Gravity Model of Trip Distribution: an Empirical Contribution

Jorge Cabrera-Delgado*

June 18, 2010

Abstract

Les approches agrégées dans la modélisation des déplacements urbains ignorent souvent, dans les exercices de prospective, les évolutions démographiques et comportementales pouvant avoir lieu dans les périodes étudiées. À partir des trois dernières enquêtes ménages déplacements réalisées sur l'agglomération lyonnaise (1985, 1995 et 2006), nous explorons les changements dans, d'une part, la structure de la population de l'agglomération et, d'autre part, les pratiques de mobilité quotidienne.

1 Introduction

The main goal of the research presented hereby is to evaluate the impact of changes in the urban system[1] on gravity model of trip distribution (GMTD) parameter values.

In the practice of urban planning, the GMTD is a well established tool, as a part of the classic four step sequence of transport demand modeling . Arguably, its longevity owes much to its simplicity — the analogy with the law of universal gravitation has served well as an explanatory tool even before more solid theoretical justifications (Wilson, 1967; Snickars and Weibull, 1977; Anas, 1983) had actually been introduced — and to its relatively good performance in terms of forecasting, even in comparison with newer approaches like the use of Neural Networks (Mozolin et al., 2000).

And precisely, when it comes to forecasting, it is common practice to calibrate the models on cross-sectional data (mostly on data from the last mobility survey available) and then make the hypothesis that the parameter(s) of the model are constant over time. In theory, it is quite possible to use data from several surveys, when these are available. However, the exercise is not undertaken because it is difficult and costly to model transportation networks at different stages in time. This reconstruction is therefore conducted once for a baseline year, which is frequently the year of completion of the last mobility survey, or at least the year for which most data is available. It is this same reason which has, to the best of our knowledge, made difficult to properly put

---

*Laboratoire d'Économie des Transports – Université Lumière Lyon 2 – École Nationale des Travaux Publics de l'État

[1] As described by Bonnafous and Puel (1983), the urban system is formed by three subsystems: location of activities, transportation network and social interactions

to the test the hypothesis of parameter stability over time (Bonnel, 2004). In fact, Southworth (1979, 1983) showed that the parameters of the gravity model are highly sensitive to the measure of time or distance used in calibration: the consequence of this is that, in the absence of a network modeled consistently over time, the shifts in the values of the parameters do not reflect solely the changes in mobility behavior, as they should, but instead they merely reflect the inconsistencies between time measures (distance measures are obviously more consistent as they generally are straight-line, *i.e.* euclidean, measures between zone centroids, so they basically don't change over time). It is well known, for instance, that declared travel time from surveys is highly inconsistent from one survey to another.

In this paper, we argue that comparing A quick review of the literature on the GMTD is useful to highlight an interesting fact: most of the research deals with *theoretical* and *methodological* considerations — like its theoretical foundations , different calibration techniques (Hyman, 1969; Evans, 1971; Kirby, 1974; Williams, 1976) or the deterrence function form (Evans and Kirby, 1974) — while *empirical validation* ones have been given relatively much less attention — it is important, however, to acknowledge the contributions of Openshaw (1976), Southworth (1979, 1983), Duffus et al. (1987) and Mikkonen and Luoma (1999), among others. A quick review of the literature ait véritablement été proposée. des notamment par les approches agrégées de modélisation des déplacements urbains utilisées font que l'on ignore souvent les évolutions démographiques et comportementales pouvant avoir lieu dans les périodes étudiées. Le modèle à quatre étapes classique continue d'être le principal outil dès lors que l'on se place dans un horizon temporel de d'étude de moyen terme (environ 10 ans) et que l'on cherche, par exemple, à prévoir le trafic d'une nouvelle infrastructure (dimensionnement ou évaluation) ou pour simuler l'évolution des déplacements sur un territoire dans le cadre d'études de planification (PDU, DVA, SCOT).

2 A disaggregated trip distribution model

We estimate the doubly constrained tour based GMTD, with a Tanner (1961) deterrence function, given in equations (1) to (3):

$$T_{ij}^p = A_i^p O_i^p B_j^p D_j^p c_{ij}^{\alpha^p} \exp\left(\beta^p c_{ij}\right) \tag{1}$$

$$A_i^p = \left(\sum_j B_j^p D_j^p c_{ij}^{\alpha^p} \exp\left(\beta^p c_{ij}\right) \right)^{-1} \tag{2}$$

$$B_j^p = \left(\sum_i A_i^p O_i^p c_{ij}^{\alpha^p} \exp\left(\beta^p c_{ij}\right) \right)^{-1} \tag{3}$$

where

T_{ij}^p is the number of *tours* (Axhausen, 2000; Bonnel, 2004, chapter 6) — we define a *tour* as the sequence of trips related to the activities made by a person, from the moment she leaves home to the moment she returns — originating from zone i and having the *dominant* purpose p located in zone j. The *dominant* purpose of a *tour* is defined in a hierarchical way, from the "most constraining" activity to the "least constraining" one. Here, $p \in$ {Work, Elementary School, Junior & High School, College, Shop, Escort, Recreation, Other} (given in order);

O_i^p is the number of *tours* originating from zone i, for the *dominant* purpose p;

D_j^p is the number of *tours* having j as the location of the *dominant* purpose p;

c_{ij} is the cost of traveling from i to j (for this study we use the generalized by car travel time).

α^p and β^p are the cost decay parameters, to be estimated.

A preliminary study lead us to discard the use of a GMTD with an exponential deterrence function. The use of a single parameter would have, *a priori*, made easier the interpretation of the changes over time, but the calibration showed poor results on the distribution of the number of tours for each generalized time range: an underestimation of the number of short tours (around 5 min.) and an overestimation of the number of average-long ones (around 30 min.).

3 Data sources, perimeter and zoning definition

The calibration of a GMTD requires two types of data: the first type, which we could call *demand data*, is used to define the observed (or reference) O-D matrices $\left(N_{ij}^p\right)$. The second type, which we could call *network performance data*, defines the distance, the generalized time of travel here, between the origins and destinations of the study area, *i.e.* the (c_{ij}) matrix.

3.1 The demand data

The main *demand data* sources for this study are three mobility surveys (1985, 1995 and 2006) conducted in Lyon: the *enquêtes ménages déplacements* or EMD, which are one of the essential tools for the study of mobility patterns and their evolution in France. Furthermore, the compliance of the surveys to the "standard Certu" (Certu, 2008) method ensures comparability of the results over time. The principles of the "standard Certu" method are:

- The EMD aim at retracing the mobility patterns of the region inhabitants for a regular business day, excluding holidays.
- They are made at people's residences, by interviewers specially trained for this type of collection. All the persons over five years living in the dwelling are interviewed.
- All the trips made the day before the survey by each person surveyed are identified. The characteristics of each trip — purpose, modes, origin, destination, time of departure and arrival — are collected.
- The EMD cover a representative sample of households, drawn randomly by area of residence from a housing data base. The housing must be

	Nb. of persons		Nb. of trips		Nb. of tours	
EMD/Perim.	surv.	total	surv.	total	surv.	total
85/85	11,449	1,017,893	39,999	3,541,776	16,773	1,483,328
95/85	13,579	1,138,161	51,562	4,433,941	20,791	1,769,452
95/95	13,997	1,195,190	53,213	4,659,783	21,463	1,861,249
06/85	13,586	1,146,520	48,461	4,150,497	19,713	1,691,085
06/95	14,523	1,209,987	52,292	4,406,054	21,135	1,786,625
06/MAL	20,302	1,537,593	74,736	5,685,298	30,061	2,291,483
06/06	25,656	1,839,251	96,250	6,905,183	38,383	2,765,611

Table 1: Source : EMD Lyon 1985, 19985 et 2006.

designated as the primary residence of the household. The sample size is determined so as to ensure a minimum reliability of the results for an analysis by sector.

With every new EMD, the the scope of the survey has been extended. The perimeter of the 1985 EMD contains 71 municipalities, that of 1995 EMD has 99 and that of 2006 includes 453. The metropolitan area — defined by INSEE[2] as “a set of municipalities with no enclave or discontinuity, formed by an urban center surrounded by rural or urban units (suburban crown), with at least 40 % of their residents having a job in the center or in the municipalities attracted by it” — of Lyon (MAL) has 294 municipalities[3]. Each individual in the survey is given a sample-to-universe expansion coefficient, on the basis of which an estimation can be made of the total average daily trips or tours. Table 1 contains information on the size and mobility of the population surveyed, by EMD and perimeter.

For legal reasons, the EMD are not geocoded and they have different zoning systems. As shown by Openshaw (1977) and (Briant et al., 2010), among others, the calibration results of the GMTD are sensitive to the zoning system used. To get comparable results over time, it was necessary then to establish a common zoning system: we used the smallest common zoning system available, which is incidentally formed by the municipalities around Lyon and the *arrondissements* inside of it. Thus, the 1985 EMD perimeter has 79 traffic analysis zones (TAZ), the 1995 EMD perimeter has 107 TAZ, the MAL has 302 TAZ and the 2006 EMD perimeter has 461 TAZ[4] (figure 1).

By examining the temporal evolution of the population within a given perimeter, we can observe that the annual population growth rate inside the perimeter of the 1985 EMD is of approximately 1.12 %, between 1985 and 1995, and 0.07 %, between 1995 and 2006. In the scope of the 1995 EMD, population has grown at a rate of 0.11 % between 1995 and 2006. This suggests that the population increases faster in the suburbs than it does downtown.

---

[2] Institut National de la Statistique et des Études Économiques (National Institute of Statistics and Economic Studies)

[3] Actually, the MAL includes 296 municipalities of which 294 are also within the perimeter of the 2006 EMD. The remaining two districts, Dragoire and Tartaras, belong to the “departement” Loire (42) and are thus outside the perimeter of the 2006 EMD. In what follows, any reference to the perimeter of the MAL will refer to the 294 municipalities also included in the 2006 EMD.

[4] As we’ll see later on, the *network data* only covers the MAL perimeter so the rest of the 2006 EMD perimeter is not treated by our study.

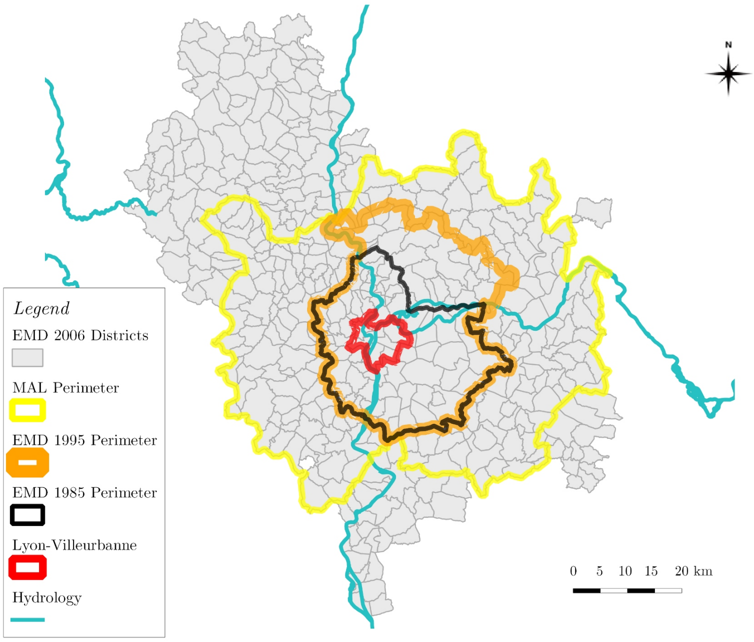


Figure 1: Municipalities and "arrodissements" which were used as TAZ, by perimeters of the different EMD's and MAL.

In terms of mobility, we observe temporal and spatial stability: taking into account only the residents of the 1985 EMD perimeter, the number of trips per person in 1985 was 3.5, 3.9 in 1995 and 3.6 in 2006 whereas the number of tours per person was 1.5 in 1985 and 2006, and 1.6 in 1995. For the 2006 EMD, the number of trips per resident of the 1995 EMD perimeter was 3.6 and 3.7 for the residents of the MAL; the number of tours per person is steady over space at 1.5.

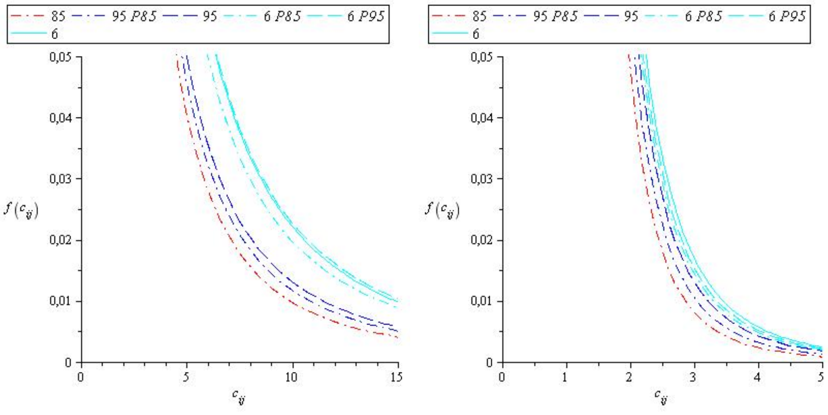
3.2 The network performance data

The *network performance data*, *i.e.* the (c_{ij}) matrices, was made available to us by a land-use transport interaction research program developed in the Lyon conurbation[5]. The generalized time matrices were established for transport in individual vehicles (cars) by assigning demand matrices — containing the total traffic of the study area, including trucks and exchange traffic with the exterior of the perimeter, constructed using data from the EMD and traffic counts — for 1985, 1995 and 2006 in modeled networks of the corresponding years (Godinot et al., 2008).

The generalized time matrices used were actually "averaged" from the matrices we were given (which were established for a much more detailed zoning system with 777 TAZ covering the MAL perimeter). As the procedure of aver-

---

[5] SIMBAD is a project developed by the Transport Economics Laboratory. It aims at evaluating the long term (25 years)impacts of different transport policies, from a sustainable development perspective.



(a) Work

(b) Shopping

Figure 2: Tanner deterrence function calibration results by trip purpose. $f(c_{ij}) = c_{ij}^{\alpha} \exp(\beta c_{ij})$.

aging matrices is the same for all years, it should not impact on the study of temporal evolution.

4 Calibration results

The estimation is done using a maximum likelihood (ML) approach. As shown by Kirby (1974), and later Erlander and Stewart (1990) and Sen and Smith (1995), ML estimates of the GMTD decay parameters can be obtained by ensuring that the constraints given by equations (4) to (7) are satisfied:

$$\sum_j T_{ij}^p = O_i^p, \tag{4}$$

$$\sum_i T_{ij}^p = D_j^p, \tag{5}$$

$$\sum_{ij} c_{ij} T_{ij}^p = \sum_{ij} c_{ij} N_{ij}^p, \tag{6}$$

$$\sum_{ij} \log(c_{ij}) T_{ij}^p = \sum_{ij} \log(c_{ij}) N_{ij}^p \tag{7}$$

where N_{ij}^p is the number of observed *tours* originating from zone i and having the *dominant* purpose p located in zone j. In order to solve the non-linear system described by equations (4) to (7), we use the “Modified Scoring Procedure” proposed by Sen and Smith (1995) (chapter 5).

Following Southworth (1983), it seemed interesting to start by showing some obvious contrasts between the variables of the model for all the trip purposes studied (table 2).

Il est interessant de noter que les relocalisations des forces generatrices de boucles ont été relativement peu importantes entre 1985 et 2006

Here, we need to be more precise about the magnitude in the increase of travel distance. Figure 3 shows that, if we consider the tours made in the same

Perimeter:	1985			1995
	85–95	85–06	95–06	95–06
Variable: O_i^p				
Work	0.977	0.962	0.977	0.979
Elementary School	0.855	0.855	0.907	0.910
Junior & High School	0.952	0.933	0.936	0.939
College	0.841	0.772	0.897	0.902
Shopping	0.960	0.936	0.962	0.964
Escort	0.875	0.918	0.938	0.941
Recreation	0.968	0.938	0.976	0.978
Other	0.697	0.706	0.722	0.741
Variable: D_j^p				
Work	0.976	0.913	0.931	0.935
Elementary School	0.886	0.835	0.910	0.914
Junior & High School	0.950	0.925	0.918	0.920
College	0.931	0.787	0.881	0.888
Shop	0.957	0.884	0.878	0.886
Escort	0.882	0.926	0.911	0.916
Recreation	0.964	0.828	0.887	0.898
Other	0.918	0.756	0.718	0.730
Variable: c_{ij}	0.966	0.952	0.984	0.985

Table 2: Simple correlation coefficients between 1985, 1995 and 2006 GMTD input variables.

perimeter (origin and destination), it appears that there has been an increase in t. It appears

5 Conclusion

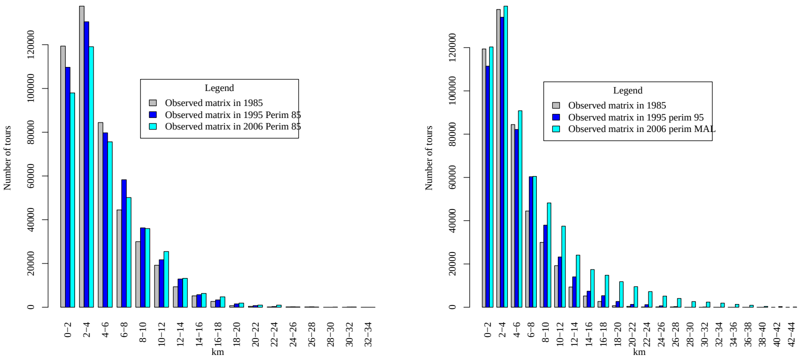
It appears then that the increase in travel distance could find an explanation in both:

(i) the improved performance of the network, which, as we kept constant the value of time, is the main source of the decreased mean generalized travel time between 1985 and 2006, within the perimeter of the 1985 EMD

(ii) and an increased willingness to spend time in the transport system, reflected by the shift to the right of the curve representing the deterrence function.

However the results on the influence of the second element results need to be tempered by the fact that, when it comes to forecasting, the use of the parameters of the GMTD calibrated for 1985 gives, paradoxically, better results in terms of forecasting of the mean travel distance for 2006 than the use of the parameters from a calibration with 2006 data. If the model predicted perfectly the number of tours for every O-D, then the observed mean generalized travel time should be equal to the predicted one, but this should also be the case of the observed and predicted mean travel distances.

Not everything is bad though. Because what matters in practice is to predict accurately travel distances (which ultimately reflect better the travel pattern shifts as distance between O-Ds remains constant over time), it seems that the



(a) Constant perimeter of 1985

(b) Varying perimeter

Figure 3: Evolution of the distribution of observed tours, for p =Work, by euclidean distance in km.

parameter stability hypothesis which is generally associated with the use of the GMTD is not a bad option for long term forecasting.

References

Anas, A. (1983). Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological*, Vol. 17 (No. 1), pp. 13–23.

Axhausen, K. W. (2000). Definition of movement and activity for transport modelling. In Hensher, D. A. and Button, K. J., editors, *Handbook of Transport Modelling*, pages 271–284, Amsterdam. Pergamon. ISBN 0-08-043594-7.

Bonnafous, A. and Puel, H. (1983). *Physionomies de la ville*. Éditions ouvrières, Paris. ISBN 2-7082-2304-6.

Bonnel, P. (2004). *Prévoir la Demande de Transport*. Presses de l'École Nationale des Ponts et Chaussées, Paris, 1st edition. ISBN 2-85978-395-4.

Briant, A., Combes, P.-Ph., and Lafourcade, M. (2010). Dots to boxes: Do the size and shape of geographical units jeopardize economic geography estimations? *Journal of Urban Economics*, Vol. 67 (No. 3), pp. 287–302.

Certu (2008). *L'enquête ménages déplacements "standard Certu" : Guide méthodologique*. Éditions du Certu, Lyon. ISBN 978-2-11-097161-6.

Duffus, L. N., Alfa, A. S., and Soliman, A. H. (1987). The reliability of using the gravity model for forecasting trip distribution. *Transportation*, Vol. 14 (No. 3), pp. 175–192.

Erlander, S. and Stewart, N. F. (1990). *The Gravity Model in Transportation Analysis: Theory and Extensions*. Topics in Transportation. VSP, Utrecht, the Netherlands. ISBN 90-6764-089-1.

Evans, A. W. (1971). The calibration of trip distribution models with exponential or similar cost functions. *Transportation Research*, Vol. 5 (No. 1), pp. 15–38.

Evans, S. P. and Kirby, H. R. (1974). A three-dimensional Furness procedure for calibrating gravity models. *Transportation Research*, Vol. 8 (No. 2), pp. 105–122.

Godinot, C., Bonnel, P., and Nicolas, J.-P. (2008). Phase d'affectation : mise en forme du réseau routier sur l'aire urbaine de lyon. LET, Rapport intermédiaire SIMBAD, 173 p., Lyon.

Hyman, G. M. (1969). The calibration of trip distribution models. *Environment and Planning*, Vol. 1 (No. 1), pp. 105–112.

Kirby, H. R. (1974). Theoretical requirements for calibrating gravity models. *Transportation Research*, Vol. 8 (No. 2), pp. 97–104.

Mikkonen, K. and Luoma, M. (1999). The parameters of the gravity model are changing — how and why? *Journal of Transport Geography*, Vol. 7 (No. 4), pp. 277–283.

Mozolin, M., Thill, J. C., and Lynn Usery, E. (2000). Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. *Transportation Research Part B: Methodological*, Vol. 34 (No. 1), pp. 53–73.

Openshaw, S. (1976). An empirical study of some spatial interaction models. *Environment and Planning A*, Vol. 8 (No. 1), pp. 23–41.

Openshaw, S. (1977). Optimal zoning systems for spatial interaction models. *Environment and Planning A*, Vol. 9 (No. 2), pp. 169–184.

Sen, A. and Smith, T. E. (1995). *Gravity Models of Spatial Interaction Behavior*. Springer, Berlin. ISBN 3-540-60026-4.

Snickars, F. and Weibull, J. W. (1977). A minimum information principle: Theory and practice. *Regional Science and Urban Economics*, Vol. 7 (No. 1-2), pp. 137–168.

Southworth, F. (1979). Spatial structure and parameter disaggregation in trip distribution models. *Regional Studies*, Vol. 13 (No. 4), pp. 381–394.

Southworth, F. (1983). Temporal versus other impacts upon trip distribution model parameter values. *Regional Studies*, Vol. 17 (No. 1), pp. 41–47.

Tanner, J. C. (1961). Factors affecting the amount of travel. Road Research Technical Paper No. 5. Her Majesty's Stationery Office, London.

Williams, I. (1976). A comparison of some calibration techniques for doubly constrained models with an exponential cost function. *Transportation Research*, Vol. 10 (No. 2), pp. 91–104.

Wilson, A. G. (1967). A statistical theory of spatial distribution models. *Transportation Research*, Vol. 1 (No. 3), pp. 253–269.