MEUSE:

AN ORIGIN-DESTINATION MATRIX ESTIMATOR THAT EXPLOITS STRUCTURE

M. Bierlaire and Ph.L. Toint

Report 93/7(revised) May 26, 1994

Abstract. This paper proposes an improvement of existing methods of origin-destination matrix estimation by an explicit use of data describing the structure of the matrix. These data can be namely obtained from parking surveys. The new model is applied on both illustrative and real examples, and the results are discussed. Comparisons with the results obtained with SATURN/ME2 and the generalized least-squares method are also presented.

Transportation Research Group

Department of Mathematics,

Facultés Universitaires ND de la Paix, Namur, Belgium

Keywords : OD matrix, optimization, parking surveys, transport modelling

1 Introduction

Among the many methodological problems faced by transportation planners, the question of determining the trip demand is one of the most delicate and pervasive. Indeed, in almost all practical applications of transportation modelling, the collection of origin/destination (OD) data is difficult, expensive and therefore very often incomplete. Exploiting all available and reasonably reliable data is hence crucial in this area of activity (see Ben-Akiva, (1987)). It is thus not surprising that many authors have considered the question of estimating an origin/destination matrix from a variety of data sources. These range from home or roadside surveys to historical data and traffic counts, preference being usually given to the latter sources because of their more affordable price.

It is the purpose of this paper to propose an OD-matrix estimation method, called MEUSE, that can explicitly take into account detailed information about the structure of certain columns of the matrix. This is motivated by the availability, in actual practical studies, of surveys giving the origins of vehicles parked in some urban areas from an analysis of their registration plates. The fact that traditional methods ignore and even destroy this information has led us to investigate new alternatives.

After introducing the data in Section 2, we propose in Section 3 a model to take the structure in the data into account. Section 4 is then devoted to practical aspects of solving this model. Its application to an illustrative example is covered in Section 5; results obtained in the framework of a real application are presented and discussed in Section 6. Section 7 contains a brief sensitivity analysis of the model. Section 8 then outlines some possible extensions of the ideas discussed while a final conclusion is suggested in Section 9.

2 Data for OD-matrix estimation

The purpose of this section is to discuss the type of data that is often used for estimating OD-matrices. We distinguish between data commonly considered in most published studies and parking survey data, which is the basic motivation of our proposal.

2.1 Usual data types

A large number of travel demand studies are based on two types of data: a priori matrices and traffic counts.

An *a priori* matrix can be built from past similar studies and from the results of home- and roadside surveys. Such an a priori matrix is unfortunately often unreliable because of the disparity in quality of its parts, resulting from possibly several successive and partial updates. Furthermore, it is commonly the case that assignment of this matrix cannot reproduce observed flows.

Traffic counts are also used in many applications, because they are often already available and because the cost of additional counts is relatively low. It is well-known that even a large number of traffic counts is not enough to determine a unique OD-matrix (see for example Bell, (1983); Cascetta and Nguyen, (1988); Robillard, (1975); Van Zuylen and Willumsen, (1980)).

2.2 Parking surveys

In some recent traffic studies in Belgium (for the cities of Brussels, Charleroi, Liège and Namur), car parks within the city centre were surveyed and the registration plates of vehicles parked therein were sampled (with a typical sampling rate of 20%). Given the necessary administrative permission (in order to preserve parking users' privacy, formal guarantees must be explicitly given on the confidentiality of treatment), it is possible to obtain the addresses of each one of the vehicle owners. This information can then be used for estimating the number and, most importantly, the *spatial distribution* of trips whose end is the considered car park.

The use of this data is however conditional to the following assumptions.

1. Vehicles are driven from the address of their owner directly to the considered car park.
2. The centroid associated with the car park must be such that most trips actually end in the car park, which means that private transportation mode is largely dominant for this centroid. This last condition is of course easier to satisfy if only one transportation mode is studied. In particular, it is automatically satisfied when the only mode considered is the private car.
3. The spatial distribution of arriving vehicles is time independent.

Guarantees that these assumptions are verified for a given car park are of course outside the survey itself, but can be obtained from other sources. For instance, one may know that 80%, say, of trips using private transportation in a city are home-based, as is the case for Namur. When car parks are associated with enterprises, the proportion of trips using alternative modes (such as public transport) can be separately available. It is of course good modelling practice to check these assumptions as much as possible, but this is not the subject of this paper. We will only assume below that parking surveys can be used in the framework just discussed.

There is an additional difficulty in using the data obtained by the parking surveys. The measurements indeed give the vehicles present in the sampled car park at survey time, and not the vehicles that effectively arrived in the parking during the *estimation period*, that is the interval of time for which the OD-matrix is being computed. Typically, parking surveys were performed between 10 and 12 a.m. for an estimation period of one hour within the morning peak period (7h30 to 9h). As a consequence, not every vehicle observed in the survey is relevant to the estimation, but only the *fill-up proportion*

$$f_j = \frac{\text{number of vehicles arriving in car park } j \text{ during the estimation period}}{\text{number of vehicles present in car park } j \text{ at survey time}} \tag{1}$$

of observed vehicles. For example, if a parking is filled in 2 hours and estimation period covers only 1 hour, then the corresponding fill-up proportion should be $\frac{1}{2}$. Using the convention that rows of the OD-matrix are associated with origins and columns with destinations, we note that the parking survey data fixes the *relative* magnitude of the matrix entries in columns corresponding to sampled car parks. The fill-up proportion f_j then determines the absolute values of entries in the j-th column from the number of observed vehicles in car park j at survey time. The fill-up proportions f_j are often unknown and cannot always be collected in the field without substantial effort. We therefore suggest to estimate them when necessary.

3 The MEUSE model for OD-matrix estimation

Given the data and problem, many methodological choices have been proposed. Among the most popular ones, we note the class of log-linear models (entropy maximization or information minimization), as analyzed for instance, in Bell ((1984)) and Van Zuylen and Willumsen ((1980)), Bayesian estimation techniques (Maher, (1983)), maximum likelihood methods (Spiess, (1987)), multi-objective analysis (Brenninger-Göthe, Jörnsten and Lundgren, (1989)) and generalized least-squares algorithms (Bell, (1991); Cascetta, (1984)) (see Bierlaire, (1991) for a survey of these and other techniques). The model that we propose falls in the class of generalized least-squares estimators and is built to handle the parking survey data. It is called MEUSE, for Matrix Estimation Using Structure Explicitly.

3.1 A priori matrix and traffic counts

Part of our proposal follows the classical formulation, where one minimizes a combination of the distances from the new OD-matrix to its a priori estimate and from the observed traffic counts to the assigned flows. More formally, the objective function that we wish to minimize includes terms of the form

$$\sum_{i\in O, j\in D} w_{ij}^t (T_{ij} - t_{ij})^2 + \gamma \sum_{a\in A} w_a^v (V_a - v_a)^2 \tag{2}$$

where

O is the set of potential trip origins in the network,

D is the set of potential trip destinations,

A is the set of arcs for which traffic counts are available,

T_{ij} is the desired entry of the OD-matrix T giving the estimated number of trips from the i-th origin to the j-th destination,

t_{ij} is the a priori known number of trips from the i-th origin to the j-th destination,

V_a is the flow on arc a resulting from the assignment of matrix T on the network,

v_a is the observed flow (traffic count) on arc a,

w_{ij}^t is the relative confidence one has in the value of t_{ij},

w_a^v is the relative confidence one has in the value of v_a,

γ is the global relative weight accorded to traffic counts compared with the a priori OD-matrix t.

Note that the values of $\{V_a\}_{a \in A}$ are determined from the value of T by using the assignment equation

$$\sum_{i \in O, j \in D} p_{ij}^a T_{ij} = V_a \tag{3}$$

for all $a \in A$, where the coefficients p_{ij}^a represent the proportion of the flow from origin i to destination j using the arc a. The coefficients p_{ij}^a are usually obtained by applying assignment techniques. It is assumed, in this context, that they are error free and flow independent. We realize that this assumption is somewhat restrictive and discuss in Section 8 of the paper some possible extension of our methodology to more general situations. Of course, the variables T_{ij} and V_a must be non-negative for our problem to make sense. Formulation (2)–(3) is reminiscent of proposals by Cascetta ((1984)).

3.2 Parking surveys

We now introduce in our model suitable objective function terms whose purpose is to take parking survey data into account. We start by considering the simple case where the spatial distribution of origins for a given car park can be recovered without error. We note that, in this case, the entry T_{ij} of the OD-matrix (where destination j is associated with a surveyed parking) is given by

$$T_{ij} = f_j t_{ij}, \tag{4}$$

where the unknown fill-up proportion f_j is defined by (1) and where we assume that the number of vehicles collected in the relevant car park have been introduced in the j-th column of the matrix t. One often has an estimate of the ratio (1) from the practical organization of parking surveys. This estimate $\tilde{f}_j$ can then be introduced in the model by replacing the first term in (2) by

$$\sum_{i \in O, j \in D \setminus S} w^t_{ij}(T_{ij} - t_{ij})^2 + \sum_{j \in S} w^s_j (f_j - \tilde{f}_j)^2,$$

where S is the index set of the destinations where parking surveys were conducted.

We next note that the situation described by equation (4) is of course idealized. In practice, errors and limited sampling in the survey data make it unlikely that all entries T_{ij} in the same column can be expressed using a single fill-up proportion f_j. More realistically, the relative magnitude of the nonzero entries in column j only approximates the idealized structure, and zero entries should not be taken as strict constraints because they might result from insufficient sampling. As a consequence, we partition the j-th column ($j \in S$) in two sets

$$P_j \stackrel{\text{def}}{=} \{i \mid i \in O \ \text{ and } \ t_{ij} > 0\} \ \text{ and } \ Q_j \stackrel{\text{def}}{=} \{i \mid i \in O \ \text{ and } \ t_{ij} = 0\},$$

which we consider separetely.

Examining the entries with their origin in P_j first, we have that

$$T_{ij} = f_{ij} t_{ij}, \ \text{ with } \ f_{ij} \approx f_j$$

where individual fill-up proportions f_{ij} have now been introduced, but whose values should be reasonably close to the ideal f_j. We have chosen to express this constraint by first adding in the objective function terms of the form

$$w^f_{ij}(f_{ij} - f_j)^2 \quad (j \in S, i \in P_j, w^f_{ij} > 0), \tag{5}$$

while imposing the constraint that the f_{ij} average to f_j in column j, that is

$$f_j = \frac{1}{n_j} \sum_{i \in P_j} f_{ij}, \tag{6}$$

where n_j is the number of entries in P_j. Substituting (6) into (5) shows that we in fact minimize the weighted variance of the positive fill-up proportions f_{ij} around their (idealized) mean f_j.

3.3 The MEUSE model

Gathering all terms, we obtain our new MEUSE model, whose objective function is

$$\begin{array}{llll} \displaystyle\sum_{i\in O,\, j\in D\setminus S} w^t_{ij}(T_{ij}-t_{ij})^2 & + & \displaystyle\gamma \sum_{a\in A} w^v_a (V_a - v_a)^2 & + \\ \displaystyle\sum_{j\in S} w^s_j (f_j - \tilde{f}_j)^2 & + & \displaystyle\sum_{j\in S, i\in P_j} w^f_{ij}(f_{ij} - f_j)^2 & + \\ \displaystyle\sum_{j\in S, i\in Q_j} w^t_{ij} T^2_{ij} & & & \end{array} \tag{7}$$

where

$$V_a = \sum_{i\in O, j\in D\setminus S} p^a_{ij} T_{ij} + \sum_{j\in S, i\in Q_j} p^a_{ij} T_{ij} + \sum_{j\in S, i\in P_j} p^a_{ij} f_{ij} t_{ij}, \tag{8}$$

$$f_j = \frac{1}{n_j}\sum_{i\in P_j} f_{ij} \quad (j\in S), \tag{9}$$

$$T_{ij} \geq 0. \tag{10}$$

4 Solving the model

Once stated, the model must of course be solved. We note that, after substituting (8) and (9) in (7), (7)–(10) represent a large-scale convex quadratic program subject to bound constraints. The solution of such a model is therefore conditional to the availability of numerical software capable of handling large nonlinear optimization. In the experiments described below, we have used LANCELOT, a Fortran package by Conn, Gould and Toint. This package is not specialized for quadratic programs, but aims at solving general nonlinearly constrained problems. It uses an augmented Lagrangian algorithm combined with trust region and specialized data structures. Conjugate gradients are applied in inner iterations in order to (approximately) solve Newton's equations. The reader is referred to Conn, Gould and Toint ((1992b)) for more detail.

4.1 Determination of the weights

Ideally, the choice of each of the weights w^t_{ij}, γ, w^v_a, w^s_j and w^f_{ij} in (7) should reflect the relative confidence one has in the associated data items. We now consider how these weights can reasonably be chosen in practice.

We first note that some of these choices have already been considered in the literature. Indeed the presence of the weights w^t_{ij}, γ and w^v_a in (2) (and thus in (7)) is classical. An attractive approach to determine their values is to identify the weights with elements of the inverse of a dispersion matrix of both the a priori information and traffic counts. We refer the reader to Cascetta ((1984)) for a description of this technique.

Consider now the choice of the weights w^s_j indicating the confidence in $\tilde{f}_j$, the a priori value of the fill-up proportions for car park j. If these values result from actual measurements (countings at the car parks' entrances, for example), the choice of w^s_j may be handled as that of the w^v_a. But it may happen that such measurements are unavailable and that little is known about the true values of the fill-up proportions. In this case, we suggest to choose a relatively low weight value for w^s_j.

We now turn to the problem of choosing a value for the weights w^f_{ij} of (7). At variance with those considered in the previous paragraph (reflecting the confidence in collected data), these weights instead reflect the confidence one has in a *structural assumption*, namely that the parking surveys provide correct information on trip origins. Unfortunately, the authors are unaware of any statistical technique that mixes both types of confidence (on data and assumptions), which implies that a specific procedure should be suggested. If we denote by m_{ij} the proportion of trips from i to j such that the three basic assumptions of Section 2.2 hold and by p the sampling rate in the parking surveys, we may then define

$$\hat{b}_{ij} = m_{ij} t_{ij} = \frac{m_{ij}}{p} t'_{ij}, \tag{11}$$

an estimator of b_{ij}, the true number of trips from i to j satisfying our basic assumptions. In this equation, t'_{ij} is the observed number of vehicles in car-park j registered at origin i. We assume that this data is extracted from an hypergeometric distribution

$$t'_{ij} = \mathcal{H}(D_j, t^*_{ij}, pD_j) \tag{12}$$

where t^*_{ij} is the ij-th element of the (unknown) true OD-matrix and where $D_j = \sum_{i \in O} t^*_{ij}$ is the (observable) number of vehicles present in car park j. Defining now m_j the proportion of trips to car park j such that our basic assumptions hold, i.e.

$$m_j = \frac{\sum_{i \in O} m_{ij} t_{ij}}{\sum_{i \in O} t_{ij}},$$

we may view

$$\hat{f}_{ij} = \tilde{f}_j \frac{\hat{b}_{ij}}{m_j t_{ij}} \tag{13}$$

as an estimate of the individual fill-up proportions f_{ij} taking into account the dispersion of these fill-up proportions around their mean and the a priori approximation of this mean. We then suggest to choose the weights w^f_{ij} as the inverse of the variance of this estimate, that is, using (11), (12) and (13),

$$w^f_{ij} = \frac{m_j^2 \, p \, t_{ij}(D_j - 1)}{\tilde{f}_j^2 m_{ij}^2 (1-p)(D_j - t_{ij})}. \tag{14}$$

This proposal seems adequate when more detailed information is not available: it includes the effects of sampling, the confidence in the assumptions and the dispersion around the assumed distribution of origins.

4.2 Underdeterminacy

One of the difficulties that appears in the practical solution of (7)–(10) is the fact that, for some otherwise unconstrained entries T_{ij}, the a priori value t_{ij} is unknown. Let us denote $Z = \{(i,j) \mid t_{ij} \text{ is unknown } \}$. If the terms indexed by Z in the first term of (7) are neglected, the values of these entries are "floating" and the minimization problem is structurally singular. This causes convergence of LANCELOT to be very slow. To circumvent this problem, we have chosen to take a value of 1 for each unknown t_{ij} with a relatively low associated weight w^t_{ij}. This choice is similar to that made by entropy methods for these entries, where the typical term (see Ortúzar and Willumsen, (1990)) $T_{ij}\ln(T_{ij}/t_{ij}) - T_{ij}$ is minimized for $t_{ij} = 1$.

4.3 Problem scaling

When applying minimization software like LANCELOT, it is useful to scale all variables and constraints such that their sizes are comparable. The technique we have used to achieve this goal is to use the scaled variables $\hat{T}_{ij} = T_{ij}/t_{ij}$, $\hat{V}_a = V_a/v_a$ and $\hat{f}_j = f_j/\tilde{f}_j$. The weights are suitably adjusted.

5 An illustrative example

Before presenting the results obtained with MEUSE model on real data, we first discuss a small illustrative example.

We consider the network of Figure 1 and assume that the true OD-matrix is given by Table 1. The a priori matrix is defined as a multiple ϵ of the true matrix. This choice ensures that the relative sizes of the cells are correct. Traffic counts are computed without error from the assignment of the true matrix on the network. It is assumed that an all-or-nothing assignment is performed, so that 0 and 1 are the only possible values for the coefficients p^a_{ij}. The flow on each arc and the list of non zero p^a_{ij} are listed in Table 2.

We first apply the MEUSE model without parking surveys, that is with $S = \emptyset$. In this case, the model is a classical generalized least-squares (GLS) estimator. We have chosen the weights in

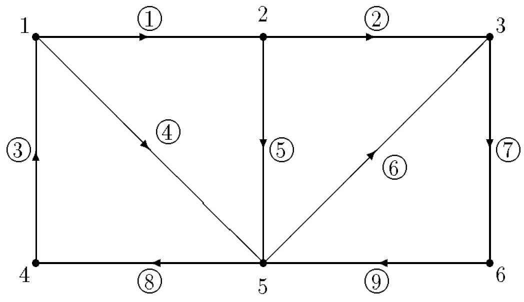

Figure 1: The network for the small example

	1	2	3	4	5	6
1		120.0	100.0	50.0	20.0	25.0
2	100.0		90.0	70.0	30.0	30.0
3	240.0	200.0		60.0	20.0	70.0
4	60.0	510.0	80.0		60.0	150.0
5	180.0	90.0	300.0	60.0		20.0
6	280.0	160.0	90.0	40.0	20.0	

Table 1: The true OD-matrix (T_{ij})

Arc	Flow	(i,j) such that $p_{ij}^a = 1$
1	1460	(1,2) (1,3) (1,6) (3,2) (4,2) (4,3) (4,6) (5,2) (6,2)
2	500	(1,3) (1,6) (2,3) (2,6) (4,3) (4,6)
3	2110	(2,1) (3,1) (3,2) (4,1) (4,2) (4,3) (4,5) (4,6) (5,1) (5,2) (6,1) (6,2)
4	130	(1,4) (1,5) (4,5)
5	200	(2,1) (2,4) (2,5)
6	410	(5,3) (5,6) (6,3)
7	840	(1,6) (2,6) (3,1) (3,2) (3,4) (3,5) (3,6) (4,6) (5,6)
8	1530	(1,4) (2,1) (2,4) (3,1) (3,2) (3,4) (5,1) (5,2) (5,4) (6,1) (6,2) (6,4)
9	1110	(3,1) (3,2) (3,4) (3,5) (6,1) (6,2) (6,3) (6,4) (6,5)

Table 2: Nonzero assignment coefficients

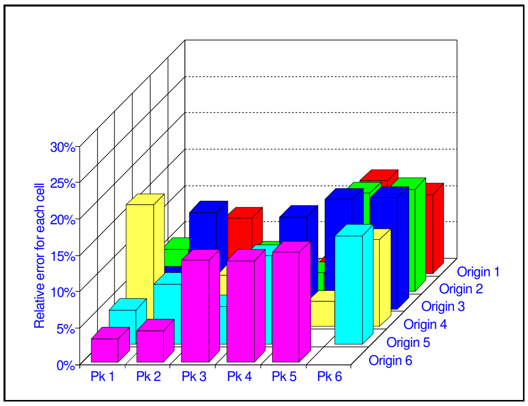

Figure 2: Errors with GLS estimator

the model reflecting our accurate knowledge of the perturbations to the true data. More precisely,

$$
\begin{aligned}
w_{ij}^t &= 1/t_{ij} \quad ((i \in O, j \in D),\\
\gamma w_a^v &= 1000 \quad (a \in A),
\end{aligned}
$$

for a value of $\epsilon = 1.15$. Indeed, since the perturbation of the a priori matrix is uniform, the weight should be inversely proportional to its value. However, the countings are exact, which theoretically imposes to choose infinite weights. We have chosen 1000 to avoid severe numerical difficulties. The relative errors between the true values of Table 1 and the results obtained are shown in Figure 2.

We immediately note the lack of structure in the errors. This is expected because the model used first aims at reproducing the traffic counts. As each cell is allowed to vary independently, the structural information present in the a priori matrix is lost.

We next apply the MEUSE model with parking surveys at nodes 1, 2 and 3. This is done without modifying the t_{ij} (the first three columns of the matrix then represent the parking surveys), but by defining $S = \{1, 2, 3\}$. The weights w_{ij}^f are set to 1000, for reasons identical to those described above for γw_a^v. We have furthermore assumed that the priori fill-up proportions are unknown: we have chosen an arbitrary value of $\tilde{f}_j = 1$ $(j \in S)$ with a very low weight. Again, this weight should theoretically be chosen as zero, but this choice would generate a singular

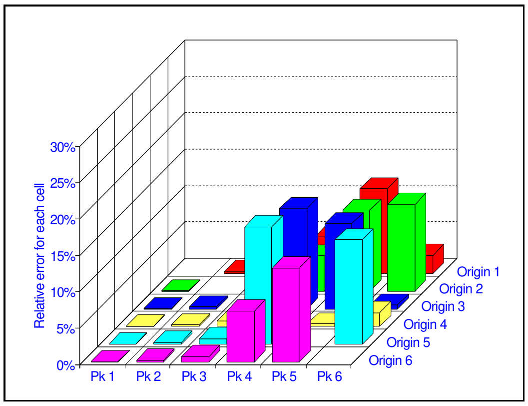

Figure 3: Errors with MEUSE (parking surveys at nodes 1, 2 and 3)

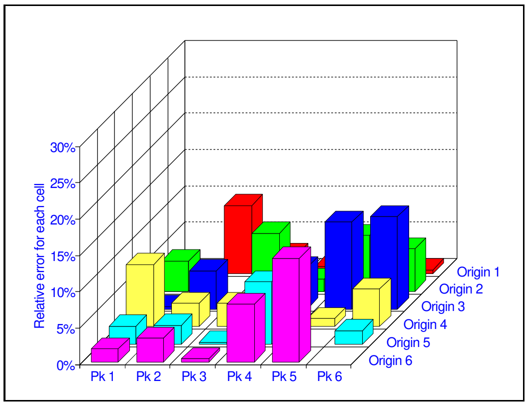

Figure 4: Errors with SATURN/ME2 estimator

estimation problem. We have set $w_j^s = 0.001$ ($j \in S$) to avoid this difficulty. The relative errors obtained are presented in Figure 3.

The situation has clearly improved. As anticipated, the errors corresponding to the destinations 1, 2 and 3 almost vanish. The other errors also decrease significantly.

We finally apply the matrix estimator SATURN/ME2 (Van Vliet, (1982)) to this example. The corresponding relative errors are plotted in Figure 4. As is the case for the GLS estimator, the structure of the matrix is lost. The MEUSE model therefore seems an attractive alternative, at least on this small example.

As a last exercise, we compute the cross-sensitivity of the total error as a function of the number of columns in S and perturbation of the a priori matrix. We apply the MEUSE model

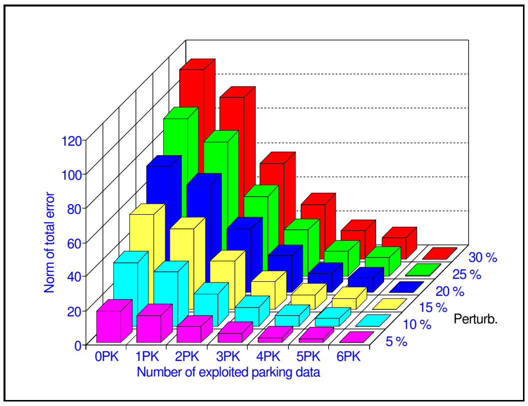

Figure 5: Global error as a function of the amount of structure used and perturbation size

for the choices $\epsilon = 1.05$, 1.10, 1.15, 1.20, 1.25 and 1.30 (corresponding to perturbations of 5, 10, 15, 20, 25 and 30%, respectively), each time using from 0 (i.e. the GLS estimator) to 6 columns in S. The resulting values for the error

$$\sqrt{\sum_{i,j=1}^{6} (T_{ij} - t^*_{ij})^2}$$

are illustrated in Figure 5, where we clearly see that the impact of a given perturbation level decreases with the amount of underlying structure used.

6 Application to a real case study

We next describe the application of the MEUSE model within a practical OD-matrix calculation for the city of Namur (Belgium). As for the illustrative example, we also ran the GLS and SATURN/ME2 estimators on this problem for comparison.

We emphasize that our purpose, in this comparison, is to show that all three models behave differently. This is therefore not a complete application exercise, where several iterations between demand estimation and assignment would typically be performed.

6.1 The problem

The network under study has 106 centroids, all situated in the city centre (the "Corbeille"). Available data for this estimation consists in

1. a partial a priori matrix obtained from populations and traffic counts on the boundary of the studied area,
2. a set of 63740 coefficients p_{ij}^{a} resulting from the equilibrium assignment of the a priori demand t on the network, using the SATURN model,
3. a set of 146 traffic counts, both from automatic cable counters and manual data collection,
4. a set of 60 parking surveys.

This data was collected and prepared for the Wallonie Regional Government by STRATEC, a specialized consultancy firm, and the obtained results were of direct interest to both STRATEC and the Regional Government. We note that, according to STRATEC, little confidence can be put in the a priori matrix and fill-up proportions.

6.2 Results for MEUSE

Incorporating this set of data in the MEUSE model yields a bound constrained minimization program with 11276 variables (10542 T_{ij}, 146 V_a and 588 f_{ij}). In order to reflect the reliability of our data sources for the problem, the various weights were chosen as $\gamma w_a^v = 1/v_a$, $(a \in A)$ and $w_{ij}^t = 0.001$ $((i,j) \in Z)$. The first of these choices corresponds to assuming a Poisson distribution on the flows and the second to the suggestion of Section 4.2. The w_{ij}^f $(i \in O, j \in S)$ are chosen according to (14) with the values $m_j = m_{ij} = 0.8$ and $p = 0.2$ obtained from external sources. Finally, the suggestions of Cascetta ((1984)) could not be applied to the a priori matrix because it does not result from an OD survey, and we have chosen to reflect the relatively poor quality

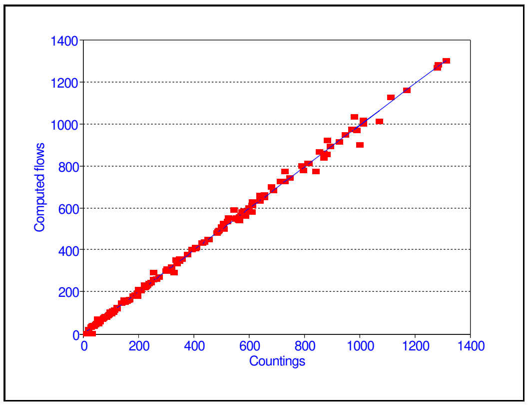

Figure 6: Flow fit for the MEUSE model on Namur

of this data by setting $w^t_{ij} = 1/t^2_{ij}$ $(i \in O, j \notin S)$. A similar choice was made for the fill-up proportions, where $w^s_j = 1/\tilde{f}^2_j$ $(j \in S)$.

Of course, the true values of the matrix entries are unknown, and we can only measure accuracy of the result indirectly: we illustrate, in Figure 6, the fitting of the flows resulting from the assignment of the computed demand on the network with the observed traffic counts. In this figure, the abscissa of each square corresponds to an observed flow while its vertical coordinate is the corresponding computed value. The ideal situation would be to have all squares on the diagonal. The small deviation allowed by MEUSE in order to take measurement errors into account explain the slight dispersion around this ideal curve, but the results are very satisfactory.

Figure 7 illustrates the distribution of the fill-up times associated with the 60 parkings considered in this study. It shows realistic fill-up times for more than 70% of the cases. Indeed, it is known from other sources that the majority of fill-up times should fall between 0 and 3 hours. The 15% of outliers are easily explained because the model's assumption do not hold well for the corresponding surveys: in particular, they involve on-street parking for which sampling is more difficult.

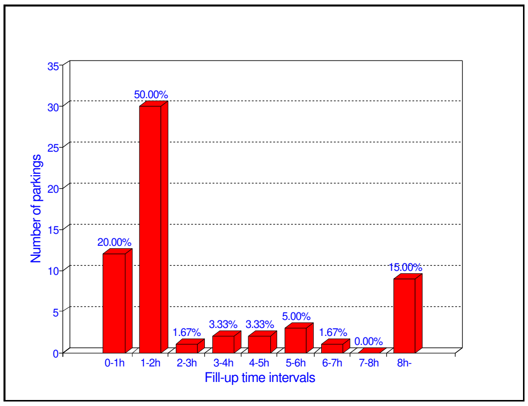

Figure 7: Fill-up times for the 60 parkings in Namur

6.3 Comparison with GLS and SATURN/ME2

As for the illustrative example, we also tested SATURN/ME2 and GLS on our real application. At variance with the example, not all entries in the a priori matrix for the real case are non-zero. Since ME2 is a multiproportional estimator, it provides a facility to "seed" zero entries in order to allow them to leave zero. We have tested ME2 on our problem, both with and without this facility (the seed value was chosen to be 1.0). Needless to say, both GLS and ME2 (in both versions) provide adequate fit between estimated and observed flows.

We first analyze the total number of trips computed by the four estimators. This number is pictured in Figure 8, where a further disaggregation between parking related and other entries is also shown.

Interestingly, ME2 gives, in its two variants, the lowest number of trips, while the largest is produced by GLS, the difference being mostly for matrix entries associated with the parking surveys, that is entries in the set $P \stackrel{\text{def}}{=} \{(i,j) \mid j \in S \text{ and } i \in P_j\}$. MEUSE and GLS produce substantially different results, although they are based on a similar philosophy: they indeed mostly differ by the special treatment applied within MEUSE to the 588 entries (5.2%) of T. This is apparent when examining, in Figure 8, the distribution of trip numbers for these entries and for the rest of the matrix.

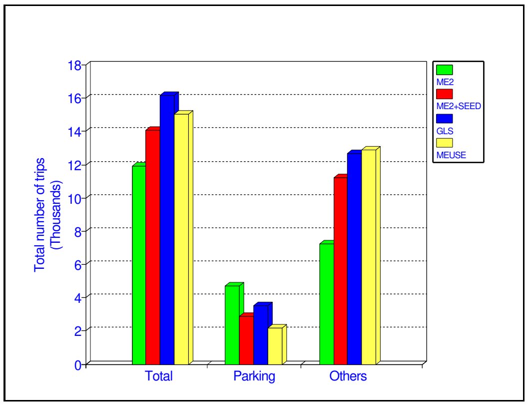

Figure 8: Number of estimated trips

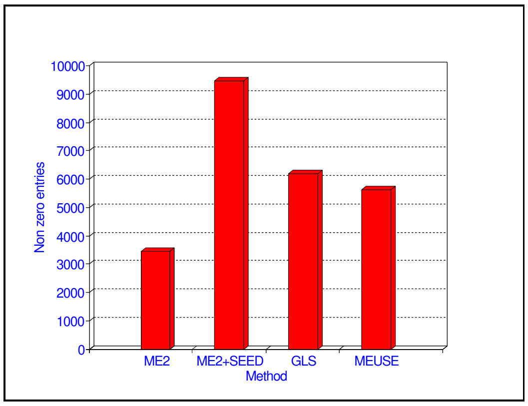

Figure 9: Number of nonzero entries

We finally examine the aggregated numbers of nonzero entries in the matrix. The results are shown in Figure 9 for the four estimators. The two versions of ME2 present an extreme behaviour, which can easily be understood by the multiplicative nature of these technique and the presence/absence of seeds in zero entries of the a priori matrix. GLS and MEUSE appear to provide a compromise between these extremes.

7 Sensitivity

Due to the large amount of model data and parameters, it is also important to examine the stability of the proposed method with respect to variations in these parameters. We thus carried out some tests whose purpose is to measure this sensitivity. Various classes of model parameters and data were successively perturbed by a random value drawn from a Gaussian distribution of zero mean and variance equal to 10% of the perturbed quantity. The results of the MEUSE model were recomputed for each perturbed problem. The relative differences between the results corresponding to the perturbed and unperturbed problems were then measured, both in ℓ_2 and ℓ_∞ norms. They are reported in Table 3. More precisely, the values quoted in this table are defined by

$$\frac{\|T-T^*\|_p}{\|T\|_p} = \begin{cases} \frac{\sqrt{\sum_{ij}(T_{ij}-T^*_{ij})^2}}{\sqrt{\sum_{ij}T^2_{ij}}} & \text{if } p=2 \\ \\ \frac{\max_{ij}(|T_{ij}-T^*_{ij}|)}{\max_{ij}T_{ij}} & \text{if } p=\infty \end{cases}$$

where T is the matrix estimated with the original model and T^* is the matrix computed with the perturbed problem.

We immediately note the relatively low sensitivity of the model with respect to its internal parameters, as indicated in the first five lines of Table 3. The fact that MEUSE is more sensitive to perturbations in its data than in its parameters is a good feature of this model. As expected,

Perturbed quantities	$p = 2$	$p = \infty$
γ	$1.09\ 10^{-2}$	$6.31\ 10^{-3}$
$w_j^s \quad (j \in S)$	$1.35\ 10^{-2}$	$1.12\ 10^{-2}$
$w_{ij}^f \quad (j \in S, i \in P_j)$	$3.97\ 10^{-3}$	$3.54\ 10^{-3}$
$w_{ij}^t \quad (i \in O, j \in D \setminus S, t_{ij}$ known)	negligible	$1.03\ 10^{-4}$
$w_{ij}^t \quad (i \in O, j \in D \setminus S, t_{ij}$ unknown)	negligible	$1.03\ 10^{-4}$
$\tilde{f}_j \quad (j \in S)$	$2.84\ 10^{-2}$	$2.13\ 10^{-2}$
$v_a \quad (a \in A)$	$3.49\ 10^{-1}$	$2.70\ 10^{-1}$
$t_{ij} \quad (j \in S, i \in P_j)$	$1.66\ 10^{-2}$	$1.14\ 10^{-2}$
$t_{ij} \quad (i \in O, j \in D \setminus S)$	$9.59\ 10^{-2}$	$1.40\ 10^{-1}$

Table 3: Results of the sensivity analysis for 10% perturbations

the results are most sensitive to variations in the traffic counts.

We conclude this section by noting that the application of MEUSE on both the illustrative example and the real problem indicates its potential and seems to assess its applicability.

8 Perspectives

Amongst further extensions of our model, we also note the following possibilities. A more general OD matrix structure than that arising from parking surveys could also be exploited by the same approach. For instance, one could consider on-street parking whose associated destination is the *set* of neighbouring centroids.

Although theoretically possible, the inclusion of flow dependent path flow proportions p_{ij}^a in the model leads to an extremely large minimization problem, which is (at least for now) unrealistic. As is the case with other demand estimators based on fixed path flow proportions, one can instead iterate between demand estimation and equilibrium assignment. The convergence of this scheme seems possible because equilibrium assignment is continuous as a function of travel demand (see, for instance, Fiacco, (1983)), but should be confirmed by a dedicated analysis.

Improving the numerical algorithm for our model solution is also of considerable interest. Although the use of LANCELOT is suitable for exploratory purposes, this tool is far too general for the problem at hand. A specialized algorithm is then expected to bring substantial efficiency gains. We could, for instance, exploit the fact that MEUSE results in a large-scale sparse minimization problem with quadratic objective subject to simple bounds, and apply special purpose techniques (see, for example, Bierlaire, Toint and Tuyttens, (1991) or the Harwell Subroutine VE14 by Gould based on Conn, Gould and Toint, (1992a)).

9 Conclusions

We have introduced MEUSE, a new OD estimation method that can take matrix structure into account, in particular when this structure is obtained from parking survey data. The behaviour of the model has been analyzed both on a simple illustrative example and in a real application. The results obtained are coherent with what can be expected for the method and indicate the nature of the methodological improvement. They also show that the new approach can effectively be applied in realistic contexts.

Some extensions of the MEUSE model have been pointed out, including further use of the OD matrix structure and improved computational procedures. They are the subject of ongoing research and will be reported on elsewhere.

More globally, the method and results presented in this paper indicate that parking survey data can indeed be taken into consideration when estimating travel demand in an urban context.

10 Acknowledgments

The authors are indebted to STRATEC, in particular to H. Duchâteau and M.A. Cousin, for bringing the problem to their attention as well as for providing real data and the facilities for the tests with SATURN. Useful comments of two referees and helpful discussions with M. Remon were also appreciated. The financial support granted to the first author by the AGIR program of the Région Wallonne is finally gratefully acknowledged.

References

Bell, M. G. H. (1983). The estimation of an origin-destination matrix from traffic counts, *Transportation Science* **17**(2): 198–217.

Bell, M. G. H. (1984). Log-linear models for the estimation of origin-destination matrices from traffic counts: an approximation, *in* J. Volmuller and R. Hamerslag (eds), *Proceedings ninth international symposium on transportation and traffic theory*, VNU Science Press, Utrecht, Netherlands.

Bell, M. G. H. (1991). The estimation of origin-destination matrices by constrained generalised least squares, *Transportation Research, Series B* **25**(1): 13–22.

Ben-Akiva, M. (1987). Methods to combine different data sources and estimate origin-destination matrices, *in* N. H. Gartner and N. H. M. Wilson (eds), *Transportation and traffic theory*, Elsevier, New-York, pp. 459–481. Proceedings of the Tenth International Symposium on Transportation and Traffic Theory, July 8–10 1987, MIT, Cambridge, Massachusetts.

Bierlaire, M. (1991). Evaluation de la demande en trafic : quelques méthodes de distribution, *Annales de la Société Scientifique de Bruxelles* **105**(1-2): 17–66.

Bierlaire, M., Toint, P. L. and Tuyttens, D. (1991). On iterative algorithms for linear least squares problems with bound constraints, *Linear Algebra and Applications* **143**: 111–143.

Brenninger-Göthe, M., Jörnsten, K. O. and Lundgren, J. T. (1989). Estimation of origin-destination matrices from traffic counts using multiobjective programming formulations, *Transportation Research, Series B* **23**(4): 257–269.

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data : a generalised least squares approach estimator, *Transportation Research, Series B* **18**(4/5): 289–299.

Cascetta, E. and Nguyen, S. (1988). A unified framework for estimating or updating origin/destination matrices from traffic counts, *Transportation Research, Series B* **22**(6): 437–456.

Conn, A. R., Gould, N. I. M. and Toint, P. L. (1992a). A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds, *Technical Report 92/07*, Department of Mathematics, FUNDP, Namur, Belgium.

Conn, A. R., Gould, N. I. M. and Toint, P. L. (1992b). LANCELOT*: a Fortran package for large-scale nonlinear optimization (Release A)*, number 17 in *Springer Series in Computational Mathematics*, Springer Verlag, Heidelberg, Berlin, New York.

Fiacco, A. V. (1983). *Introduction to sensitivity and stability analysis in nonlinear programming*, Academic Press, New York.

Maher, M. J. (1983). Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach, *Transportation Research, Series B* **17**(6): 435–447.

Ortúzar, J. D. and Willumsen, L. (1990). *Modelling Transport*, J. Wiley and Sons, Chichester (England).

Robillard, P. (1975). Estimating an O-D matrix from observed link volumes, *Transportation Research* **9**: 123–128.

Spiess, H. (1987). A maximum likelihood model for estimating origin-destination matrices, *Transportation Research, Series B* **21**(5): 395–412.

Van Vliet, D. (1982). SATURN, a modern assignment model, *Traffic Engineering and Control* **23**: 578–581.

Van Zuylen, H. J. and Willumsen, L. G. (1980). The most likely trip matrix estimated from traffic counts, *Transportation Research, Series B* **14**: 281–293.