanson (1978). We replace eqn (22) by

$$W'(T_{ij}, t_{ij}) = T! \prod_{ij} q_{ij}^{T_{ij}} / \prod_{ij} T_{ij}! \tag{31}$$

where

$$q_{ij} = t_{ij} \Big/ \sum_{ij} t_{ij}. \tag{32}$$

Then, following steps analogous to those leading from eqn (22) to eqn (30) we find

$$T_{ij} = t_{ij} \prod_a X_a^{p_{ij}^a} \tag{33}$$

and

$$X_a = \left(\sum_{ij} t_{ij}\right)^{1/L} \cdot e^{-\lambda_a} \tag{34}$$

and L is the number of counted links.

Of course, wherever no *a priori* information is available regarding the trip matrix model (33) reverts to model (30).

7. SOME PROPERTIES OF THE MODELS

The main difference between the models in eqns (23) and (31) resides in the exponents, $p_{ij}^a / \sum_a p_{ij}^a$ for Van Zuylen's model and simply p_{ij}^a in Willumsen's. These have the form of weights to be associated to observations on link a. The similarity is not surprising as the close links between entropy maximising and information minimising have long been recognised. In fact, it is possible to show that the basic difference between the models resides in what is considered to be the unit of observation or the relevant meso-state: Van Zuylen's model uses a counted vehicle and Willumsen's a trip.

One way of showing this is to derive Van Zuylen's model by means of maximum entropy considerations.

Each trip from i to j is counted $g_{ij} = \sum_a p_{ij}^a$ times on average. Then in general

$$\tilde{T} = \sum_{ij} g_{ij} T_{ij} \tag{35}$$

vehicle counts are assigned to origin–destination pairs. Following Wilson (1970) again the number of possible ways of doing this may be used to determine the most likely trip matrix. The *a priori* probability that a counted vehicle actually moved from i to j is

$$q_{ij} = g_{ij} t_{ij} \Big/ \sum_{ij} g_{ij} t_{ij} = g_{ij} t_{ij} / \tilde{t} \tag{36}$$

The problem is then to maximise

$$\tilde{S} = \log_e \tilde{T}! \prod_{ij} \frac{(g_{ij} t_{ij} / \tilde{t})}{(g_{ij} T_{ij})!} \tag{37}$$

$$\tilde{S} \simeq \tilde{T}(\log_e \tilde{T}/\tilde{t} - 1) - \sum_{ij} g_{ij} T_{ij} (\log_e T_{ij}/t_{ij} - 1). \tag{38}$$

A further maximisation of $\tilde{S}$ subject to eqn (1) using Lagrangean multipliers gives again eqn (21).

It can be argued whether the assignment of traffic counts to origin–destination pairs or the determination of the most likely O–D trip matrix consistent with the observed flows best represents the real problem. Both S' and $\tilde{S}$ give a measure of the difference between a trip matrix $\{T_{ij}\}$ and an *a priori* matrix $\{t_{ij}\}$. As Murchland (1978) suggests, there is little or no theoretical argument to prefer any of such measures. Practical issues like computational problems and ease of use are likely to be more important.

Both models result in a multi-proportional problem whose treatment will be discussed in the next section. It is also easy to see that it is in no way necessary to have counts on all links in the network. Of course, one would expect a more complete set of counts to produce a better estimation of the O–D trip matrix.

8. SOLVING THE MULTI-PROPORTIONAL PROBLEM

Murchland (1977) studied the multi-proportional problem and pointed to its equivalence to a convex programme. He suggested an algorithm for its solution which can be easily adapted to our problem. The algorithm requires the recurrent modification of an initial trip matrix so that modelled flows are made equal to the observed ones. It requires setting up a list of links with counts $\hat{V}_a$ from $a = 1$ to L. Here and below, a is to be interpreted as a single label in the list of counts. The algorithm for the model in eqn (33) is as follows:

(i) Obtain, using a suitable assignment method, the values for p_{ij}^a and set the number of iterations $n = 0$.

(ii) Set $X_a^n = 1$ for all links.

(iii) Set $a = 0$.

(iv) Increase the link counter a by one. Take link a and calculate the modelled (estimated) flows.

$$V_a^n = \sum_{ij} t_{ij}(\pi_a X_a^{nP_{ij}^a}) \cdot p_{ij}^a$$

and make $X_a^{n+1} = X_a^n \cdot Y_a^{p_{ij}^a}$ where Y_a is to be obtained by solving

$$\hat{V}_a = \sum_{ij} p_{ij}^a t_{ij}(\pi_a X_a^{nP_{ij}^a}) \cdot Y_a^{P_{ij}^a}$$

(v) If the link list has not been exhausted ($a < L$) proceed to step (iv) otherwise check for convergence in step (vi).

(vi) If all the modelled flows V_a^n are suitably near to the observed ones (say within $\pm 5\%$) stop, otherwise increment n by 1 proceed to step (iii) and process the link list again.

Although Murchland (1977) has shown that the multi-proportional problem has a unique solution the ultimate convergence of this algorithm has not been satisfactorily proved yet. Nevertheless, a large number of experiments with artificial data has failed to show any case in which convergence was not attainable. It can be seen that in the limit the solution will always reproduce the observed volume counts.

9. TESTS WITH ARTIFICIAL DATA

In order to gain a better understanding of the strength and weaknesses of these new approaches one of the authors (L. Willumsen) tested these models on an artificial network with 15 centroids and 36 two-way links as depicted in Fig. 1. Four different artificial O–D trip matrices, were loaded (all-or-nothing) to it and the resulting flows recorded.

These flows and the network data were then used to estimate the original "observed" O–D trip matrix. Thus, no use was made of prior information, i.e. all $t_{ij} = 1$.

The four artificial "observed" O–D matrices were generated using different methods. The first O–D trip matrix contained simply 50 trips in each cell. A very good match between "observed" and modelled O–D matrix was expected for this type of trip matrix.

The second and third matrices were generated using a singly constrained gravity model

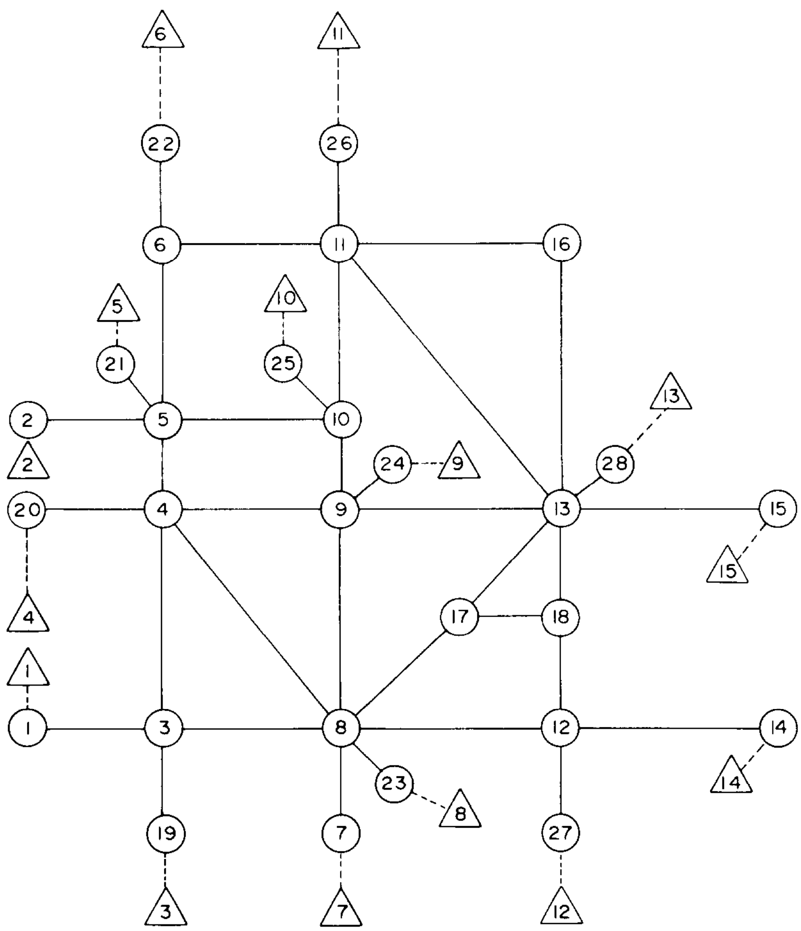


Fig. 1. Test network.

assuming certain trip generation and trip attraction values for each centroid (but these were not used in the estimation process). Trip matrix number two was generated using the network in Fig. 1. This case can be deemed to represent the problem of estimating an O–D trip matrix in a free standing town. Trip matrix number three modified the original network by increasing the lengths of all links feeding into the central area. For example the length of link (1, 3) was increased from 120 to 1200 m. The link flows thus obtained were used in conjunction with the *original* network descriptions to estimate an O–D matrix. This case can be deemed to represent the problem of estimating an O–D matrix in a central area where only local counts are available.

The fourth O–D trip matrix was generated by filling each cell with a random number between 1 and 100. This case could represent a situation in which the trip making behaviour reflects almost no perceivable regularity.

The models were run until all modelled link flows were within $\pm 2\%$ of the "observed" link flows. The results of these tests are summarized in Table 1 and Figs. 2 and 3 depict "observed" against modelled trips per O–D pair for Case 3.

As expected the models performed well in Case 1. The good performance in Cases 2 and 3 can be explained by the common origin of both the gravity and entropy maximising models. It should be noted that in Network 1 trip end totals are also "counted". The results in Case 4 are not so good but could certainly be improved with better *a priori* information.

In these tests the model suggested in Section 6 seems to have a slight edge. The information minimising model took on average 2.52 sec per iteration and the entropy maximising model only 1.42 sec. This difference is mainly due to the non integer exponent in the first model and it should disappear with other types of proportional assignment. Both models required about the same number of iterations and yielded similar. Root mean squared error figures for cases 1 and

Table 1. Results of tests with synthetic data. Comparison of estimated against "real" trip matrices

	Information Minimising Model						Entropy Maximising Model					
	Running time (secs)*	Number of iterations	Average RMSE**	Relative RMSE(%) 'real' trips ranges 0-4	5-10	10-∞	Running time (secs)	Number of iterations	Average RMSE	Relatavie RMSE(%) 'real' trips ranges 0-4	5-10	10-∞
CASE 1 (50)	37	14	1.2	0.0	0.0	2.4	16	11	1.1	0.0	0.0	2.1
CASE 2 (Gravity)	36	15	21.3	68.0	44.3	25.0	22	16	1.7	18.2	4.4	2.0
CASE 3 (Part Grav.)	34	13	16.3	59.5	55.0	18.9	21	15	1.7	20.1	0.0	2.0
CASE 4 (Random)	44	18	25.8	1888.1	365.7	43.5	16	11	25.5	1870.0	362.0	43.1

* In ICL 1906A at Leeds University including INPUT/OUTPUT

** Average RMSE calculated as $\text{RMSE} = \sqrt{\dfrac{\sum_k (E_k - O_k)^2}{n}}$

where n = number of cells (210)

E_k = Estimated number of trips for the kth origin destination pair

O_k = 'Observed' or 'real' number of trips for the kth O/D pair

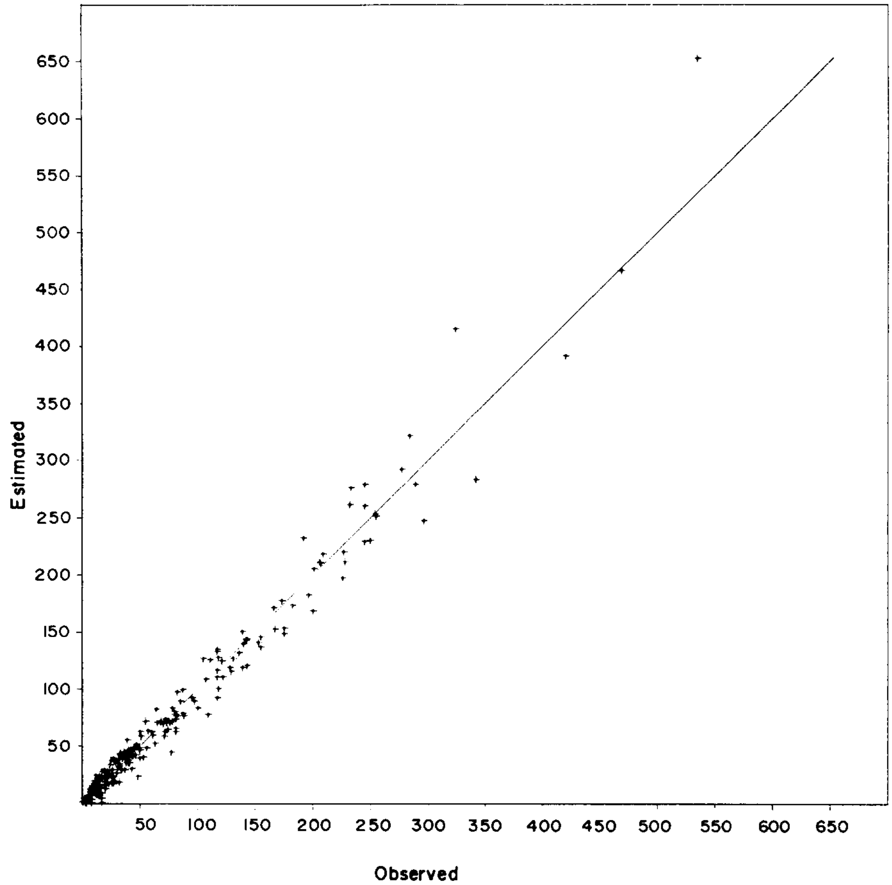


Fig. 2. Information minimising model. "Observed" vs estimated trip matrix for Test Case 3.

4. The O–D matrices estimated by the entropy maximising model were closer to the original ones for cases 2 and 3. Nevertheless, further tests with real data would be required to settle the question of which model is superior.

Tests were also made with the O–D matrix in Case 3 to compare the results of the entropy maximising model under different criteria for convergence. These results are displayed in Table 2 and Fig. 4.

As expected more stringent convergence criteria produce better results but these improvements take place mainly for cells with 5 or more trips. It appears desirable to aim at replicating the observed flows at least within 5% as the running times do not increase dramatically with more stringent convergence criteria.

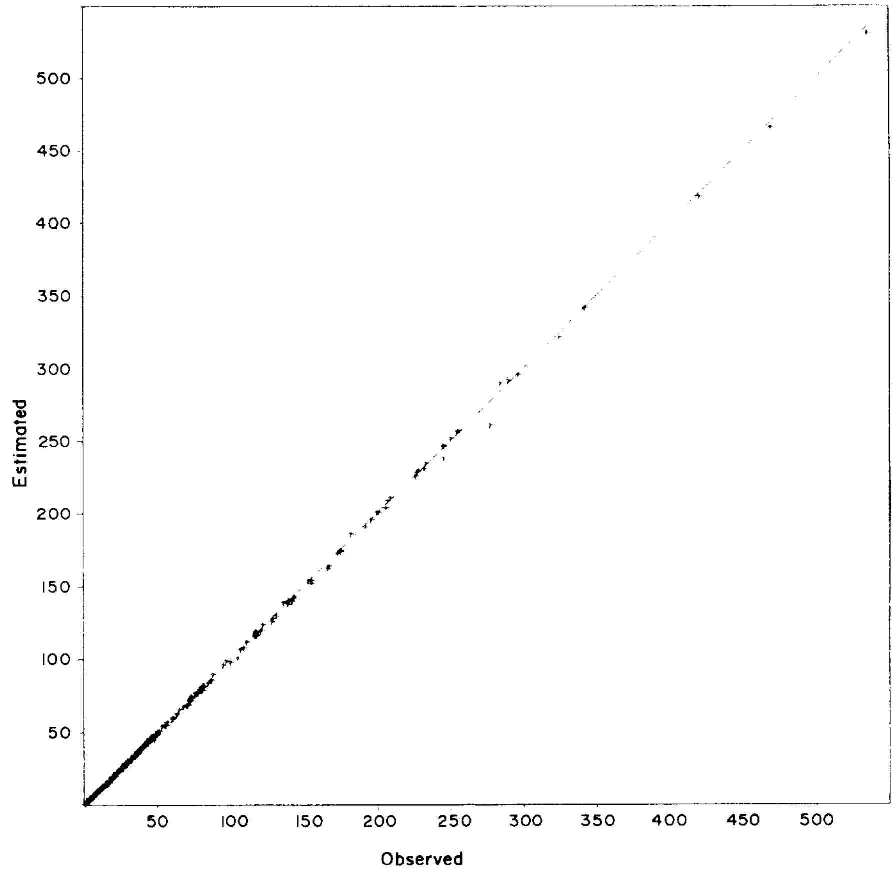
10. CONCLUSIONS

This is still an early stage of research to draw definitive conclusions about these models. The authors think they have shown the promise of this type of approach to the problem of estimating a trip matrix from traffic counts. Its main advantages, in principle, may be summarised as follows:

(a) The models make full use of the information contained in the observed counts. They produce an unbiased estimate on this available information.

(b) The inclusion of a prior estimation for the O–D matrix facilitates the inclusion of information different from traffic counts into the estimation problem. For example an old O–D trip matrix could be used to this end.

(c) The models appear particularly appropriate in areas where it is difficult to justify the hypothesis of trip making behaviour based on the gravity or similar models, e.g. small areas and partial networks.



ved"Fig. 3. Entropy maximising model. "Observed" vs estimated trip matrix for Test Case 3.

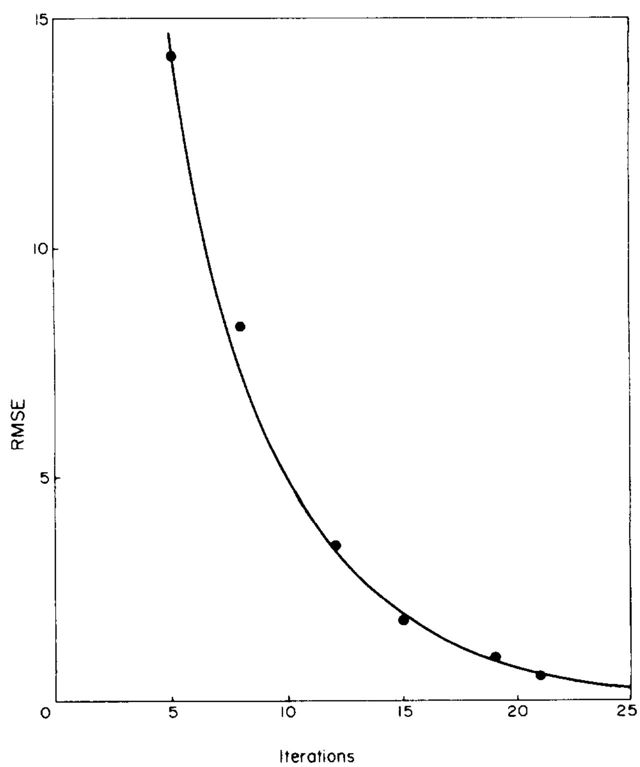


Fig. 4. RMSE against number of iterations for Entropy Maximising Model, Case 3.

Table 2. Comparison of estimated against real O–D matrix for Case 3, entropy maximising model

Error in the modelled link flows	Running time (secs)	Number of iterations	Average RMSE	Relative RMSE(%)		
				0-4	5-10	10-∞
15%	10	5	14.2	20.1	10.8	16.5
10%	13	8	8.3	20.1	7.3	9.6
5%	17	12	3.4	20.1	0.0	4.0
2%	21	15	1.7	20.1	0.0	2.0
1%	22	18	0.9	20.1	0.0	1.0
0.5%	25	22	0.5	20.1	0.0	0.6

(d) As pointed out in Willumsen (1979) the model may be extended to equilibrium assignment conditions but still greater work is required in this direction.

Research is currently under way attempting to validate these approaches with real data.

Acknowledgements—Most of this research has been carried out under a grant by the British Science Research Council. The authors also wish to thank their colleagues Dirck Van Vliet and Huw Williams for their advice and help.

REFERENCES

Batty M. and March L. (1976) The method of residues in urban modelling. *Environ. Planning* **A8**, 189–214.

Brillouin L. (1956) *Science and Information Theory*. Academic Press, New York.

Gur Y. *et al.* (1978) Determining an origin–destination trip table based on observed volumes. *ORSA–TIMS Annual Meeting*, New York, May 1978.

Hamerslag R. and Huisman M. C. (1978) Binaire calibratie. *Verkeerskunde*, pp. 166–168.

Högberg P. (1976) Estimation of parameters in models for traffic prediction: a non-linear approach. *Transpn Res.* **10**, 263–265.

Holm J. *et al.* (1976) Calibrating traffic models on traffic census results only. *Traffic Engng and Control* **17**, 137–140.

Low D. (1972) A new approach to transportation systems modelling. *Traffic Quart.* **26**, 391–404.

Murchland J. (1977) The multi-proportional problem. University College London, Research Note JDM 263 (unpublished).

Murchland J. (1978) Applications, history and properties of bi-and multi-proportional models. Unpublished Seminar at London School of Economics, London.

Nathanson M. (1978) Information minimisation Markov transition and dynamic modelling. *Environ. Planning* **A10**, 879–888.

Nguyen S. (1977) Estimating an O–D matrix from network data: a network equilibrium approach. Centre de recherche sur les Transports, Université de Montreal, Publication 87.

Robillard P. (1975) Estimating an O–D matrix from observed link volumes. *Transpn Res.* **9**, 123–128.

Symons J. *et al.* (1976) A model of inter-city motor travel estimated by link volumes. *ARRB Proc.* **8**, 53–65.

Van Zuylen H. (1978) The information minimising method: validity and applicability to transport planning. In *New Developments in Modelling Travel Demand and Urban Systems* (Edited by G. R. M. Jansen *et al.*). Saxon, Farnborough.

Wilson A. G. (1970) *Entropy in Urban and Regional Modelling*. Pion, London.

Willumsen L. (1978a) Estimation of an O–D matrix from traffic counts: a review. *Institute for Transport Studies Working Paper* 99, Leeds University.

Willumsen L. (1978b) O–D matrices from network data: a comparison of alternative methods for their estimation. *Proc. of PTRC Summer Annual Meeting* 1978 *Seminar in Transport Models*, PTRC Education Research Services Ltd., London.

Willumsen L. (1979) Estimating the most likely O–D matrix from traffic counts. *11th Ann. Conf. of Universities Transport Studies Group*, University of Southampton, January 1979.