SSD

SCIENCE FOR A SUSTAINABLE DEVELOPMENT

BEHAVIOUR AND MOBILITY WITHIN THE WEEK

"BMW"

M. CASTAIGNE, E. CORNELIS, F. WALLE
R. FREDERIX , C. TAMPERE , P. TOINT, F. VITI

ENERGY

TRANSPORT AND MOBILITY

AGRO-FOOD

HEALTH AND ENVIRONMENT

CLIMATE

BIODIVERSITY

ATMOSPHERE AND TERRESTRIAL AND MARINE ECOSYSTEMS

TRANSVERSAL ACTIONS

.be

SCIENCE FOR A SUSTAINABLE DEVELOPMENT
(SSD)

Transport and Mobility

FINAL REPORT

BEHAVIOUR AND MOBILITY WITHIN THE WEEK "BMW"

SD/TM/03

Promotors

Philippe Toint
Facultés Universitaires Notre-Dame de la Paix (FUNDP - GRT)

Ben Immers & Chris Tampère
Katholieke Universiteit Leuven (KULeuven)

Kay W. Axhausen
ETH Zurich

Authors

Marie Castaigne, Eric Cornelis, P. Toint & Fabien Walle
Facultés Universitaires Notre-Dame de la Paix (FUNDP - GRT)

Rodric Frederix, Chris Tampere, Francesco Viti
Katholieke Universiteit Leuven (KULeuven)

D/2011/1191/5
Published in 2011 by the Belgian Science Policy
Avenue Louise 231
Louizalaan 231
B-1050 Brussels
Belgium
Tel: +32 (0)2 238 34 11 – Fax: +32 (0)2 230 59 12
http://www.belspo.be

Contact person: Georges Jamart
+32 (0)2 238 36 90

Table of content

1. INTRODUCTION

Analysing mobility demand leads more and more to researches taking into account behaviours. One of the most serious approaches is activity-based models. The main idea in this promising line of thought is to view traffic not as a stand alone phenomenon, obeying its own logic, but rather as a derived effect of activity patterns. Yet, most of the models (with rare exceptions like Mobidrive) are built on the paradigm that mobility is essentially linked to work and therefore exhibits daily cycles. Even if useful, this emphasis on the daily horizon contradicts the intuitive knowledge that a substantial fraction of people and household activities are repeated from week to week, not from day to day. Moreover surveys have shown that other purposes (shopping e.g.), more relying on weekly cycles, also induce an important part of mobility. This context is the motivation for this project.

Therefore, aiming at a better knowledge of weekly cycles in mobility, the BMW project deals with two complementary views on this weekly mobility: the longitudinal disaggregate behavioural aspects over the week and the transversal aggregate measure of traffic for each successive day of the same week.

The objectives of the project could be detailed as follows:

- collect data to validate the project view that weekly cycles are important in the household mobility decision;
- propose a descriptive analysis of the resulting weekly activity patterns and their impact on day to day variations in travel demand;
- reconcile these variations with observed variations measured in the field;
- enrich both activity-based demand models and dynamic origin/destination traffic models to include weekly cycles;
- disseminate the obtained conclusions with special attention given to policy implications and readability for non-specialists.

To fulfil these objectives, the project has eventually collected two new data sets:

- a multi-day travel survey and
- traffic data from countings.

Since this project is a first attempt in Belgium to collect data about mobility behaviours over a week, this available information will be a fruitful source of data for all the Belgian research teams involved in mobility behaviours and transport demand topics.

Moreover the descriptive analyses achieved on these data sets are of interest for Belgian policy planners and decision makers. They demonstrate the importance of taking into account the weekly rhythms in mobility behaviours for sustainable transport policies.

This report will start with a presentation of the analysis found on this topic in the literature (state of the art). Then the methodology of each data collection will be detailed, before explaining the results of our researches.

2. METHODOLOGY AND RESULTS

I. STATE OF THE ART

I.1. Preliminary analysis of mobility over a week

Our skills in mobility surveys led us to consider studying mobility patterns as chains of trips. Indeed, we consider that the daily trips scheme of an individual cannot be restricted to the simple sum of trips he achieved during this day, but well that most of the trips are linked together and form organization systems, resulting from more complex activities patterns, notably what we call «tours» (see Cirillo et al. 2001).

Beyond these daily organizations, we would like to analyse how activities are organized on a longer period: the week. Researches were already undertaken in this area (among others by K. Axhausen), but, up to now, the lack of data has not allowed us to prove the existence of such weekly organisation schemes for Belgium. Actually, no Belgian survey was conducted concerning households or individuals displacements over a whole week.

However, databases regarding household's daily trips are available (e.g. MOBEL). Moreover these data cover the whole week (including week-ends) since the data collection protocol spread the reference days (i.e. the day during which the trips had to be reported) on all the days during a year. Therefore we thought that a preliminary analysis could be achieved establishing how several mobility indicators vary according to this "day of the week" parameter. Nevertheless, the results of such an analysis must be read keeping in mind that each daily trips pattern occurred for a different respondent and that the used dataset does not allow to take into account correlations between days (as it would be the case in our new survey where all the daily trips schemes over a week will be related to a same person).

So we exploit the MOBEL survey dataset for such analysis, aiming to test, for several mobility indicators, their sensitivity, to the "day of the week" factor. Even if the reader has to keep in mind the here above caveat, nevertheless the achieved analyses highlight certain trends, reinforcing us in the idea that the activities schemes over week could provide us a better understanding on how the households or individuals organize their mobility patterns.

The number of trips achieved by a person seems significantly different only on Sundays (Figure 1), but if we consider only trips made by non-workers (Figure 2), we point out that on Fridays and Saturdays, their average number is higher.

Trips duration and distance (Figures 3 and 4) vary in opposite direction: if the persons achieve fewer trips on Sundays, average travelled distances are longer (as well as duration).

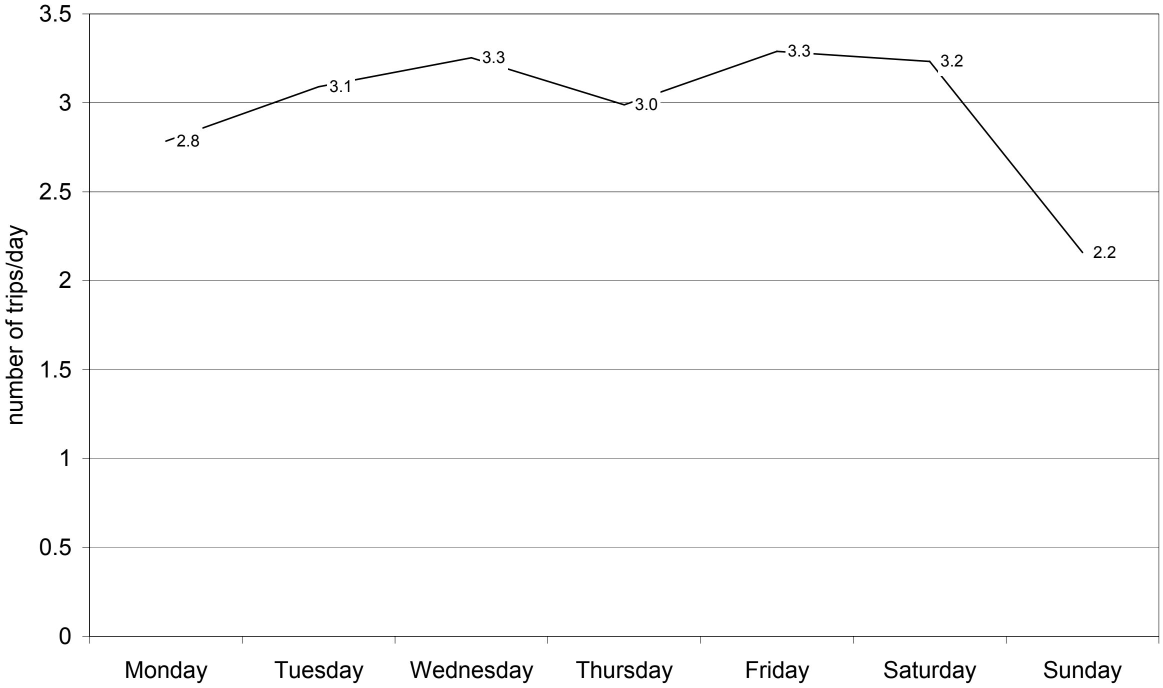


Figure 1: average number of trips according to the day of the week (MOBEL)

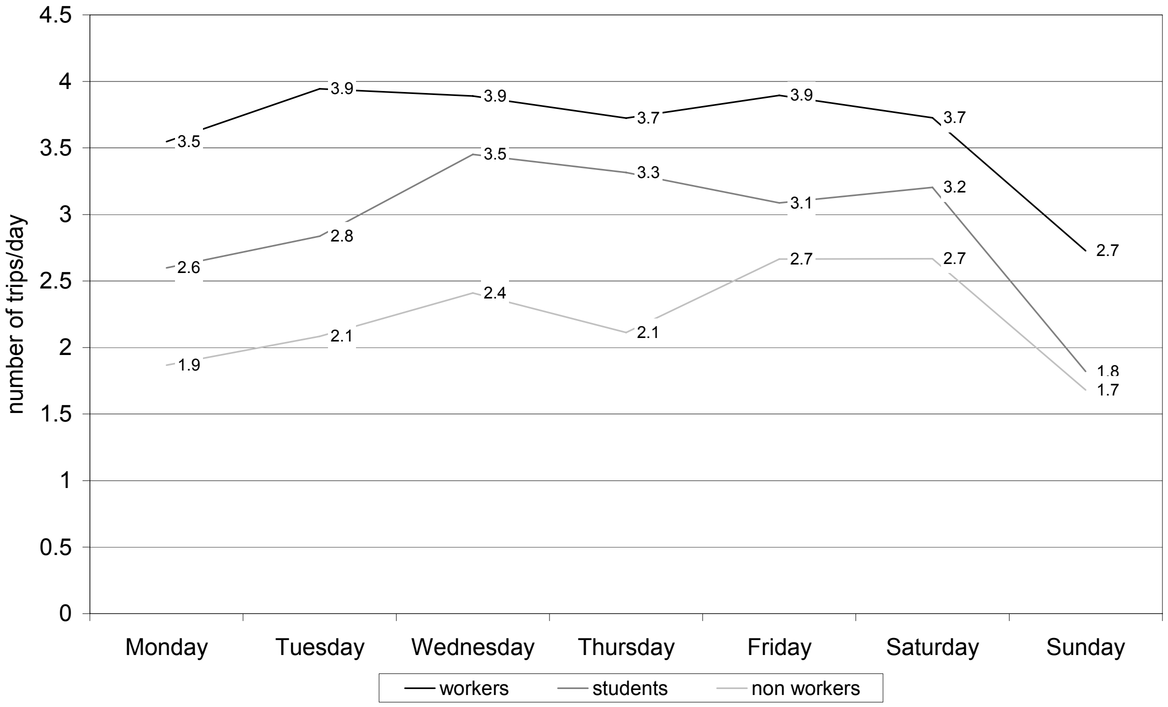


Figure 2: average number of trips according to the day of the week and the status (MOBEL)

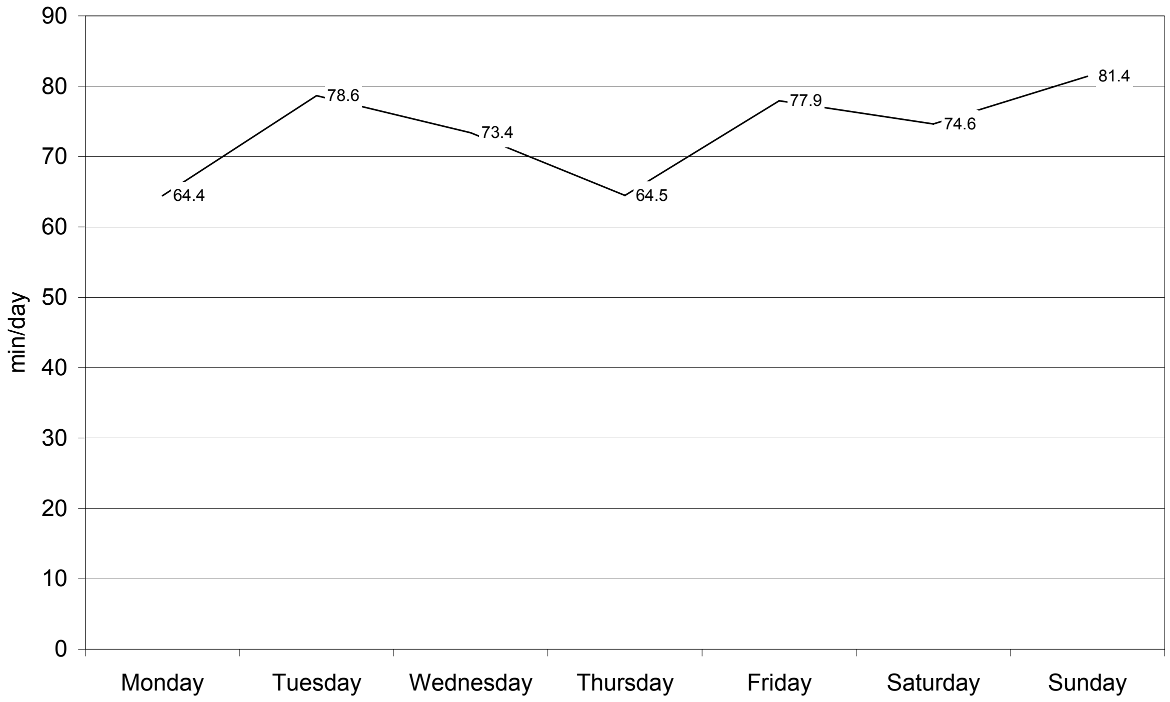


Figure 3: time budget for trips according to the day of the week (MOBEL)

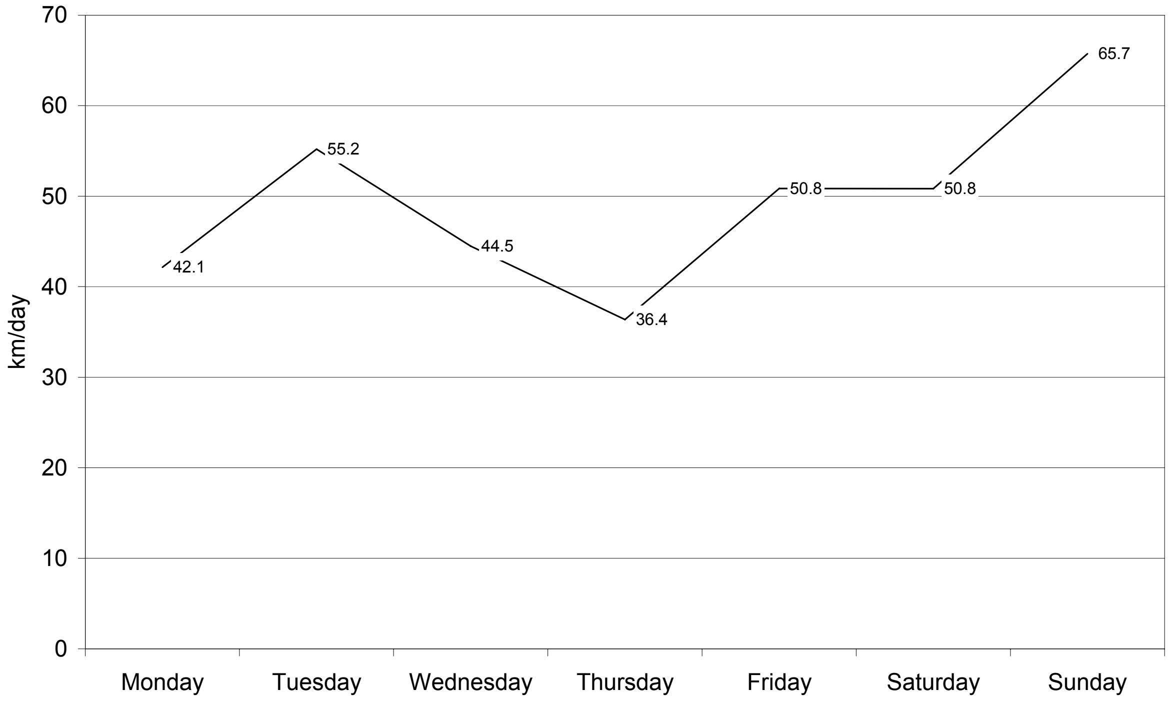


Figure 4: average length of trips according to the day of the week (MOBEL)

The spreading of trip purposes over a week also shows interesting elements (Figure 5). Differences, as we could expect it, are especially marked between working days and week-end days. Nevertheless some differences could also be drawn among working days: Mondays and Tuesdays are days when we work most, while leisure and meal taken outside are more important at the end of the week.

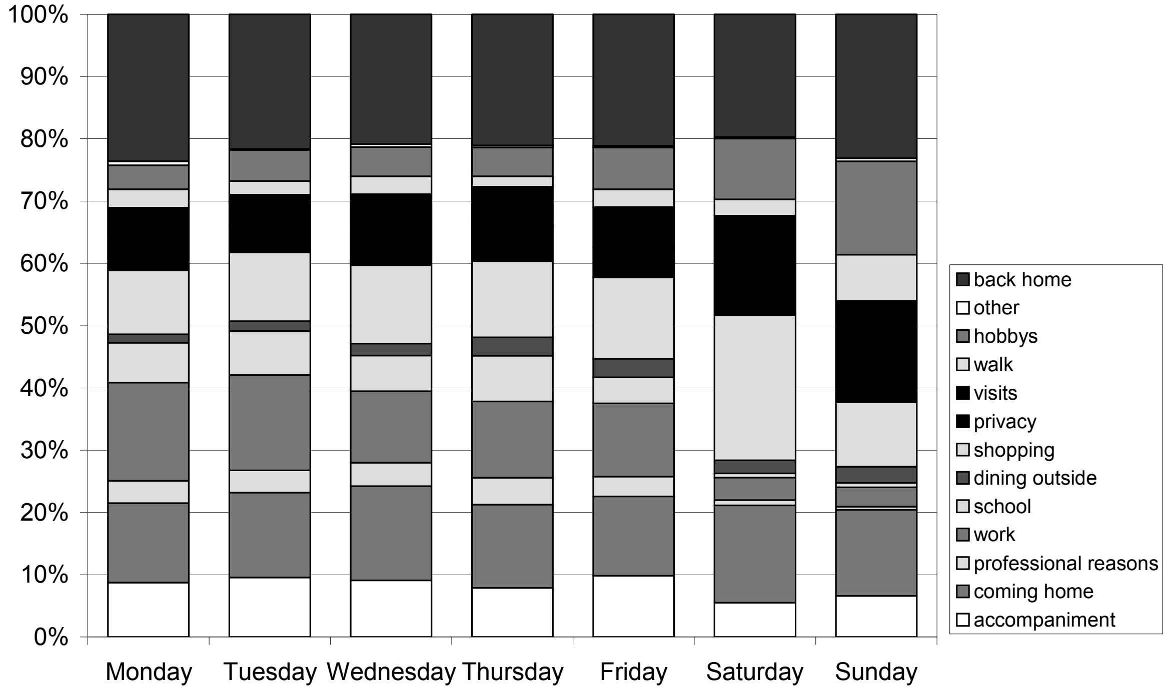


Figure 5: trips purposes for working days according to the day of the week (MOBEL)

We also analysed the occurring of given types of activities according to the different days of the week. We spotted the number of trips performed by every individual from our database for different purposes, during the reference day (for instance: number of trips accomplished for shopping, for leisure …). (Figures 6 to 9)

We point out in the following graphs the importance of shopping on Saturdays, visits to the family or to friends during week ends (mostly on Saturdays), leisure also during week ends (especially on Sundays). The opposite appears for purposes related to work: a decrease of trips to work on Saturdays and Sundays which seems natural according to the usual pattern of working activities.

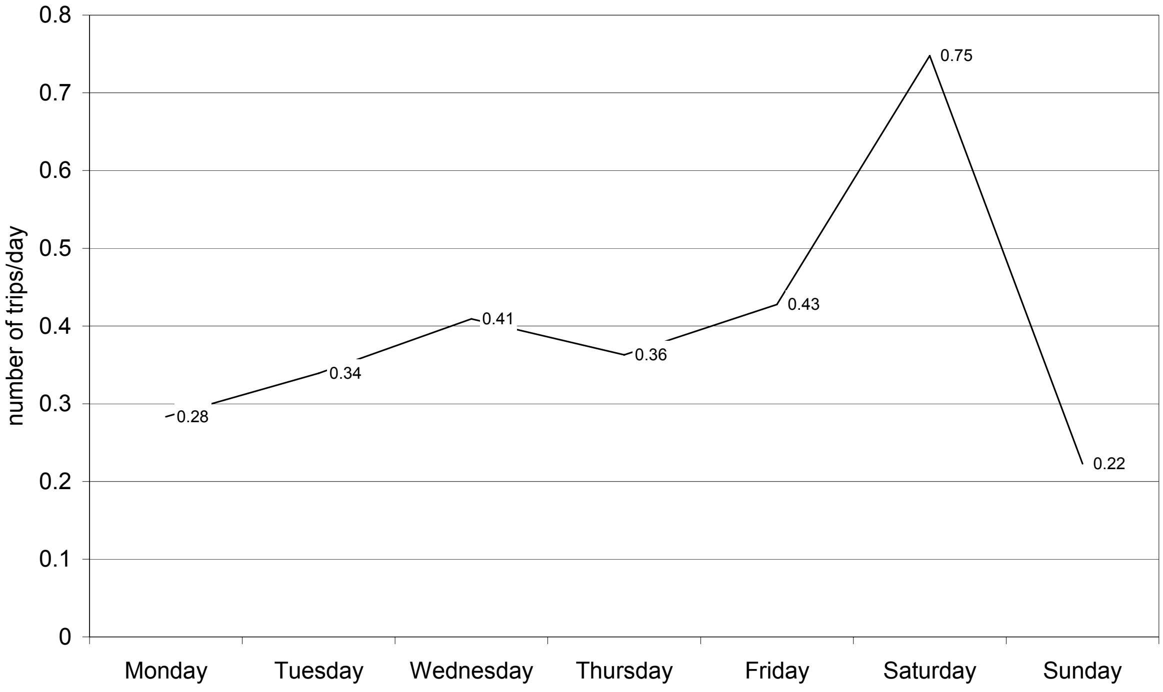


Figure 6: average number of trips to do shopping, according to the day of the week (MOBEL)

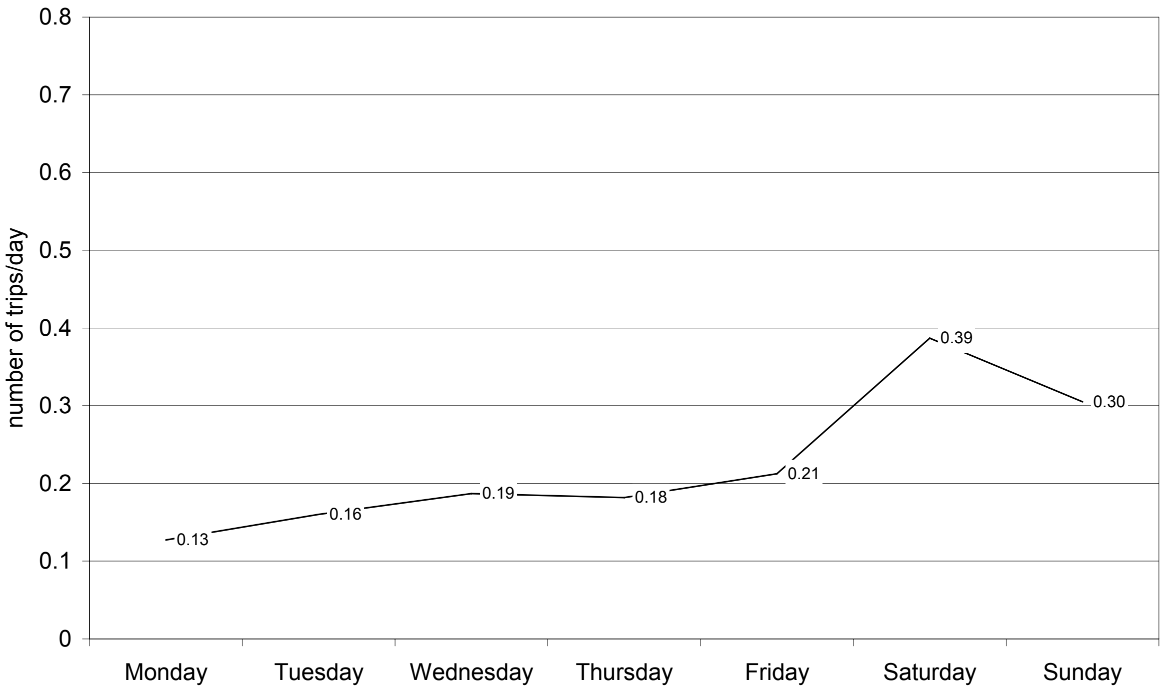


Figure 7: average number of trips to visit family or friends, according to the day of the week (MOBEL)

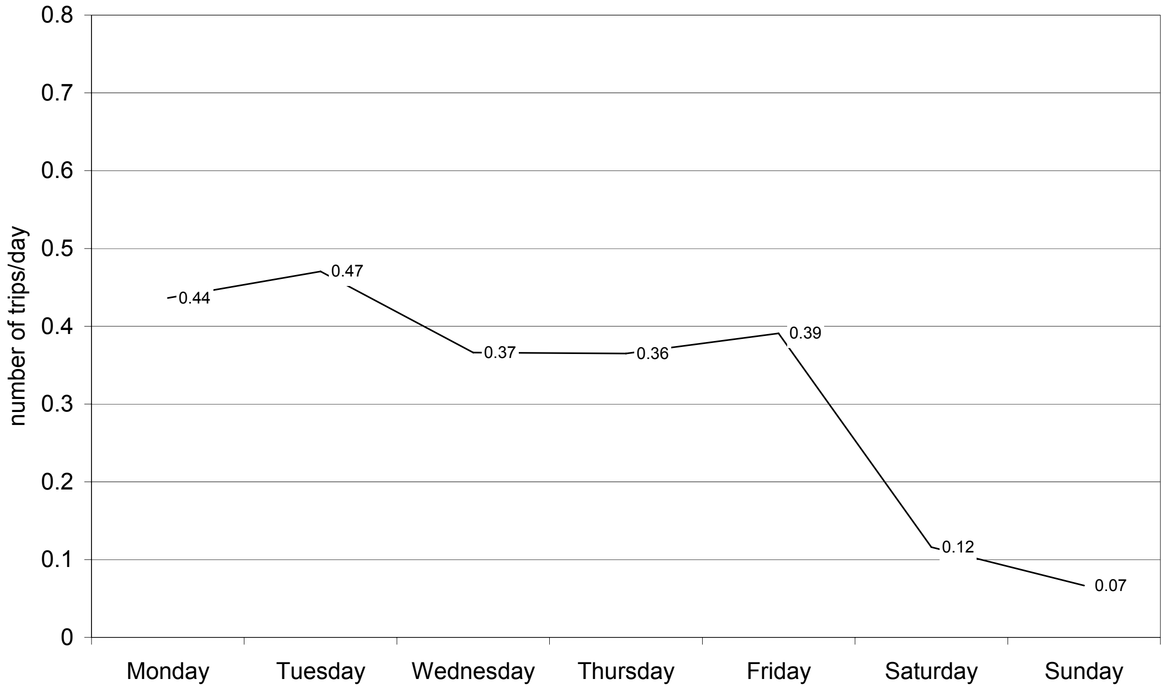


Figure 8: average number of trips for work, according to the day of the week (MOBEL)

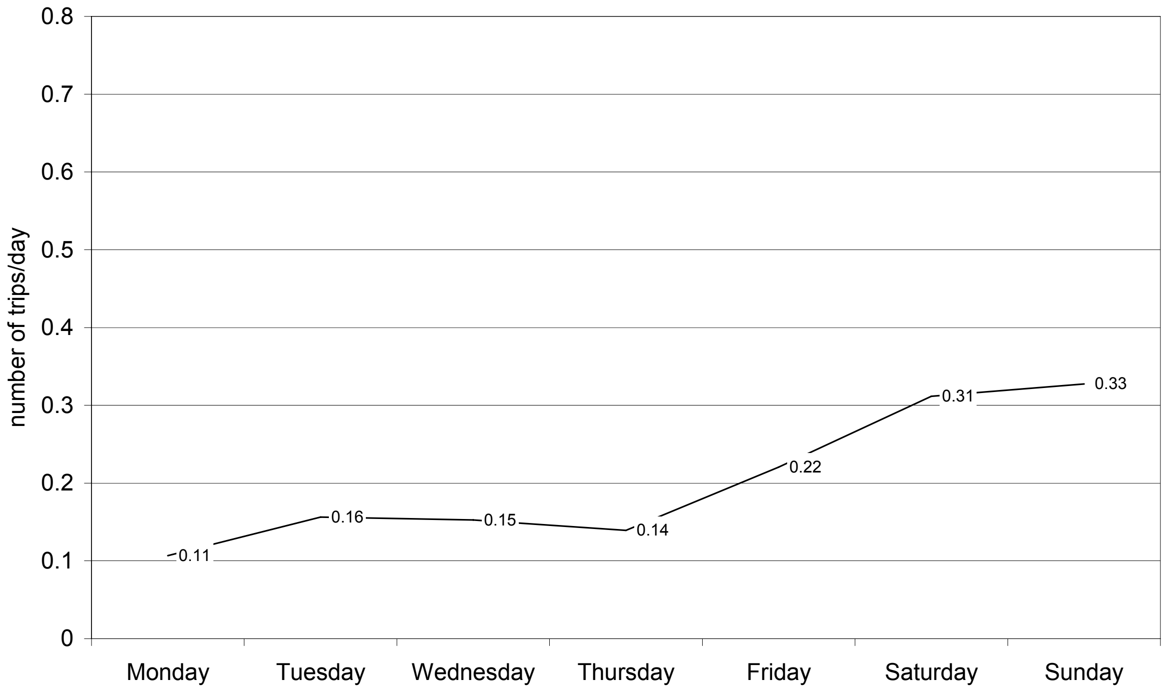


Figure 9: average number of trips for leisure according to the day of the week (MOBEL)

According to the results obtained from these analyses, it sounds fruitful to investigate deeper the weekly organising schemes of individuals and the analysis of their mobility behaviours over a week through the planned survey.

I.2. Literature Review on O-D estimation from traffic counts

This section gives an overview of the traditional methods found in literature for determining the relationship between traffic counts collected at road sections in (discrete) time intervals and the travel demand that has (most likely) generated this data. This problem is framed within the more general Origin-Destination matrix estimation problem, i.e. given information on traffic and/or travel data what is the most likely distribution of trips in a traffic network. For the purpose of our study, reference is made exclusively, in this section, on past research on OD estimation methods that use traffic counts as input, while no regard is given to approaches based on other information sources or different measures than the traffic counts (e.g. travel time data from floating cars, mobile phones, cameras, or flow and speed data taken from double loop detectors, travel survey data, etc.).

The main research questions addressed in this section are therefore the following:

1. Given positions and number of traffic counts, what is the most likely (time-dependent) O-D matrix that explains the traffic counts and how large is the uncertainty around this estimate;
2. Inversely, given the network topology and characteristics, what is the best location and number of detectors in order to maximize the information on the network and to estimate at best O-D trips and route flows along the network.

We will only briefly explain the methods and their main characteristics, leaving all the mathematics to the original papers. Moreover the bibliography is not exhaustive, while only a few representative papers have been selected for each approach.

We start by presenting static OD estimation methods, which have strongly inspired dynamic approaches and are still widely used because of their simplicity and interpretability. Later we discuss the dynamic models, which incorporate dynamics both at the operational level (i.e. in the spatio-temporal propagation of flows in the network), and at the decisional level (i.e. in terms of route and departure time choice). We focus here on car traffic only, thus no review is given on other modes. The total generated demand is also assumed given and inelastic in the problem. Finally, according to the bounds of our study, we will not consider en-route decisions, such as rerouting.

According to the second research question, we will also give an overview of the traditional methods to identify the optimal locations where to measure traffic counts.

I.2.1. Static O-D estimation: proportional assignment models

Originally O-D matrices have been estimated in two ways: using travel surveys (or direct sampling estimation) and model-based estimation, i.e. the O-D matrix is estimated by applying a system of models -physical and/or behavioral- that compute the approximate number of journeys made with a certain mode during a certain period of time. The linking between O-D trip matrices and traffic counts has instead relatively recent origins (late ’70s). 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 in which traffic is monitored.

The estimation of trip matrices from traffic counts and the problem of finding the optimal locations for counting vehicles have been considered dual problems since their origin. The common ground, which explicitly links these two problems, is the assignment of origin-destination trips on each (used) route alternative connecting an O-D pair and therefore on each link in the network. Mathematically speaking, traditional static OD estimators can be formulated in a general way as follows:

$$\mathbf{T} = \arg\min_{\mathbf{T}} \left[\sum_{rs} f_1(T^{rs}, \hat{T}^{rs}) + \sum_{a} f_2(U_a, \hat{U}_a) \right]$$

Where f_1 and f_2 are performance functions controlling the performance of the OD estimator, $\mathbf{T}$ is the estimated OD matrix of elements T^{rs} for each origin-destination pair (r,s), $\hat{T}^{rs}$ is the corresponding element in the target matrix, and respectively U_a and $\hat{U}_a$ are the estimated and the measured link counts. By definition, the vector of link flows U_a satisfies the relationship:

$$U_a = M_a^{rs} T^{rs}$$

Where M_a^{rs} is the so-called assignment matrix, which controls the fraction of flows from OD pair (r,s) which uses link a.

Different forms of the above optimization problem can be found where the assignment relationship is used as constraint to the problem. In both modeling approaches the problem shifts from the estimation of OD trips to the estimation of the assignment matrix. Originally, two types of assignments have been proposed: 1) proportional assignment, and 2) equilibrium assignment. Here we briefly describe the main static approaches based on these two assignment rules.

I.2.1.1. Maximum Entropy/Minimum Information approach

As already mentioned, the main difficulty of estimating O-D pairs from traffic counts is the under-specification of the problem, i.e. multiple solutions exist, which map a set of link counts to O-D flows. This issue has been highlighted by Robillard (1975), who proposed to overcome this problem by using Generalized Gravity model, which is simply a balance between travel costs for an O-D pair and link flows measured at traffic counts. The main criticism according to Van Zuylen and Willumsen (1980) is that this model forces the trip matrix to follow a gravity-type of pattern and it does not make full use of the information contained in the counts. This issue has been solved by introducing an a priori matrix, or a target matrix and by using Information Theory. We quote the works of Van Zuylen and Willumsen (1980) and Bell (1983) as representatives of trip matrix estimation in case of proportional assignment, and Nguyen (1977) and LeBlanc and Fahrangian (1982) and Fisk (1988) for the equilibrium assignment case.

Both Van Zuylen's information minimization and Willumsen's and Bell's maximum entropy maximization have been developed with the aim of getting an a posteriori matrix starting from a target matrix, for instance an outdated O-D matrix. Van Zuylen and Willumsen have pointed out that the two approaches they developed almost contemporarily lead to basically the same results. The main philosophy that these approaches share is simple: one can associate for each link flow a proportion, which originates and ends at one O-D pair, and therefore calculate the probability that this portion of flow comes from a specific O-D; the second task is then to find the a posteriori trip matrix that maximizes the overall probability of generating the observed link flows, starting from an a priori matrix. The need for the a priori matrix is obvious given the multiple solutions to this optimization problem and the need for an initial starting solution. This yields to different formulations for the OD estimator. The following form is often used in practice:

$$\mathbf{T} = \arg\min_{\mathbf{T}} \sum_{rs} (T^{rs} \ln(\frac{T_h^{rs}}{\hat{T}^{rs}}) - T^{rs}) + \sum_{a} \phi_a (U_a \ln(\frac{U_a}{\hat{U}_a}) - U_a)$$

Where $\mathbf{T}$ is the estimated OD matrix of elements T^{rs} for each origin-destination pair (r,s), $\hat{T}^{rs}$ is the corresponding element in the target matrix, and respectively U_a and $\hat{U}_a$ are the estimated and the measured link counts. Finally, ϕ_a is a weight assigned to each link count as function of its degree of reliability.

Maher (1983) pointed out that Maximum Entropy/Minimum Information (ME/MI) methods suffer from one main issue: the a priori matrix is simply used as an initial condition and the philosophy is to use as little information as possible from this initial guess. The largest importance is therefore given to the traffic counts, which are in the model assumed error-free. Usually the target matrix comes from survey or past studies, but in general it can have some degree of error due to e.g. inference errors or to changes in the OD flows due to changes in activities, modal shifts, population growth etc. On the other hand, traffic counts contain also errors, e.g., missing counts, incomplete data etc. Moreover, the proportional assignment principle, which links target matrix and traffic counts, simplifies and approximates the real flow fractions at each link, and does not apply in congested networks, where small speeds and low flow rates are observed. This implies that considering the a priori matrix as the least "trustable" information in this process is not obvious. Fisk (1988) extended the entropy model of Van Zuylen and Willumsen to the congested case by introducing the user-equilibrium conditions as constraints. Smith (1979) showed that the model of user-optimal behavior can be expressed as Variational Inequality Problem. The proposed model has a Bi-level structure that maximizes the entropy on the upper level and solves a user-equilibrium problem on the lower level.

I.2.1.2. Statistical inference approaches

Statistical inference approaches originate from Maximum Likelihood (ML) estimation techniques, and have the advantage of using information retrieved from traffic counts and from the target O-D matrix, searching for the most likely solution that satisfies both information sources. It is assumed in this approach that the elements of the target O-D matrix are obtained as observations of a set of random variables. This optimization problem was firstly considered by Spiess (1987), who provided also a descent algorithm to calculate the minimum of the likelihood function.

Cascetta (1984) extended the ML approach by proposing to use the Generalized Least Square (GLS) method to combine target trip matrix, model prediction and traffic counts within a single framework. The OD estimator proposed by Cascetta has the following mathematical formulation, in its most simplified case:

$$\mathbf{T} = \arg\min_{\mathbf{T}} \left[\sum_{rs} (T^{rs} - \hat{T}^{rs})^T (T^{rs} - \hat{T}^{rs}) + \sum_{rs} (\hat{M}_a^{rs} T^{rs} - \hat{U}_a)^T (\hat{M}_a^{rs} T^{rs} - \hat{U}_a) \right]$$

Thus, the above estimator minimizes the quadratic deviation between estimated OD trips and link flows with respect to respectively the target matrix and the measured link counts. This method has been shown to outperform the Maximum Entropy maximization approach in predicting the true trip matrix using a toy-network scenario. The same author does not guarantee on the other hand that this will hold true in real applications, since non-negativity constraints may be violated and solutions may fall outside the feasible set of solutions. Later Bell (1988) corrected this method by binding the output to overcome these undesirable outcomes. GLS estimation was also adopted by Bierlaire and Toint (1994) to develop a model able to incorporate congestion effects.

Maher (1983) proposed the application of Bayesian statistical inference to combine the reliability of information gathered from the a priori trip matrix and the one gathered from link counts. The method is valid for O-D estimation problems as well as for estimating turning flows at intersections. The main characteristic of this approach is to assign an a priori set of weights for the a priori trip matrix, which can depend on the degree of reliability around this information. This method has therefore two desirable properties: 1) it extends the Maximum Entropy/Minimum Information criterion, being equal when the weight for the prior estimates are equally distributed among OD pairs, and 2) it can potentially balance the information of link traffic counts with many other sources of information, not only with the a priori trip matrix.

Although statistical inference is a rather sound method to distribute well the value of the various sources of information, the proposed methodologies are still based on a "less sound" criterion, i.e. the proportional assignment approach. These methods can lead to relatively small errors if congestion levels are low, but, as Bell (1988) points out, since route choice and congestion phenomena are strongly correlated, the relationship between OD flows and link flows should not be linear. This limitation has motivated the introduction of equilibrium-assignment based approaches, as explained in the following section.

I.2.2. Static O-D estimation: equilibrium assignment models

In case of congested networks drivers choose which routes to take in proportion to the current (and past) information of route travel times. If this effect is considered explicitly, the O-D matrix estimation problem becomes more complicated. The route choice proportions will depend in fact on the current traffic situation (measured with e.g. link or route travel times or flows), which in turn depends on the O-D matrix. Thus, the relationship between route proportions and the O-D matrix can only be implicitly defined. Traditionally, approaches in literature have followed a bi-level structure to combine trip-distributions with route choice distributions, as described in the following of this section.

I.2.2.1. OD estimation approaches formulated as Bi-level problems

Bi-level O-D estimation models commonly consider the equilibrium assignment as lower level problem and finding the most likely O-D trips consistent with both traffic counts and traffic equilibrium at the upper level. Within this approach equilibrium is commonly considered as a "constraint", while the distance between measured counts and estimated counts is minimized.

I.2.2.2. Lower level: Traffic Assignment

Nguyen (1977) and LeBlanc and Fahrangian (1982) stressed the importance of using a non-proportional assignment in order to properly catch the effects of congestion. The main difference is that in the proportional method there is no regard of the capacity of each link and the relationship between link and O-D flows remains linear. Nguyen formulated the problem as a constrained optimization problem, whose constraints are the conservation of vehicle equations (where traffic counts are explicitly playing a role), and the problem is equivalent to the first Wardrop's equilibrium principle, i.e. flows are distributed along the shortest paths and therefore in equilibrium.

I.2.2.3. Upper level: OD estimation

The OD estimation in the Bi-level approach is done traditionally using two methods: 1) Minimum Euclidean distance and 2) Generalized Least Squares. LeBlanc and Fahrangian (1982) proposed to formulate the upper level as minimization of the Euclidean distance between the solution matrix and the target matrix. This approach has recently been used by Lindveld (2003) and implemented in an algorithm which aims at solving OD matrix estimation with Dynamic Traffic Assignment (DTA) for large scale networks and that explicitly takes into account drivers' behavior in terms of route and departure time choice.

O-D estimation problems formulated as Bi-level problems have the advantage of considering both target O-D matrix (in the upper level) and the traffic counts (in the lower level), and to deal with the dynamic effects of congestion. However, they are characterized by a high complexity and solution algorithms able to solve efficiently these problems are not readily available, but only heuristic solution algorithms are currently used. Even accounting for the simplifications of a deterministic assignment, Bi-level problems are normally non-convex, thus the solution is not guaranteed to be unique.

I.2.2.4. OD estimation solved with gradient-based solution methods

To solve the non-convexity issues, one-level structures are usually preferred to Bi-level formulations. One-level formulations that combine trip-distribution with assignment commonly require the computation of the marginal functions between these two estimation parameters. In gradient-based solution techniques, the target O-D matrix is taken simply as an initial solution to the O-D matrix estimation problem. The target O-D matrix is "adjusted" to reproduce the traffic counts by iteratively calculating directions based on the gradient of the objective function. The link volumes are implicit functions of O-D flows and obtained by the assignment procedure and the problem is transformed into a one-level problem, with the advantage of having a larger spectrum of solution approaches.

In the past, many different solution methods were proposed for solving this problem. Spiess (1990), Drissi-Kaitouni and Lundgren (1992), Yang et al. (1992), Florian and Chen (1993) and Chen (1994) are only few examples. For instance Florian and Chen (1993) reformulated the Bi-level problem into a single level problem using the concept of marginal functions. Chen (1994) proposed an augmented Lagrangean approach, which can be shown to converge to a stationary point. This approach, however, requires that all used paths in each O-D pair are known, and it is thus applicable only to very small networks. As mentioned, an advantage of this approach is its computational tractability. Spiess (1990) presents applications to several large scale problems. The problems include an urban application of Bern, Switzerland, with about 2700 links and one interregional application to the road networks of Finland with about 12500 links. In the method of Spiess proportional assignment is assumed to hold locally and the method does not necessarily converge to a solution of the optimization problem. With this assumption the gradient of the objective function becomes easy to compute, attainable from the solution of two equilibrium assignment problems. The results indicate that the computations involved are reasonable. The emphasis is on the quality of the search directions and not on attaining a computationally efficient O-D matrix estimator.

I.2.2.5. Stochastic OD estimation formulations

In all above estimation procedures it is assumed that the assignment is made according to the Deterministic User Equilibrium (DUE) assumption, for sake of mathematical tractability. A more realistic approach could be to allow differences in cost perceptions, and for the heterogeneity among travelers' choices. This can be done using the Stochastic User Equilibrium (SUE) principle. In literature, there are very few studies using SUE in the development of O-D estimation methods. Examples are the methods proposed by Liu and Fricker (1996), Cascetta and Postorino (2001), Clegg et al. (2001), Maher et al. (2001), and Yang et al. (2001).

Liu and Fricker (1996) proposed a two-stage iterative method to estimate both the OD matrix and the dispersion parameter of the Logit model. However, the authors used the observed link flows to calculate the link costs in the Logit models, and hence could not solve the inconsistency problem created by the congestion effects in the link flows. Yang et al. (2001) improved this approach by using, in the cost function, the link flows obtained from the SUE traffic assignment and estimated O-D flows. They also proposed an iterative quadratic-programming algorithm to solve the simultaneous model. This procedure requires the computation of derivatives of the objective function and the SUE constraints. Although they used a general expression for the objective function, no reference was made to the statistical properties of the observations and estimators; in fact, only a simple sum of the squares of errors was used as the objective function in their example. This objective function is heuristically reasonable, but it does not make full use of the statistical components contained in the data.

For SUE traffic assignment, the most commonly used route choice model is the multinomial Logit model. In the multinomial Logit model, the route choice probability is a function of the dispersion parameter □, which may be interpreted as a measure of travelers’ sensitivities to the route costs. This parameter can also be interpreted as the degree of information available to network drivers. In many previous studies, □ is considered as a predetermined fixed constant. However, according to the aforementioned interpretations, its value should be related to traffic conditions of the network under study. Based on this consideration, Lo and Chan (2003), instead of arbitrarily assigning a value for □ or estimating the value based on historical data, proposed to estimate this parameter simultaneously with the O-D matrix from observed data. Results from a numerical study using a hypothetical network have shown that models allowing □ to be estimated simultaneously with the O-D matrix from observed data perform better than models with a predetermined fixed □. The authors showed also that the proposed algorithm is quite robust towards inaccuracies in the survey data and in the number of observed links used.

I.2.3. Overview of static OD-estimation methods

The table in Figure 10 gives an overview and a classification of the most important models developed in the past (picture taken from Abrahamson, 1998) and presented in this section. As one can see the ME/MI models of Van Zuylen and Willumsen, and the GLS estimation of Cascetta are the simplest models, not requiring information on the target matrix. However, they are also incapable of capturing congestion effects if proportional assignment is assumed.

The methods described so far are strongly bound to be static and therefore applicable to only, e.g., planning and design problems, while doubts remain on their applicability in a dynamic context, i.e. in presence of non-negligible congestion phenomena, or when dynamic effects are important to be caught to obtain a more realistic assessment of the effects of a management plan (e.g., dynamic traffic management problems, route guidance). This is only partly accounted in equilibrium-based O-D estimation methods where drivers’ response to changes in the system can be explicitly modeled.

	Target OD matrix	No target OD matrix
Congestion	LF 82, YSIA 92, Wil 84, Spi 90, DL 92, EJL 84	Fis 88, Fis 89, JN 79, ENS 79, JN 83, KLH 92
No congestion	Mah 83, Bel 91, Spi 87, BT 95, KS 87, TW 89	ZW 80, Cas 84

Figure 10: relevant models for static OD-estimation (source: Abrahamsson, 1998)

Apart from the relative reliability of each source of information available (survey, link counts, models), question is whether the above models cope with the relative quality of information within the same traffic counts. Some link counts can "tell" more than others, e.g. they are highly capacitated and highly demanded links, or they connect several routes and therefore tell information over more O-D-pairs. On the other hand, the more the routes overlapping on a specific link, the less this information may reliably be transferred to upstream locations (and also downstream in a related route flow prediction problem). Moreover, two link counts can partly give the same information and therefore be redundant in static models, but not in dynamic O-D estimation (Yang and Zhou, 1998). We give an example: we can consider two detectors placed on two links, which are solely used by one route; thus we expect that in terms of (static) O-D estimation one of the two detectors should suffice. The usefulness in dynamic estimation is not that easy to understand, since the position of these counters may help at tracing and monitoring the dynamics of the system. Looking at the above methods it seems that none has explicitly considered this issue

Finally another point of interest (which involves also the currently available DODE problems, is to know what happens when more sources of information of the current state of traffic are available than simply the traffic counts (e.g. cameras, floating car data, etc.), and whether they can be used to correct or give estimate of the reliability of link counts. One can refer for example to Van der Zijpp (1996) to have an overview of the problem and how it has been analyzed for matching link counts with camera detections. The increasing use of e.g. floating car data or electronic toll collection systems suggests that in the near future one will have more than simply traffic counts to estimate trip distributions, and specifically route choices and travel times will be party available. To correctly exploit this information and the combination of two or more of these data sources Dynamic O-D Estimation methods will be necessary.

I.2.4. OD-Matrix estimation from link counts: dynamic models

As we observed earlier, the estimation of OD matrices using a static approach has limited applicability in the estimation of OD matrices in a dynamic context, especially when the effects of congestion are non-negligible. A time-dependent (dynamic) model must consider the influence of traffic conditions in a certain time period to any succeeding time period (Peterson, 2007). Since the estimation of these OD tables is needed, among other reasons, to obtain the prediction of how flows propagate on all the links, including those where traffic counts are not available, we easily understand that an error in the estimation of OD trip tables is carried over, causing often incorrect predictions of traffic parameters on these links. It becomes therefore necessary to include time-dependency in the OD flows and to shift to Dynamic OD Estimation (DODE).

The number of applications where a time dependent OD-matrix is required has grown rapidly in the last decade, mainly as a result of the increasing computing possibilities, the introduction of Dynamic Traffic Management and more generally Intelligent Transportation Systems, the new techniques for supplying interactive information via internet, variable message signs and so forth. Time-dependent OD-matrices are used both for strategic and operational purposes. In the strategic area the aim is to model the normal traffic situation as good as possible. Such OD-matrices are used for evaluating the time-dependent effect of different scenarios, for example for generating plans of actions in case of, e.g., incidents. Developing plans of actions for different emergency situations, and studying the effect of time-varying road tolls, are examples where time-dependent models are applied for strategic planning. In the operational planning, time-dependent models running in real-time can be used for providing information to variable message signs, or controlling traffic-lights and other traffic facilities, etc. Time-dependent OD matrices are also used to describe the traffic conditions in the operational management. This type of operational models are used to produce, e.g., travel time forecasts, which in turn are essential for different kind of information systems or route guidance systems (Peterson, 2007).

The main difficulty characterizing DODE problems is, as already mentioned, that traffic is observed as single commodity, i.e. O-D flows for different routes, O-D pairs and/or time intervals are intermixed and observed as aggregated. This implies that all parameters contributing to traffic variations (e.g., daily and weekly activity schedules, heterogeneity in travel choices and in driving behavior) are shown by traffic counts as one single value at each counter and at each time interval, thus as a single, aggregated, entity. The largest effort in this research field has been therefore to providing solution methods to reconstruct the multi-commodity nature of traffic.

Therefore, we can subdivide the research developed on DODE onto three main aspects:

1. Models that incorporate dynamics in travel behavior;
2. Models that incorporate dynamics in the traffic flow propagation;
3. Models that incorporate and aggregate both dynamics.

I.2.4.1. OD estimation incorporating dynamics of travel choice behavior

The class of OD estimation methods that incorporate explicitly the dynamics due to travel choice behavior is considered as inverse of Traffic Assignment problems (Bierlaire (2002)). In line with this analogy, and the way static traffic assignment models have been extended to the dynamic context, also DODE models have been initially developed as adaptation of the available static OD estimation methods, presented in the previous sections. Several attempts to extend Wardrop's equilibrium principles (Wardrop, 1952) for a time-dependent model have been in fact proposed in the past. In the context of OD-matrix estimation the majority of studies have incorporated explicitly route choice (or route proportions), while less importance has been given to other decision levels. Only recently research is made on incorporating departure time choice, as it was indicated to be a very sensitive parameter to congestion effects. The importance of this decision level in a dynamic OD-estimation problem is almost straightforward. Despite the importance of this choice level, in practice route and departure time choices are estimated simultaneously in dynamic OD estimation by using the Space-Time Extended Network (STEN), or hypernetwork, representation, where different departure time choices are graphically represented as alternative routes and the problem does not differ significantly from traditional OD estimation with route choice. Examples of such approach are given by Janson and Southworth (1992) and van der Zijpp and Lindveld (2003).

The first extension of static models to the dynamic context was made by Willumsen (1984). He proposed an extension of the ME/MI static model, where time-dependent OD flows are generated in sequence from the sequence of traffic counts. Since it is based on the proportional assignment, also the dynamic version of the ME/MI model requires time-independent route choices, which is clearly a strong assumption in dynamic systems. Davis and Nihan (1991) developed a maximum likelihood estimator, which can be viewed as a development of the method proposed by Spiess (1987) for the static case. Davis (1993) extended the ideas to a general Markov model, for which it can be shown that consistent OD-matrix estimates can be derived from link flows, also under relatively weak conditions. Bell et al. (1991) made assumptions on the travel time distribution and thereby they accounted for different flow propagations in different time periods. This improvement is important for large networks, where the assumption of equal travel times might be too rough. Hereby, the model becomes dynamic both in flow propagation and in route choice. Cascetta et al. (1993) developed a model for a general two-objective form of the problem, a simultaneous and a sequential estimation problem. In their numerical tests a general least-square estimator has been used, which can be viewed as an extension of the static GLS model developed by Cascetta (1984). This approach offers computational advantages, since it reduces a large optimization problem into a number of smaller ones and gives the possibility of using the estimates for an interval as a priori estimates of subsequent ones.

The most commonly used model in OD-estimation applications incorporating route choice is the multinomial Logit model. Logit-based stochastic methods have well-known weakness, such as their inability to take proper account of overlapping or correlated paths, as they require the assumption of ''independence of irrelevant alternatives''. Nevertheless, due to its simple structure and ease of use, the Logit model has enjoyed much attention (Lo and Chan, 2003). In the multinomial Logit model, the route choice probability is a function of the dispersion parameter ▯, which may be interpreted as a measure of travelers' sensitivities to the route costs. This parameter can also be interpreted as the degree of information available to network drivers. In many previous studies, ▯ is considered as a predetermined fixed constant. Exceptions are the works of Liu and Fricker (1996), Yang et al. (2001), and of Lo and Chan (2003). Liu and Fricker (1996) proposed a two-stage iterative method to estimate the OD matrix and the dispersion parameter of the Logit model. However, the authors used the observed link flows to calculate the link costs in the Logit models, and hence could not solve the inconsistency problem created by the congestion effects of the link flows. Yang et al. (2001) improved this approach by using, in the cost function, the link flows obtained from the SUE traffic assignment and estimated OD flows. They also proposed a successive quadratic-programming algorithm to solve the simultaneous model. However, this procedure requires the computation of derivatives of the objective function and the SUE constraints and it is valid for uncongested networks.

In case the network is congested, and the routes are chosen with respect to the current travel times, the OD-matrix estimation problem is more complicated. The route proportions depend on the current traffic situation (travel times/link flows), which in turn depends on the OD-matrix. Cremer and Keller's System Dynamics (Cremer and Keller, 1987) represent maybe the first attempt to develop dynamic O-D matrices by considering congestion-dependent OD flows, although these methods are especially developed for estimating route flows at small networks (e.g., intersections). They proposed four different methods for this problem: 1) cross correlation, 2) constrained optimization, 3) recursive estimation and 4) Kalman filtering approach. Common feature of these methods is that traffic flow through a traffic facility is considered to be a dynamic process in which O-D flows and exit flows are time-dependent variables, which depend by causal relationships upon the time-variable patterns of entrance flows. It was shown by treating the problem with system dynamic methods that more information can be obtained from volume measurements when collected as time-variable sequences. Cremer and Keller's method has been developed to determine route flow proportions on measured intersections, where all incoming and outgoing flows are measured (this is often done in practice). However, this method has some inherent limitation for application on larger traffic network systems, as it is founded on the conservation of vehicles principle, which has relatively less importance in simplified networks used in OD estimation, where flows virtually originate from fictitious points, i.e. the centroids, whilst in reality they are more widespread. Lo and Chan (2003) proposed a procedure for the simultaneous estimation of an origin–destination (OD) matrix and link choice proportions from OD survey data and traffic counts for congested networks. The main contribution of their work with respect to, e.g., equilibrium-based models is that difference in route cost perceptions may change with respect to the traffic context, e.g., whether it is congested or uncongested period.

Hazelton (2000) proposed a method which can also make use only of link counts, but it requires explicit path enumeration and is therefore practically strong time-requiring for large-size networks. Changes in travel choice modeled in a within-day time horizon should correctly account for congestion effects. However, as pointed out by and Hazelton (2003), a promising research development deals with considering time-series link counts (e.g. referred to several days) as a key aspect for improving the reliability of OD matrix estimation. Day-to-day variations due to varieties in travel behavior may be equally important in some application, for instance to connect the weekly schedule of activities of road users, or to predict the effects of management measures. Cascetta and Cantarella proposed the use of a doubly dynamic Markov model to account for both day-to-day and within-day fluctuations, Cascetta and Cantarella (1990)). Hazelton (2008) considered the problem of estimating time-varying OD matrices from sequences of traffic counts taken over a given observational period from day-to-day and proposed also a Markov Chain Monte Carlo simulation approach to solve this problem, but his approach is still of small practical value as its computational complexity makes it inapplicable to real sized networks.

I.2.4.2. OD estimation incorporating dynamics in traffic flow propagation

In a time-dependent traffic assignment model not only the route choice and the resulting route flows must be described, but also the interaction in time between vehicle streams. Beside the equilibrium assignment rules, which essentially are time dependent extensions of the rules for a time-independent case, there must be a model for the flow propagation in the network. This model is often controlled by a special subroutine, called the Dynamic Network Loading (DNL) procedure (Peterson, 2007).

DNL models are generally classified into two main categories. The first group consists of analytical models, which describe the average behavior of traffic with macroscopic traffic flow variables such as inflow rates and travel times. The traditional analytical approaches fail in capturing the spatio-temporal effects of congestion in a network. This has motivated the development of models based on the Theory of Kinematic Waves, which has the advantage of capturing the effects of congestion more realistically. Well known examples of such models are the Cell Transmission Model (Daganzo, 1994) and the Link Transmission Model (LTM), developed contemporarily by Gentile et al. (2007) and by Yperman (2007). Frederix et al. (2008) showed that dynamic OD estimation results are considerably different when LTM is used instead of a more simplified model of congestion based on either vertical or horizontal queuing.

A second group contains the simulation-based models, which keep track of individual vehicles, or vehicle packets, at each time step. Such models describe certain traffic phenomena more accurately, though at a higher computational cost. Widely used models of this type in OD estimation are DynaMIT (Ben-Akiva et al., 1998), or Dynasmart (Jayakrishnan et al., 1994). Examples of OD estimation approaches using these models and traffic counts are respectively the ones of Antoniou et al. (1997) and Balakrishna and Koutsopoulos (2008) for DynaMIT and Zhou et al. (2003) for Dynasmart. A more detailed overview of the different simulation-based DNL models and their properties can be found in (Peeta & Ziliaskopoulos (2001) and Zhang & Nie (2005)).

I.2.4.3. Dynamics of OD estimation from traffic count variations

Relationships between link flows and OD demand when both are allowed to vary across intervals of the reference period are far more complicated than in the previous approaches (Cascetta, 2002). If on the one hand a correct modeling of traffic flow dynamics is fundamental for a correct estimation of travelers' costs over the day and in turn to capture drivers' travel behavior strategies as time-dependent effects both in a within-day and in a day-to-day time horizon, on the other hand the problem becomes mathematically challenging. It is for this reason that in practice all approaches, which incorporate these two dynamic factors, follow either a simultaneous or sequential, iterative approach, where route-departure time choices are preliminarily determined by fixing the travel costs and later these costs are updated to become consistent with route choices, which are assumed invariant at this stage. The process becomes therefore a sequential correction towards a consistent solution where the measured time-varying traffic counts are justified by time-dependent route and departure time choices represented by time-varying OD tables.

Examples of such models are those of van der Zijpp (1996), Okutani (1987), Chang and Wu (1994), Ashok and Ben-Akiva (1993). Van der Zijpp (1996) proposes an approach based on time-space trajectories, Chang and Wu (1994) use a random walk model, Okutani (1987) described the dynamics through an auto-regressive formulation capturing serial formulation across OD flows of subsequent time intervals. Ashok and Ben-Akiva (1993) propose an auto-regressive process as well, but based on the deviations between actual and historical OD flows.

I.2.5. Network Sensor Location Problem (NSLP)

The above methods aim to deduct OD trip tables assuming that link counts are given together with their positions. Therefore these methods have been developed to obtain as much information as possible from existing counting points, while no interest has been given on the relative quality and usefulness of each detector and thus each piece of information obtained from it. This research question was highlighted only from the 90's by Lam and Lo (1990) and later by Yang and Zhou (1998).

Since the OD estimation problem based on traffic counts has been devoted initially to existing network monitoring systems, the measure of the reliability of information has been always determined in a final task of the whole procedure, i.e. it has been used as performance measure or to derive confidence intervals. Numerical simulations have been used to check the sensitivity of estimates to the input data but no regard was given to real quality of the set of links monitored. In this section we deal with the inverse problem, i.e. finding the minimum number and the position where to locate traffic counting points. This problem has been addressed as the Network Sensor Location Problem (NSLP). Despite the extensive literature found on OD estimation given link counts, the research on NSLP problems is relatively small.

I.2.5.1. Maximal Possible Relative Error method

Yang and Zhou (1998) formulated a rigorous mathematical framework for the NSLP problem, which inspired most of the research thereafter, which is mainly based on the rules set in this article. The final objective in Yang and Zhou's approach is to find the set of link count locations and the minimum number that minimizes the Maximal Possible Relative Error (MPRE). The developed framework is founded on the following (intuitive) rules:

1. *OD-covering rule*: the traffic counting points on a road network should be located so that a certain portion of trips between any OD pair will be observed;
2. *Maximal flow fraction rule*: for an OD pair, the traffic counting points should be located at the links so that the flow fraction between this OD pair out of flows on these links is as large as possible;
3. *Maximal flow-intercepting rule*: within a set of links, the ones to be monitored should intercept as many flows as possible;
4. *Link independence rule*: the traffic counting points should be located on the network so that the resultant traffic counts on all chosen links are not linearly dependent.

Constraints can be assumed for the upper number of detectors for e.g. budget reasons, as later Chung (2001) pointed out. Since the framework is independent of the selection of the type of assignment (proportional, static or dynamic assignment) it is potentially combinable with any of these techniques, as a sort of bi-level problem. The same authors proposed solution algorithms to solve this problem.

Although its apparently sound methodology, two main pitfalls can be shown in this criterion. The first one is that the MPRE method selects the locations in order to minimize the error from an a priori matrix (or from known link flow proportions from the different routes that use that link) and assumes that link counts are error-free, which means that if one between the a priori matrix, the route choice model, or the link counts is wrong the locations will also be wrong. The second is the "static" nature of the method. This method, since it is strongly dependent on the selected a priori matrix or route split portions, will certainly be a non-optimal solution in "non-ordinary" conditions, such as large road work areas, or special events, where route splits are certainly affected. However, the method can be used to find the location of the additional locations for extra detectors or cameras in order to cope with this new distribution pattern.

Ehlert et al. (2006) started with the work of Yang and Zhou (1998) and developed a software tool, based on Mixed Integer Programming techniques, which solves the complex problem formulated only mathematically by the previous authors. In this software also the extension of Chung's budget constraint and a set of weight, for ranking OD pairs by importance, were included. Moreover, the software tool can solve two new elements of the NCLP:

- *Second-best solution*: if old detectors are present already in the system it calculates the place where other detectors should be installed;
- *Weighting rules*: some OD-pairs may be of greater importance and interest than others for the manager's viewpoint and be "empirically" favored. In general, even if there is no rule-of-thumb opinion, some links carry on more informative data than others. A relative weight is also assigned to the OD-flow, which takes into account that the reliability of information is sensitive to flow split rates among the different OD routes. The form of the weight function is chosen from Information Theory.

The first new feature considers the case where detectors are pre-existing (and therefore the solution of best detectors to add to the existing ones might be sub-optimal). The second feature is probably the main contribution of the study. By applying a weight that takes into account the proportion of a certain OD flow, which can be explained by a specific link flow portion, one can control the relative importance of that particular part of the flow for the overall OD-flow. The software is shown in the paper to well perform in realistic networks and to outperform Yang's approach.

Bianco et al. (2001) and Gentili (2002) have recently developed solution algorithms for the NSLP that, preliminarily to obtaining the most likely OD-matrix estimation, aim to extend the information gathered from the traffic counts to the whole network. To do so, they adopt a different approach than Yang's, since they consider sensors to be located at traffic nodes instead of links. In this sense, their approach is more suited for OD estimation on small networks, where route fractions may be estimated with relatively small uncertainty.

I.2.5.2. Simulation-based Sensor Coverage

Traditionally, the sensor problem has been dealt with analytically, i.e. OD flows and vehicle counts have been interrelated in a system of equations. Yang (1998) proposed a risk-averse solution in order to minimize the error in traffic count information while maximizing the information inference power. Yang's solution is strongly dependent on the proportional assignment approach adopted, and, as said in the case of OD-matrix estimation problems, DTA-type of approaches are not yet fully developed. Fei et al. (2007) among other studies adopted a simulation approach to solve the problem. The simulation software used was Dynasmart. Adopting from Yang the 4-rule criterion, they propose to solve the problem with the GLS approach as in Cascetta (1984) and a Kalman filtering method to match real traffic counts and the ones simulated with the DTA.

I.2.6. Conclusions on the literature review

This section has presented two facets of the classical OD estimation problem: 1) the inference of information, available from traffic counts, to the OD-flows and -its inverse problem- 2) the optimal location and number of detection points to obtain the true OD-flows.

I.2.6.1. Conclusions on OD estimation based on available traffic counts

Three main approaches have been selected from literature to solve the OD estimation problem: 1) proportional assignment approach, 2) statistical inference approach, and 3) equilibrium assignment approach, the latter being further developed using both static and dynamic equilibrium assignment solutions. The proportional assignment approach (e.g., Maximum Entropy estimator) seems to give full trust to traffic counts, assumed error-free, and the solution found is the closest to an a priori solution that explains these counts. Moreover by assuming proportional assignment, it is bounded to be static and not to catch the dynamics of traffic. The same holds for the statistical inference approach (e.g., Bayesian inference), which nonetheless assumes traffic counts not necessarily error-free. The DTA approach should outperform the previous two methods, but analytical solutions that are successfully implemented are still under development. Simulated DTA is up-to-date the largest applied methodology in the traffic practice, but it is time consuming as it requires several simulations since their solutions are strongly affected by random effects.

Apart from lack of efficient DTA-based solution algorithms, the main question that still affects the estimating and predicting power of OD estimation methods is the uncertainty of the error present in the three main elements of this problem: the a priori matrix, the link counts and the way the estimated OD-flows are projected onto the network. Statistical methods like GLS and Bayesian inference partly solve this problem, heuristics have been also proposed to refine these techniques for more practical problems but the solution of this problem is yet bound to a certain degree of error, hard to be traced back and corrected.

I.2.6.2. Conclusions on the Network Sensor Location Problem

From the inverse problem point of view, the Network Count Location Problem, one major stream of research has been developed from Yang's 4-rule criterion at an already advanced stage of the OD estimation problem from existing traffic counts. Solution procedures have been developed based on these rules and extended for practical issues, such as budget constraints, pre-existing detectors, weighted importance of link locations for specific OD flows. Analytical solutions are stuck at the same point as OD estimation problems and only statistical-Kalman filtering solutions have been applied in realistic networks. This topic shares the same limitations that affect its inverse problem, mainly that it gives a static and deterministic solution. Moreover the 4-rules are largely accepted and intuitive, but very general, while heuristics and rule-of-thumb criteria become often more important in practical applications.

Recently, we proposed an alternative approach to Yang's method, which explicitly looks for sensor locations that can better explain the traffic flow variations (Viti et al., 2009). We will explain later in this document the basics of this approach, which has been used to identify the positions where we proposed to install extra detectors on our study area.

I.2.6.3. Conclusions in view of the BMW project

The various approaches described in this literature review have inspired the OD estimation method used in this project. Despite the many methodologies proposed in the past and mentioned in this document, most of the available approaches are based or inspired on the traditional static OD estimator, and in particular on the Maximum Entropy and the Generalized Least Square estimators. The main advance in the dynamic OD estimation can be appointed to the application of these estimators in combination with Dynamic Traffic Assignment procedures, which explicitly controls the time-dependency of route and departure time choices as function of route travel times, and that model traffic flow propagation by means of Dynamic Network Loading models. This is the approach we adopt in our study.

Mathematical and computational tractability are in real-sized networks, like the one studied and presented in this document, of main priority. On the other hand we need to assure that the approach adopted correctly deals with the time-dependency of flows and network states both in a within-day and in a day-to-day time horizon, to identify the route and departure time choices of road users and link these with the scheduling of activities at destination. The approach adopted is therefore more in line with the OD estimation using both dynamics in the travel choice process and in the traffic flow propagation. In accordance, the solution algorithm adopted uses a simultaneous estimation of OD flows using traffic counts during the whole day, and weekly changes in traffic patterns are identified by estimating different OD tables for each day of the week. According to the approaches described in the previous sections, the within-day dynamics are ensured by adopting a proper Dynamic Network Loading model, while the day-to-day dynamics are controlled by the OD estimation procedure, which follows a statistical inference approach (Generalized Least Square estimation). We will describe more in detail the chosen methodology further in this document.

II. PRESENTATION OF DATA COLLECTIONS

II.1. Common ground: survey sites evaluation and selection

Clearly, to study the weekly mobility from the two different approaches proposed in this project, namely the activity-based approach and the traffic count-based approach, a common study area had to be chosen, which could be interesting for both sides. A list of practical requirements to candidate survey sites which allowed selecting a suitable case study was outlined for this purpose. This list of requirements included the following issues:

- availability of maps, traffic detection, registration of road works and incidents;
- availability of traffic models (network description, traffic demand matrices);
- availability of previous analyses on traffic patterns or activity surveys;
- complexity/regularity of the traffic situation;
- geographical spread of activities, residential areas etc.;
- absence of potential disruptions during the data acquisition period (special large events, road works);
- willingness of regional authority to cooperate (among other: for selecting sample of respondents);
- willingness of regional authority to support additional traffic measurements;
- potential willingness of regional authority to do or fund additional analyses.

On the basis of the above requirements the area around the city of Ghent was selected for our analysis. This site is rather compact and centered around the city, as it is described more in detail in the next section, has many data and traffic models available. And the municipality was willing to collaborate and accepting both to select a sample of respondents for the survey from its records and to share both traffic data and models.

It should be pointed out that although the study requires a common study area, the boundaries of this area are not necessarily the same for the two approaches. In fact in the OD estimation from traffic counts, as explained earlier, a consistent part of the estimation errors are due to the network and modeling simplifications. In particular, the edges of this area suffer particularly of these errors (we will refer to these errors as *boundary issues*). It was therefore necessary to study a reasonably more extended area than for the survey case, to be sure that these boundary issues were confined to an acceptable level. In the following section we give a description of the study area defined for the OD estimation from traffic counts.

Although the municipality of Ghent could provide us with a detailed model of the network it was necessary for our purposes to simplify this network, for the sake of computational tractability of the OD estimation problem. This simplification has certainly yielded an extra degree of error in the problem, which we will acknowledge and consider in the analysis of the results.

Another important aspect we want to highlight is that although the area is partly equipped with loop detectors, which cover most of the highways and the main provincial roads connecting Ghent with its surrounding cities, they were not sufficient to obtain a reliable estimate of OD trips, as large uncovered areas could be identified.

We will discuss these issues more in detail in the next sections.

II.1.1. Description of the Ghent network

Ghent is the capital and biggest city of the East Flanders Belgian province. The municipality contains a population of about 250000 inhabitants, mostly contained within the urban ring road R4, while many satellite towns are centered around the city, forming a metropolitan area with total population of about 600000 inhabitants widely distributed in mid-sized towns, the largest being Aalst, Deinze, Dendermonde, Eeklo, Oudenarde, Lokeren, Beveren, Sint-Niklaas and Temse.

It is a busy city with a port and a university, and a number of research oriented companies are situated in the central and southern part of the city. As the biggest city of East-Flanders, Ghent has many hospitals, schools and shopping streets. Moreover, the city is an important touristic attraction. Therefore, it is easy to understand that Ghent acts as an important attractor for daily activities for the whole province.

Figure 11 shows the road transportation system centered on Ghent. By car the city is accessible by two of the country's main roads:

The E40/A10: connects Ghent with Bruges and Ostend to the west, and with Brussels, Leuven and Liège to the east.

The E17/A14: connects Ghent with Sint-Niklaas and Antwerp to the north, and with Kortrijk and Lille to the south.

Moreover, alternative connection with Antwerp is the Expressway A11 located at north of the city, near the border with the Netherlands.

In addition Ghent also has two ring-ways:

The R4: connects the outskirts of Ghent with each other and the surrounding villages, and also leads to the E40 and E17 roads.

The R40: connects the different downtown quarters with each other, and provides access to the main avenues.

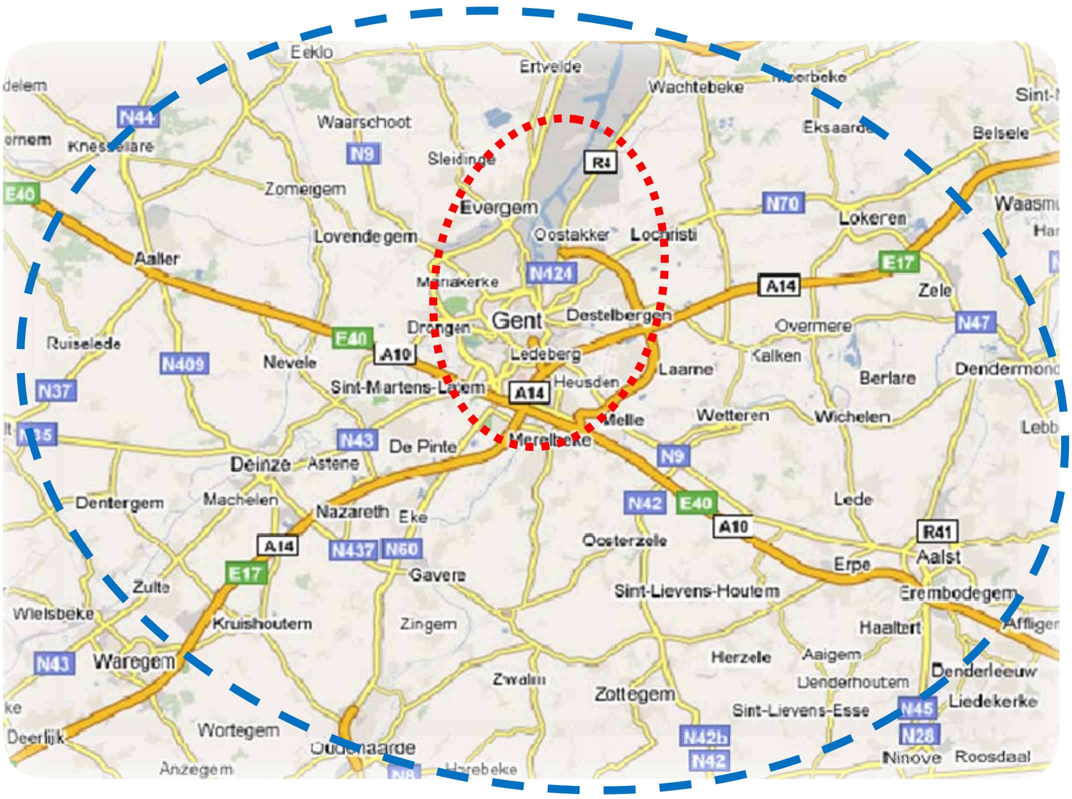


Figure 11: Google Maps™ view of the study area. In red dots the area where surveys have been collected is identified, while in dashed blue the area used for OD estimation from traffic counts is identified

Apart from the main accesses constituted by the above-mentioned motorways, the metropolitan area is richly supplied with provincial roads, connecting Ghent with all towns in the province. From the modeling viewpoint we need to define the relevant aspects of this area for representing the supply system. This is fundamental for a correct estimation of the demand system, represented by the generated OD flows and their distribution on the network.

The supply model has two main functions: the first one is to enable one to simulate the performance of the transportation system in terms of, e.g., level-of-service, travel times and costs, while the second is to contribute through these performance measures to the assignment of flows on the network on all routes connecting each OD pair. Given the complexity of both functions and especially of the algorithms to calculate these parameters simplification of the network is necessary. The supply system is therefore determined mathematically in two steps:

1. *Zoning*: the geographical area is subdivided in a number of sub-areas, to cluster the origin and destinations of each road user in a limited number of access-egress points;
2. *Graph representation*: the road system is simplified using Graph Theory, to represent the main topological structure of the supply system, which is relevant for our study.

We examine and describe the two steps in the following sections.

II.1.2. Zoning

In real networks, individual trips can originate and end in a much dispersed way. However, for the sake of mathematical and computational tractability and for an easier interpretation of the results, it is necessary to simplify it by discretizing these origin and destinations to a limited number of points in the area delimited by traffic zones. Trips from one zone to another are commonly referred to as interzonal movements, while those originating and ending within a zone are called intrazonal movements. The latter are often neglected in practical works, and accordingly they will be neglected in this study.

The subdivision of the study area into traffic zones is a very important step for the OD estimation process, as a significant part of the modeling approximations are due to this simplification. There is no systematic way of defining the way an area should be subdivided into zones, as it may differ considerably depending on the type of problem analyzed and the purpose of the study, the geographical and social layout of the area, the network topology, etc. In some studies zones it can confine one or many urban areas, for instance in regional planning applications, but it can also consists of a few blocks of houses, for instance in the case of urban planning applications.

Each zone is represented by a point where all trips start and end, usually called centroid. This point is connected to the network through connectors, which are fictitious links usually characterized by infinite capacity and arbitrary length. A number of criteria and rules-of-thumb are typically used to define these zones. Here we outline a few of these criteria:

- The total number of zones N should be delineated by the complexity and mathematical tractability of the problem. In the case of OD estimation problems this number is important not only to keep the number of estimated parameters to a limited extent (the number of OD flows per time period is straightforwardly N^2), but to keep the number of routes to a tractable extent, which is normally much larger than the number of OD pairs.

- Since every zone is represented by a centroid, there should be some physical reason to place the boundaries of a zone around this point. The easiest case of course is when a major city is located in the center of the zone.
- Edges of a zone can be identified by physical separations, where interactions between two points of an area are less likely. For instance rivers or cliffs, can be natural separators, but also railways, or motorways can be human-made separators.
- Administrative and census areas can also be useful to define a zone, as they are also based on similar criteria used in transportation problems. Moreover, this choice can also be beneficial to identify important statistical parameters available at municipalities, such as number of inhabitants and their characteristics.
- Different levels of detail can be used to define different zones, depending on the study and the purpose of the analysis. It is for instance the case of our study where we chose to use a different level of detail for the zones within the ring road of Ghent and the outside area.
- A zone should represent a portion of geographical area with a high degree of “homogeneity”. For instance it is recommendable to separate zones where the main activities differ considerably (e.g., residential area, industrial areas, etc.).

Based on the above criteria we have defined 26 zones within our study area, respectively 9 zones for the city of Ghent and 17 for the outside, and 6 zones outside of our study area, thus a total of 32 zones. Figure 12 and Figure 13 show these zones.

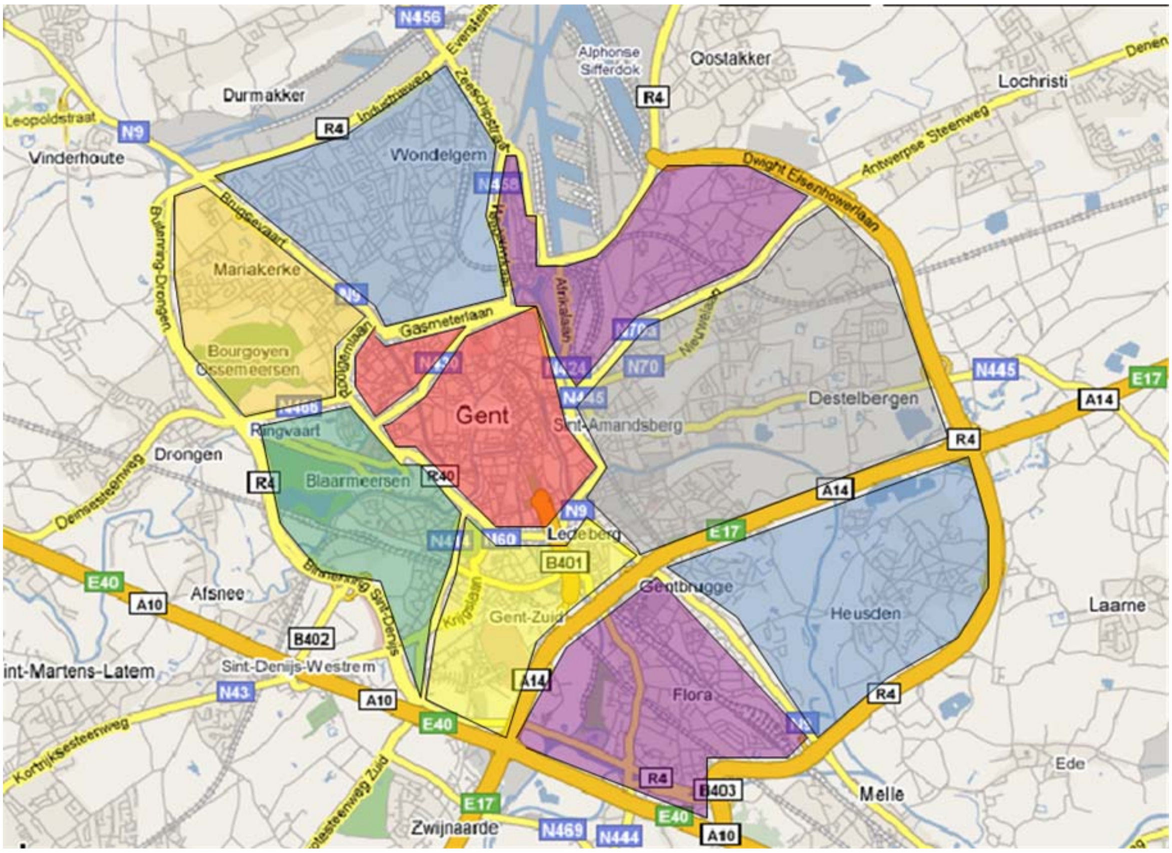


Figure 12: Zoning within the ring of Ghent

The zones within the city have been chosen mainly based on the census zones defined by the municipality of Ghent, while outside of the city the main criterion used was to center the zones to the main satellite cities.

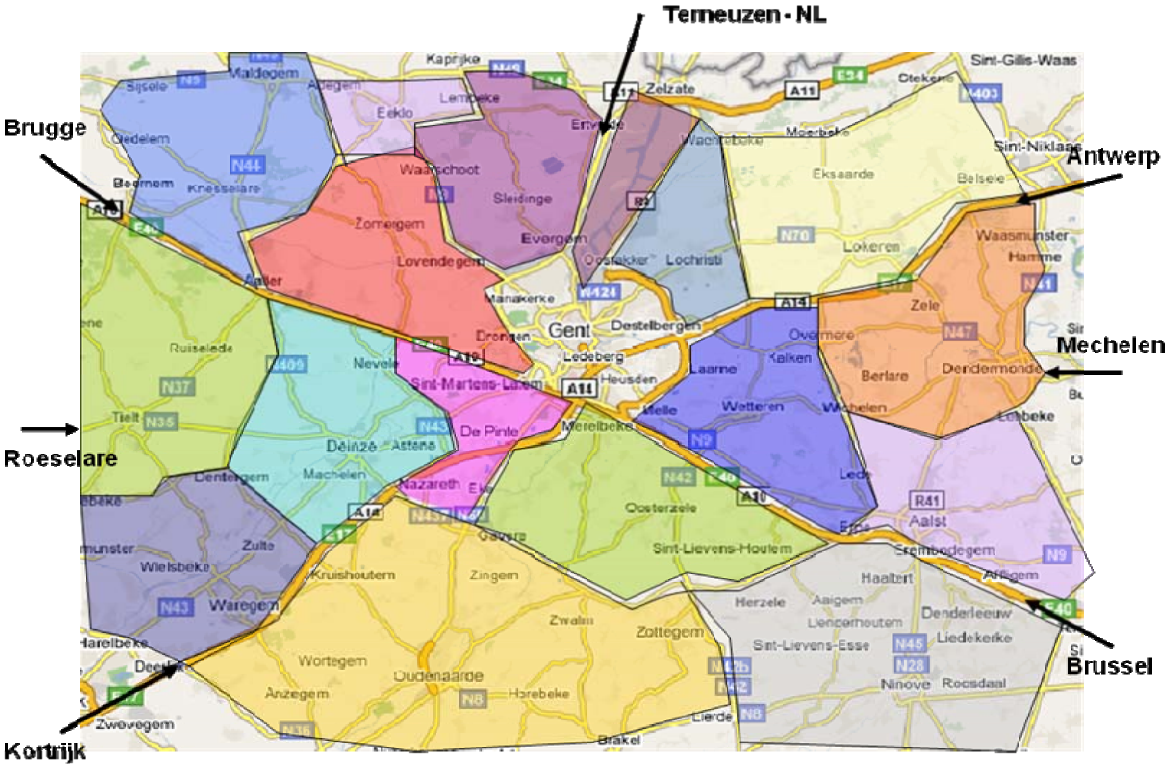


Figure 13: Zoning outside of the city

In Figure 13 we also show the access to the area from the main cities, respectively Brussels, Mechelen, Antwerp, Terneuzen, Bruges, Roeselare and Kortrijk. We assumed that all traffic in the region mainly originates from the defined zones and from these cities.

II.1.3. Graph representation

A network is traditionally modeled as an oriented graph represented by a set of nodes N connected by a set of links L. Each link is normally unidirectional so a road with two directions is commonly represented by two parallel links. According to graph theory the characteristics of a link are determined by a number of parameters. The following parameters are necessarily assumed for our study:

- The *length*, representing the nominal distance traveled by a vehicle. In a graph representation is not necessary to represent the link in detail, e.g., with curves.
- The *speed* at which vehicles drive onto the link. This is normally a function of the speed limit and some other correcting factors (e.g., slope, presence of parking lots and garages, condition of the road surface etc.).
- The *capacity*, which represents the maximum number of vehicles that can pass through a section within a certain time period. This is usually determined by the same speed, in combination with the size of the section (number of lanes, width, etc.). It is commonly assumed that a link is *homogeneous*, i.e. the capacity is invariant at any transversal section of the link, and normally it is assumed equal to the capacity of the section with the smallest capacity (usually called *bottleneck*).
- The *direction* on which vehicles move from a starting node to an ending node. Thus a link can be identified by either an ordinate number or by a specific pair of nodes.

Any sequence of connected links that start from a specific starting node r (or origin centroid) and ends to a specific ending node s (or destination centroid) constitutes a *route* connecting the pair (r,s). In the methodology part of this document we will formalize mathematically the relationship between links, routes and OD pairs in a network.

An important step for determining the graph representation of the network is the selection of the roads to be represented. Analogously to the simplifications in the zoning process, also for representing the transportation system it is usually not recommendable to represent all roads in the network, again for sake of computational tractability of the problem. Again the criteria with which these roads are selected depend on the purpose of the study, and the way the area has been subdivided in zones. In principle, all roads used for interzonal movements should be considered, while all those used exclusively by intrazonal movements could be left out. It is easy to understand therefore that zoning and construction of the graph are closely connected processes, and that normally they are defined in a sort of iterative process.

Figure 14 shows how we simplified the network for our analysis. On the first picture the network with all national and provincial roads is represented, while in the second one we show our simplified network. Figure 14(a) shows the complete road network system around the city of Ghent consisting of national and provincial roads. Thus local roads have not been printed for easing visual illustration. A large part of these roads (mainly those denoted by three digits by the Belgian national road authority) serve mainly towns within the same zone, thus they are used for mainly intrazonal movements. For this reason, the large majority of these roads have been neglected in our simplified network, reprinted in Figure 14(b).

Figure 14(a): original network with all national and provincial roads in the network

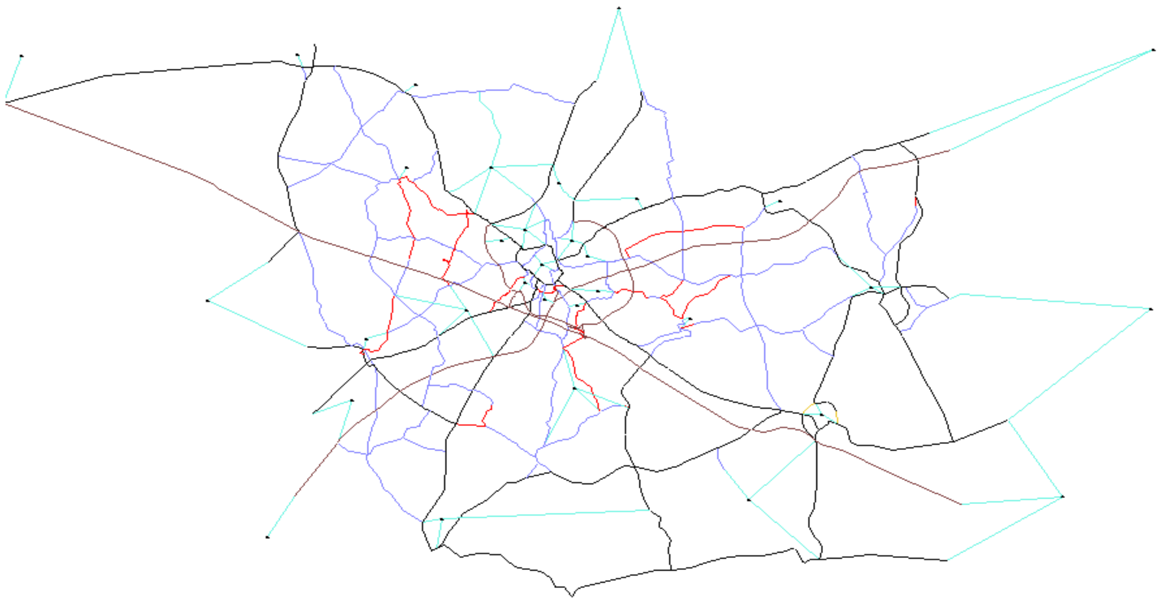


Figure 14(b): simplified network used in our study

Figure 14 : Original road network provided by the municipality of Ghent (Fig. 14(a)) consisting of all national and provincial roads around the city of Ghent, and simplified network (Fig. 14(b)) used for our study

In the bottom picture, different colors have been used to represent the different road levels, respectively with purple we denoted the national roads, in black the provincial roads of the first level defined by one digit (e.g., N9), or regional roads, in blue those of the second level, i.e. defined by two digits (e.g., N49) and in red the ones of tertiary level, identified by three digits (e.g., N409). Moreover, in the picture we also printed the position of the centroids with black dots and the connectors of centroids to the network, identified with cyan lines. As one can see, most of the local roads have been left out from the analyzed network, except for a number of roads, which are clearly used for interzonal movements.

II.2. Expertise from international partner

This project has benefited from the expertise of Professor Kay Axhausen from ETH Zurich, especially in the start-up of the project. A one-day work meeting was organised on January 11th 2008. Most issues of the project were discussed :

- regarding the traffic data : estimation of the Origin-Destination matrices from the available traffic counts, reliability of O-D Matrix estimates from traffic counts (models, sources of error, …), dynamics of the O-D matrices, sensor location problem for extra sensors, importance to know the freight (goods transportation) as well as the share of through traffic;
- regarding the survey : duration and period, motivation of the respondents, survey protocols, survey realization (self versus subcontractor), metadata. A first version of the questionnaires was also discussed in depth from both form and content sides.

Later, Kay Axhausen gave us advices on budgeting the survey and improving the final version of the questionnaires (before and after pre-test).

II.3. Behavioural survey: methodological aspects

As mentioned in the introduction, a deeper knowledge of mobility behaviours over a week, which is our goal within this project, needs conducting specific mobility survey. Nevertheless some preliminary results (see II.1) were already drawn from previous data collected for Belgium, namely from the MOBEL survey, but these analyses were evidently limited by the fact that this survey didn't collect trips of individuals on a week but on one day. Therefore the insights came only from separate views of the different days of the week, but not from a wide picture of individuals' weeks. Hence a specific "weekly" survey appears really necessary to improve the knowledge of mobility behaviours on a week.

Such a survey requires an appropriate protocol, because it is clearly not easy to survey people all along a week. That is why we relied, among others, on the skills developed at ETH Zurich (Prof. K. Axhausen), which has already conducted such a kind of survey (Mobidrive) in the past. In the framework of this collaboration, a complete review of the planned protocol and a deep discussion on how to solve outstanding questions were carried on.

II.3.1. Survey protocol

The main objective for the survey is clearly gathering trips over a whole week. Such a goal has an impact on the protocol to be taken into account. So, regarding the sampling method, we decided that the surveyed sample would be drawn from individuals. Indeed, it seems that surveying on a week all household members would be a too heavy task and will be a serious drawback for the success (response rate) of our survey. Phone calls (which are planned to be a way for conducting the survey or, at least, part of it) are also less appropriate for surveying several people since a proxy bias would be present (if one individual answers for the whole family) or the call would be too long or very difficult to conduct (if each household member has to be present and to answer the questions, one after the other) leading to a dramatic decrease in the response rate.

However, surveying only individuals does not allow us to take into account the household internal discussions and agreements on how the global household's activities patterns are spread not only over the days of the week but also among the household members. We agree that this is clearly a drawback but taking also this parameter into account would have meant a more huge survey which could not fit within the time and budget constraints of this project. Therefore we focused on an individuals' survey which would be a first step for understanding how people organize their trips schemes over the week (but, as said here above, without taking into account more generally the global household's organization).

Even if trips are the main focus of the survey, we could not avoid also collecting information on socio-economic characteristics of the individual and his/her household since such data cannot be ignored for analysing mobility determinants. Therefore a part of the questionnaire has been designed for such a data collection. However, we had in mind to keep this questionnaire as "light" as possible since recording trips over a week is already a quite heavy task for the respondent. This socio-economic part of the survey has therefore focused on the main variables which would be of interest for our analyses (especially the ones which could explain differences in weekly trips patterns). Moreover to decrease the burden for the respondents, we apply this part of the questionnaire through a phone call (recruitment call). This phone call was designed to last 20 minutes at the maximum.

On the other hand, the "trips part" of the questionnaire was filled in with paper and pencil (with postal free return) or through a dedicated web interface according to the respondent's best convenience. It is worthwhile to mention that whichever method has been chosen by the respondent, he/she receives a paper questionnaire which could then be used as memory jogger if web is preferred. During the survey, we had to notice that people having firstly chosen the web to answer moved to paper and pencil because once having recorded the journeys on the memory jogger they decided to avoid a second input trough the net (even if, in that case, they had to pay for sending back the filled questionnaire by post) .

Concerning a web protocol, this option has been quite discussed with the members of the follow up committee for our project. Since we thought that a phone support remains quite helpful for this kind of survey and that such a protocol (phone+paper) also allowed us to reach some categories of people who do not have Internet access or are not comfortable with computers (old people, households with low incomes ...). Therefore we claimed that a survey exclusively on the web wouldn't fit our objectives. However the web remains an interesting alternative, allowing another way for answering the survey for people preferring this solution. That is why we eventually decided to use both ways of surveying: phone/paper, and phone/web.

Even if the Internet protocol for surveying avoids the time consuming task of encoding answers, it must not be understood as really an easier protocol. Indeed the design of the dedicated web site is also a huge effortful work. Dynamic pages need to be built to guide the respondent according to his/her previous answers, automatic checking processes need to be coded to avoid, as much as possible, incoherencies in answers, "on the fly" archiving of answers must be forecast to avoid losing too much data in case of crashes or other incidents, confidentiality and restricted accesses must be taken into account and opportunities to spread answers in different sessions (e.g. each day you input your travels for this specific day) must be implemented. Moreover the interface must be kept as user friendly as possible even if the questionnaire is quite long. All these aspects have to be coded and validated before allowing respondents to access the site.

II.3.2. Recruitment

The response rate and thus the recruitment of individuals always being a critical problem for such burdensome surveys, we decided to offer incentives for respondents: 10 euros if the person has filled a complete questionnaire, it means having recorded the displacements for the whole week. This (small) amount was fixed as a balance between what is offered in other countries for such surveys (e.g. survey led by ETH Zurich) and the budget which was available for this project.

II.3.3. Sample

The planned goal was to gather 600 validated questionnaires. This figure seemed a good compromise between the needed amount of answers for ensuring the statistical significance of our analyses, and the available budget.

Nevertheless, we had to adapt this figure according to the tenders of subcontractors in charge of achieving the work in the field, if we wanted to fit our budget. Hence we dropped our goal to 500 respondents. Fortunately the response rate was much better than expected and the chosen subcontractor discounted additional answers so that we were able to receive 717 validated questionnaires for the available budget (including some financial effort from the scientific team's side since we thought it would be a pity not to benefit from valuable supplementary data). This figure does not include questionnaires (76 respondents) from the pre-test conducted to assess the protocol and the questionnaire.

Drawing of our sample first means choosing the surveyed area. As explained before, the municipality of Ghent seemed favourably encountering several necessary criteria:

- matching with the area on which the countings will be performed for the « traffic data « part of the project, allowing comparative analyses between both types of data
- willingness to give access to the population register
- willingness for collaboration from authorities
- possibility, as we would like, to restrict interferences with other ongoing surveys (on the Flemish region).

A random sample has therefore been drawn from the Ghent population register. Sampling was achieved on individuals from 12 to 75 years, with stratification according to the household size (single vs. other household types since it is well known that the response rate is quite lower for one person households, see e.g. the pre-test realized for the MOBEL survey, BARBIERI et al., 1998), the gender and the age. It must be kept in mind that our stratification strategy is limited by the only variables which are available in the population register. Moreover the bounded size of our sample does not allow too much strata without disturbing the statistical relevance of the analyses.

As a part of the survey was planned by phone (socio-demographic part of the questionnaire), we cut our sample to people for whom we were able to find a phone number. We were aware that this could introduce a bias in our sample, but the heaviness of our survey imposed phone contact to help and motive the respondents, as well as to collect a part of the data.

All the individuals who were finally retained in the sample were then warned through an official letter, signed by an authority (Ghent mayor) and invited to take part to the survey. An element worthwhile to be noticed is that we mentioned, in this letter, the phone number we have found and which we planned to use to contact the respondent. Many people let us know that this number was wrong (and gave the correct one) or that they preferred to be reached by GSM (also giving their mobile phone number). Such a small trick also improved the response rate.

II.3.4. Questionnaires

As mentioned above, "Household" and "Individual" parts of the questionnaire (i.e. socio-economic characteristics), have been lightened compared with questions classically used in trips surveys (e.g. MOBEL) but remained enough accurate for the analysis needs regarding mobility determinants. This was essential to avoid, as much as possible, burdening the respondents. So the "household" part was mostly limited to a few questions concerning the household composition, the owned vehicles and the home. The "individual" part brought us information on the usual mobility behaviours especially the "home-work" or "home-school" ones.

On the other hand, the "trips" part constitutes the core of our survey. We first thought about collecting activities rather than trips. Such an idea was motivated by the fact, often reported in the literature (see e.g. ARMOOGUM et al., 2005), that fewer trips are forgotten by this method, and that it is easier to point out the gaps in consecutive activities (and thus the forgotten trips) through such a methodology. However, we changed our mind after having tested the questionnaires, and seeing that reporting every activity of a day is quite heavier than reporting trips. Thus the gathered information covers the destination, the purpose of the trip, its starting and ending times, the used transport modes, the trip distance and its duration, and some other variables (accompaniment, park fees,…). Let us remember that individuals were invited either to encode their answers to this questionnaire or directly via the web, or with paper and pencil on the provided memory jogger to be then sent back by post (free) to the firm in charge of the survey.

Copies of the questionnaires are provided as annexes to this report.

II.3.5. Pre-test

As one week surveys are quite new in Belgium, the methodology is, for this project, as important as the expected results. Since there were remaining questions about the methodology, we decided to start with a pre-test, first to verify if the options we took (trips or activities, web-survey, comprehensibility of the questions,…) and the hypothesis we rely on (e.g. response rate) were the good ones, and second to test some topics we didn't resolved (mainly in the layout of the questionnaire).

We planned to survey 50 individuals in this phase where we tested 2 layouts, to examine the impact of several changes in the presentation.

The differences between trip agendas are presented here below:

Version 1	Version 2
- four pages per day for recording trips	- every trips are recorded one after the other, and the day has to be noted on the top of each trip
- amount of professional trips is asked at the beginning of each day	- amount of professional trips are asked at the beginning of the questionnaire with a reminder at the end
- if there is a day without trips, it is asked why at the end	- no questions about the reasons of days without trips, to avoid "soft refusal"
- order of the questions : destination, goal, time of departure, …	- order of the questions : day and time of departure, goal, destination, ...

Table 1: comparison both tested diaries

The pre-test phase lasted two weeks and 500 individuals were contacted. Eventually we received complete and validated answers from 76 individuals. The obtained response rate (14.6%) was over our expectations (10%) as shown in the Table here below

Sent announcement letters	500

agreements	124	24.8%
refusals or unsuccessfull contacts	376	75.2%
	500	100.0%

sent back diaries	76	61.3%
not sent back diaries	48	38.7%
	124	100.0%

Table 2: response rate for pre-test

Such a good response rate (having in mind the complexity of the survey) allowed us to reduce a bit the amount of persons to contact in order to obtain our 500 final valid forms. Due to our expectations of a 10% response rate, we would have to contact 5000 individuals, but we dropped this figure to 4000 (3333 would have been sufficient with a response rate of 15%, but we preferred to remain cautious and keep a margin, in the case of a lower response rate for the "real" survey).

The pre-test phase also allowed us to note that there were only slight differences between the two trips agenda designs (7 "one-day agendas" or 1-week agenda with weekday to tick on each trip); e.g. there was no significant difference in the amount of trips per day, as shown in the tables below.

Number of trips, version 1 : 4 pages/day (42 returned forms)								
	Day 1	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**	**Week**
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.00
Mean	4.12	3.74	3.60	4.26	3.29	3.83	3.64	26.48
Max	8.00	10.00	11.00	11.00	9.00	10.00	10.00	45.00

Number of trips, version 2 : 20 pages/week (34 returned forms)								
	Day 1	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**	**Week**
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	12.00
Mean	3.82	4.09	3.68	3.53	3.53	3.18	3.65	25.47
Max	10.00	11.00	11.00	9.00	12.00	8.00	11.00	52.00

Table 3: comparison of the number of trips per day according both versions of questionnaires

For easiness, we decided to keep the version with fewer pages i.e. the 1-week agenda.

One other result we could deduce from those figures is that the amount of trips was not decreasing along the weekdays of the reference week. Whilst we were afraid, before starting this survey, of having a high amount of trips during the very first reference days and then a decrease in recording trips due to fatigue or tediousness, we could draw from the pre-test that it was not the case. Moreover the average number of trips per day was rather high, comparatively with other Belgian one-day mobility surveys (conducted with the same kind of methodology).

From the returned forms, we also noticed unequal response rates according to age classes: we found that people born between 1975 and 1984 had a lower response rate than the others. Going deeper in the analysis, we could infer that such a situation is not really to be attributed to reluctance of these people to answer the survey, but is mainly due to a greater difficulty in finding a phone number for people in this age class. Even if the global phone finding rate seems quite low (45%), it becomes dramatically smaller (23%) for people of 20 to 30 years. A possible explanation would be that these young adults are more mobile, are more often single and therefore have less interest for a «fixed telephone». Illustrating this problem, we give here below the phone finding rate for the different age classes:

Age classes (date of birth)	Phone finding rate
1933-1944	60.3%
1945-1954	50.1%
1955-1954	45.5%
1965-1974	43.7%
1975-1984	23.4%
1985-1996	43.5%
global	45%

Table 4: phone finding rate per age class

According to these figures, we decided to modify the stratification scheme for the sample in order to obtain a set of respondents closest to the real population: for example, we increased the amount of young people in our sample, in a way to also obtain, after finding of phone numbers, a good representativity of this part of the population in our sample.

II.3.6. Data collection

The data collection phase began on September 1st (2008), and ended in December (2008). The selected market firm Phonecom realized all the field work, except the web site, realized by the university team.

How took place the survey in concrete terms?

- 2 weeks before their "first day of reference", selected individuals received a formal letter (signed by the Mayor of the city of Ghent, the Director of the GRT, and the Chairman of Belgian Science Policy Office), presenting the objectives of the survey and guarantying the seriousness and the respect of the privacy (a declaration has been made to the privacy policy commission for this survey in this aim). In this letter, we also mentioned the phone number found to reach the person, and we invited people to call us or send us a mail if this number was wrong or not the appropriate number to call the person. At our great surprise, we received many messages correcting this phone number.
- 5 days before their "first day of reference", selected individuals received a phone call, asking them if they agreed to participate. If yes, individuals and household characteristics were collected at that time by phone. Then they were asked if they prefer to fill in the trip part of the survey by paper or by web. If by web, we noted an e-mail address. In both cases, the paper questionnaires were sent to people who accepted to answer the survey, for the "paper-method-ones" with a pre-paid envelope to be sent back when filled, and for the "web-method-ones" as memory-joggers.
- 2 days before the reference week, all agreeing individuals should have received the questionnaires by post, and "web-method-ones" an e-mail with the URL of the website, and their private login(in order to be able to connect the web-site as many times as they want but also to restrict access to only sampled individuals).
- 2 days after the starting of the reference week, individuals are called for motivation (to see if everything goes right, to provide answers to potential questions,...). For those who are filling the agenda via the Internet, motivation and recall e-mail were sent, and if these mails had no effect, the survey firms called them back by phone to remind them to answer the survey.

- Some checks were undertaken once the questionnaires are sent back, and the firm called the respondents with suspicious ones to verify their answers. We asked the firm for a minimum of 25% checking calls.
- Individuals who did not send back their agenda were also called back.
- Once the survey completed, all the respondents with complete and validated questionnaires were rewarded through a bank transfer of 10 euros.

Remark about the web method:

We received a very low number of agreements to fill in the agenda via Internet, and a large part of those people changed their mind during the reference week, and finally sent us their paper forms back by post. This can be explained by the fact that the dedicated website was not so easy to use (particularly for one week!), but also because people received paper forms by post, to help them to remind trips made during the week if they were not encoded immediately, and so, once the work was made on the paper, it was easier for the respondent to send the paper back rather than starting to encode all the trips on the website.

In total, we received 30 completed web forms (in addition to the 717 paper collected). Because the merging of both methodologies in a common database should have raised questions that we could not solve with such a little sample (it is difficult to set up statistical tests to check the coherency of both methodologies with only 30 people for one of those methodologies), we decided to left those 30 web forms.

However, we did some checking on the web forms, mainly to examine if the number of trips was not lower than for paper forms. In the limits of the figures we have, we did not see significant differences for this indicator according to the methodology. It can be due to the fact that it was more an "encoding operation" (copying what was on the paper form) than really responding on the website.

II.3.7. Response rate

From the 4000 individuals contacted with the announcement letter (all with a phone number found), we get 717 valid returned questionnaires, which means a global response rate of 18%. The table here below details the lost of respondents at different steps of the survey:

Announcement letter sent	3992

Phone contact	2140	53.6%
Error or unsuccessfull (2x3 attempts)	1852	46.4%
	3992	100.0%

Agreement	1238	57.9%
Refusal	902	42.1%
	2140	100.0%

Sent back questionnaires	760	61.4%
No answer, etc.	478	38.6%
	1238	100.0%

Validated questionnaires	717	94.3%
Incomplete questionnaires	43	5.7%

	760	100.0%

Table 5: response rate for the survey

II.3.8. Data cleaning, geocoding and weighting

II.3.8.1. Data cleaning

Before starting the analyses, a data cleaning step has been necessary.

Although some corrections have to be done in the individuals' database (socio-economics variables), most errors appear in the trips database. The main reason is that it was easier answering to socio-economics questions than completing the weekly diary. We can also note that these questions were asked on the phone so that they had not to be read and then the answers encoded (process which increases the risk of encoding errors).

In the trips database received from the firm conducting the fieldwork, we observed two kinds of errors:

- encoding errors which are due to the encoders, i.e. they did not exactly input what is written on paper. Such errors are unavoidable even if we control the encoders through the survey firm, give them more advices, and if needed, ask the firm to exclude some "bad" encoders;

- respondents errors : these mistakes can be careless mistakes or due to misunderstanding the instructions. Most of errors of this kind should have been detected by the survey firm so that they have called back the respondents to ask missing information or to correct incoherent answers. Nevertheless some are remaining.

The cleaning process means two steps: first detecting the errors and then, as much as possible, correcting them.

The used detection criteria have been missing data and incoherences. Incoherences can be tested inside a single trip or across trips. An example of a test for detecting the first type of incoherence is the speed computation from trip duration and distance (eventually also according to the main transport mode) allowing to detect errors in these variables. The correct sequence of departure and arrival times for successive trips is a typical test for detecting incoherences of the second type.

The corrections can be automatic, on a case by case basis, or, as a last resort, done thanks to a paper diary form checking.

Some automatic corrections are undertaken when missing data or exact values of incorrect data can be guessed from other variables. For example: if the arrival time is missing, we can compute it from the departure time and the trip duration. More sophisticate automatic corrections, taking into account more variables (also from previous or following trip), have been implemented. It is nevertheless difficult to generate a general code valid for all similar incoherences as different corrections are possible according to the case. We therefore need to remain cautious with such automatic correcting process since it could add maybe coherent but however incorrect information.

Most corrections are thus ad-hoc ones achieved by hand, looking in the trips database for incoherences. Checking the previous or the following trip is often very helpful (e.g. for times, return mode, distance, etc.).

Finally, if the needed correction sounds not evident, it is possible to correct encoding errors by taking the paper diary form out and looking at it. But as it is excessively time-consuming, we only have recourse to this option as a last resort. For example, in the case of a probable missing trip (e.g. if no return to home), it is impossible to guess all the fields (purpose, departure time, destination, etc.) for this trip. Then, we check the respondent's diary form: if it is an encoding error, we can correct it; but if such lack of data is due to a probable respondent error, it is too late to call back the respondent so that it is not possible to (totally) correct it.

Globally, the main variables which have been tested for consistency are:

- coherence between duration, distance and transport mode,

- coherence between duration, departure and arrival times,

- coherence between successive purposes (e.g. two successive identical purposes (such as «home») are an indication of a possible encoding or respondent's error),

- distance (e.g. 2 km 5m in place of 2km 500 m).

As the sequence of trips along the week is even more crucial in a one-week survey and for the analyses we were considering, we focused on the consistency between the successive trips:

- coherence between day of the week and date (day, month),

- coherence between order of the trips and sequence of the date/time variables (both departure and arrival), including trips terminating after midnight.

When incoherence is detected in the trips sequence, the main difficulty is to find which corrections is the right one: changing the order of the trips or correcting the date/time of one or more trips. Indeed some respondents add trips they had forgotten at the end of their diary, so that some of their trips are in the wrong order. A sorting of the trips according to the departure times is not a suitable solution because of mistakes appearing in these times (e.g. times written in 12-hour notation rather than 24-hour notation) and in the days (especially for trips after midnight). For such incoherences, automatic correction is thus impossible and a one-by-one checking is required to adapt corrections to each case.

Finally, another important part of the cleaning relates to the trip destinations (addresses), as a geocoding is necessary in order to compare with OD matrices computed from traffic counts. There the difficulties mainly come from the fact that these are «open variables» (not pre-coded), handwritten, and often not precisely or rightly known by the respondents. The method used for cleaning these variables will be sketched in the geocoding part.

II.3.8.2. Geocoding of the trips

Why a geocoding?

In order to compare with the OD matrices computed from the traffic counts, it is necessary to use common zones. These zones have been described in III.1.2 (see Figure 12 and Figure 13).

In the diary, respondents had to note the addresses of their destination as precisely as possible (city, zip code, street name and street number). This information is not directly usable to know in which zone places are located. Data transformation implies first a correction of the addresses (as people often do not exactly and precisely know them) and then their matching with a geocoding database (to code the information).

How was undertaken the geocoding?

First we recall that, in the diary, respondents could input their destination as the number of a previous trip destination rather than write down again the address (of a place where they had already gone). The purpose was twofold: shortening the time to answer for respondents (and so avoid them get tired of it), but also decreasing errors or misspellings in the addresses. This process also leads to a decreased amount of (identical) addresses to be geocoded.

The geocoding of the locations (origins and destinations of trips) was carried out in different steps.

At first stage, all the different addresses were stored in a separate file where each of them received a unique ID. Hence, all the trips recorded in the trips database had their destinations encoded only through their ID. So addresses do not appear anymore in the trip database which is a main concern for privacy reason. Only statistical sectors and higher level zoning characterize the locations in this database.

Due to this methodological issue, the geocoding had to be undertaken only in the ad-hoc addresses file.

The statistical sectors were chosen as the most disaggregated spatial level taken into account for this geocoding exercise.

The process for going from the plain address to the statistical sector could be detailed through the following steps:

- test of consistency between the municipality name and the postal code and, if needed, necessary corrections performed;
- automatic and by hand corrections of the street name (taking into account the postal code) in order to match with the geocoding database (as the street name must be written exactly with the same spelling in both databases);
- from the postal code, the street name and the street number, automatic determining of the statistical sector of the address;

At an additional step, a zone (covering several statistical sectors) number could be added for compatibility purpose with data from traffic sensors and countings.

At the end, a geocoding file is finally provided associating to each place ID, a statistical sector, a postal code, and an intermediate zone. Thanks the ID, these geocoding data could then be incorporated in the trips database both for the destinations and the origins.

II.3.8.3. Weighting of the observations

The observations have been weighted according to the stratification variables: gender, age (3 classes) and type of household (single or not) and taking into account the margins drawn from the National Register.

II.4. Traffic data : Overview of collected data

II.4.1. Initial matrix: static matrix

As we stressed out in the literature review, a main important parameter for the calculation of the dynamic OD matrix is the initial or target matrix.

The municipality of Ghent provided us with 2 static OD matrices, one for the morning peak (between 08.00h and 09.00h), and the other one for the evening peak (between 17.00h and 18.00h). These two static OD matrices were estimated in a previous project by the firm TRITEL starting from socio-demographic and socio-economic data, on the level of statistical sectors, or else, if not available, on the level of cities. They used this data to make trip matrices for every activity (using the socio-economic survey (SEE 2001)) and mode of transport, using a modal split model based on a Multinomial Logit model. Finally the model was calibrated using traffic count data, and then again a modal split model was used to correct the matrices for the other modes of transport.

The static OD matrices are provided for a wider area than the one used in this study, and contain a considerably higher level of detail for the zoning and the graph representation with respect to our analysis. Therefore a preliminary operation for using this matrix in our study is to further simplify it based on our aggregation criteria. By doing so, we summed up all trips originating from each origin and ending at each destination within our zones, and we excluded all trips originating and ending within the same zone. Moreover we used only passenger car traffic.

Figure 15 gives a visual impression of the production-attraction of trips from each zone in our network. Figure 15(a) shows the case of the morning rush hour while Figure 15(b) shows the one for the evening peak. It is easily observed from these graphs the inversion of tendency for most of the zones, i.e. there is an analogous number of trips originated from a zone in the morning and attracted in the afternoon, and vice versa. This is straightforwardly due to a large portion of commuting traffic observed at these times of the day. The same picture is presented aggregating all traffic starting and ending within the city of Ghent in Figure 16. The very large number of trips originating and ending in the city suggests that there is a very high interaction between satellite towns and the city.

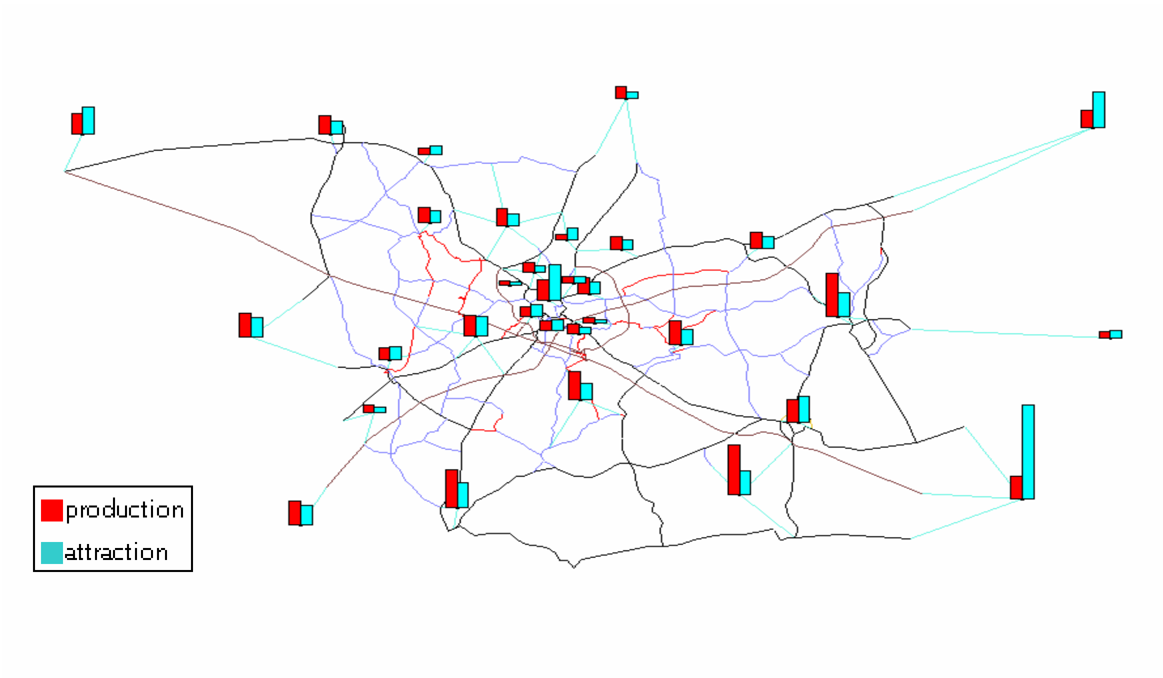


Figure 15 (a): production and attraction of trips for all zones and for the morning peak hour

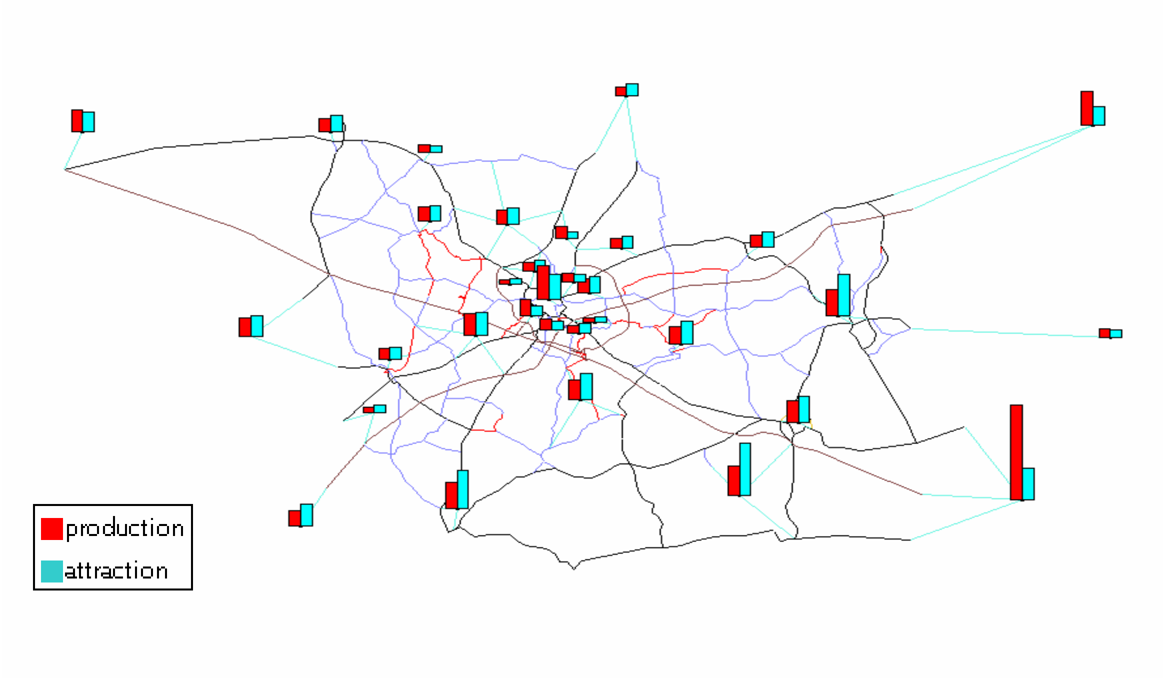


Figure 15 (b): production and attraction of trips for all zones and for the evening peak hour

Figure 15: Production and attraction of trips for all zones in the studied network

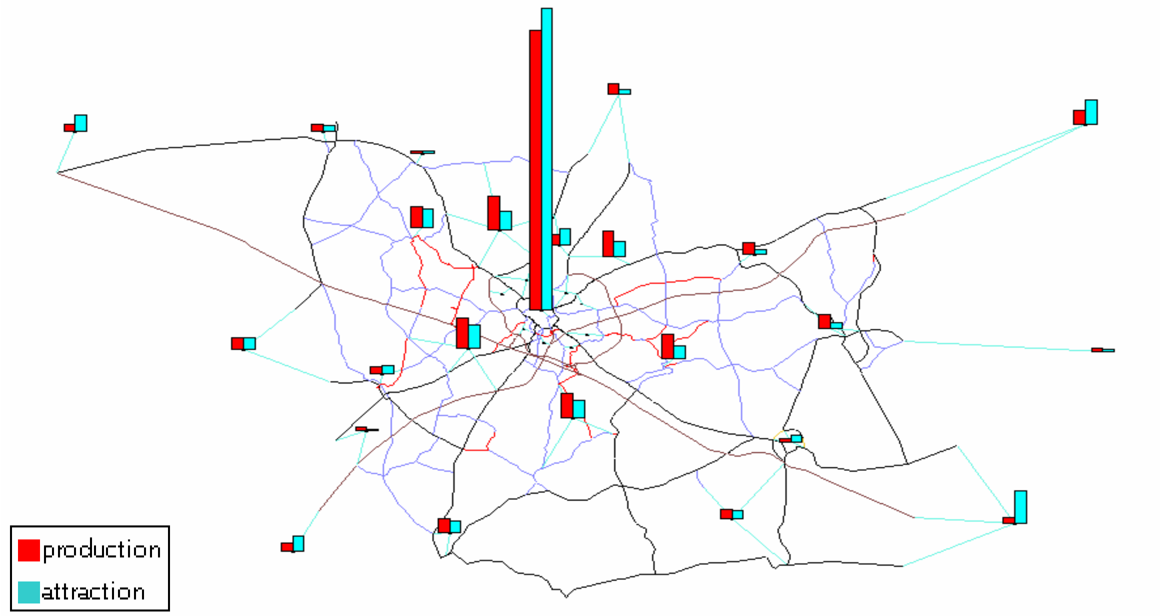


Figure 16 (a): production and attraction of trips in and out the city of Ghent and for the morning peak hour

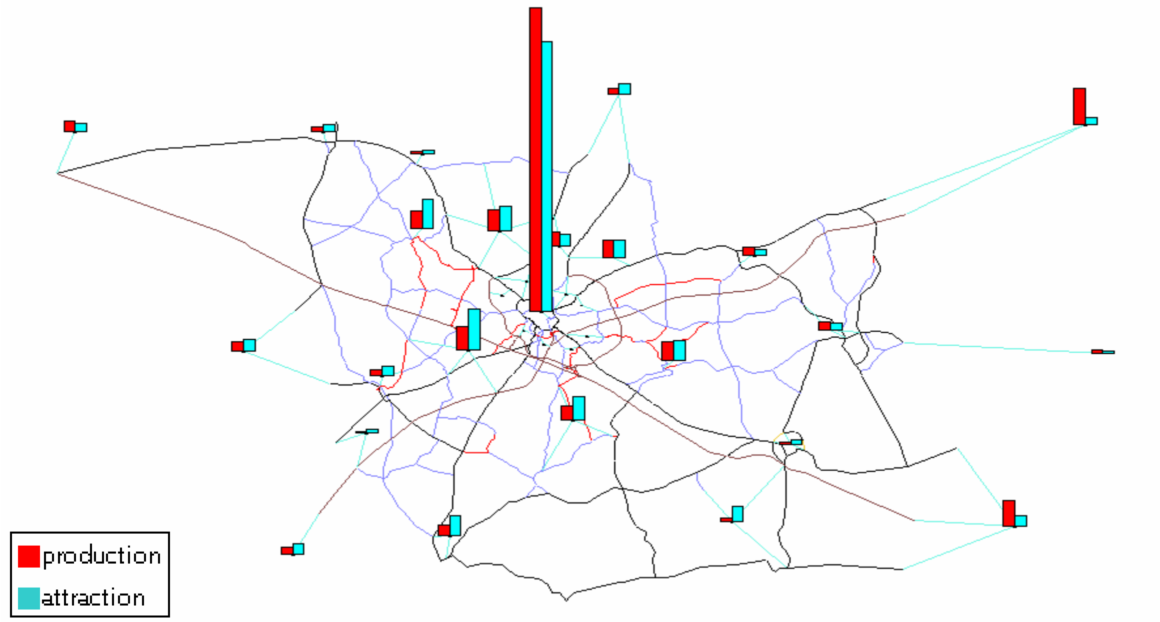


Figure 16 (b): production and attraction of trips in and out the city of Ghent and for the evening peak hour

Figure 16: Production and attraction of trips originating and ending in the city of Ghent

II.4.2. Traffic count data

For this project 3 sources of data were available. The first one consists of data from the Start Sitter database («FOD Mobiliteit en Vervoer - Samenwerkingsakkoord Federale Overheid - Gewesten - START/SITTER»). Secondly there are a number of fixed loop detectors installed near the intersections between provincial roads. The third source consists of data from tube detectors that were installed for this project. We describe more in detail these three datasets in the following of this section, together with the criterion used to locate the tube detectors.

II.4.2.1. Start Sitter data

The Start Sitter database provides all kinds of information regarding the traffic on most of the Belgian highways. The majority of this information is measured by single loop detectors. For this project traffic counts for the A10, A14 and R4 were used, and also speed measurements were used implicitly, and this during the time period between the 1st of September till the 7th of October 2008. This data is available on a one-minute aggregation level. In Figure 17 an overview of the detected links is displayed.

Figure 17: Detected highway links taken from the Start Sitter system

II.4.2.2. Fixed loop detectors

On the secondary road network a number of fixed loop detectors are installed permanently. They provide traffic counts on a one-hour aggregation level. The data was retrieved during the time period between the 1st of September until the 7th of October 2008 to overlap with the data taken from the tube detectors. In Figure 18 an overview of the links with fixed loop detectors is displayed.

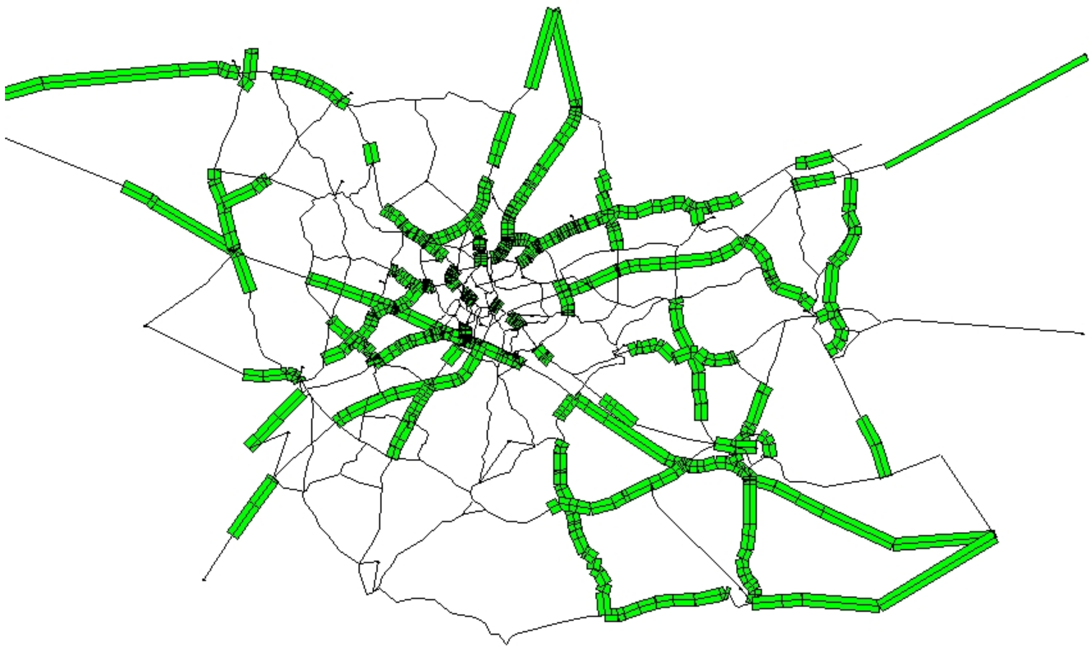


Figure 18: Fixed loop detectors already installed in the network

As one can see by looking at both Figure 17 and Figure 18 there is already a significant portion of links in our study where traffic count data is available. However, there is also part of the network that is not covered by any loop detector. As we discussed in the literature review, this can yield to a considerable error in the OD estimation. It was believed that a number of extra detectors should have been added to these permanent loop detectors.

We describe in the next section how we identified the links where to install a number of extra detectors.

II.4.2.3. Tube detectors

To identify the most significant positions whereupon installing the extra detectors we used two criteria. We solved the NSLP problem using the classical Yang's algorithm (Yang and Zhou, 1998), and more specifically the modification of Yang's algorithm proposed by Ehlert et al. (2006) to account for the presence of pre-installed detectors. In addition, we used a new method to solve the NSLP, which was recently developed by the KU Leuven group (Viti et al., 2009). These methods will be described here after.

II.4.2.4. Yang's Maximum possible Relative Error method

The objective in Yang and Zhou's proposed methodology is to find the minimum set of link count locations that minimizes the error between the estimated trip matrix from link counts and the true one. Within this scope, the authors developed 4 general rules, discussed in the literature review and here recalled:

1. *OD-covering rule*: the traffic counting points on a road network should be located so that a portion of trips between any OD pair will be observed;

2. *Maximal flow fraction rule*: for an OD pair, the traffic counting points should be located at the links so that the flow fraction between this OD pair out of flows on these links is as large as possible;
3. *Maximal flow-intercepting rule*: within a set of links, the ones to be monitored should intercept as many flows as possible;
4. *Link independence rule*: the traffic counting points should be located on the network so that the resultant traffic counts on all chosen links are not linearly dependent.

Yang and Zhou's approach is therefore developed with the scope of covering traffic under specific requirements, i.e. catching as many OD-pairs as possible (Rule 1), or as many traffic flows as possible (Rule 3). Constraints have been thought to discard links with scarce information content (Rule 2) or information redundancy (Rule 4). The same authors propose two basic solution algorithms, which guarantee respectively 1) full OD-coverage, regardless of the total amount of flow caught for each pair and 2) route flow coverage, thus regardless of the actual OD coverage. Since the first approach lacks in considering an important characteristic such as the network flows while the second can give solutions where many ODs are not covered at all, Yang and Zhou proposed a greedy heuristic algorithm to obtain a solution that combines them. The algorithm is characterized by an iterative procedure, here briefly outlined:

- Solve the OD coverage algorithm using Linear Programming and find the minimum number of sensors needed to cover all OD-pairs;
- Find the sensor among the OD coverage solution that monitors the largest link flow and remove this link flow and the portion of flow from all routes that converge to this link;
- Re-compute the OD coverage algorithm and find the new link with largest flow monitored;
- Stop when a maximum value of sensors has been placed, or a minimum route flow percentage has been caught.

Ehlert's modification of Yang's algorithm accounts for the presence of existing detectors by simply giving a zero weight to those detected links, so that these will never be considered in the solution of the Linear Programming problem.

The OD coverage algorithm is based on the concept of Maximum Possible Relative error. Thus, given an oriented network $G(N,A)$, let a denote a link within the network, *A* the complete set of links in the network and $w \in W$ an OD-pair from the complete set W; let also denote by v_a, t_w, t_w^* and p_{aw} respectively the measured link flow, the estimated and the true OD-flows and the split ratio (i.e. the fraction of OD-flows from *w* that passes through *a*). Due to the conservation of vehicles equation, it holds:

$$\sum_{w \in W} p_{aw} t_w = v_a$$

Assuming that link counts are error-free, this equation can be calculated for all links a that are monitored. Thus, for these links the following equation should be satisfied:

$$\sum_{w \in W} p_{aw} (t_w^* - t_w) = 0$$

If we denote by $\lambda_w = (t_w^* - t_w)/t_w$, the MPRE is defined by the following equation:

$$MPRE = \max(G(\lambda)) = \max\left(\sqrt{\sum_{w \in W} \lambda_w^{\ 2} / m}\right)$$

where m is the number of OD-pairs. The solution of the NSLP should therefore be found so that it minimizes the MPRE.

As we already said, Yang and Zhou's method suffers from a number of shortcomings. Specifically for this project, the main limitation is the lack of handling with spatial coverage of the links. This means that although the match between measured and estimated links can be good, the estimate of flows on unmonitored links may be poor. For this reason we reformulated the MPRE using link flows instead of OD flows, as here briefly described.

II.4.2.5. Proposed new NSLP algorithm

The statistical concept of MPRE can be easily adapted to account better spatial coverage. We formulated for this scope the MPRE by using link flows as target values. Let denote by v_a^* the estimated flow on (any) link *a*, while true flow measured by loop detectors be denoted by v_a. We can redefine the MPRE in an alternative way with the following formula:

$$\sum_{a\in \tilde{A}} (v_a^* - v_a) = 0$$

$\tilde{A}$ denotes the total set of links in the network (network-wide MPRE), or any subset may be used (subnetwork-wide MPRE). This formula is therefore proposed with the assumption that the flow calculated using the measured flows should be equal to the flow measured by loop detectors. The estimated link flow can be calculated using any traffic model. In our proposed approach we used the same Dynamic Network Loading model later used for dynamic OD estimation. By doing so we argue that the positions of the detectors will minimize the relative error in the state estimation using that particular traffic model. For a more detailed description and discussion of this approach one can refer to Viti et al. (2009).

By using estimated link flows in the objective function this approach can take into account link capacities, since they are input in traffic models. Moreover, by adopting a dynamic network loading model for computing the travel times on all links one can account for the temporal variation of flows more correctly.

Based on (a combination of) the above criteria, a number of tube detectors were installed specifically for this project. They provide traffic counts on a 15-minute aggregation level. The detectors were not all installed during the same time period. The earliest date a detector was installed was the 1st of September 2008, and the last detector was removed the 7th of October 2008. In Figure 19 an overview of the links with tube detectors is displayed.

Figure 19: Tube detectors installed for this project

As one can see, the extra detectors were placed mainly on those areas where no permanent loop detectors are present, more specifically in the area around Oudenaarde, Dendermonde, Zomergem, between Lokeren, Lochristi, Waregem and the E17, and a large part was installed inside the city of Ghent and near the western part of the ring R4.

II.4.2.6. Data cleaning and manipulation

Before starting with the estimation of dynamic OD matrices for this study we need to clean and manipulate the data that has been provided to us. The three traffic count datasets needed to be aggregated, synchronized and checked for consistency, and finally cleaned from unreliable counts.

II.4.2.7. Traffic count aggregation

For solving the OD estimation problem the traffic counts collected at 15 minutes intervals are aggregated to hourly data. All of the above mentioned data sources do however contain missing and inaccurate data due to various reasons. For the tube detectors around 23.5% of the data was missing. For the fixed loop detectors this percentage was 5.5, while the highway detectors had 7.3% of their data missing. The large percentage of missing data for the tube detectors is due to the fact that many tube detectors were installed only during certain parts of the considered time period, and thus no data was available for the remaining parts. One has to account for these missing data when aggregating the traffic counts. To this end the following procedure was performed for the tube counts: when less than 3 out of 4 15-minute-counts were available within an hour, the counts were disregarded. If only one 15-minute count was missing, the missing count was replaced by the average of the other 3 counting periods.

For the highway counts another procedure was carried out. The traffic counts were used together with the speed measurements as input for the so-called Helbing filter (Treiber and Helbing (2002)), which uses concepts from shockwave theory to correct measured link flows on highways. Missing data for the traffic counts was thus replaced by the filtered data.

II.4.2.8. Synchronization and data cleaning

All data sources were available in different formats, and therefore special attention was given to a correct synchronization of the traffic counts. It appeared that there was a synchronization error between the tube detectors and the fixed loop detectors. There seemed to be a time lag of exactly one hour, as can be seen in Figure 20. It seems reasonable to conclude that the counts of the fixed loop detectors are registered one hour earlier. Therefore the loop detector counts were shifted forward one hour in time.

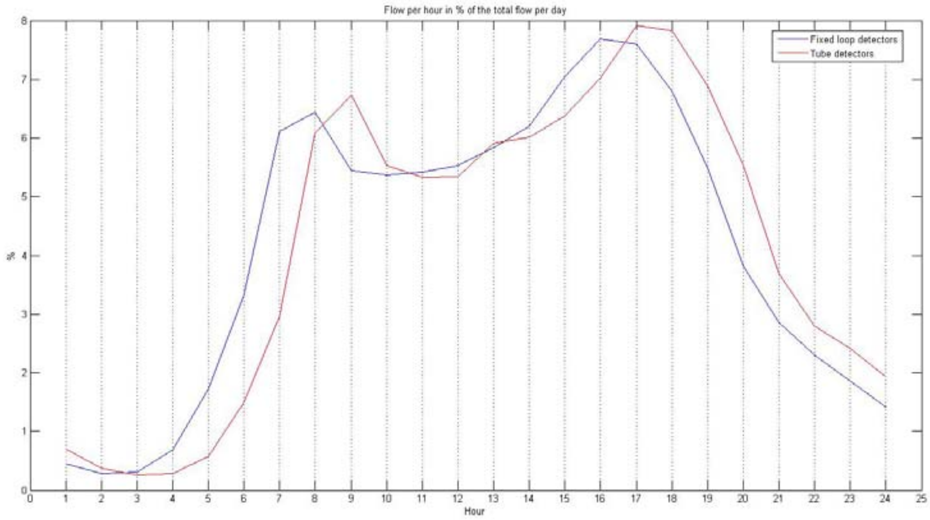


Figure 20: Time profile for loop and tube detectors

Finally some traffic counts had to be deleted from the dataset for a number of reasons:

1. Data containing clear faulty values (e.g., zeros during the peaks)
2. Traffic counts placed near the connectors, since due to the aggregation traffic flows in the model are usually higher than in reality
3. Counts near the limits of the study area, as these counts contain also traffic that may originate and end outside of the study area.

II.4.3. Deriving a dynamic target OD matrix

An important input for the OD estimation process is the dynamic starting OD matrix or target matrix. It is used in the estimation process itself, but it is also essential for providing an estimation of the route fractions. However no dynamic OD matrix for the region around Ghent was available, but as already said only a static OD matrix for the morning peak and one for the evening peak were provided.

To derive a starting dynamic OD matrix that partly describes the actual daily fluctuations of the demand we start by assuming that link flows can be seen as a linear transformation (if we neglect congestion effects) of the OD flows: the temporal pattern of a link flow is therefore the weighted sum of the temporal patterns of several OD flows. As the temporal pattern of the link flows is known, one can try to reverse this relationship. The question remains how to associate a certain OD flow with the link flows.

The following approach was chosen to associate the link flow pattern and the OD flow pattern and thus find an initial estimate of the dynamic OD matrix:

1. the link flows are first categorized according to ratio between the total amount of traffic passing before noon and the total amount of traffic passing in the afternoon. This is done to determine whether a link direction experiences a higher morning peak or a higher afternoon peak. Note that only non-highway data is used, in order to minimize the amount of congestion in the dataset. Large congestion effects would in fact invalidate our previous assumption of a linear transformation.
2. For each of these categories the average temporal pattern was calculated. This was done for every type of day (Monday, Tuesday, etc.). In Figure 21 the temporal patterns for a Monday are depicted for different categories. Also the ratio between the flow between 08.00h and 09.00h and the flow between 17.00h and 18.00h was calculated for every category. Note that in all following figures a certain value at for example 18.00h actually means that this value holds for the time period between 17.00h and 18.00h.

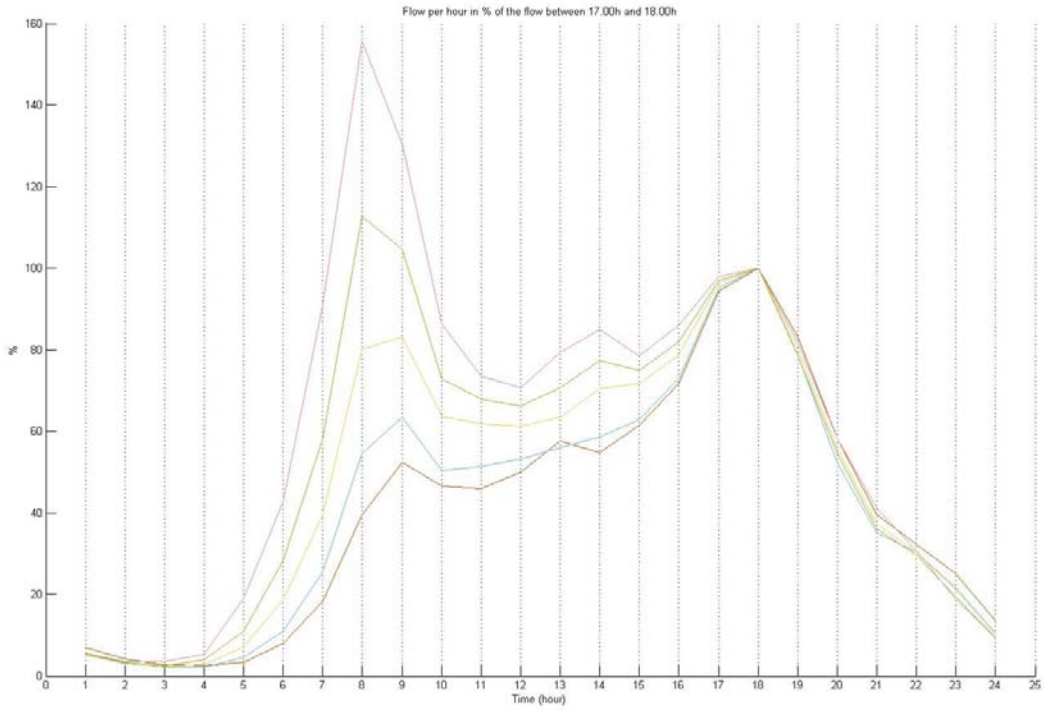


Figure 21: Temporal flow pattern for different categories (normalized to the flow at 18.00h)

3. Next for each OD pair the same ratio is calculated from the static OD matrices for 09.00h and 18.00h, and each OD pair is subdivided in the best matching category, and we assume that the temporal pattern of the OD pair corresponds to the temporal pattern of that category.
4. We now have the temporal pattern for each OD pair, but this pattern still needs to be translated to absolute flow values. This can be done by using the static OD flows, but the question remains which flows to use: those of morning peak or those of the evening peak? The two choices would result in two (slightly) different dynamic OD flows. This difference is a consequence of the fact that the ratio of the static OD matrices does not always match exactly with the ratio of a certain category.
5. The final dynamic OD flow is then calculated as the average of these two dynamic OD flows. It is also possible to take the average of these OD matrices to get a dynamic OD matrix for an average workday.

This procedure works fine for the workdays, but it should not be applicable for the weekend days, because the static OD matrix is only specified for a typical workday and the pattern is expected to be significantly different during the weekend. However, we do not have any information about how this patterns change, as well as the magnitude of OD pairs compared to each other. Therefore we choose to use the dynamic OD matrix of the average workday as a starting point in our OD estimation process for every type of day, leaving the correction of this error to the OD estimation procedure. In this way differences between the final estimated OD matrices for each day are solely due to differences in the counts, and not to using different starting matrices for each day.

Finally a special procedure had to be undertaken for OD pairs with destination Brussels. This was done for two reasons. The first one is because the most severe congestion in the network, namely on the E40 highway between Aalst and Brussels in the morning peak, is caused by these OD flows. Because a relatively small error in these OD flows can cause high levels of congestion in the rest of the network, it is important to have a decent starting estimate. A second reason can be found in a previous study dealing with the traffic on the E40 (Tampere et al., 2007) indicates that there are two 'departure waves' in the morning peak: one that departs between 6h and 7h, and another that departs between 8h and 9h. While the demand peak between 8h and 9h can easily be observed on highway segments between Ghent and Aalst, the early peak between 6h and 7h is less notable on these segments. This is because the demand peak between 8h and 9h towards Brussels is accompanied by a demand with intermediate destination. This is not the case for the demand between 6h and 7h. This demand peak can only be observed on the detector after Aalst, because at this point all traffic with intermediate destinations has already exited the highway. Therefore one could simply use the temporal pattern of this detector for deriving the dynamic OD flows with destination Brussels. However an additional problem is the fact that there is often congestion spillback on this detector, and in congestion regime the detected flows give an underestimation of the actual demand.

Our approach to derive a representative temporal pattern is as follows: we sum up all traffic that passes the detector after Aalst during the whole period in which we detect congestion. For this time period we use the time profile from a detector further upstream where there is no congestion, and multiply this (normalized) time profile with the previously calculated sum. In this manner the total amount of traffic passing in the congestion period remains correct. The time pattern will not be entirely correct, but this method provides a good estimate. Next we normalize this flow pattern (so it becomes a time pattern again), and make a distinct category characterized by this time pattern. All OD pairs towards Brussels are divided in this category. The dynamic OD flow is then calculated similar as above. Finally we multiply all these OD flows with a certain ratio such that the demand towards Brussels causes congestion with a correct queue length.

In conclusion to derive dynamic OD matrices on a weekly time horizon in this study we dispose of two information sources. On the one hand we obtained a very rich set of traffic count data, which covers a significant part of the studied network links. This large number allows us to reduce as much as possible the undeterminedness of the OD estimation process. On the other hand we disposed of two static OD matrices to derive a target or initial matrix. However, to derive a daily traffic pattern we had to use both information from the above mentioned static matrices and the same traffic counts (which can give us an indication of the daily fluctuations), observing that OD flows can be seen as a linear transformation of link flows in uncongested networks.

III. ANALYSES ON WEEKLY EFFECTS

III.1. Behavioural survey

Before presenting the main results related to the mobility behaviours over a week, we will first describe the surveyed population and give a socio-demographic picture of it. But if we mention explicitly the contrary, all following analyses are made on weighted data.

III.1.1. Description of the surveyed population

Households of the respondents

Let us first have a quick look at the households the respondents come from.

	N	Mean	Maximum
Household size	716	2.92	9
Number of children (under 6)	717	0.18	3
Number of kids (6 to 11)	717	0.20	4
Number of teenagers (12 to 17)	717	0.29	3
Number of workers	717	1.39	5

Table 6: household's characteristics

18% of our population lives alone, 26% in a 2 persons household, 20% in a 3 persons household, 37% in a 4 or more persons household, which is comparable to the structure of the "real" population of the individuals aged from 12 to 75 in Ghent (as given by INS). This is expected since the household size is a weighting variable (even if only 2 categories were retained for the weighting: isolated and non isolated). However we can mention here that in our non weighted sample, the percentage of isolated households was slightly lower.

37% of the individuals live in a household having at least one person under 18 (13% in a household having at least one child under 6, 14% in a household having at least one kid from 6 to 11, 21% in a household having at least one teenager from 12 to 17).

21% of the individuals live in a household without worker, 30% in a household with one worker, 49% in a household with 2 workers or more. Do not forget that we are talking here about individuals, and not about households. So if the percentage of individuals belonging to "2-workers-households" appears high, we cannot compare it to the proportion of "2-workers-households".

	Household size				
Number of workers	**1**	**2**	**3**	**4**	**5 ou +**
0	36.5%	47.5%	6.7%	2.6%	0.0%
1	63.5%	24.9%	28.9%	19.4%	12.4%
2	0.0%	27.7%	52.9%	65.6%	60.1%
3 ou +	0.0%	0.0%	11.6%	12.4%	27.5%
total	100.0%	100.0%	100.0%	100.0%	100.0%

Table 7: Number of workers according to household size

Socio-demographic description of the individuals

75% of our individuals are head or spouse of the head of their household, 24% are children (let us recall that we focused on individuals from 12). 27% of the individuals are younger than 30, and 19% older than 59. We do not have any respondent older than 75 in this survey (it is because there was none of them in our sample, not because a bad response rate from their part).

If we except respondents still going to school, 1% of our population does not have any degree, 5% have a primary school degree, 35% a secondary school degree, and 58% has a high school or university degree. As already said, these results are drawn from weighted data If we use non weighted data, we observe that 53% of our respondents aged 15 years or more have a high school or university degree, which is a very high rate of well instructed persons, as the rate of high school or university degrees for all Belgium (for people over 15) in 2008 is of 25% (Source: Direction Générale Statistique et Information économique - Enquête sur les forces de travail (http://www.statbel.fgov.be)). However in the results from 2001 socio-economic survey, we can see that the number of high degrees in Ghent is quite high (Cortese et al., 2006, p51.), but the individuals with low or no degree are also over-represented in the city (Op. Cit, p.45.). So we have to recognize that the specificity of the city of Ghent (university town) does not explain entirely those figures, and we have to draw attention to this bias identified in our survey: the survey has been more answered by high educated people, which can be explained by the complexity and the heaviness of it.

Concerning the socioprofessional status, we can mention that we have 56% of active people in the population. If we consider the non weighted rate of working people in our sample, among the 15-65, we obtain 65%, which is comparable to the figures of employment in Belgium (and in Flanders in particular) for 2008 (Source : Enquête sur les forces de travail, op. cit.).

Gender	
Men	50.3%
Women	49.7%
Total	100.0%

Position in the household	
Head or spouse	74.9%
Child	23.9%
Other	1.3%
Total	100.0%

Age (classes)	
12 to 19	10.8%
20 to 29	16.1%
30 to 39	21.3%
40 to 49	18.9%
50 to 59	14.1%
60 to 69	11.6%
70 to 75	7.2%
Total	100.0%

Degree	**All**	**Without schoolboys (girls) and students**
No degree	1.1%	1.1%
Primary school	9.9%	4.7%
General secondary school	19.5%	16.6%
Technical secondary school	11.4%	12.6%
Professional secondary school	5.2%	5.6%
Special secondary school	0.3%	0.3%
High school, 2 or 3 years	24.2%	27.6%
High school, 4 or 5 years	6.0%	6.9%
University	21.6%	23.9%
Other	1%	0.7%
Total	100.0%	100%

Socio professional status	
Schoolboy (girl), student	15.2%
Housewife, -husband	3.7%
Between jobs	3.3%
(Pre)Pensioner	17.6%
Invalid	0.3%
Workman in the private sector	7.0%
Employee in the private sector	30.4%
Freelance	4.3%
Liberal profession	1.5%
Teacher	5.6%
Civil servant	7.0%
Other	4%
Total	100.0%

Table 8: socio-demographic characteristics of individuals

Practices for trips

40% of the population is entitled to a reduction on public transport. Almost 60% of the people have a season ticket for public transport (50% for De Lijn, 20% for SNCB). 21% of Ghent citizens have a free season ticket for the De Lijn.

Price reductions on public transport	
Child, kid, young	7.8%
Senior	11.7%
Large Family	5.6%
OMNIO_BIM (ex-VIPO : widower, invalidates, pensioner, orphan)	1.5%
Employee of a public transport company, Belgacom, Post, ...	7.3%
Other	6.6%
No reduction	59.6%
Season ticket	57.8%
Season ticket De Lijn	49.6%
Season ticket TEC	0.4%
Season ticket MIVB	2.5%
Season ticket SNCB	20.2%
Free season ticket De Lijn	20.8%
Free season ticket other company	8.9%

Table 9: reductions and season tickets for public transport

82% of the population holds a driving licence, and for 93% of those, it is a car driving licence (licence B).

97% of the population affirms to have no difficulties with the use of transport modes.

Work and school

Round 70% of our population is working or going to school. 85% of them has a fixed (or steady) work place or school place (7% has a fixed work place at home, 8% has no fixed place to work or study). 90% of them never carpool to go to school or to work.

The table below presents the work or school places regarding to their location (from the zip codes) :

Place of school or work	
Gent	66.7%
Provincie Oost Vlaanderen (outside Gent)	13.6%
Other Flemish province	7.9%
Brussels	8.5%
Other	3.3%

Table 10: distribution of school or work place

Regarding the workers, 75% are funded back (totally or partially) by the employer for their home-work trips. As we can see in Table below, parking on work place is not so a crucial point for our population.

Parking on the work place	
Yes	62.4%
No	37.6%

Facility to find a parking place near the work place	
Without problem	70.8%
With some problems	15.4%
Very difficult	13.8%

Paying parking near work place	21.9%
Free parking near work place	78.1%

Table 11: issues related to parking near work place

45% of the workers never need to travel during their work hours, 39% need to travel sometimes, and 16% have to do many trips for their job.

Here below are presented some characteristics of the workers and their job: 14% make teleworking (at least sometimes), 85% work during the day, ¾ of the workers have to follow work hours fixed by the employer, more than ¾ of the workers are working full time and 87% of the workers work at least 80% of a full time. Almost half the workers are employees.

Teleworking	
No never	83.6%
Yes, at home, more than 90%	0.5%
Yes, at home, from 20 to 90%	4.6%
Yes, sometimes (less than 1 day/week)	10.2%
Yes, other	0.11%

Time of work	
Day work	84.5%
Night work	1.6%
Shifts, without night	3.8%
Shifts, with nights	5.5%
Other	4.6%

Work hours	
Fixed hours, imposed by employer	53.3%
Fixed hours, chosen by the worker	14.6%
Variable hours, imposed by the employer	21.0%
Variable hours, chosen by the worker	11.1%

Full time workers	77.5%
Part time workers	22.5%

Professional status	
Workman	10.1%
Employee	47.3%
Executive	1.9%
Freelance	6.5%
Liberal profession	2.9%
Teacher	9.8%
Civil servant	13.5%
Other	7.1%

Table 12: jobs characteristics

Housing

¾ of the individuals have a garage at home, and parking close to the house is free in 84% of the cases. 89% of our population have an Internet connection at home. This figure can appear high, but let us remind that our sample was designed on individuals for whom we found a phone number, as explained in the methodology section. This can have an impact on this variable "Internet connection at home".

Available vehicles for the household

The average number of kid bikes per household is 0.6, it is 2.5 regarding adult bikes, 0.1 for motorbikes, and finally 1.3 for cars.

Number of kids bikes	
0	73.0%
1	9.6%
2	9.4%
3 ou +	8.1%

Number of adult bikes	
0	10.6%
1	14.7%
2	34.5%
3 ou +	40.2%

Number of motorbikes	
0	88.9%
1	9.2%
2	1.5%
3 ou +	0.4%

Number of cars	
0	12.5%
1	51.8%
2	30.3%
3 ou +	5.4%

Table 13: vehicles per household according to the type of vehicle

Incomes

The tables here below present the distribution of personal and household incomes in our population. Let us remark that we have to be cautious with this variable because of the huge level of non response[1].

Personal income	With non responses	Without non responses
0 tot500/month	11.8%	17.1%
500 to 1000/month	5.8%	8.4%
1000 to 1500/month	17.1%	24.8%
1500 to 2000/month	19.7%	28.6%
2000 to 2500/month	10.3%	14.9%
2500 to 3000/month	2.2%	3.2%
3000 to 4000/month	1.6%	2.3%
4000 to 5000/month	0.1%	0.1%
more than 5000/month	0.4%	0.6%
no answer	31.0%	
Total	100.0%	100.0%

Household income	With non responses	Without non responses
0 tot500/month	0.3%	0.5%
500 to 1000/month	1.8%	2.8%
1000 to 1500/month	7.1%	11.1%
1500 to 2000/month	10.4%	16.2%
2000 to 2500/month	10.8%	16.8%
2500 to 3000/month	7.1%	11.1%
3000 to 4000/month	16.3%	25.4%
4000 to 5000/month	7.3%	11.4%
more than 5000/month	3.0%	4.7%
no answer	35.8%	
Total	100.0%	100.0%

Table 14: individuals and households incomes

To conclude this descriptive part of our work, we will insist on the fact that our population does not seem to reflect quite faithfully the Ghent population, as we have a high graduated population, a high Internet connections rate. Those indicators can make believe that we are facing a wealthy population, possibly being more mobile,... We are here confronted to one limit of such a survey, very long and complex, and conducted by phone. The following results must therefore be considered with the greatest caution.

III.1.2. Burden effect

[1] If we compare to other non response rates about income questions, our non response rate is very high, others being usually rather comprised between 15 and 27% (See Yan et al., 2006).

A well known bias observed in mobility survey is the burden effect. When asking people to describe their trips for more than one day, one can observe that the number of trips decreases with the consecutive days (see Flemish OVG : Hajnal et al., 1996, Nuits et al., 2001; German MOP : Zumkeller et al., 1997 and 2002, Chlond et al., 2003; Swiss Mobidrive: Axhausen et al., 2002, Schlich et al., 2003, Cirillo et al., 2006). For example, in the first OVG, Flemish respondents were asked to describe their trips for two consecutive days. On average, the second reported day showed a quite lower number of trips (Nuits et al., 2001). It is mainly explained by the fact that people get tired of answering and therefore skip some trips.

We were afraid of suffering of such an effect in our study, since we asked trips for a whole week. Fortunately it was not the case as we could deduce from the analysis here after.

Remember that each person who agreed to participate had been attributed one week (included in the 3 months of the survey, i.e. from September to November 2008). This week could start on any day of the survey period so that the first day can be any of the seven days of the week. However this date could be changed for the respondent's convenience (e.g. : if he forgot noting down his trips or was not in Belgium during the assigned reporting week). But the newly attributed week should have started, as much as possible, on the same day of the week as the initial reporting week. [Let us point out that some respondents switched their dates by themselves, for their convenience or by mistake, or maybe because they had not made any trips on the first day. In this case, they usually do not follow the here above rule].

Even if we formally must face a uniform distribution for the days where the reported diaries start, we remark, for people whose trips agendas were valid, i.e. for agendas which have been used for the analyses, that there are a bit less respondents whose reporting week starts on a Sunday, and more on a Monday as it can be seen from Table here below.

First reference day (Day 1)	Number of respondents
Monday	121
Tuesday	103
Wednesday	99
Thursday	99
Friday	107
Saturday	99
Sunday	89

Table 15: number of respondents according to the first reference day

.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Week
% motionless	4.60%	6%	5.72%	5.72%	6.14%	6.83%	8.51%	0%
Number of trips								
Minimum	0	0	0	0	0	0	0	2
Mean	3.96	3.9	3.91	3.83	3.97	3.65	3.65	26.86
Standard deviation	2.37	2.33	2.52	2.27	2.54	2.32	2.43	10.13
Maximum	14	21	18	13	15	17	14	58

Table 16: Statistics about number of trips per individuals from Day 1 to Day 7 of their reporting week
(Rem : the means include people with 0 trips whose percentage is also indicated in the Table)

The average percentage of motionless people per day is about 6%. It is quite lower on the first day and higher on the seventh reporting day. It could partially be explained by the fact that we have a bit more Mondays as first day, and by the way more Sundays as seventh day and, as we will see later, Sunday is traditionally a day with less trips. As we do not see any significant evolution (both for the percentage of motionless people and for the average number of trips) from day 2 to day 5, we think that the small inequality in the distribution of the first reporting days is the main explanation of these small differences observed for the immobiles' rates.

As we will see later, the difference in the percentage of motionless people between a Sunday and any weekday is quite higher than between "day 1" and "day 7": as it rises up to 15% for a Sunday but only to 5% for weekdays. These figures tend to confirm our hypothesis.

So we are fortunately not embedded within a burden effect bias. Such a "good" situation was already reported for the pre-test in June 2008.

It remains quite hazardous to highlight why it is so. The interest shown by contacted people for a mobility survey, the incentives, the bias in our sample (well educated people), a more strict respect of instructions from Flemish population have perhaps played some role but this would need deeper investigations to be sure which conditions need to be repeated in future surveys for assuming the same lack of burden effect.

III.1.3. Mobility indicators

The here presented results are drawn from the 717 valid travel diaries sent back by post by the respondents. The raw data have been weighted according to the variables used for stratification of the sample: age classes, gender and type of households (one person vs. 2 and more person households) before analyses. It is worthwhile to notice that even if diploma sounds as the most correlated variable for mobility behaviour, it has not been used in the weighting process since margins regarding this variable are not available from the National Register (used both for sampling and weighting).

The maximum observed ratio between weights is no more than 2.9.

Finally, trips longer (or equal) to 200 km (52 trips in a total of 19400 trips) have not been taken into account for analyses related to distance.

III.1.3.1. Motionless people [Statistics per person]

We do not have any respondent who was motionless during the whole reporting week but evidently it occurs that some people do not achieve any trip during one of the reported days.

Here a day is defined as starting at 4 a.m. and ending at 4 a.m. the day after. It means that a trip made on Sunday between midnight and 4 a.m. is attributed to Saturday. Doing so allows considering as motionless on Sunday someone who do not make any trips on Sunday except his return back home at 1 o' clock after his Saturday evening activities.

If we analyse the number of such motionless people according to the day of the week, it gives the following Figure.

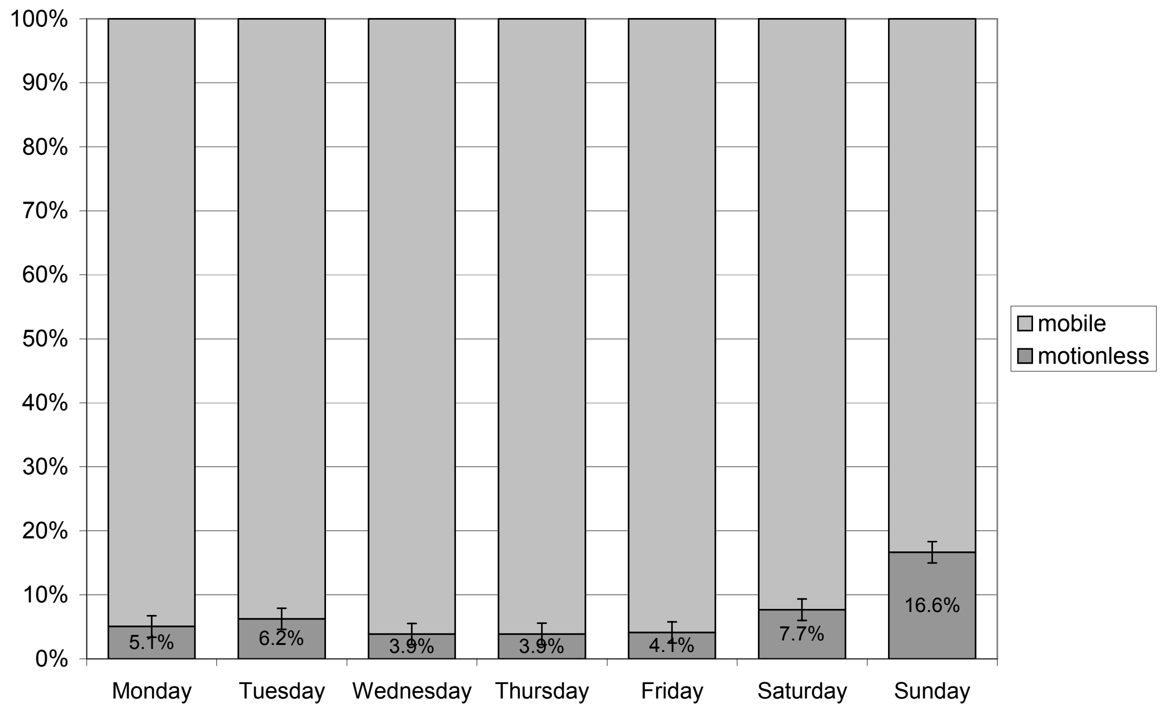


Figure 22: percentage of motionless people per day

There are clearly more motionless people on Sunday (16%) than on the other days of the week (5% on average). More generally week ends are periods where more people do not travel at all.

The motionless rate is undoubtedly but unexpectedly much lower than in previous Belgian mobility surveys even if those only covered one day travels diaries. For example, in Mobel, the percentage of motionless reached 24%, which is certainly overestimated due to soft refusers (people who reported that they did not move at all because it shorts the survey and therefore the answer time). It seems to prove that this one-week diaries survey does not suffer of this "soft refusal" effect.

Once more it seems difficult (as for the lack of burden effect) to accurately determine the reasons why we observe such a difference in motionless rate. However some hypotheses could perhaps been drawn:

- the trips agenda was the only part of the survey that respondents had to answer with paper and pencil since all the socio-demographic stuff was collected in a previous phone call; therefore respondents were less burden than in previous surveys where they had first to answer to a lot of questions about their household and their mobility habits before filling the diary. So they were perhaps less likely to decrease their amount of trips;
- the survey only covers an urban area (the city of Ghent) where people are more mobile but also more aware about mobility problems;
- as there was a whole week to be filled, maybe people choose to do it seriously or not to do it at all… Whilst for one day, it is more difficult to make the difference between those who did not move and those who did not correctly answer, for one week, a "no-trip-week" would have been more suspicious.

III.1.3.2. Number of trips per day [Statistics per person]

Note that further statistics still include all the 717 people, i.e. even those whose number of trips for a day is zero.

The average number of trips per week is 26.7. This amount corresponding to an average of 3.81 trips per day is quite high compared with the results from MOBEL (2.97 trips per day) but we have to keep in mind that this previous survey is 10 years old and covered the whole country, as well urban as rural areas. The motionless people, less numerous in the present survey, are also included in those means.

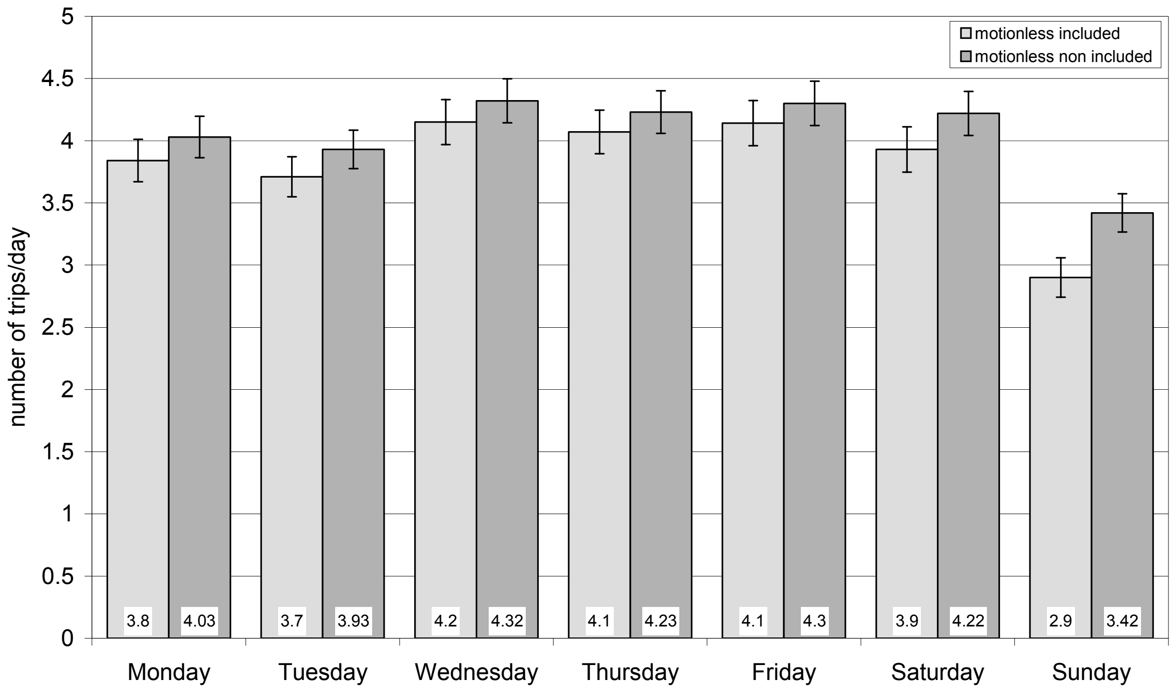


Figure 23 : Average number of trips per day

Regarding the average number of trips per day, we observe some significant differences, according to the Tukey test which compare the means, and as it can be seen with «95% error bar»[2]:

- with 2.9 trips per person, Sunday is by far the quietest day (but we have to keep in mind that it includes the 15% of motionless); other days have on average 4.0 trips per day per person;
- the number of trips per person on Tuesday (3.7) is lower than on Wednesday (4.2), Thursday and Friday (4.1); it is a bit surprising since Tuesday is often taken as a reference "full " day for traffic model, but the differences are not huge;
- other differences are not statistically significant.

III.1.3.3. Time budget [Statistics per person] :

As we already mentioned let us recall that for time budget and total distance per day and per week, we only considered trips shorter than 200 km in order to skip exceptional trips (to foreign country by plane, etc.). By doing this, 52 trips were not taken into account. It avoids that the means are too much influenced by such outliers.

[2] Tukey's test is a statistical test generally used to find which means are significantly different from one another.

The time that individuals spend in travelling is 9 hours (and 3 minutes) per week. On average, it means 1 hour and 18 minutes per day, which seems consistent with MOBEL data but also in accordance with Zahavi's conjecture[3].

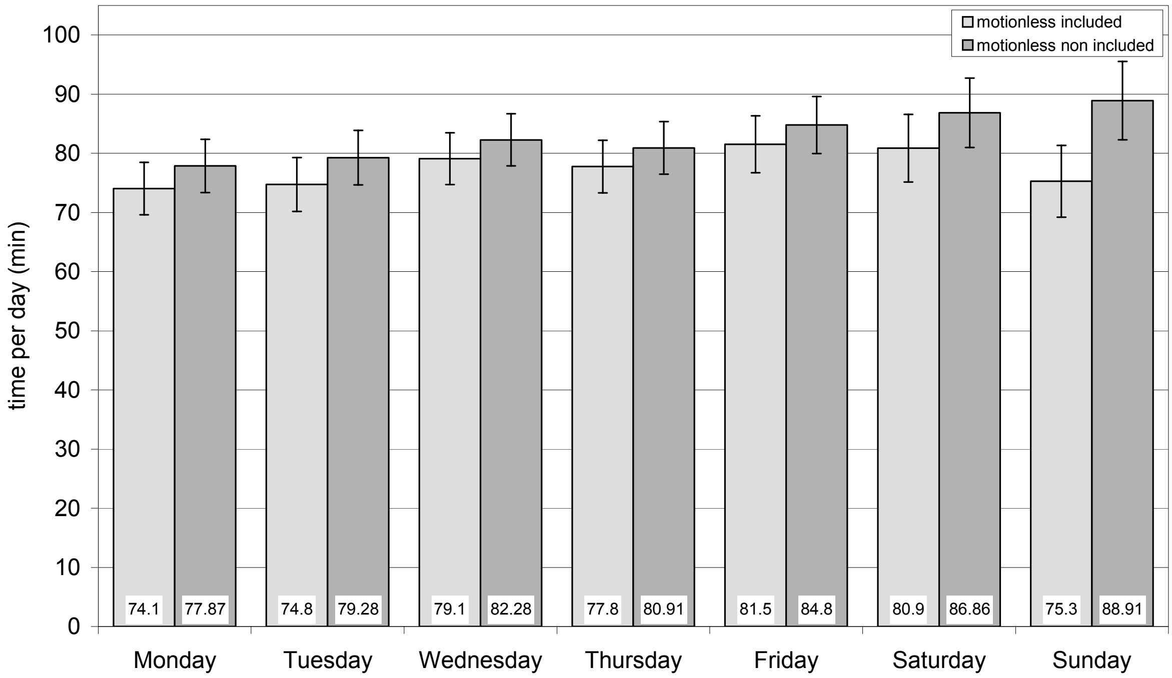


Figure 24 : Average individual daily travel time budget

There are no statistically significant differences amongst days of the week, even for Sunday. The time budgets we got for Monday and Tuesday are even shorter (but not significantly) than the one for Sunday. It means that if the number of trips is lower on Sunday, the trips are however longer (see below), so that the time budget amongst days is equal on average.

III.1.3.4. Distance per day and per week [Statistics per person]

During a week, each individual travels on average 278.5 km; it means, 39.8 km per day.

Sunday highlights the highest average distance. But, even if the differences amongst days are more important than for time budget, they also are not significant. The main reason is that the standard deviations are much larger for distances than for durations. On the Figure below, the error bars which represent the confidence intervals show this dispersion. This dispersion is besides 30% higher for Saturday and Sunday, which seems to indicate that travelled distances are indeed larger during weekends.

However let us recall that these distances are the ones reported by the respondents and point out that it is much more difficult for people to estimate distance they travel than time spent for a trip.

[3] "Zahavi (1977) advanced the conjecture of constant travel time budgets (and constant travel expenditures as a percentage of income) in the development of the UMOT model of travel. Under the assumption of constant travel time budgets, an individual will allocate a fixed amount of time to travel; thus, if travel speed improves, then the time saved will be used to travel more or further, while if congestion worsens, then people will make fewer trips, choose faster modes, and/or choose closer destinations. This controversial notion of constancy in travel time budgets is essentially the sole behavioural paradigm that has been applied to the issue of induced/suppressed demand » (Pas et al., 1997)

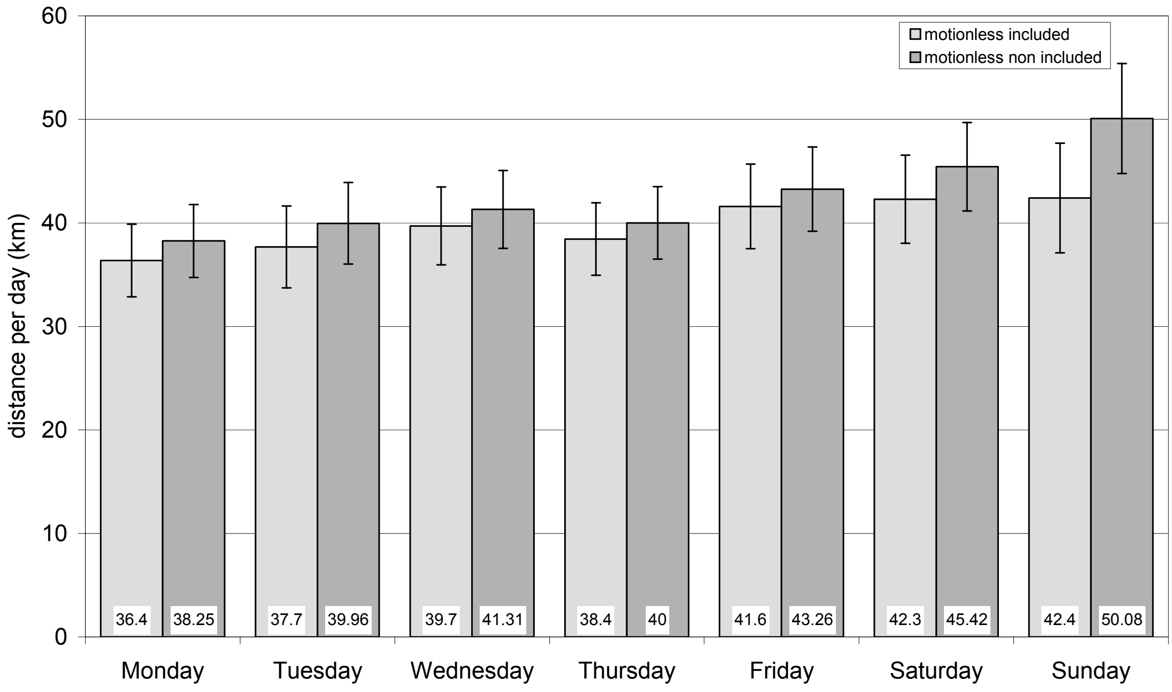


Figure 25 : Average individual daily travelled distance

III.1.3.5. Distance [Statistics per trip]

If we compute the average distance per trip, we can see that it is really higher on Sunday with more than 13km per trip versus less than 10 km on weekdays. The Tukey test also shows that the distance per trip on Saturday is significantly longer than on Wednesday and Thursday.

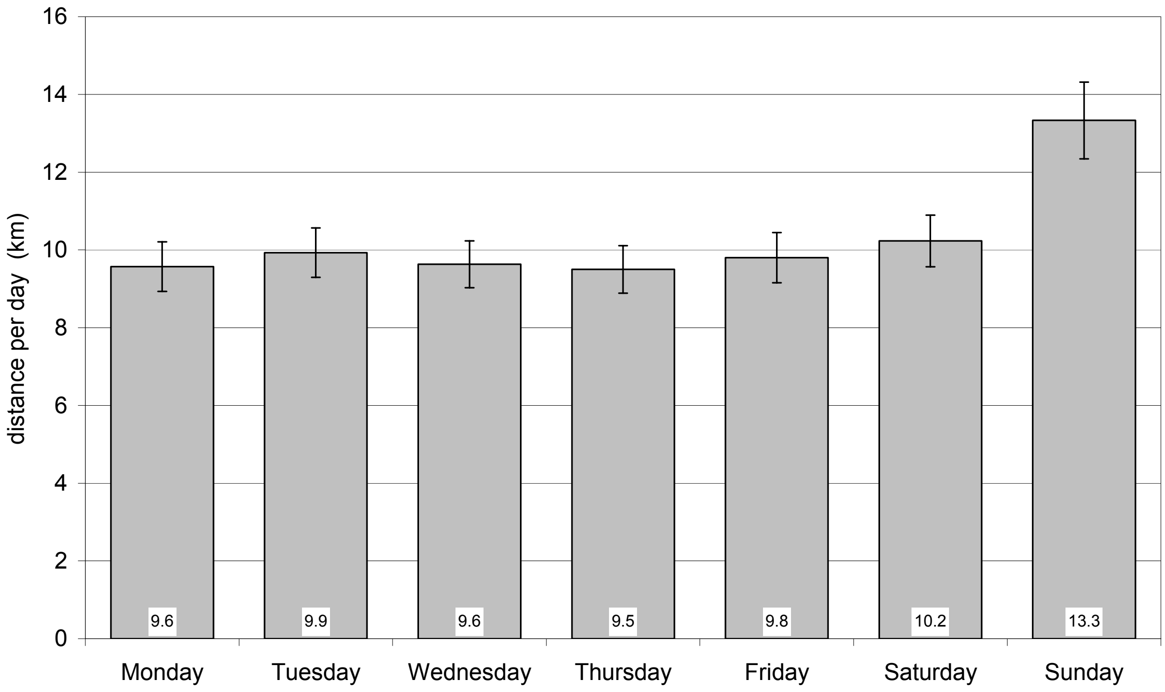


Figure 26 : Average distance per trip for each day

III.1.3.6. Purpose distributions [Statistics per day]

In the following graphs, the trips purposes will be group together in the following way, for more legibility:

- the reason «leisure» put together the original reasons «leisure» and «walk» ;
- the reason «shopping» put together the reasons «daily shopping» and «long term shopping» ;
- the reason «other» put together the reasons «eat», «personal business» and «other».

In the tables, the original reasons (non grouped) will be preserved.

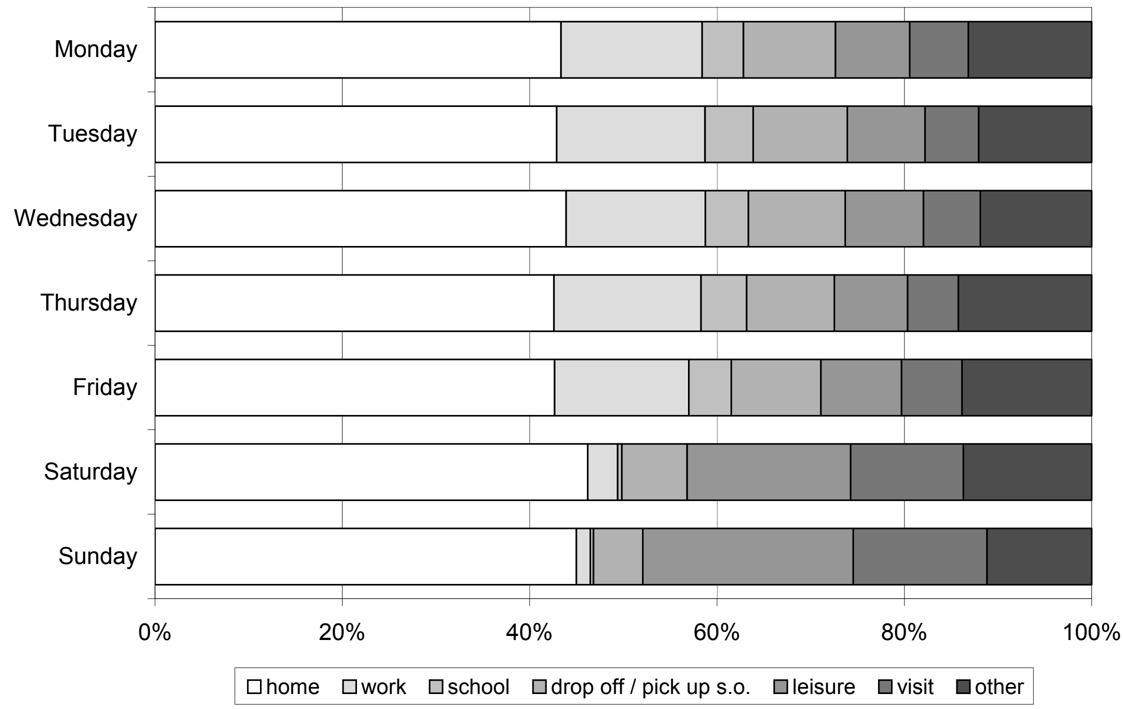


Figure 27: purposes distribution according to the day of the week

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
drop off / pick up s.o. (1)	8.60%	8.90%	9%	8.30%	8.20%	5.60%	4.80%
Home (2)	38%	38%	38.30%	37.80%	36.50%	37%	40.70%
Work (3)	13.20%	14%	12.90%	14%	12.30%	2.60%	1.40%
School (4)	3.90%	4.50%	4%	4.30%	3.90%	0.40%	0.30%
Eat (5)	2%	2%	1.60%	2.40%	2.50%	2.90%	3.20%
daily shopping (6)	9.10%	9.20%	9.70%	8.70%	11%	13.50%	7.60%
long-term shopping (7)	3.30%	2.20%	3.10%	2.40%	3.40%	6.30%	1.80%
personal business (doctor, bank) (8)	4.80%	3.80%	4.30%	4.40%	4.10%	2.30%	1.90%
visit to family or friends (9)	5.50%	5.10%	5.30%	4.80%	5.50%	9.70%	12.90%
walking, riding, etc. (10)	3%	2.60%	2.30%	2%	2.10%	4.60%	8.80%
leisure, sport, culture (11)	3.90%	4.80%	5%	5%	5.30%	9.40%	11.60%
Other (12)	4.70%	4.90%	4.50%	5.90%	5.20%	5.70%	5%

Table 17: purposes distribution according to the day of the week

Table below present the confidence intervals (at 95%) for the distribution of purposes, for each day of the week. Purposes are coded with a number, see in Table here above.

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
(1)	[7.6 ; 9.6%]	[7.8 ; 10%]	[8 ; 10%]	[7.3 ; 9.3%]	[7.2 ; 9.2%]	[4.8 ; 6.4%]	[3.9 ; 5.7%]
(2)	[36.2 ; 39.8%]	[36.1 ; 39.8%]	[36.6 ; 40%]	[36.1 ; 39.6%]	[34.8 ; 38.2%]	[35.3 ; 38.8%]	[38.6 ; 42.9%]

(3)	[12 ; 14.5%]	[12.7 ; 15.3%]	[11.8 ; 14.2%]	[12.7 ; 15.2%]	[11.1 ; 13.5%]	[2 ; 3.1%]	[0.9 ; 1.9%]
(4)	[3.2 ; 4.6%]	[3.8 ; 5.3%]	[3.3 ; 4.7%]	[3.6 ; 5%]	[3.2 ; 4.6%]	[0.1 ; 0.6%]	[0.1 ; 0.5%]
(5)	[1.5 ; 2.6%]	[1.5 ; 2.5%]	[1.1 ; 2%]	[1.8 ; 2.9%]	[2 ; 3.1%]	[2.3 ; 3.6%]	[2.4 ; 4%]
(6)	[8 ; 10.2%]	[8.2 ; 10.3%]	[8.6 ; 10.7%]	[7.7 ; 9.7%]	[9.9 ; 12.1%]	[12.2 ; 14.8%]	[6.5 ; 8.7%]
(7)	[2.6 ; 4%]	[1.6 ; 2.7%]	[2.5 ; 3.7%]	[1.9 ; 3%]	[2.7 ; 4%]	[5.4 ; 7.2%]	[1.2 ; 2.3%]
(8)	[4 ; 5.6%]	[3.1 ; 4.5%]	[3.6 ; 5%]	[3.7 ; 5.1%]	[3.4 ; 4.8%]	[1.8 ; 2.9%]	[1.3 ; 2.5%]
(9)	[4.6 ; 6.3%]	[4.2 ; 5.9%]	[4.5 ; 6.1%]	[4 ; 5.6%]	[4.7 ; 6.3%]	[8.6 ; 10.7%]	[11.5 ; 14.4%]
(10)	[2.4 ; 3.7%]	[2 ; 3.2%]	[1.8 ; 2.9%]	[1.5 ; 2.5%]	[1.6 ; 2.6%]	[3.8 ; 5.4%]	[7.6 ; 10%]
(11)	[3.2 ; 4.6%]	[4 ; 5.6%]	[4.2 ; 5.7%]	[4.2 ; 5.7%]	[4.5 ; 6.1%]	[8.3 ; 10.5%]	[10.2 ; 13%]
(12)	[3.9 ; 5.5%]	[4.1 ; 5.7%]	[3.8 ; 5.2%]	[5 ; 6.7%]	[4.4 ; 6%]	[4.9 ; 6.6%]	[4.1 ; 6%]

Table 18: confidence intervals for purposes distribution, for each day of the week

Let us first point out that there are more trips to home on Sundays. It means that the number of trips per chain (successive trips between two home returns) is lower on Sunday, i.e. that people combine fewer activities when they go out.

It is easier to compare the spreading of trips purposes on Figure 28, where trips to home have been omitted.

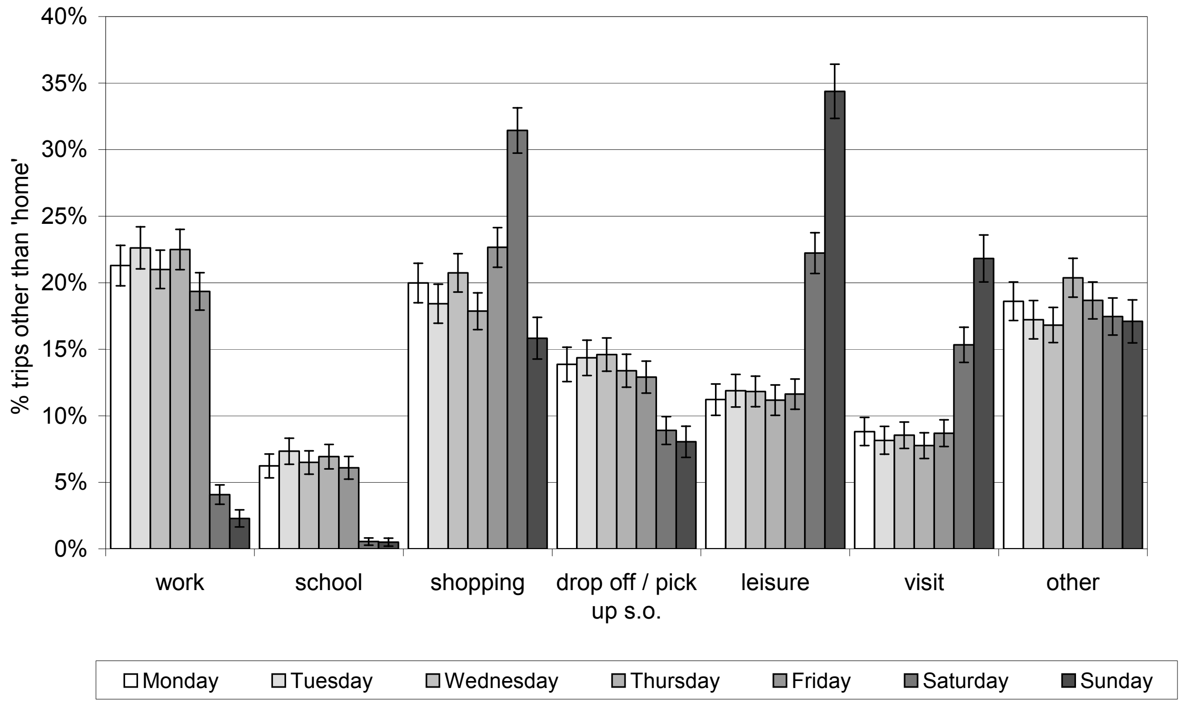


Figure 28 : Trips purposes per day (returns to home have been omitted)

The purposes «visit» and "leisure" (this last one grouping «walk, ride» and «leisure») of course considerably increase during week-ends, with a little less than twice more trips on Saturday and more than twice on Sunday. Saturday is a special day for shopping (both daily and long-term) which pertains to more than 30% of the trips, while this purpose only reaches between 15 and 20% on weekdays and around 13% on Sundays.

The purposes which fall on week-ends are, as expected, «school» and «work» which gather together more than 25% of the trips on weekdays but less than 5% on the weekends. «Personal business (doctor, bank)» also decreases, along with the «drop off / pick up» trips. In this last case, it probably must be related with the fact that there is no school on these days; but the «drop off/pick up» purpose keeps still around 7% during the weekends, while it reaches 13% on weekdays.

Finally, we can find out that the short activities which are fewer on Sundays (and that could explain the less numerous trips per chain) are well the «drop off/pick up», but not really the daily-shopping which remains nearly as important on Sunday as on weekdays. Maybe the higher number of shops opening on Sundays can explain this phenomenon. The fall of «personal business» trips is more representative. We can also suggest that people probably combine these activities with their work trips, what they cannot do on the week-ends.

III.1.3.7. Purposes during the week [in terms of number of trips per person]

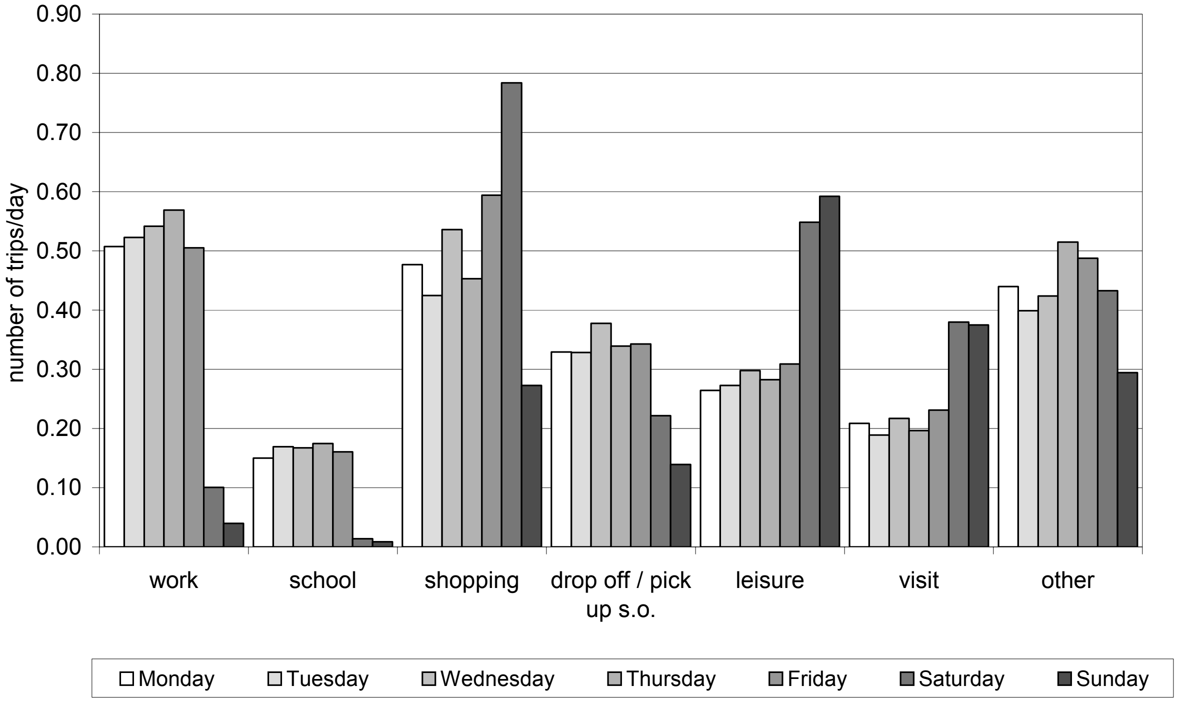


Figure 29 : Average number of trips per purpose according to the day (returns to home have been omitted)

Figure 28 showed the % of trips per purpose. The Figure here above is different from the previous one as it shows the average number of trips achieved for various purposes (e.g. on Monday people achieve on average 0.5 trips to work).

We could see differences for "work" activities in the previous graph (in percentages) according the different types of days, these appear less clear when we speak about number of trips. We still have fewer "work" trips on Friday and Monday, but Wednesday is in a medium position. The highest day for work is Thursday, while Tuesday knows little fewer trips for work than the other days.

From another point of view, if we look at daily shopping during the week (not in the week-end, where we clearly have different behaviours: the highest on Saturday, the lowest on Sunday), this is undoubtedly less important on the days when it is more worked (Tuesday and Thursday), and higher on Friday. Long term shopping seems to be foreseen mainly on Saturday.

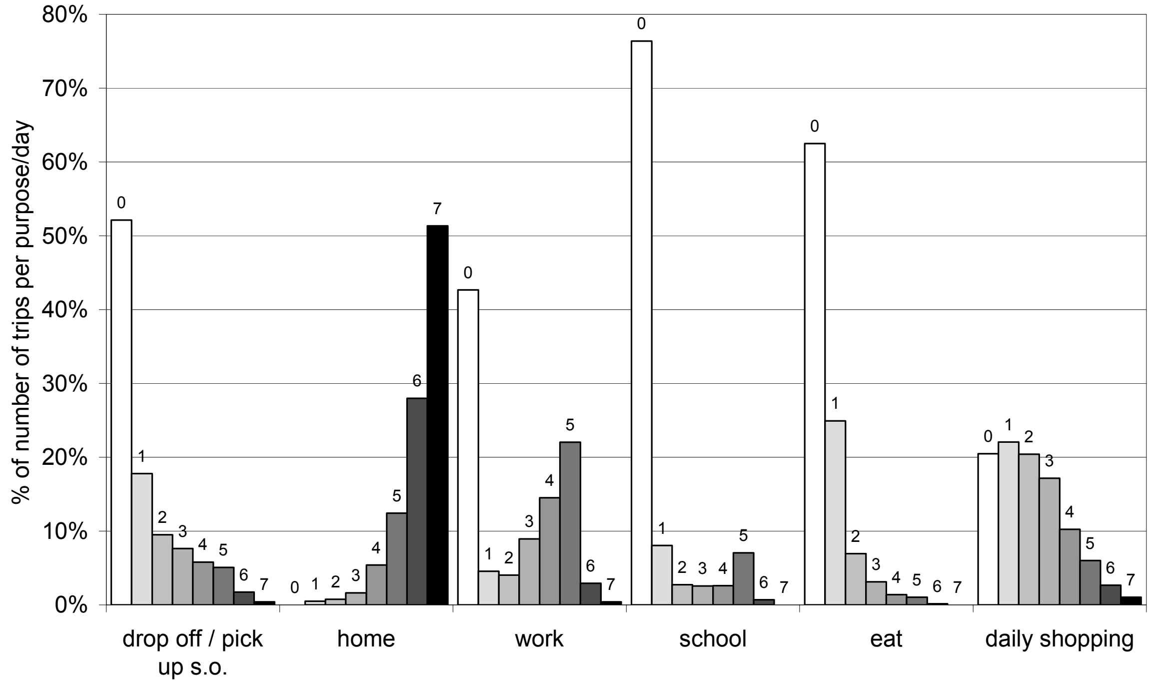
As in the previous graph, we can find that leisure and visits to friends and family are more week-end activities. Their numbers increase a lot on Saturday and Sunday.

Concerning the "drop off/pick up" purpose, we can see the peak on Wednesday, day traditionally used by children to have some activities, where they are dropped off and picked up by their (grand) parents.

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
drop off / pick up s.o.	0.33	0.33	0.38	0.34	0.34	0.22	0.14
home	1.46	1.4	1.59	1.54	1.51	1.45	1.18
work	0.51	0.52	0.54	0.57	0.51	0.1	0.04
school	0.15	0.17	0.17	0.17	0.16	0.01	0.01
having a meal	0.08	0.07	0.07	0.1	0.1	0.12	0.09
daily shopping	0.35	0.34	0.41	0.35	0.46	0.53	0.22
long-term shopping	0.13	0.08	0.13	0.1	0.14	0.25	0.05
personal business	0.19	0.14	0.18	0.18	0.17	0.09	0.05
visit to family or friends	0.21	0.19	0.22	0.2	0.23	0.38	0.38
walking, riding, etc.	0.12	0.1	0.1	0.08	0.09	0.18	0.26
leisure, sport, culture	0.15	0.18	0.2	0.2	0.22	0.37	0.33
other	0.18	0.18	0.18	0.24	0.21	0.22	0.15

Table 19: Average number of trips per purpose according to the day

A one-week survey allows going further in the analyses. We could e.g. see how many days a week a given purpose implies to travel.



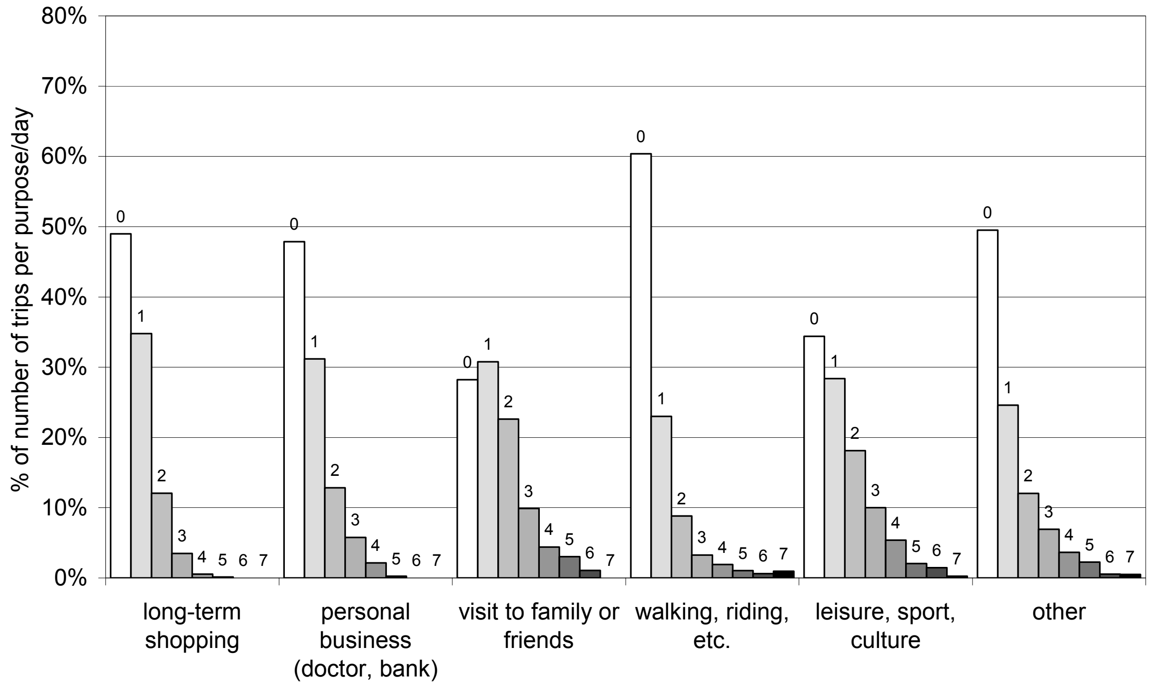


Figure 30: How many days per week are people travelling for each purpose (% of individuals)

	number of days	
	mean	**CI 95%**
drop off / pick up s.o.	1.2	[1.09 ; 1.33]
home	6.2	[6.1 ; 6.26]
work	2.3	[2.12 ; 2.44]
school	0.7	[0.6 ; 0.82]
having a meal	0.6	[0.52 ; 0.67]
daily shopping	2.1	[1.96 ; 2.21]
long-term shopping	0.7	[0.66 ; 0.79]
personal business (doctor, bank)	0.8	[0.76 ; 0.91]
visit to family or friends	1.4	[1.35 ; 1.55]
walking, riding, etc.	0.7	[0.64 ; 0.83]
leisure, sport, culture	1.4	[1.26 ; 1.47]
other	1	[0.92 ; 1.12]

Table 20: Average number of days per week when people travel for each purpose, and confidence intervals

As it could be expected, most people go 7 days per week back home. Less than 10% of the respondents go home less than 5 days a week. It seems quite realistic, as some can be on holidays for the week-end, even if we fear that some respondents omit to note some of their return trips to home.

On average, people go to work 2.3 days a week. More than 40% never goes to work; but it includes people who work at home, people on holidays, as well as non-workers. If we only consider working people, the mean goes up to 3.7 days. Similarly, if we only consider school-boys / -girls and students, they go on average 3.7 days a week to an education establishment. Let us remark that the survey period included the one-week All Saints holidays.

The daily shopping is the purpose for which most respondents have realized at least one trip during their reporting week. This purpose occurs on average a little more than 2 days a week. As expected, the long-term shopping is less frequent for the respondents.

Those analyses allow us to make a difference between activities that are mainly performed once a week (long term shopping, personal business) and those that are rather achieved more than once a week (visits to friends and family, leisure,...).

However these analyses regarding the purposes would lead us to suggest that methodologically, one-week is still a too short period for such survey. For example, if someone goes shopping or visit family every ten days, it is possible that such purposes do not appear in his agenda for the reporting week.

III.1.3.8. Mode distributions [statistics per day]

For this analysis we define the main mode for a trip as the mode used during the longest stage (in time) of this trip.

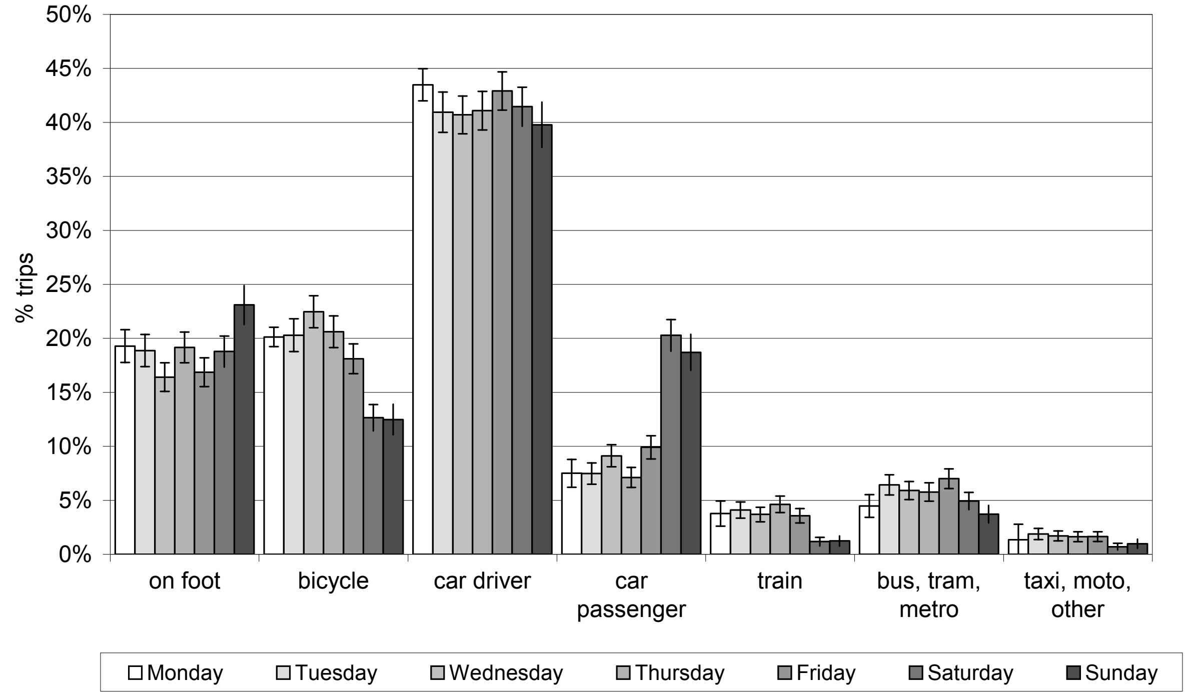


Figure 31: mode distribution for each day of the week

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
on foot	19.30%	18.90%	16.40%	19.20%	16.90%	18.80%	23.10%
bicycle	20.10%	20.30%	22.50%	20.60%	18.10%	12.70%	12.50%
car driver	43.50%	40.90%	40.70%	41.10%	42.90%	41.40%	39.80%
car passenger	7.50%	7.50%	9.10%	7.10%	9.90%	20.30%	18.70%
train	3.80%	4.10%	3.70%	4.60%	3.60%	1.20%	1.20%
bus/tram/metro	4.50%	6.40%	5.90%	5.80%	7.00%	4.90%	3.70%
taxi, moto, other	1.30%	1.90%	1.70%	1.60%	1.60%	0.70%	1.00%

Table 21: mode distribution for each day of the week

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
on foot	[17.8 ; 20.8 %]	[17.4 ; 20.4 %]	[15.1 ; 17.7 %]	[17.7 ; 20.6 %]	[15.5 ; 18.2 %]	[17.3 ; 20.2 %]	[21.3 ; 24.9 %]
bicycle	[18.6 ; 21.6 %]	[18.8 ; 21.8 %]	[21 ; 24 %]	[19.1 ; 22.1 %]	[16.7 ; 19.5 %]	[11.4 ; 13.9 %]	[11.1 ; 13.9 %]
car driver	[41.6 ; 45.3 %]	[39.1 ; 42.8 %]	[38.9 ; 42.4 %]	[39.3 ; 42.9 %]	[41.1 ; 44.7 %]	[39.6 ; 43.3 %]	[37.7 ; 41.9 %]
car passenger	[6.5 ; 8.5 %]	[6.5 ; 8.5 %]	[8.1 ; 10.1 %]	[6.2 ; 8.1 %]	[8.8 ; 11 %]	[18.8 ; 21.8 %]	[17 ; 20.4 %]
train	[3.1 ; 4.5 %]	[3.3 ; 4.9 %]	[3 ; 4.4 %]	[3.9 ; 5.4 %]	[2.9 ; 4.2 %]	[0.8 ; 1.6 %]	[0.8 ; 1.7 %]
bus/tram/metro	[3.7 ; 5.2 %]	[5.5 ; 7.4 %]	[5.1 ; 6.7 %]	[4.9 ; 6.6 %]	[6.1 ; 7.9 %]	[4.1 ; 5.7 %]	[2.9 ; 4.5 %]
other	[0.9 ; 1.8 %]	[1.4 ; 2.4 %]	[1.2 ; 2.2 %]	[1.2 ; 2.1 %]	[1.2 ; 2.1 %]	[0.4 ; 1 %]	[0.6 ; 1.4 %]

Table 22: confidence intervals for mode distribution, for each day of the week

As often, car is the most used mode, with more than 40% if we only consider car as driver and around 50% when we add car as passenger. This modal part could seem quite under the Belgian one (nearly 70% in MOBEL) but we must keep in mind that this survey covers an urban area where mobility measures have been taken to avoid too many cars in town.

The Flemish people are reputed to be regular bicycle users and this is confirmed by this survey in Ghent. Bicycle is the main mode for 20% of the trips. It is more or less at the same level as the walk which is even significantly overtaken on Wednesday.

The main and remarkable difference according the days is that car as passenger is used more than twice more on weekends than on weekdays. The modes that loose there market shares are mainly bicycle and in a lighter way train.

We could advance some explanations:

- bicycle is mainly used by children during the week to go to school or to other activities (more on Wednesday where walk slightly decreases), whilst they follow their parents during week-ends in their car for less regular activities, for family activities;
- train is mainly used for commuting.

We can also remark that

- public transport is around 5% but lightly decreases on Sunday and Monday;
- car passenger is slightly more used on Wednesday and Friday than on other weekdays. We can remark that those days are the days where it is less worked. We can venture the hypothesis that car is more needed on those days because trips are less regular (less home-work trips), and there are more "drop off/pick up" trips;
- walk is nearly 5% higher on Sunday than on the other days.

An explanation of the difference in the used modes according to the day should certainly be found in the fact that the trip purposes are also different from day to day especially during weekends.

Nevertheless, if we performed a logistic regression to explain the mode used for a trip according to its purpose and the day of the week, the day remains significant (and the purpose is evidently also significant).

If we add the trip distance in the model, then the distance is the most explanatory variable. Purpose is still significant; and so is the day.

III.1.3.9. Departure time spreading [statistics on trips]

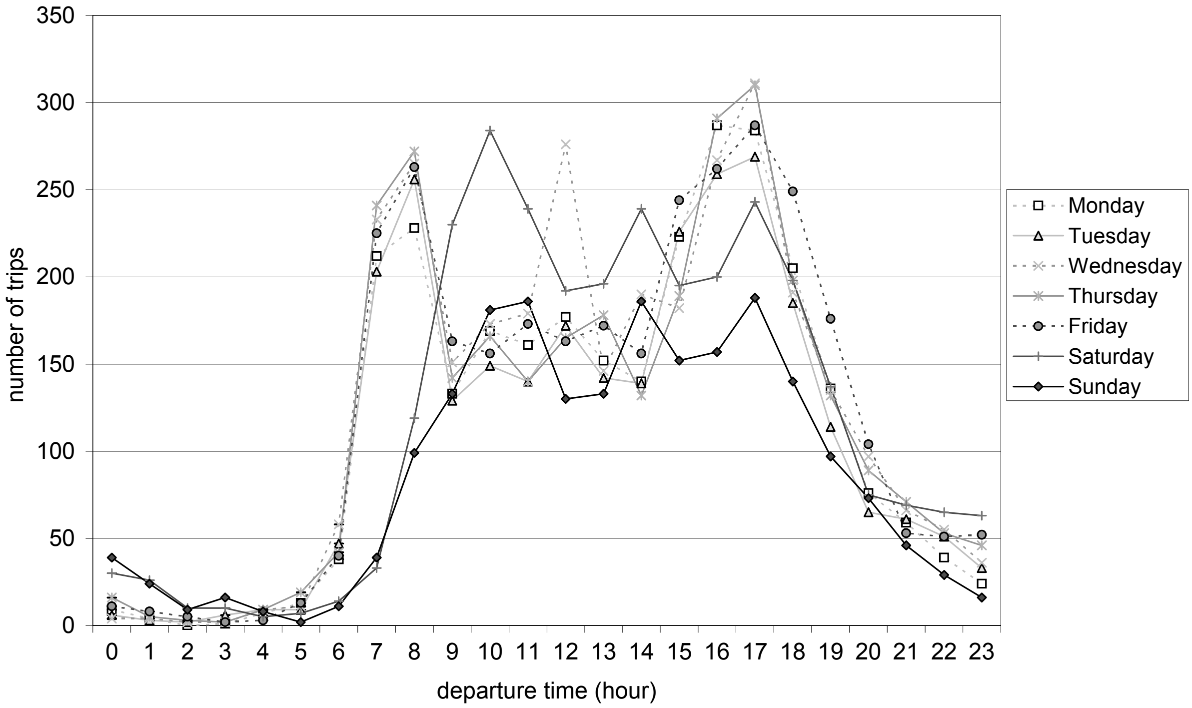


Figure 32 : Trip departure times according to the day

The morning peak hour is quite similar for all weekdays, except for Monday where it is a bit weaker. The evening peak is wider, especially on Friday. Wednesday is marked by a midday peak that can mainly be explained by the fact that there is no school on Wednesday afternoon in Belgium. Saturday and Sunday have specific hourly profiles. Saturday has an important morning peak but later than on weekdays (around 10) but also two minor peaks at 14 and 17. As we know, the number of trips on Saturday is at the same level as on weekdays but purposes are different. People travel less on Sunday, also presenting three peaks but with equal trips densities: it is nevertheless larger in the morning (between 10 and 12).

If we only consider car mode (both driver and passenger), the profiles are quite different (see Figure below). It increases the difference between morning and evening peaks on weekdays. The Wednesday midday peak has also considerably decreased, which could be expected if we suppose that it is mainly caused by trips by children and young people back from school (on foot, by bicycle). The morning and midday peaks of Saturday are really higher than on weekdays. This could perhaps be explained by the fact that most commuters do not travel by car for their home-work trips but use their car for their less usual trips achieved on Saturday (shopping, leisure, visit, etc.)

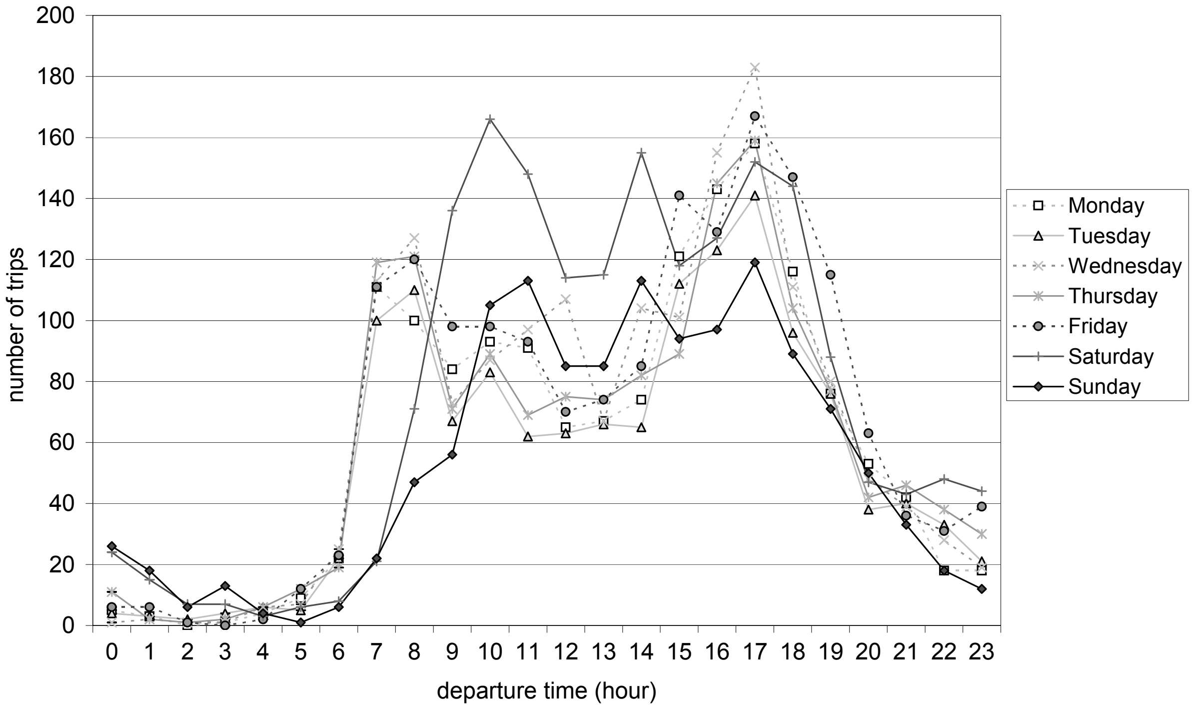


Figure 33: trip departure times according to the day (by car)

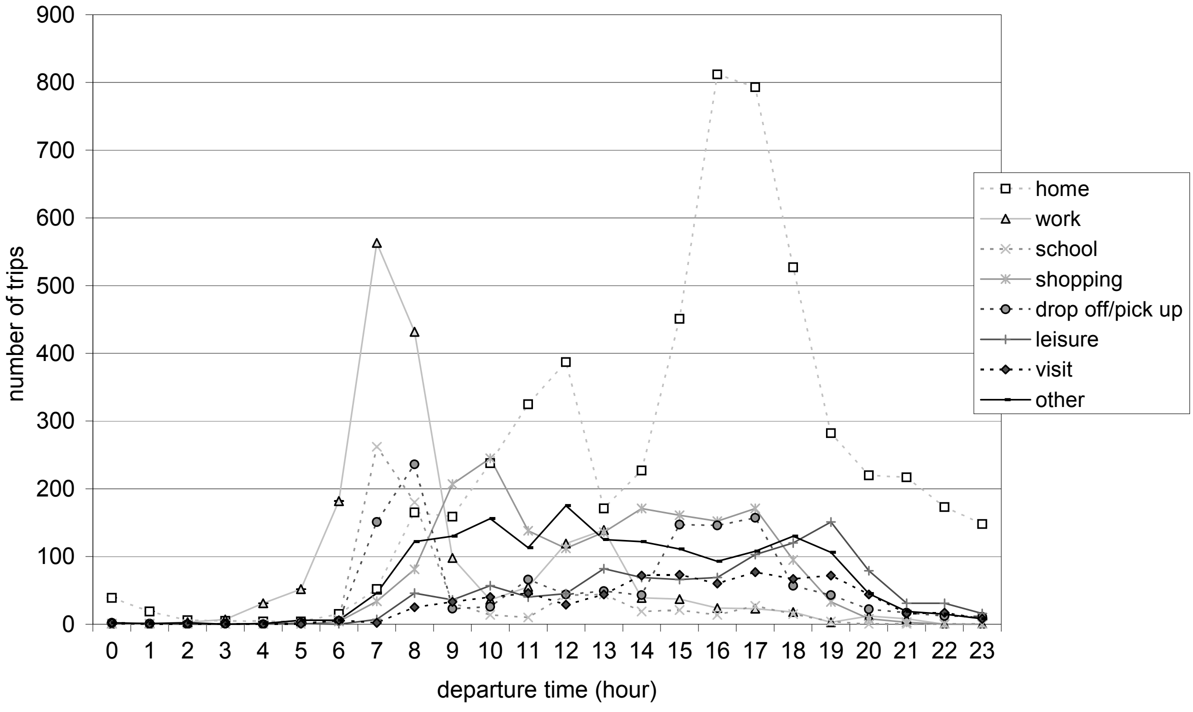


Figure 34: trip departure times according to purposes, for weekdays

While the morning peak on weekdays is mainly due to work, school and drop off/pick up, leisure and especially shopping dominate on Saturday. These trips are rapidly followed by a return to home which can explain why we observe a huge peak on Saturday morning (as two successive trips could be taken into account into the narrow window of the morning peak).

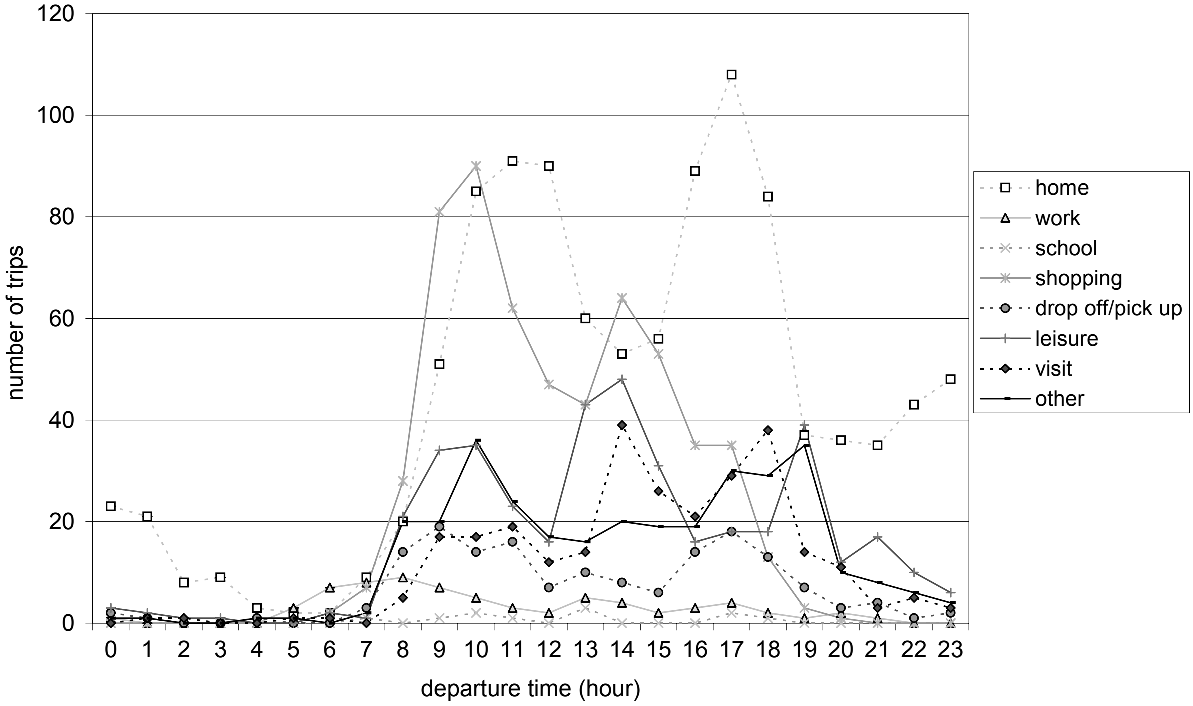


Figure 35: trip departure times according to purposes, for Saturdays

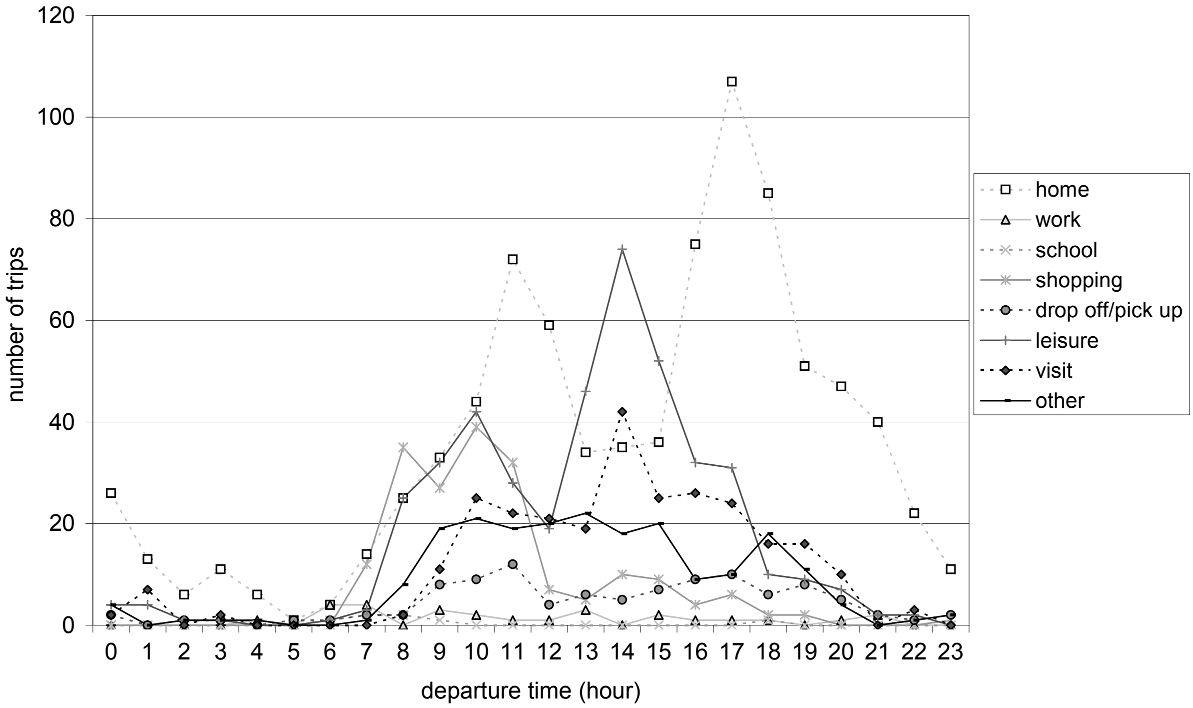


Figure 36: trip departure times according to purposes, for Sundays

III.2. Traffic data

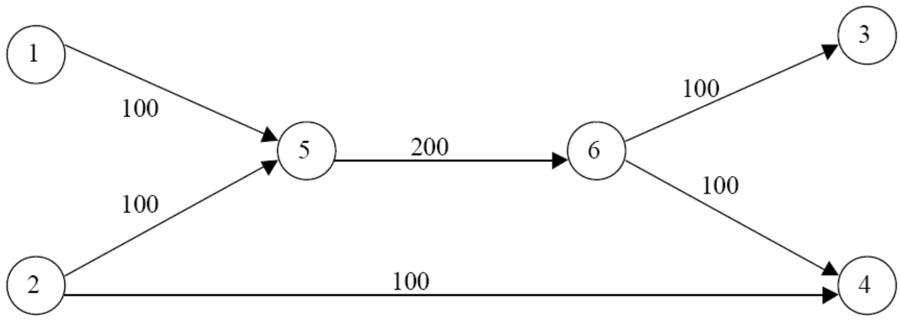
With the aim of linking measured daily traffic patterns with (weekly) activity patterns of road users, this project has undertaken two approaches. The *activity-based approach* views traffic patterns as natural derivation of activity scheduling processes, thus traffic is seen as an aggregate representation of disaggregate activity and travel choices. On the other hand weekly mobility can be estimated directly from aggregate measures of traffic (*traffic data-driven approach*), in the case of this project the traffic counts collected through loop detectors. These vary for each successive day of the week, and are argued to show similar trends on the same day along different weeks. This part of the report deals with the latter approach.

Therefore, in this part of the report we describe how we tackled the problem of estimating (time-dependent) Origin-Destination trips from traffic flow data collected in the Ghent region over multiple weeks. First, we describe the estimation problem and highlight the modeling aspects that are relevant for analyzing the weekly mobility of road users, and the issues and limitations characterizing this problem. Our approach to the OD estimation is then presented both in a theoretical and algorithmic way. Finally we show the results of the estimation, focusing in particular on the variation in traffic patterns generated in a within-day and in a weekly time horizon and we summarize the conclusions of our findings.

III.2.1. Problem description

Traffic flows, observed by means of traffic counts on a subset of all links, originate from specific traffic patterns generated from all origin-destination pairs in a network. This implies that traffic counts contain information on multi-commodity flows, i.e. they can be decomposed in several partial flows depending on the number of routes connecting any OD pair. As one can easily understand, there can be many combinations of these flow fractions that result in the same link flow values, so the problem is typically underdetermined and the set of possible solutions usually grows with the size of the network in consideration, and the travel alternatives available for each OD pair, while it usually reduces by increasing the number of detectors. Generally speaking, the under-determinedness is expected in all cases where the information used to estimate the OD flows is insufficient to determine them unambiguously, which is the most likely scenario in practice (Bierlaire and Crittin, 2003, Marzano et al., 2007).

Two clarifying examples of this assertion are given in Wu et al. (2004) and in Yang and Zhou (1998). A simple network connecting four OD pairs is represented in Figure 37. A bypass is assumed for OD pair 2-4. Link flows are assumed known for all links. The table below the graph shows two equally possible OD tables consistent with the measured link flows. This is easily understandable considering that in this problem the number of unknowns (the four OD flows) is higher than the number of independent known parameters. In fact only three link flows are needed to derive all other measured link flows.



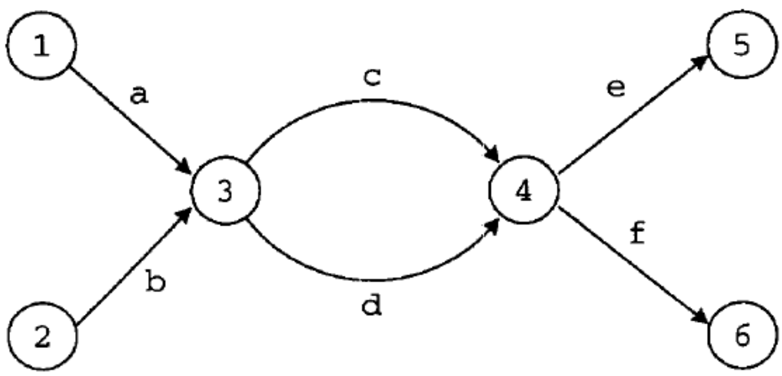
O\D	3	4
1	0	100
2	100	100

O\D	3	4
1	50	50
2	50	150

Figure 37: Two equally possible solutions for a set of link flows (taken from Wu et al., 2004)

A way to solve this under-determinedness is to add extra information to the traffic counts, traditionally in the form of a prior estimate of the OD table (e.g., an outdated OD table), or by including route choice criteria.

Multiple solutions can also be calculated from the same prior matrix and by selecting different positions for the detectors, as shown in an example by Yang and Zhou (1998) and replicated in Figure 38. The first table, below the network representation, gives all relevant input for the example, while the second table shows the results of applying Generalized Least Square Estimation. As one can observe, the solution deviates from the true OD matrix in all scenarios, and it can also differ using the same number of traffic counts in the network. In the last two columns a reliability measure is represented, which gives a measure for the degree of under-determinedness for each scenario. The choice of where to locate sensors is therefore very important in view of OD estimation.



Link flows, route fractions, prior and actual OD matrices for the above toy network

O–D pair	p_{aw}	p_{bw}	p_{cw}	p_{dw}	p_{ew}	p_{fw}	Prior matrix	Actual matrix
1–5	1.0	0.0	0.4	0.6	1.0	0.0	100	120
1–6	1.0	0.0	0.5	0.5	0.0	1.0	60	80
2–5	0.0	1.0	0.8	0.2	1.0	0.0	50	60
2–6	0.0	1.0	0.7	0.3	0.0	1.0	80	100
Observed flow	200	160	206	154	180	180		

OD estimation results and their reliability for different detector positions

Set of counting links	Total observed flow	Net observed flow	Estimated O–D matrix				MPRE(λ)	Re(T)
			t_{1-5}	t_{1-6}	t_{2-5}	t_{2-6}		
d, e	334	250	123.57	82.43	56.43	92.86	2.08	0.32
b, d	314	272	121.95	78.29	63.07	96.83	1.49	0.40
c, d	360	360	121.50	79.50	63.50	95.50	1.42	0.41
a, b	360	360	120.00	80.00	65.00	95.00	1.26	0.44
c, d, e	540	360	119.00	82.00	61.00	98.00	0.90	0.53

Figure 38: Example of multiple solutions for OD estimation on a toy network using different traffic count locations (taken from Yang and Zhou, 1998)

The above examples are limited to a static representation of traffic, i.e. link and OD flows are assumed stationary for an indefinite period of time, and links are assumed having infinite capacity, or, equivalently, they are assumed to be under-saturated. Moreover, all known parameters have been assumed error-free. The problem becomes more complex if one considers also the dynamic and stochastic nature of traffic flows, which reflects into variable route fractions measured at each link, and the limited capacity of road networks, which causes the emergence of congestion. As a consequence, traffic counts can be interpreted wrongly, as the same number can be observed under light conditions of traffic (low density and high speeds), and during heavy traffic (high density and low speeds). To resolve these issues extra information is again needed, for instance one can use information on speeds or link occupancies to identify unambiguously the state of the network, or choose a proper dynamic traffic model. This last approach will be used in this study, as it will be described later in this document, since we cannot obtain the above-mentioned extra information.

In addition to the above theoretical and methodological shortcomings, measured traffic counts are not error-free, as often datasets are disseminated with corrupted data caused by missing counts or faulty detectors. Moreover, traffic variations, caused by both traffic dynamics and by the stochastic nature of flows, are observed in an aggregate manner at the link level and result hardly separable in the OD estimation process without a proper smoothing of the data.

III.2.2. OD estimation method

In this section we describe the method developed to estimate the dynamic OD matrix from link counts. First a theoretical framework is provided, where we show the various parts of the approach and their relations, and later we describe the various parts of the method and the computational algorithm.

III.2.2.1. Theoretical framework

We search for the OD flows that reproduce our measurements as closely as possible when assigned to the network using a traffic model. This can be formulated as a minimization problem, with an objective function containing the deviations from the measurements. Because of the under-specification of the problem the deviation from the starting OD matrix (or target matrix) is also included in the objective function. The OD estimation problem can therefore be formulated as follows:

$$\min_{x} \left(Z_1\left(x,\hat{x}\right) + Z_2\left(y(x),\hat{y}\right)\right)$$

where x is the OD matrix to be estimated, $\hat{x}$ is the target matrix, y are the simulated link flows, and $\hat{y}$ are the traffic counts. The simulated link flows are the outcome of a traffic model. As already mentioned this model requires an initial OD matrix as input. It should be also stressed out that the relationship between the link flows and the OD matrix can become rather complex when using a dynamic traffic model. If every OD pair is in fact defined for *n* periods of time, and every OD flow is defined for *m* periods of time, then the relationship between link flows and OD flows can be expressed by using an assignment proportion $a_{i,m}^{j,n}\left(x\right)$, representing the proportion of OD flow *j* departing in time period *n* that passes link *i* during time period *m*. This relationship can be written down as follows:

$$y_i^m = \sum_{n=1}^{m} \sum_{j} a_{i,m}^{j,n}\left(x\right) x_j^n$$

If we rearrange the OD matrix and the link flows into vectors, then these assignment proportions can be grouped together in a matrix, which we will call the assignment matrix A. This assignment matrix can be decomposed further by introducing the path flows. Let $h^{r,n}$ be the path flow departing during time period *n* following path *r* between OD pair *j*. These path flows can thus be expressed as the product of a certain OD pair x_j^n and the proportion of travelers of that OD pair using path *r* during time period *n*:

$$h^{r,n} = p^{r,n} x_j^n$$

We also introduce the relationship between the path flows and the link flows:

$$y_i^m = \sum_{n=1}^{m} \sum_{r} b_{i,m}^{r,n} h^{r,n}$$

where the summation is done over all paths passing link *i*. In this expression $b_{i,m}^{r,n}$ is the crossing proportion, which is the proportion of path flow $h^{r,n}$ passing link *i* during time interval *m*. This proportion thus defines the spatio-temporal propagation of the OD flows over the network.

The above equations can be combined to find the decomposition of the assignment matrix:

$$a_{i,m}^{j,n} = \sum_{n=1}^{m} \sum_{r \in j} b_{i,m}^{r,n} p^{r,n}$$

or in matrix notation:

$$A = BP$$

where A is the assignment matrix, B is the crossing matrix, and P is the route choice matrix.

This decomposition makes it possible to allocate the two sources of influence that the OD flow has on the assignment matrix. Both are due to congestion effects: an increased OD flow can cause increased travel times, and therefore a part of this (or another) OD flow will arrive in a later time period than before. This means that the spatio-temporal distribution, described by crossing matrix *B*, will alter. Furthermore these increased travel times can have an effect on route choice behavior, altering the proportions in route choice matrix *P*.

Because of this interdependency this problem is a fixed point problem. As mentioned in the literature overview of algorithms for solving this fixed point problem generally an iterative approach between two levels is used: an upper level in which the OD flows are estimated, and a lower level in which these OD flows are assigned using a DTA model. In the next subsections we will go into detail about each level, and present the computation algorithm.

III.2.3. The dynamic OD estimation method

In this section the models chosen for the upper and lower levels are justified and explained. Paragraph IV.2.4 will then deal with specific implementation details in a computation algorithm.

III.2.3.1. Upper level: OD estimation problem

The major components of the upper level consist of the objective function and the optimization technique. A first issue is the choice of a proper objective function, which consists of a deviation function presented in symbolic form in the previous section. The Maximum Entropy function does not seem appropriate here, because we do not wish to use the target matrix solely as a starting point (see literature review): the structure of the final OD matrix should resemble the structure of the target matrix. The Maximum Likelihood and Bayesian estimators consider the target matrix and the traffic counts as stochastic variables, and probability distributions for both types of variables have to be defined. Neither of both distributions is known to us, and a more general approach is more appropriate. Therefore we choose for Generalized Least Squares as deviation function in this project.

The objective function includes the deviation from the start OD matrix and the deviation from the counts. The objective function is thus stated as follows:

$$f(x) = \left(\gamma_1^T \left(x - \hat{x}\right)^2 \ + \ \gamma_2^T \left(Ax - \hat{y}\right)^2\right)$$

where A is the assignment matrix, γ_1 is a vector of weights that express the confidence one has in the starting value of a certain OD pair, and γ_2 is a vector of weights that express the confidence one has in a link count. The confidence in the target OD matrix should normally be derived by some stochastic inference, but no information on the actual reliability of this data was available. Therefore the weights were chosen such that the total amount of information from the traffic counts equaled the information included in the target matrix. Practically, γ_1 and γ_2 were taken equal to the inverse of the total number of elements in the OD matrix and the inverse of the total number of counts respectively.

For the optimization algorithm there are many different possibilities. Algorithms most used in literature in OD estimation applications include Kalman filter approaches (e.g., Ashok, 1996, Kang, 1999, Zhou and Mahmassani, 2004), gradient search methods (Maher et al., 2001), genetic algorithms (e.g., Kim et al., 2001, Kattan and Abdulhai, 2006), and more recently Simultaneous Perturbation Stochastic Approximation algorithms (SPSA) (Balakrishna and Koutsopoulos, 2008). Each of these algorithms has its own limitations in terms of applicability. The Kalman filter approach is normally used in a real-time environment, and therefore sometimes requires very crude simplifications. Therefore this does not seem as a proper approach. Genetic algorithms are not as susceptible to local minima in the objective function as gradient search and SPSA, but suffer from scalability issues as studies have reported (Henderson and Fu, 2004, Thierens, 1999). Since we are dealing with a large-scale network, we do not use a genetic algorithm. SPSA is a technique that is quite popular nowadays in OD estimation problems. It calculates a gradient approximation using only 2 objective function evaluations (a objective function evaluation is equal to one traffic simulation), in contrast with the Finite Difference method, which requires *n* function evaluations, *n* being the number of variables. The fact is however that one objective function evaluation (= traffic simulation) provides us with an assignment matrix. This assignment matrix expresses the local linear interrelation between the OD matrix and the link flows, and thus makes it possible to calculate the exact (local) gradient of the objective function. Thus the gradient search method was chosen, because it has the advantage that it is able to make full use of the output of the lower level, namely the assignment matrix.

The gradient method calculates the gradient of the objective function by deriving the objective function to each of its variables, and setting this equation equal to zero. With our objective function the gradient can be calculated as follows:

$$\nabla f(x) = \begin{bmatrix} \frac{df(x)}{dx_1} & \frac{df(x)}{dx_2} & \cdots & \frac{df(x)}{dx_n} \end{bmatrix}$$

$$\text{with } \frac{df(x)}{dx_i} = 2\gamma_{1,i}\left(x_i - \hat{x}_i\right) + 2\gamma_{2,i}\left(Ax - \hat{y}\right)^T A_i$$

The gradient points out the direction of steepest descent of the objective function. The current estimate of the OD flows x_k can thus be updated as follows:

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

where α is the step-size. The gradient now has to be recalculated for x_{k+1}, requiring another traffic simulation in the lower level to obtain an assignment matrix consistent with x_{k+1}.

III.2.3.2. Lower level: Dynamic Traffic Assignment

The lower level consists of a route choice model and a dynamic network loading (DNL) model. The input for this level is an OD matrix, the output consists of link flows, the assignment matrix, travel times, queue lengths, etc. For assigning the OD matrix onto the network first the route proportions have to be calculated. This is done by the route choice model. We used the route choice model from the program Indy (BLIEMER et al., 2004). This model consists of a route generation model and a route distribution model. In the route generation model the route alternatives for drivers are calculated. Enumeration of all possible routes between an origin and a destination is not possible for large networks as ours, so an algorithm is needed to select the most appropriate routes. Indy uses a Monte Carlo route generation, which only considers the network characteristics. This algorithm iteratively tries to find new fastest routes based on stochastic link travel times. Each link travel time consists of a constant free-flow travel time and a stochastic part. In each iteration the fastest route is computed for each origin-destination (OD) pair. If it is a new route and this new route does not overlap with the other routes for a significant part, it is added to the route set. New random numbers are drawn from a stochastic distribution function (depending on the link length) and added to the free-flow link travel times.

This yields new link travel times in each iteration and therefore potentially lead to new fastest routes to be added to the route set. The variance of the stochastic term is increased in each iteration (up to a certain maximum value), yielding an accelerated Monte Carlo approach. The higher the variance, the more likely it is to find a new fastest route.

Next, the obtained routes are used in the route distribution model. All drivers are assumed to take the route with minimum perceived travel cost c (here the travel costs consist solely of the travel time). The route proportions therefore depend on the drivers' perception. To simulate the fact that drivers have imperfect information about the travel times of different routes, an error term ε is added to the actual route costs, yielding perceived route costs c'.

$$c' = c + \varepsilon$$

The behavior of drivers is modeled as such that drivers try to minimize their perceived travel cost. Depending on the assumed form of the error term ε, the route proportions can be calculated. In Indy, it is assumed that the error terms are identical and independently distributed, resulting in a simple Multinomial Logit model to compute the proportions.

$$p^{r,n} = \frac{\exp(-\mu c_{r,n})}{\sum_{r \in j} \exp(-\mu c_{r,n})}$$

where $\mu \geq 0$ is the scale parameter of the Logit model. If μ is small, then drivers have an inaccurate perception of the travel costs on different routes, and therefore there is a large spread among the different routes. If μ is large, then a large proportion of drivers will use the shortest route. Using these route proportions the path flows can now be calculated, and thus the link also by using a traffic assignment model. We see that the path flows, and thus the link flows depend on the travel costs, but these travel costs depend on the link flows. This interdependency makes the problem difficult to solve: we are dealing with a fixed point problem again. Indy tries to solve this problem using a simple iterative procedure based on the method of successive averages, and iterations are made between the route distribution model and the DNL model. When this procedure converges and the path flows and path costs are consistent both with the route distribution model and the DNL model, the system is said to be in equilibrium, following the equilibrium law of Wardrop (1952). In this state drivers have no incentive to change from one route to another, as they cannot unilaterally decrease their travel cost by changing routes (Wardrop, 1952).

III.2.3.3. Dynamic Network Loading model

We now go into detail about DNL models. These models require an OD matrix and route proportions as input, and will load the OD flows onto the network, resulting in path flows and link flows. Basically they propagate the OD flows according to certain rules.

DNL models are often classified in microscopic, mesoscopic and macroscopic models. Microscopic models such as AIMSUN2 (Barcelo, 2002) and VISSIM (PTV, 2005) describe traffic flow on the level of individual vehicles. These models are generally too time consuming in the OD estimation framework, since several simulation runs are needed.

Mesoscopic models such as DYNASMART (Mahmassani et al., 2001) and DYNAMIT (Ben-Akiva et al., 1998) move individual (packets of) vehicles according to macroscopic traffic flow relations. They are less cumbersome computationally, but they are also less precise in the representation of traffic dynamics. In large networks, macroscopic models (moving vehicles as a continuum) have a significant computational advantage over microscopic and mesoscopic models. Examples are Indy (Bliemer et al., 2004), METANET (Messmer and Papageorgiou, 1990) and the Cell Transmission Model (Daganzo, 1994). Yperman et al. (2007) have recently presented the Link Transmission Model (LTM), a macroscopic DNL model that provides high realism in the representation of queue propagation and dissipation and high computational efficiency. It is therefore perfectly suited for simulation on large networks like the Ghent network.

LTM is consistent with simplified first order kinematic wave theory. This theory assumes a functional relation between traffic flow q and density k, captured in the triangular shaped fundamental diagram (Figure 39). From the figure, a bi-linear fundamental diagram is uniquely determined by a number of calibration parameters, respectively q_M and k_M are the so-called maximum flow and maximum density. This point in the diagram is determined also by the free flow speed measured at one link v_f. Finally the value of the jam density k_{jam}, i.e., for which the flow is assumed to become zero as vehicles are in state of complete immobility, determines the speed of the shockwave, moving in the opposite direction of the flow.

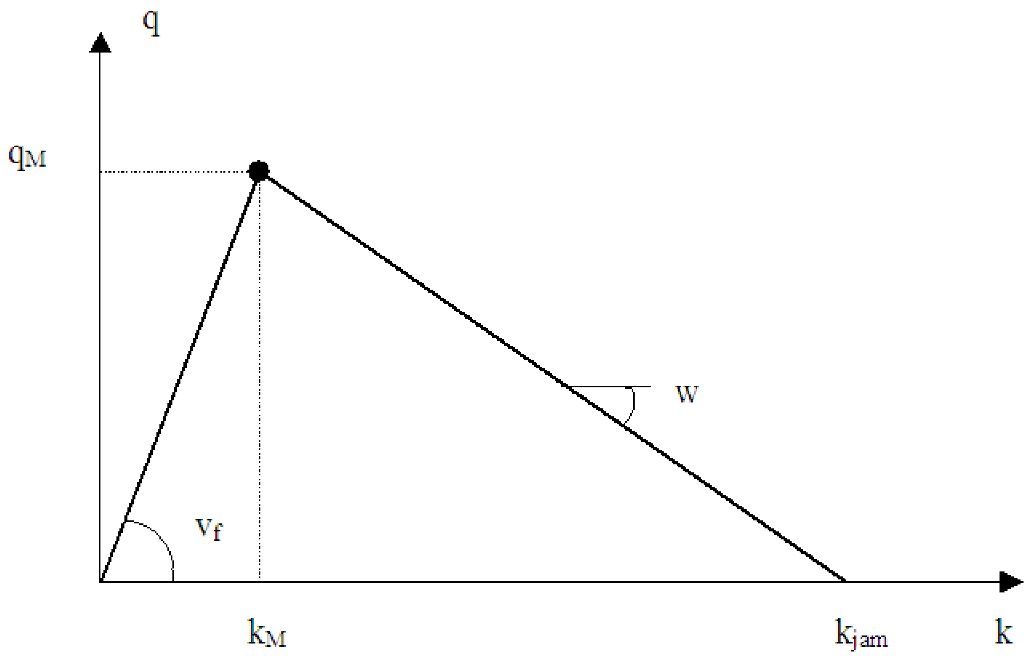


Figure 39: Triangular-shaped fundamental diagram

In free flow ($k < k_M$), vehicles travel with a fixed free flow speed v_f. Congested traffic ($k > k_M$) travels with a speed q/k. Traffic states move through the links of the network with a wave speed dq/dk. When a free flow state intersects with a congestion state, a shock wave originates that may travel upstream or downstream in accordance with the intersecting traffic states.

The LTM is a multi-commodity model, where each commodity corresponds to a specific (pre-defined) route. Since vehicles are disaggregated by route, traffic is represented by the cumulative number of vehicles $N^p(x,t)$ of route p that has passed location x by time t. These numbers are calculated and stored every time step at the upstream end x_a^0 and the downstream end x_a^L of each link a. Link volumes and link travel times are derived from these cumulative vehicle numbers.

The LTM algorithm can be divided into three steps, executed every time step:

Step 1:

For each node n, the sending flows at the downstream end of all incoming links and the receiving flow at the upstream end of all outgoing links are calculated. The sending flow $S_{ij}(t)$ at time t is defined as the maximum amount of vehicles that could leave incoming link i and enter outgoing link *j* during time interval *[t -*Δt*, t]*, assuming an infinite capacity for link *j*. The receiving flow $R_j(t)$ of outgoing link *j* at time *t* is defined as the maximum amount of vehicles that could enter link *j* during time interval *[t -*Δt*, t]*, assuming an infinite traffic demand upstream (see Yperman, 2007a for details).

Step 2:

For each node *n*, the flows that are actually transferred from every incoming link *i* to every outgoing link *j* are determined. These transition flows $G_{ij}(t)$ are calculated by the node model, which always obeys to conservation of vehicles.

Step 3:

For all link boundaries x_a^0 and x_a^L, the cumulative vehicle numbers *N(x,t)* are updated:

$$N(x_i^L, t+\Delta t) = N(x_i^L, t) + \sum_j G_{ij} \qquad \textit{for all } j \; (2)$$

$$N(x_j^0, t+\Delta t) = N(x_j^0, t) + \sum_i G_{ij} \qquad \textit{for all } i \; (3)$$

III.2.4. Computation algorithm

Now that the theoretical framework is presented and the models used in framework are selected, we will go into detail about the specific implementation issues. In Figure 40 a flowchart of the OD estimation model is presented.

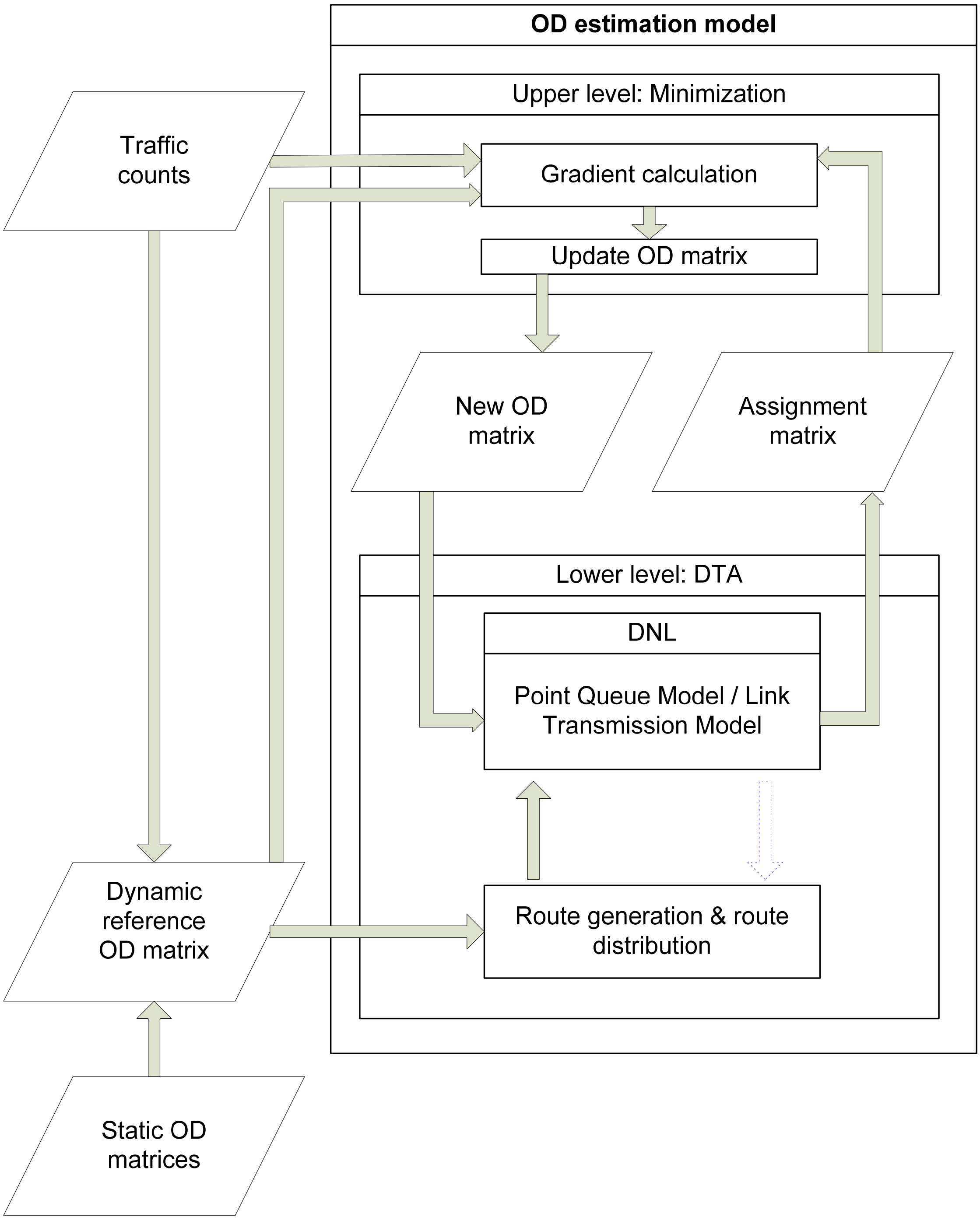


Figure 40: OD estimation flowchart

As described in Section III.3.3 of this document the dynamic reference OD matrix is calculated by making use of the traffic counts and the static OD matrices for the morning and evening peaks. This reference OD matrix is input for the upper and lower levels. In the upper level the deviation from this reference matrix is penalized in the objective function, albeit the name target matrix. For every type of day we use the same target matrix in the estimation process, because of two reasons.

First of all, it was not possible to generate a proper dynamic OD matrix for all different days. The available static OD matrices are representative for an average workday; no distinction is made between the different work days. Although the temporal pattern of the OD matrices can be deduced approximately, no distinction of magnitude can therefore be made between the different work days. Secondly, by using the same reference matrix in the OD estimation process, differences between the final estimated OD matrices for each day are solely due to differences in the counts, and not to using different starting matrices for each day. If every type of day had a different target matrix, then we would already implicitly impose differences between the days, and therefore it would not be clear whether the differences between the estimated OD matrices would be due to real differences in the traffic counts, or due to different input.

In the lower level the reference OD matrix is used as a starting point: first the different routes are generated, a process that is determined solely by the network characteristics. A maximum of 7 routes per OD pair was determined, although the actual amount of routes per OD pair was much lower, especially for long distance OD pairs, as a result of scarcity of alternative routes. In Figure 41 an example of the output of the route generation model can be seen.

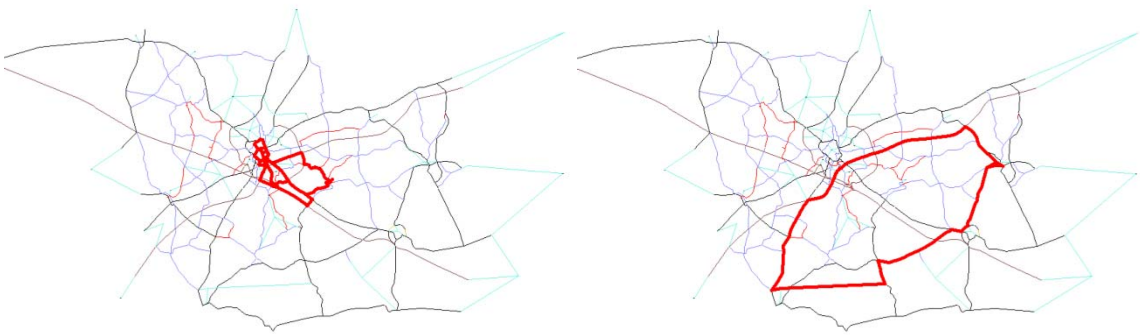


Figure 41 : Different route alternatives identified by the route generation algorithm

Next, the OD matrix is used in the DTA procedure of Indy to produce route proportions. The DNL model chosen in this procedure was a first-order model that could reproduce shockwaves, but problems were encountered using this model. The route proportions were changed too drastically from one iteration to another, causing serious congestion problems, and finally causing gridlock, a state in which a closed circuit is formed where all vehicles stand still (Daganzo, 1996). To remedy this problem we chose to use another DNL model available in Indy that uses travel time functions for propagating the OD flows. Such models cannot model congestion spillback, and are thus not recommended when dealing with highly congested networks. In the network of Ghent we do not expect serious congestion effects apart from the motorway, and even on motorways congestion is localized within specific areas. Therefore we expect only minor errors from using this simplified model.

The DTA procedure consisted of 10 iterations between this DNL model and the route distribution model. The total travel time can be considered as a measure for convergence, as drivers tend to change routes because of differences in travel time. In Figure 42 the total travel time for all vehicles in every iteration is depicted, and convergence, and thus equilibrium, is assumed at iteration 10.

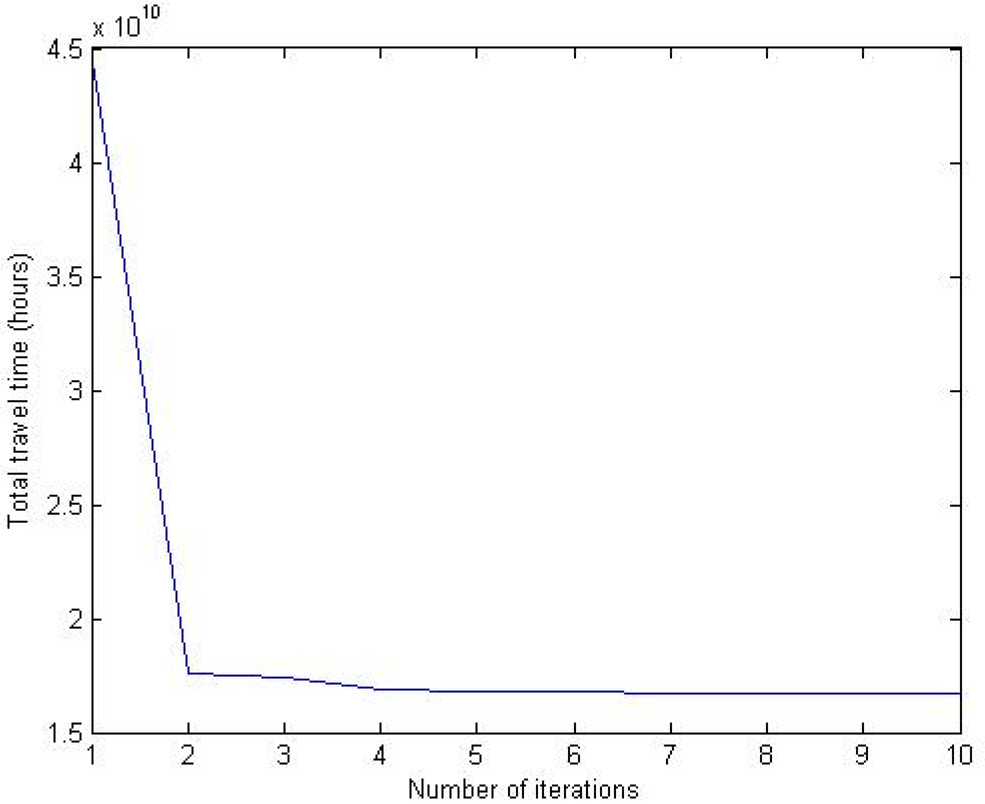


Figure 42: Total travel time in every iteration

The route proportions produced by the DTA procedure as well as the reference OD matrix are input for the DNL model. As to avoid the problem of gridlock that was encountered in the DTA procedure, we chose to replace the LTM with a point queue (PQ) model in the first iterations of the OD estimation. PQ models are not able to represent congestion spillback, as they stack all queuing vehicles in one link. For more details about the PQ model, we refer to Zhang and Nie (2005). After these first iterations with the PQ model, we assume that the estimated OD matrix is close enough to the real solution, and we use the LTM again, as we do not expect any more gridlock phenomena.

The DNL model then loads the OD flows onto the network, and calculates the link flows and assignment matrix. The assignment matrix is sent back to the upper level, and a new OD matrix is determined, which is input for the DNL model. This new OD matrix is loaded onto the network, and again the link flows and assignment matrix are calculated. As the dotted arrow in Figure 40 indicates, these link flows should be transferred to the route distribution model to recalculate the route proportions, and the new route proportions should be used to repeat the DNL process. This should be repeated until equilibrium is reached. This process would thus require many DNL calculations for every iteration in the OD estimation process. This does not seem feasible considering computation time, as the DNL calculation is the bottleneck regarding this aspect. Therefore the feedback loop represented by the dotted arrow in Figure 40 was not used, but instead the route proportions for the reference OD matrix were used throughout the estimation process.

In other words, we only perform a DTA (=route choice + DNL) in the first iteration; thereafter the DTA is replaced by one DNL. This simplification can be justified by considering the fact that the network of Ghent does not experience serious congestion, with an exception for some link on the highway. Therefore route redistribution effects can be neglected on the secondary roads, while drivers on the highway normally do not have many route alternatives, so the effects for highways are also negligible.

III.2.5. Presentation of results

III.2.5.1. Validation

In this section the outcome of the OD estimation process is evaluated and later analyzed. In Figure 43 the evolution of the objective function (upper level) during the iteration process is presented. As mentioned in the previous chapter the first iterations were done using a point queue model, while the final iterations were done using LTM. In the weekend there is almost no difference between both models, as no significant congestion occurs then. Note that in the iteration process weekend days start with a larger error, as the starting OD matrix was calibrated for an average work day. The OD estimation process however was able to reduce this error to a similar level as for work days.

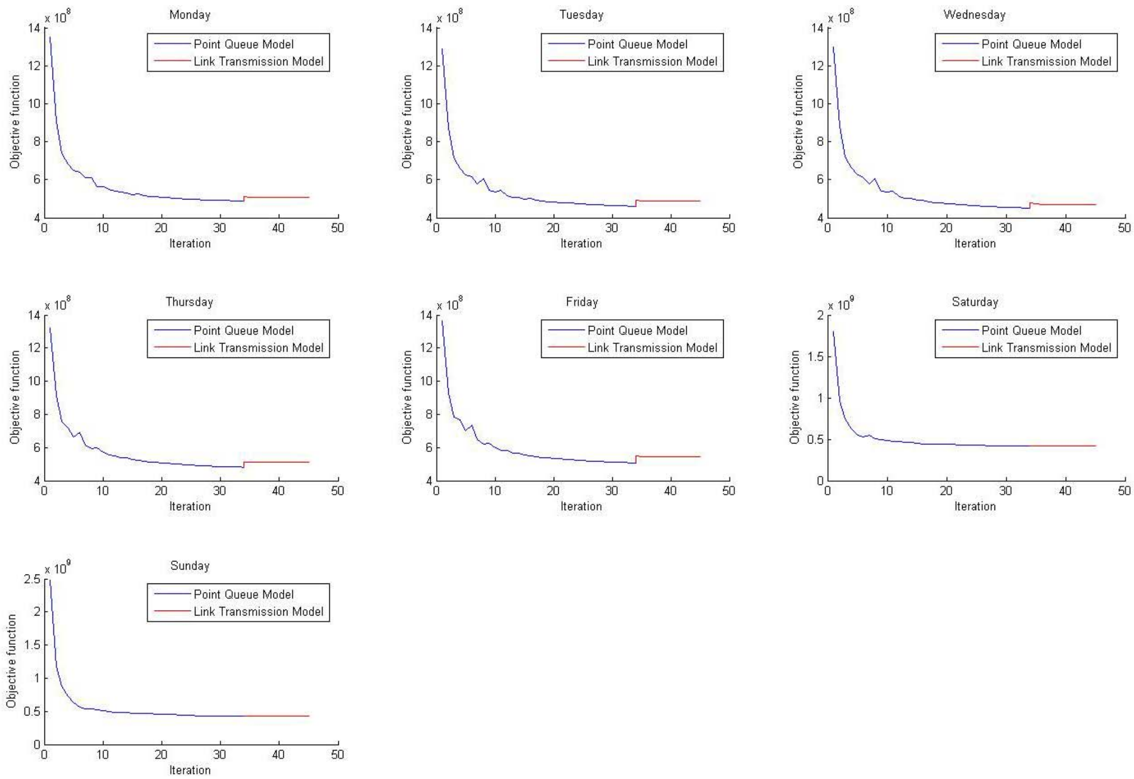


Figure 43 : Evolution objective function

As one can see, the optimization process quickly converges using the point queue model, and gets close to the final solution in less than 20 iterations. By using LTM the value of the objective function slightly increases for the work days, as this loading procedure adds the effect of congestion to the solution. However, the estimated OD matrices do not change significantly, as the value of the objective function remains invariant for the following iterations.

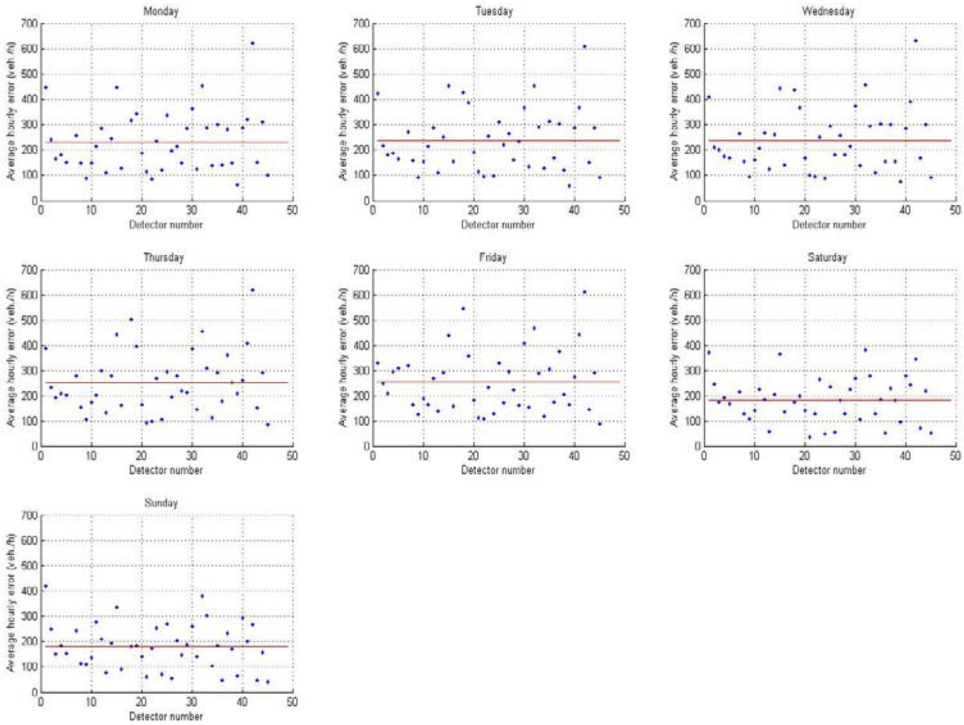


Figure 44 : Average hourly error of highway detectors

As a measure of performance of the applied OD estimation procedure in Figure 44 the average hourly deviation between the simulated link flows and the highway counts is depicted. The red line indicates the average hourly error over all highway detectors while the blue dots represent the error for a specific link flow. For work days this error is around 240 vehicles per hour, while in the weekend the error reduces to around 180 vehicles per hour.

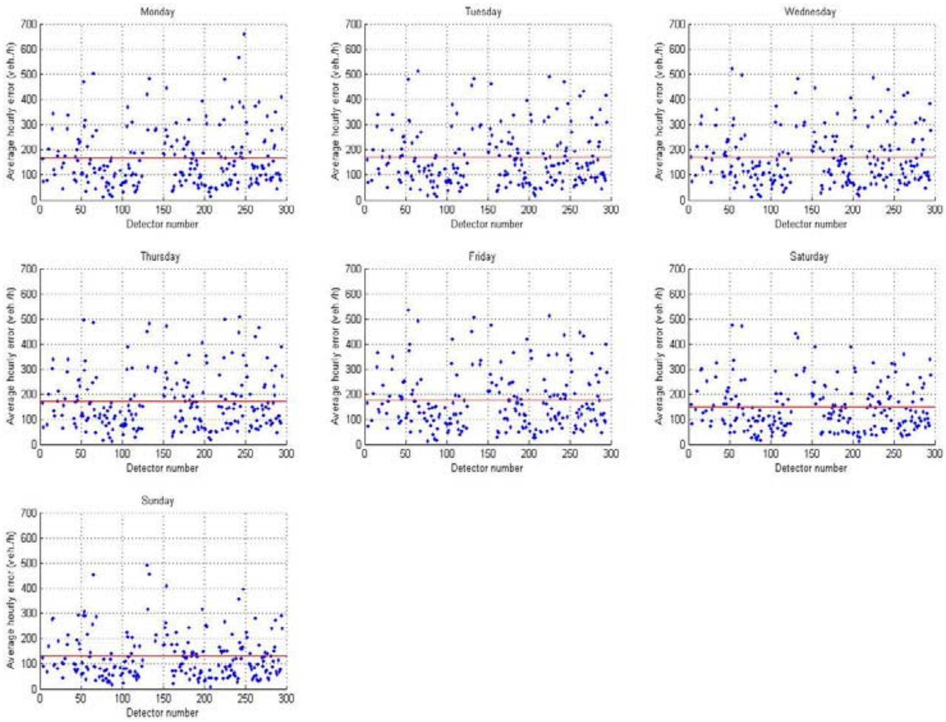


Figure 45 : Average hourly error of secondary road detectors

In Figure 45 the average hourly deviation between the simulated link flows and the counts on the secondary roads is depicted. For work days this error is around 170 vehicles per hour, in the weekend the error is around 140 vehicles per hour.

The significant reduction in the deviation between work days and weekends is probably due to the relatively smaller flows observed in the weekend, rather than a better fit of the model.

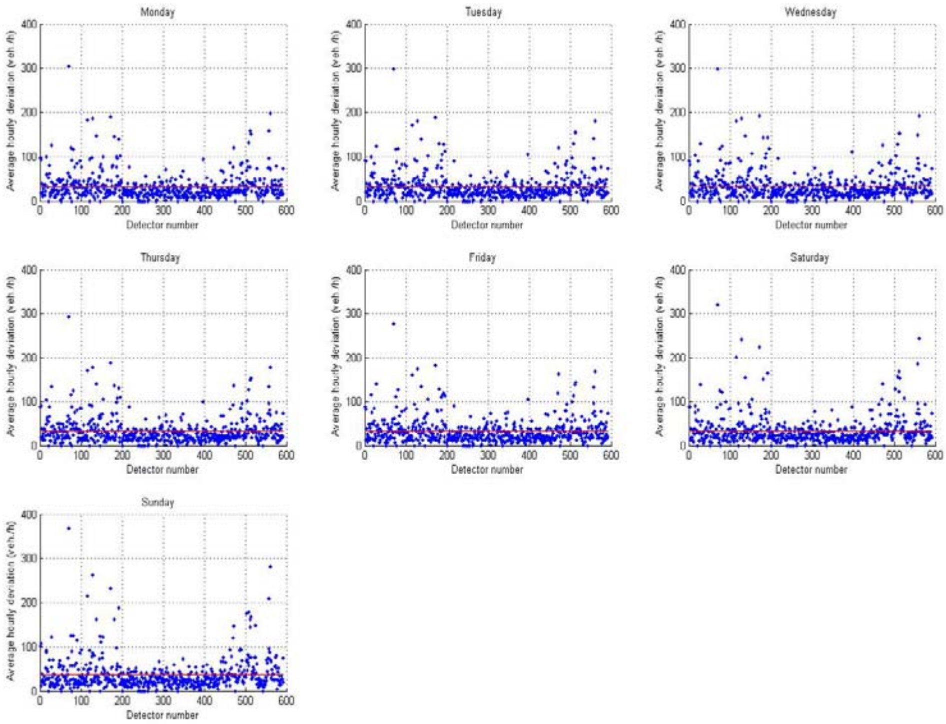


Figure 46 : Average hourly deviation from target matrix

In Figure 46 the average hourly deviation between the target matrix and the final OD matrix is depicted. Both for work days and for weekend days this deviation is around 35 vehicles per hour.

One can easily observe that the estimated OD flows are relatively closer to the values in the target matrix, if compared to the deviation between measured and estimated link flows. This is possibly due to two reasons. The first and most important is the choice of using a gradient approach, which is bound to find a local optimum in the vicinity of the initial solution. The second one is due to errors in the measured link counts.

III.2.5.2. Analysis of traffic patterns from real counts

III.2.5.3. Temporal differences

We study the average traffic patterns, inferred from all available traffic counts, for each type of day. These patterns are reflected in Figure 47. The patterns of each of the weekend days stand out, not only compared to the work days, but also compared to each other. Both weekend days don’t have a morning or afternoon peak, as expected, but experience a busy period from around noon till about 20h. Saturday is on overall a busier day than Sunday, with the only exceptions from midnight till 5h and from 19h till 22h. The traffic gets going much earlier on Saturday. Both weekend days experience minor dips in their peak period, between 12h and 14h and between 15h and 17h.

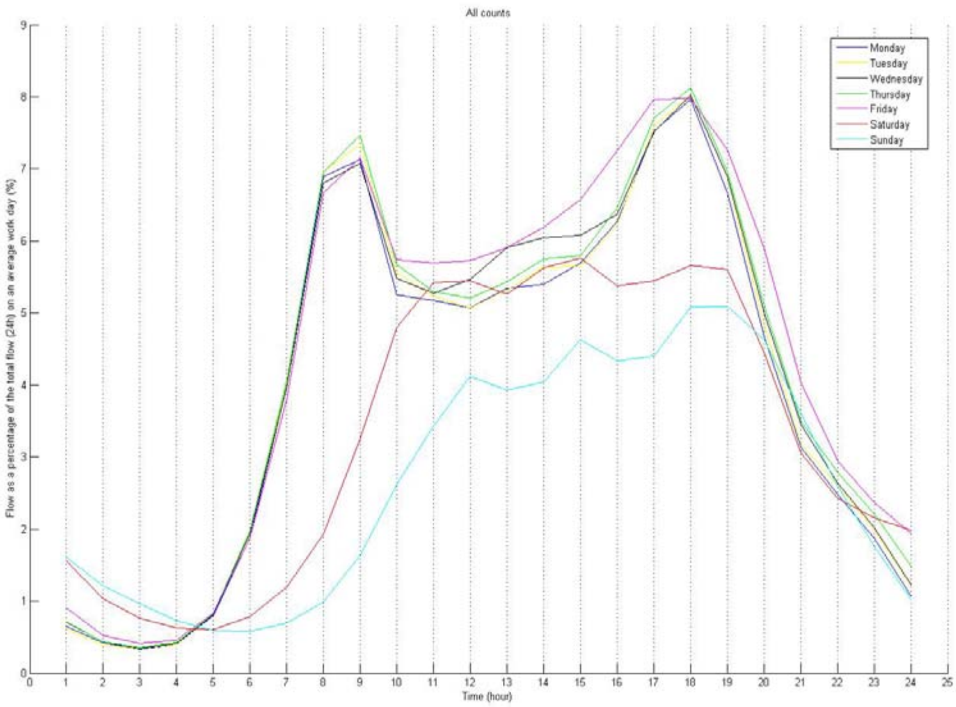


Figure 47 : Average traffic flow for each type of day

For better visualization of the differences between the work days, the difference from an average workday is represented in Figure 48 for each work day.

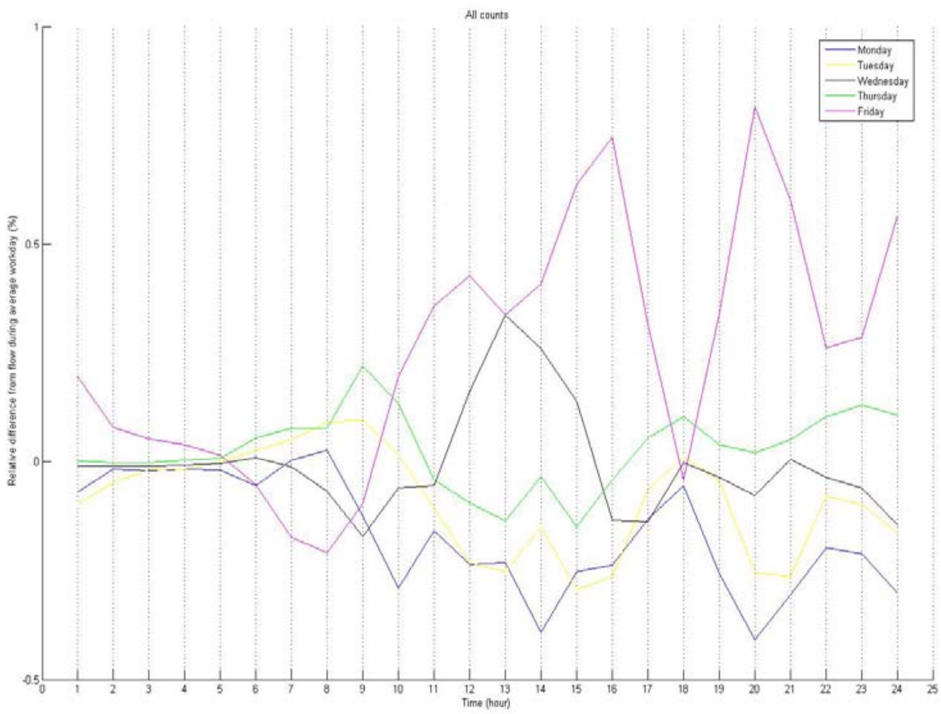


Figure 48 : Difference with average workday for each workday

From this figure it can be seen that Monday is on overall a calm day as compared to other days. Tuesday has a busy morning and evening peak, but is also a bit calmer than other days outside these peaks. On Wednesday the morning peak is less intense. There is increased traffic compared to the other days starting from 11h till 16h, with a peak between 12h and 13h. Outside these periods the traffic pattern follows the average trend. Thursday has the largest morning and evening peaks, and also during the rest of the day the traffic pattern remains at a high level. On Friday there is much more traffic on overall, but only in the off-peak periods (between 9h and 17h). The morning peak seems to be the calmest of all workdays. Also on Friday evening and night there is a serious increase of traffic as compared to other days.

From the above figures a number of conclusions can be attempted on individuals' activity patterns:

1. Work days show similar patterns. In particular Monday, Tuesday and Thursday do not seem to differ significantly, apart from a slightly lower demand on Mondays and a slightly higher demand on the Thursdays;
2. On Wednesdays around 40% more traffic than a typical working day is observed at noon while there is no significant change during morning and afternoon peaks. The peak is probably due to early closing of schools on Wednesdays;
3. On Fridays there is a systematically higher demand during the whole day, apart from the morning peak, where one can observe a reduction of around 20% with respect to a typical working day. The steep increases on the left and right shoulders of the evening peak are probably due to respectively a significant portion of workers finishing their job earlier, and users going out on Friday evening;
4. Weekend days show a completely different pattern with respect to work days, as they do not show the peaks typical of morning and evening commute. Straightforwardly, they are also rather different from each other, since shops are mostly closed on Sundays;
5. On Saturdays traffic flows seem to be rather stationary in between 10h and 19h, i.e. during the opening times of most of the shops.
6. On Sundays traffic gradually increases during the day, showing the highest peak at around 19h.

III.2.5.4. Functional differences

In this section the flows on different road types are studied. We make a distinction between highway sections and other secondary roads. We subdivide this last category in roads with an average daily flow below and above 10000 vehicles. In the highway category we find mostly long-distance trips, while on the secondary road network a larger proportion consists of local traffic. In Figure 49, the relative flow difference between an average workday and an average type of day (Monday till Friday) is plotted for the highway counts. The weekend days have been left out of the figure, because they differ too much from the workdays, and including them would outweigh the differences between the workdays.

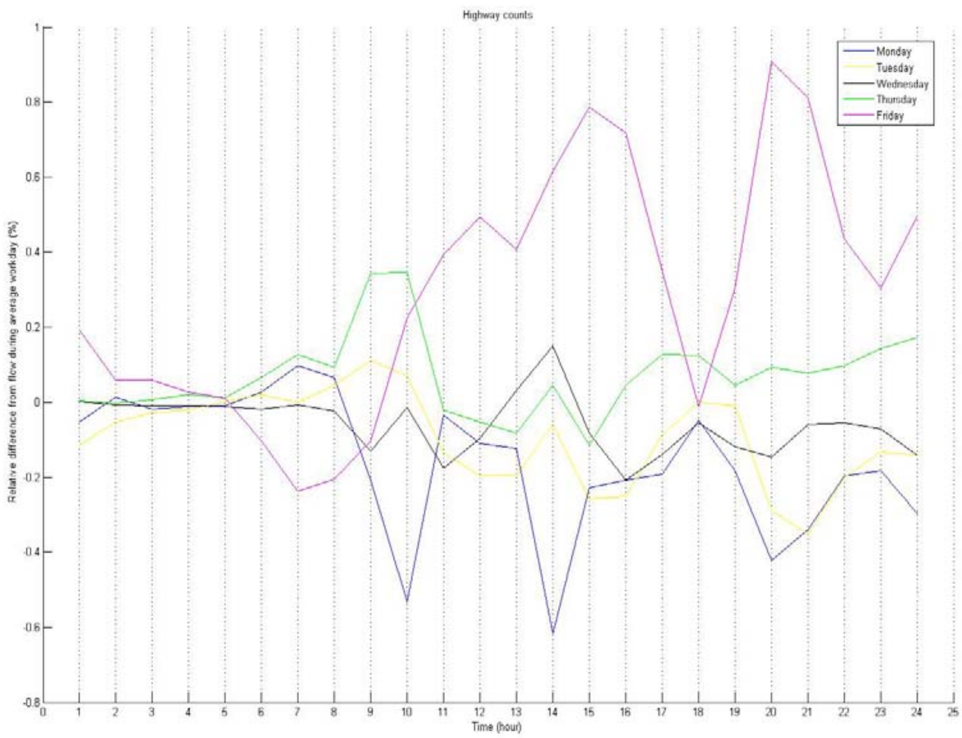


Figure 49: Relative flow difference during workdays for highways

In Figure 50 the relative flow difference between an average workday and an average type of day (Monday till Friday) is plotted for the secondary roads of type 2.

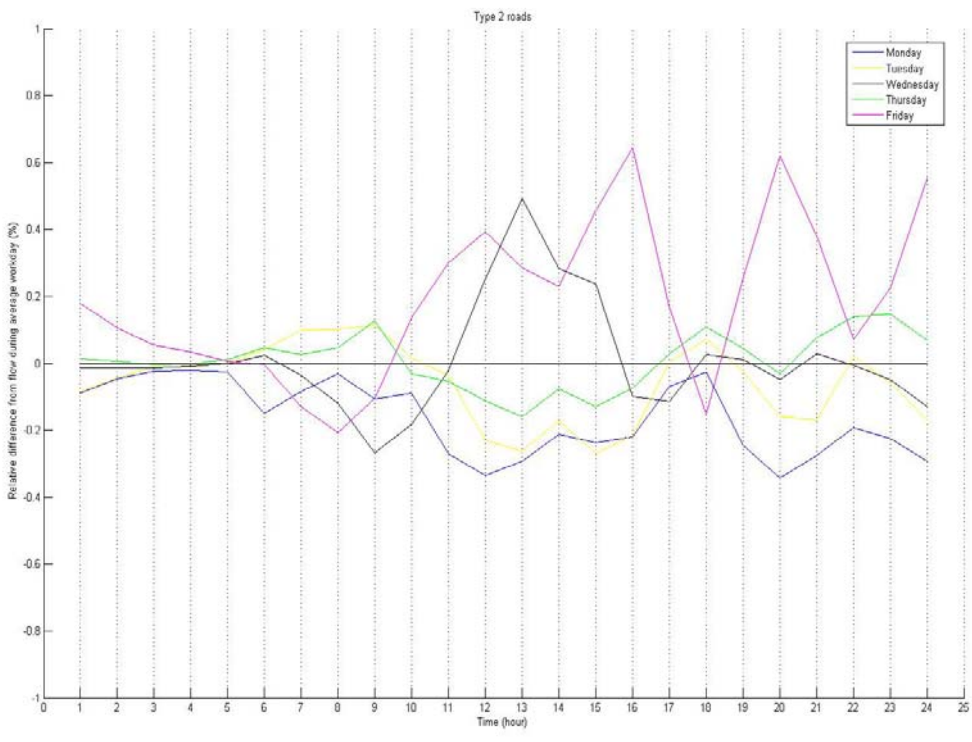


Figure 50 : Relative flow difference during workdays for secondary roads of type 2

In Figure 51 the relative flow difference between an average workday and an average type of day (Monday till Friday) is plotted for the secondary roads of type 1.

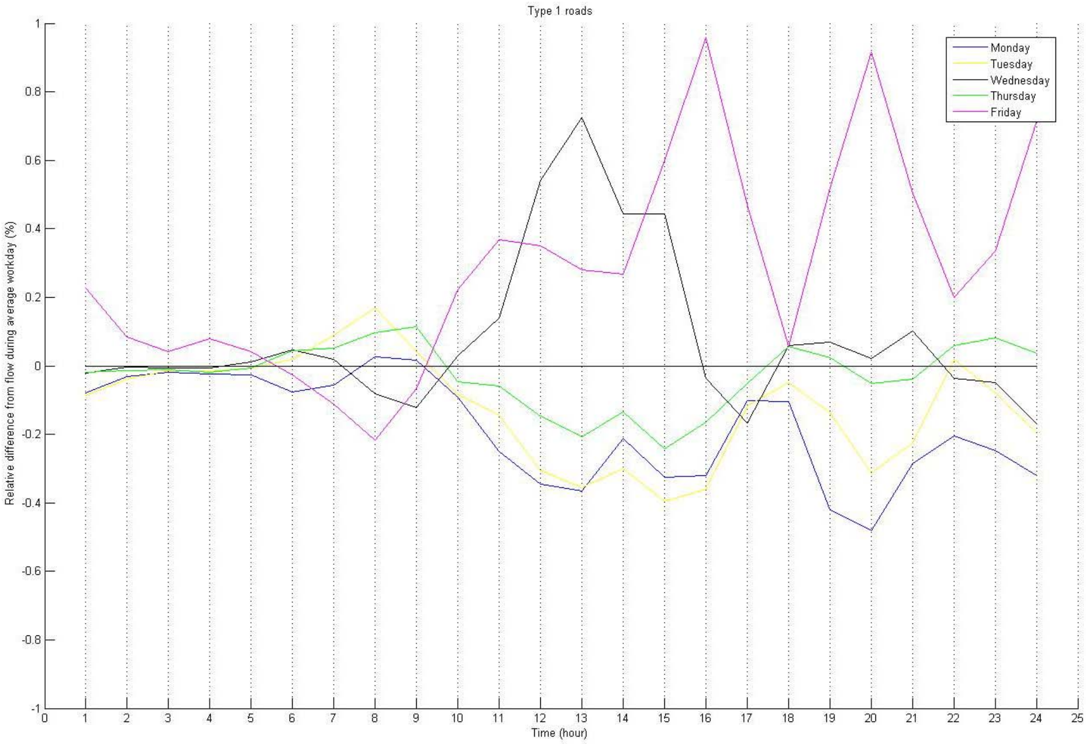


Figure 51 : Relative flow difference during workdays for secondary roads of type 1

From these figures we can observe that

- On Mondays there is a significant decrease of traffic on the highways between 9h and 10h and between 13h and 14h.
- Between 5h and 6h there is a small decrease on the secondary roads, mostly on type 2.
- The morning peak is relatively larger on the highway on Thursday compared to other days. This is not the case for the secondary roads.
- On Tuesdays between 6h and 8h there is an average amount of traffic compared to other days, while on secondary roads there seems to be a large amount.
- The major difference between the different types of road can be found again on Wednesdays: there is an increase in traffic between 11h and 13h compared to these counts, and also during the evening peak and the rest of the evening there is more traffic. It can also be seen that the differences between the different days in the morning peak are not as large as they are on the highway. The Thursday morning peak is not as large in this case.

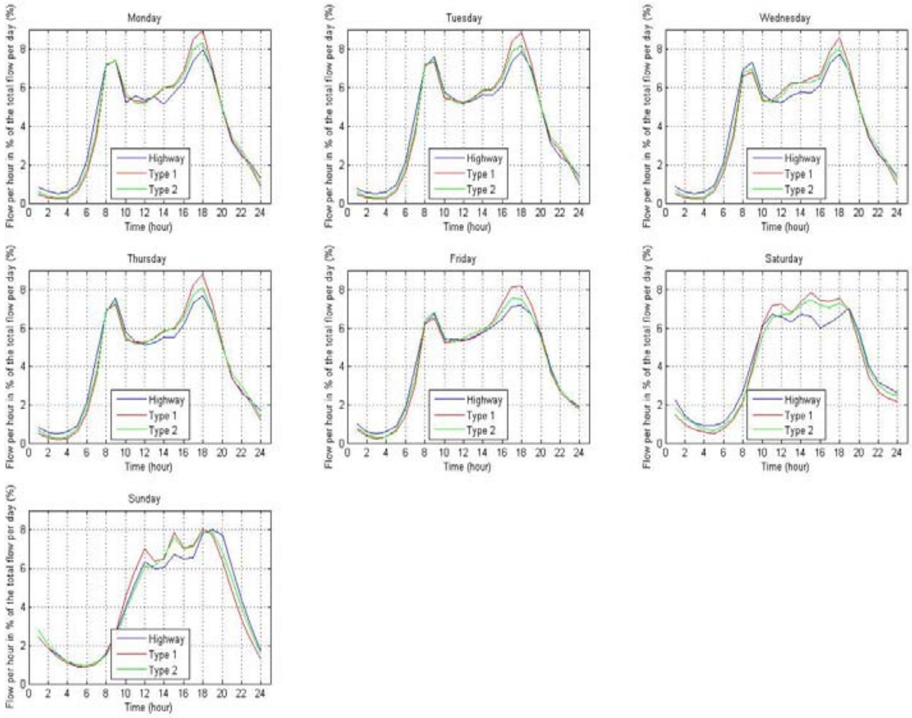


Figure 52 : Flow on highway, type 1 and type 2 secondary roads for every day

In Figure 52 the flow pattern is visualized for each type of day both on the highways as on the secondary roads. Both flow patterns are normalized to the average total flow measured on that day for those links. This figure indicates that:

- there is relatively more traffic on the highway during the night and also slightly more during the morning peak.
- From noon till the evening peak the difference is more pronounced on working days: there is relatively more traffic on the secondary road network.
- This extra share of traffic has a similar pattern as the traffic pattern on a weekend day. This can then suggest that for other trips than going to work the road users make shorter trips.

III.2.5.5. Analysis of traffic patterns from the estimated OD flows

III.2.5.6. Temporal differences

In this section we analyze the evolution of trip length distribution for different days, and make a comparison between these days. The trip length distribution during a certain moment in time is determined by first calculating the average trip length for every OD pair. Each OD pair is then categorized in a certain trip length interval, and for each trip length interval the total number of travelers is calculated from the dynamic OD matrix. This can be done for each hour and each day. Figure 53 shows these trip length distributions in percentiles for each day of the week. The 10, 25, 50 75 and 90-percentiles are used in this figure.

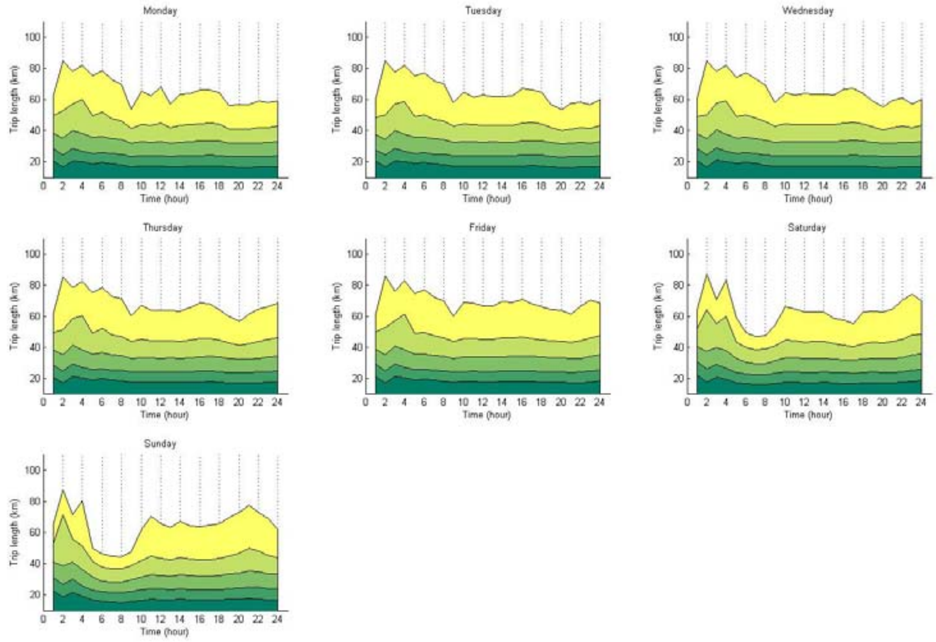


Figure 53 : Trip length distribution for all days in percentiles

The evolution of the trip length during the day does not seem to vary significantly. Also the differences between the days seem rather small (except for the difference between work days and weekend days).

III.2.5.7. Geographical differences

In this section we want to study the differences in flows in a geographical context. The network is subdivided in zones corresponding to the zoning in the OD estimation problem. This makes possible to achieve a comparison with the results from the OD estimation. The counts in a zone are then summed up, to give us an idea about the average travel pattern in a zone. Note that no highway counts are used, even if it crosses a zone, as in general only a small fraction of the highway traffic will use the zone as an origin or destination. The total flow of a zone is then normalized to the total flow during a week, so that also differences in magnitude can be studied. Next the deviation between the different days is calculated for every zone for every hour. For every hour and for every pair of days the mean and standard deviation can be calculated. These results are illustrated in Figure 54.

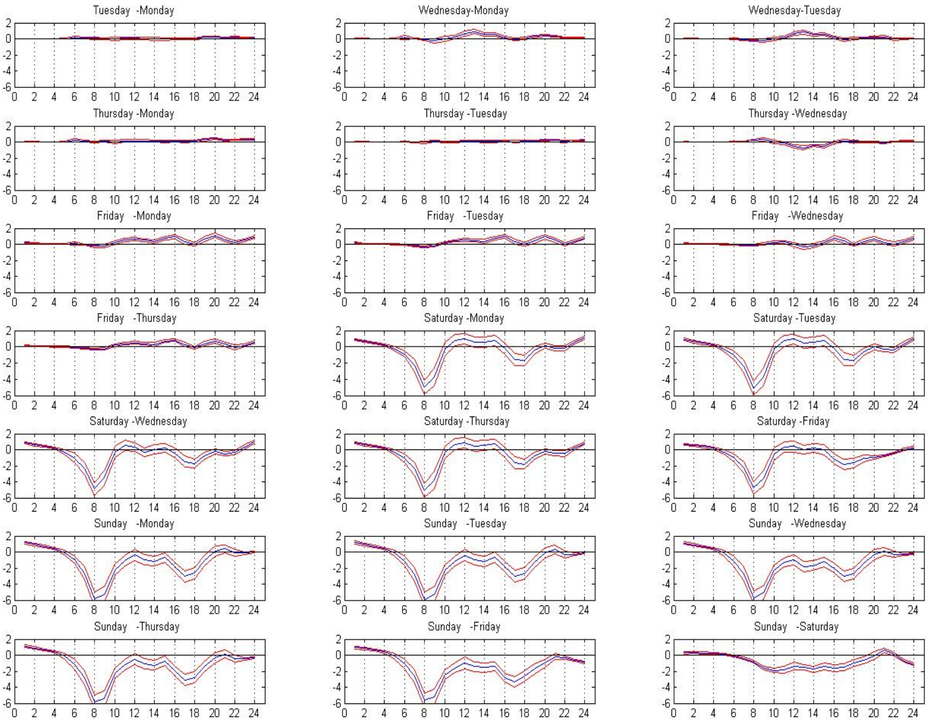


Figure 54 : Mean difference and standard deviation for pair of days

As one can easily observe, there is no significant change in the daily traffic patterns on Monday, Tuesday and Thursday, thus we can deduce again that similar activities are observed in these three days. Clear differences are however highlighted during Wednesdays and Fridays, suggesting a number of shifts in weekly activities of a significant portion of the travelers. On Wednesdays, a clear peak is observed at around 12:00 with about 100% increase of the total hourly traffic generated. On Fridays, two peaks are observed, and in general more trips are generated during the afternoon. An early peak can be associated to commuters leaving earlier the work place, while the later peak to possibly the use of the free time for other activities, e.g., shopping, leisure etc.

In Figure 55 the results are summarized in one graph, including also the weekend days. The flow differences are calculated using a typical average working day. Figure 55(a) displays the differences between each of the work days and an average workday, in Figure 55(b) the differences for all days are summarized.

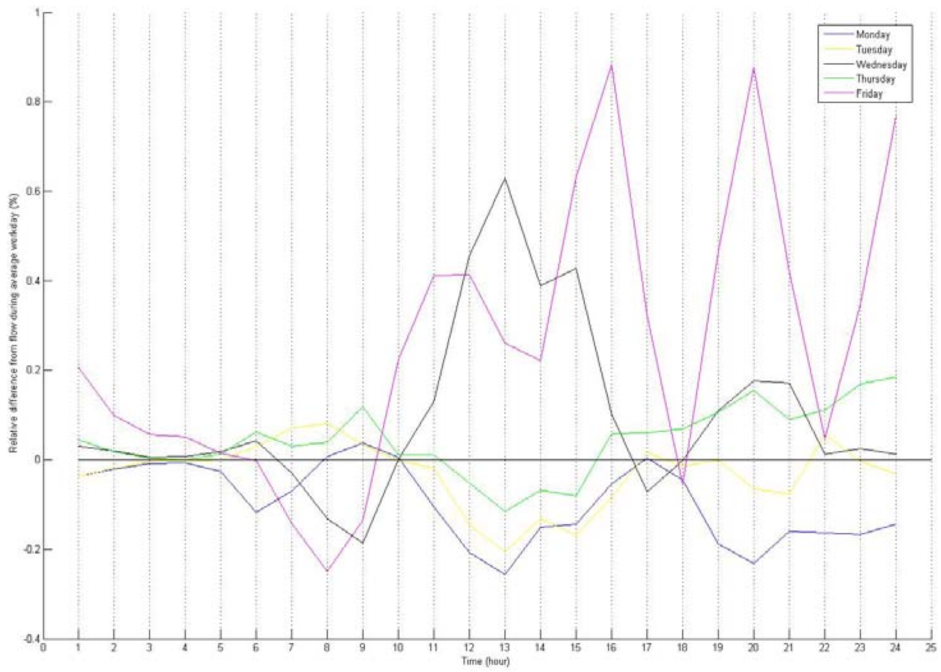


Figure 55 (a): relative flow difference on work days

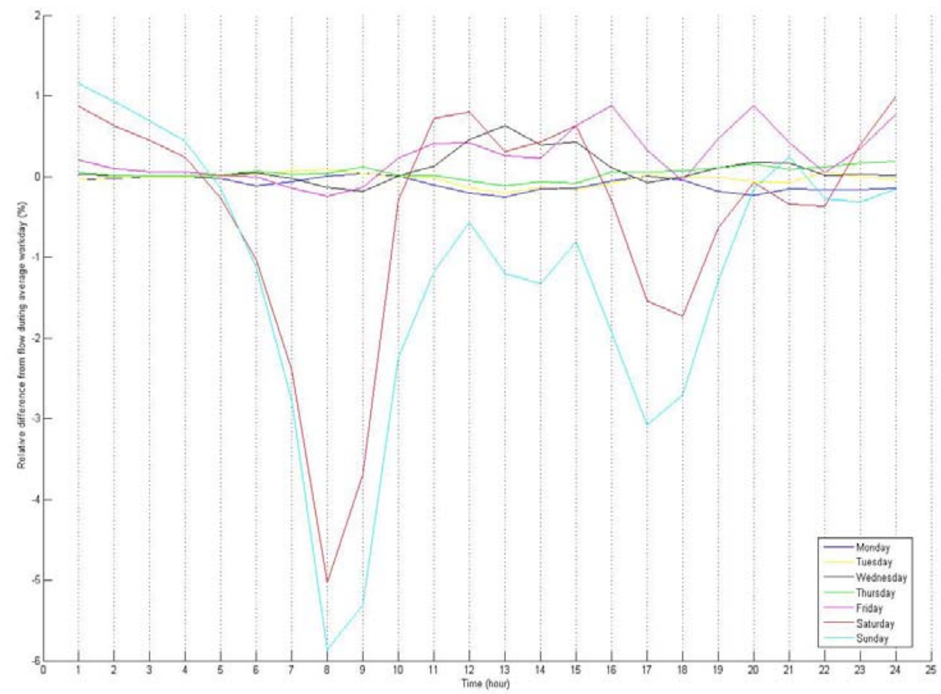


Figure 55 (b): relative flow difference on all days

Figure 55: Relative flow difference averaged over all zones

III.2.5.8. Temporal differences

In this section we further analyze the traffic patterns from the different zones. A comparison is made between the traffic patterns resulting from the counts (see section IV.2.5.2) and the traffic patterns resulting from the OD matrix. The traffic pattern within a zone is dependent on the production and attraction of that zone. We therefore calculate the average traffic pattern as the sum of the total production and the total attraction of a zone. To be able to make a comparison with the traffic pattern resulting from the counts, both patterns are normalized so that the sum over 24 hours equals one. There are 2 major differences between both approaches. The first is the fact that the pattern derived from the production and attraction of a zone does not take internal traffic into account. The second difference originates from the time lag between OD flows and link flows. This is especially important for the attraction of a zone: if for example the attraction of a zone is 1000 vehicles between 10h and 11h, then this means that 1000 vehicles departed in this time interval with the zone as destination. The time of arrival in the specified zone can however occur at a later moment in time. Therefore there is a time lag between both approaches. If we account for this time lag we get the following figure:

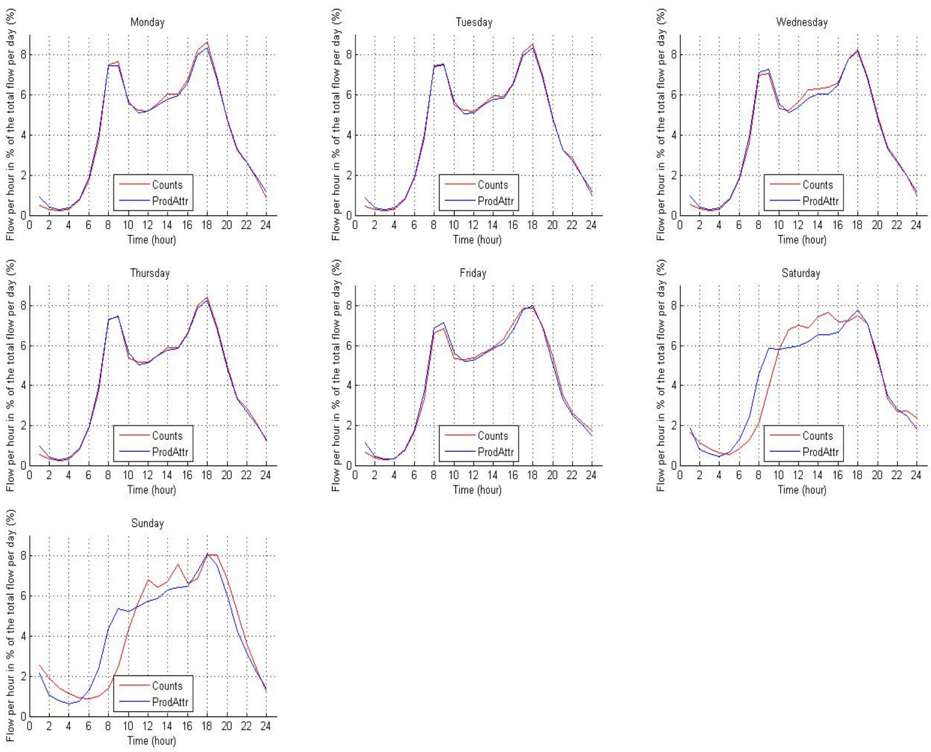


Figure 56 : Mean travel pattern resulting from counts and production/attraction

The difference for each day between the 2 curves resembles the difference between the curves from Figure 52, with the curve for the counts matching the curve for the secondary roads, and the curve for the production/attraction matching the curve for the highways. The magnitude of the difference is however not so large in Figure 56. This is due to the fact that the curve for the production and attraction not only represents highway traffic, but also traffic on secondary roads. The average pattern will thus be a sort of mean between the traffic patterns of these types of road, and the difference will not be as large. On weekend days the curves on both figures do not match as well. This is presumably due to a worse OD estimation, which itself is a result of the use of a target matrix that did not reflect the correct travel pattern on weekend days.

III.2.5.9. Analysis of traffic in and out of Ghent

The analysis has so far investigated the differences in traffic patterns identified by both traffic counts and OD flows both from a temporal and a geographical point of view. With the aim of comparing the results from the individual survey data and those from traffic counts we focus in this section on the OD flows originating or ending in the city of Ghent.

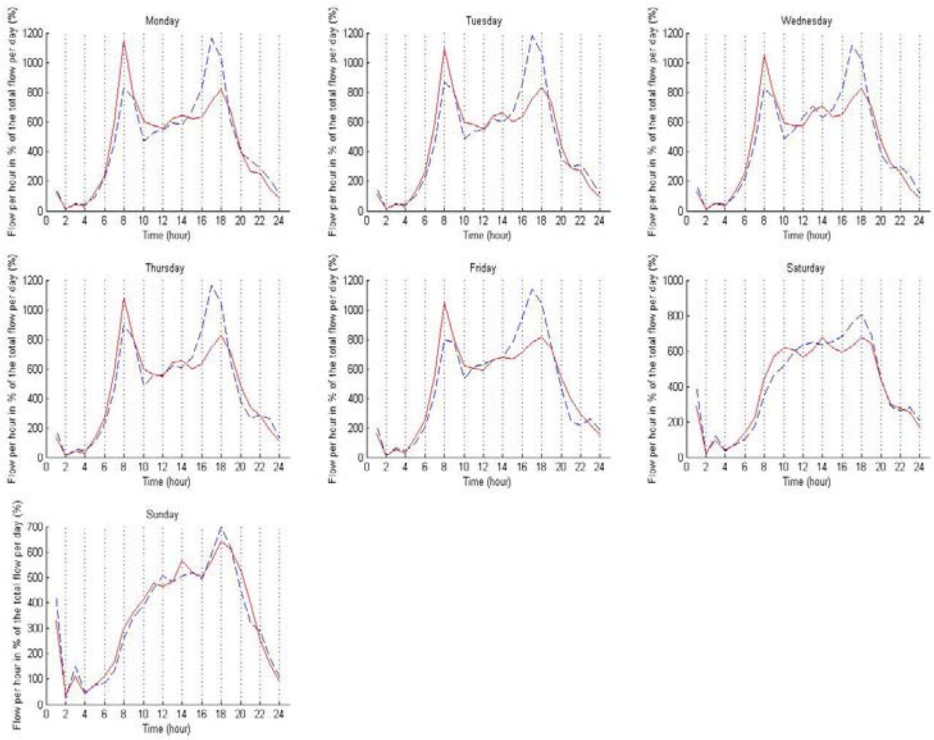


Figure 57: Production (blue, dashed line) and attraction (red, continuous) in and out from the city of Ghent

As mentioned in Chapter III.1, the city of Ghent is an important attractor for the activities of the whole East Flanders region. This is confirmed also by looking at Figure 57. It is in fact easy to observe that the morning peaks during the work days are systematically higher when the destination is any zone within the ring-road of Ghent, while it is the other way around during the afternoon peak.

III.2.6. Conclusions on the data analysis

This chapter has provided a number of insights into the relationship between traffic flows observed on the East Flanders region and the spatio-temporal distribution of trips and purposes estimated through OD flows. Here we summarize the main findings:

Direct analysis of traffic counts shows that:

- Work days show similar patterns looking at both traffic count data and the estimated OD flows. In particular Monday, Tuesday and Thursday do not seem to differ significantly, apart from a slightly lower demand on Mondays and a slightly higher demand on the Thursdays.
- On Wednesdays around 40% more traffic than a typical working day is observed at noon while there is no significant change during morning and afternoon peaks. The peak is probably due to early closing of schools on Wednesdays.

- On Fridays there is a systematically higher demand during the whole day, apart from the morning peak, where one can observe a reduction of around 20% with respect to a typical working day. The steep increases on the left and right shoulders of the evening peak are probably due to respectively a significant portion of workers finishing their job earlier, and users going out on Friday evening.
- Weekend days show a completely different pattern with respect to work days, as they do not show the peaks typical of morning and evening commute. Straightforwardly, they are also rather different from each other, since shops are mostly closed on Sundays.
- On Saturdays traffic flows seem to be rather stationary in between 10h and 19h, i.e. during the opening times of most of the shops.
- On Sundays traffic gradually increases during the day, showing the highest peak at around 19h.

Analysis of traffic on different road types shows that:

- On Mondays there is a significant decrease of traffic on the highways between 9h and 10h and between 13h and 14h.
- On Mondays, between 5h and 6h there is a small decrease on the secondary roads, mostly on type 2.
- On Thursdays, the morning peak is relatively larger on the highway compared to other days. This is not the case for the secondary roads.
- On Tuesdays between 6h and 8h there is an average amount of traffic compared to other days, while on secondary roads there seems to be a large amount.
- The major difference between the different types of road can be found again on Wednesdays: there is an increase in traffic between 11h and 13h compared to these counts, and also during the evening peak and the rest of the evening there is more traffic. It can also be seen that the differences between the different days in the morning peak are not as large as they are on the highway. The Thursday morning peak is not as large in this case.
- There is relatively more traffic on the highway during the night and also slightly more during the morning peak.
- From noon till the evening peak the difference is more pronounced on working days: there is relatively more traffic on the secondary road network.
- This extra share of traffic has a similar pattern as the traffic pattern on a weekend day. This can then suggest that for other trips than going to work the road users make shorter trips.

The above results are mostly confirmed when analyzing the results of the OD estimation procedure and by looking at the average OD flows. Two major differences have been found. The first is the fact that the pattern derived from the production and attraction of a zone does not take internal traffic into account. The second difference originates from the time lag between OD flows and link flows.

Finally analysis of OD flows originating or ending in the city of Ghent confirms that this city acts as an important attractor for the whole East Flanders region. This can be deduced by observing that the morning peaks during the work days are systematically higher when the destination is any zone within the ring-road of Ghent, while it is the other way around during the afternoon peak.

IV. COMPLEMENTARITY AND CONCLUSIONS

Understanding the relationship between traffic and activity-travel patterns is fundamental for guaranteeing a sustainable transport system. In this perspective, a weekly horizon should be envisaged, since people seem to schedule their regular activities on a weekly basis. This report has presented the results of a preliminary project aiming at analyzing this relationship through two distinguished views, i.e. the longitudinal disaggregate behavioral choices over the week and the transversal aggregate traffic measures for each day of the week. A sample of individuals has been randomly selected within a study area, and they were asked to describe their activities and movements precisely on a week. In parallel, traffic flows have been measured during the same weeks into the same study area. By comparing the two datasets, specific daily traffic patterns could be directly related to the scheduling of individuals' activities, both on day-to-day and on a within-day basis.

In this conclusion we want in a first step to go back over the main differences between both data collections, then to look at the similarities regarding weekly variations, rhythms in the week. And then this report will end by highlighting the complementarity of the two approaches and the interest of using both of them to have a complete picture of mobility, according to weekly cycles in particular.

A common ground for different methods

Two different methodologies to collect mobility data on a week were used in this project, in order to study weekly cycles as well from behavioral survey as from traffic counts. To start from same bases, we defined a common study area (the city of Ghent), and the data collections occurred during the same period, from September to December 2008. A common zoning of the territory has also been defined for the analysis.

However, we must keep in mind that some differences remain, due to the specificities of each method. Before comparing the results, we want to point out those differences:

Survey	Traffic counts
Information about all used transport modes are collected	Only motor traffic is counted
Concerning road traffic, almost only car traffic is collected (only personal trips are recorded)	Lorries are included in countings (freight trips)
Only Ghent's residents are surveyed. Students or people living in other place but coming in Ghent (for work or other reasons) are not surveyed	Transit flows are also in the countings database, people living outside Ghent and travelling into the study area are included
Short trips are also included in the databases (e.g. inside a same zone)	Only trips on axes with sensors (mainly major axes) are recorded

Table 23: main differences between the two data collections

Those differences can explain the small variations we can find in the weekly patterns, according to the used method. The reader must keep in mind these differences.

Similar spatio-temporal weekly patterns

Due to the pointed out differences between the methods, the achieved comparisons on results will mainly relate to temporal patterns during the week. Indeed, we cannot argue about modes or purposes, since the first are limited to cars for countings, and the seconds are absent from countings. In the same way, trips distances and durations will not be compared.

However, in spite of the differences between methods, we can find common diagrams of traffic distribution during the days of the week. Departure times of trips are indeed the indicator on which we can carry out parallel analyses to compare the information brought by both methods for data collection.

Some additional difficulty must be solved before being able to achieve the trips time distributions comparison: the raw data mention the departure and arrival times of trips in the case of the behavioural survey, whereas counting data mention the time of passage of the vehicles at various points on the territory, this time probably being neither the time of departure nor the hour of arrival of the trip, but well an unspecified moment during the trip. Transforming the data sets to align the two bases as well as possible was our first thought, but on second thought, it proved that this handling would likely be to deform reality rather than to improve the comparisons. Thus we deliberately chose to leave the data such as they are, knowing that the shift induced by this difference is more than probably tiny.

Once these restrictions posed, we can have a look at time distributions of trips in Ghent's area according to the measuring method.

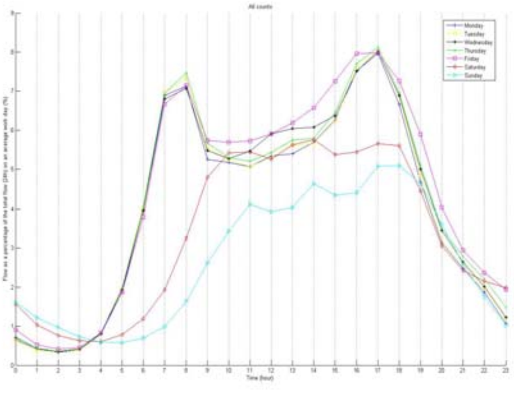


Figure 58: Average traffic flow for each type of day[4]

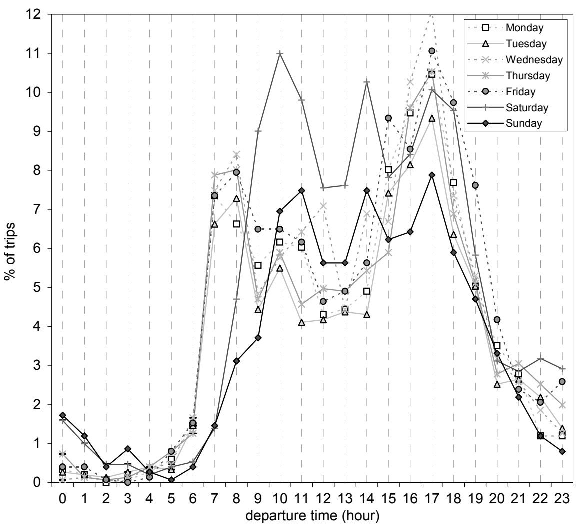


Figure 59: Trip departure time for each type of day (by car)

Figure 58 and figure 59 show these trip distributions according to both methods (for the survey data, we present here only trips made by car, in order to compare with traffic data). Data are presented as a percentage of the total amount of trips (flow) on an average workday. A first thing we can note is that we find similar peak hours in both graphs. The most important flow is between 7:00 and 8:00 in the morning, and between16:00 and 17:00 in the evening, the evening peak being wider than the morning peak.

If we can find a global common form for the curves, we can also see differences. The two major ones, which can probably be explained by the same reasons, are the "midday-trips" and the

[4] As the times are generally rounded off to the upper time unit for traffic data, and to the lower unit for survey data (a trip recorded at 7:30 is coded as 7:00 for survey data and as 8:00 for traffic data), we modified traffic data in order to use the same rounding for both databases. Figure 58 is the same graph as figure 47 before, but now with a rounding to the lower unit for countings data.

"weekend-trips". Midday trips appear in survey data, but almost not in traffic data. In a same way, trips achieved on Saturdays and Sundays seem much more numerous in survey data than in traffic data, Saturday being busier than Sunday in both data. Thus, if the survey shows that weekend or midday trips are quite real and considerable, traffic data testify to a fall of traffic on studied axes.

An analysis on destinations for the behavioural survey does not help us to understand this difference. We indeed thought that trips recorded in the survey during the weekend could be more "local" trips, using less the exit axes of the city, but it proves that weekend trips destinations are not at all statistically different from workdays trips destinations. These differences are due, from our point of view, to the fact that data collections recover non identical information. For example, professional trips are not included in the survey. However, we imagine without difficulty that those constitute a non negligible part of the traffic on workdays (let us think of the number of trucks on the roads during those days), thus inflating the figures for workdays compared to the figures for weekend. Moreover trips of non Ghent's people are not included in the survey, which can bring considerable variations compared to only Ghent's citizen trips.

This brings us to the last point of our conclusion, namely: although it is reassuring to find overall similar temporal diagrams, except for some variations, it is not so much the comparison of the two types of data acquisitions which can be an interest, that the description of the qualitative contribution which could result from a joint use of these two types data collections.

A complementarity to be taken into account

This project shows that the simultaneous use of behavioural surveys and traffic data can give a more complete vision of mobility when we want to study it on a given territory.

In the case of the study on Ghent, if we take the point of view of the behavioural survey, we see very well the contribution that the traffic data can constitute: not being informed of any trips of non Ghent's residents on the territory of the city, and having a very summarized knowledge of professional trips (we only asked the number realized per day, not the hours neither distances nor durations), it is extremely interesting to see that vehicles countings highlight a traffic, mainly during workdays, much more important than what let suppose the only interrogation of Ghent's citizens. When the studied territories are gravitational territories in terms of employment and commerce, the comings and goings on the roads of people external with the territory are an important data for urban development and land use planning.

Since the survey of people coming from all contiguous territories could be an expensive operation for a local study, and is likely well to lead to incomplete data if we do not go rather far in the attraction zones (people coming to Ghent for work or other reasons can come from rather far), road countings offer a reasonable alternative for a better knowledge of the mobility of external ones. In the same way, determining the professional traffic by the means of a survey would appear also expensive, and probably not very effective because of the heaviness of the questionnaire which should be set up. Countings are thus more interesting in this case too, to help determining the importance of the flows during the day (we would have less information in this case than a survey could bring, but maybe this information would be sufficient for the needs). Countings remain also interesting when we wish to carry out modifications on the existing road infrastructures, in order to better dimension those for example.

Conversely, survey data bring also their contribution to traffic models. Information on trips is much more precise in survey data, and that already brings considerable further information in oneself, but, as we just saw it, which can not easily be assimilated such as it is. On the other hand, the very invaluable data brought by such surveys to traffic models are the origins and destinations of all trips, those allowing determining with more precision the origin-destinations matrices, a central issue in transport modeling. In the case of our study, we tried some modelling tests with our data, but we did not get enough respondents to have significant figures.

This project thus showed that the two approaches can be used simultaneously and bring very additional information. According to the aims in view by the silent partners, one or another approach could be privileged, each one bringing its particular lighting on aspects of mobility.

Points of view on weekly cycles

We will conclude this work with a reconsideration of the importance to take into account cycles in the study of mobility. As well the survey that countings data show that people do not move in a similar way according to the various days of the week. Certain political decisions are based on data in relation to an average day, which, in the light of the figures presented in this report, does not seem to be the most relevant information. If trips remain very important on workdays, and traffic data confirm it well, we also see that weekend movements are not negligible.

The trips realized once per week, even once per month, are of a particular importance for the equipment choices of the households. For example, a person can decide to buy a car whereas she does not need any daily trip to work, just because she needs some to carry out her weekly shopping, for her semi-monthly visits to its family etc. And the choice of this equipment will have in its turn repercussions on her modal choice, because it was shown in preceding studies that the fact of owning a car strongly encourages its use (Castaigne et al., 2004).

The issue of mobility cycles of longer-term that one day thus starts to be better understood, and it is a contribution to a better knowledge of those, by various methods of analyses, that this work and this report wished to bring.

Perspectives

This report has shown as well with traffic counts as with survey data that mobility varies a lot according to the days of the week. It would be very interesting to deepen the analysis in this direction, even increasing the study duration (more than a week). A larger data collection for survey data would also allow drawing a significant origin-destination matrix, very helpful for traffic modelling.

3. POLICY SUPPORT

Any reasonable prospect, forecast or operational planning in the sector of daily mobility must relay on quantitative analysis for the demand for transportation during the day, the week or the year. It is now widely accepted that daily mobility can best be viewed as the aggregate sum of many individual demands, each demand being caused by the desire (or obligation) of an individual to participate in various activities in different physical locations. As a consequence, daily mobility demand analysis should be derived form the integrated effect of the individual's "activity chains" and their associated transportation requirements.

It is a fact that this view has gained acceptance both in research circles but also in the exercise of mobility planning in the public sector, via the growing use of disaggregated models based on activity modelling. The BMW project aims at making this trend more solid by bridging a gap between the desire for planners to consider daily mobility patterns not just on an (inexistent) "average day", but also for specific days over the week, and, at the same time, at improving the quality of the models. Indeed, in the so far more frequent models, the activity chains of individuals are collected over a sampled day, and subsequently combined in the models to attempt the description of the average day. However, few would argue with the observation that activities are often planned, at individual and household level, on more than one day, the natural common unit being the week. Modelling weekly activity patterns is therefore more natural, more accurate that modelling an "average day", and also leads to richer tools for daily mobility management.

The BMW project has been collecting data on two fronts: a household survey whose purpose was to collect weekly activity patterns and also road traffic counts, whose variations over the week also bring some light on the repartition of individual trips over various days. Specific methods were used to determine where detectors must be placed in this context and how to estimate O/D matrices from the detected flows. Ratio between traffic flows and number of trips during rush hours and off-peak hours provide a partial measure of the impact of personal mobility in the global flows. Analysis of the trip purposes, available in the survey data, has allowed the formulation of new relevant questions for further research.

These data collection exercises, beyond their use in future weekly activity based models, have also already been used by other transportation projects. Let us cite the Tijs Neutens Post-doctoral research from the University of Ghent on the analysis of spatial differences and individual disparities in space-time accessibility, and the Eurocities-DATTA research from the LET (Laboratoire d'Economie des Transports-Lyon) on the spatial, temporal, and social dimensions of activity-travel behaviour (see also validation section). Finally, the data is also part of the ongoing effort of the Namur Center for Complex Systems (naXys) in the VirtualBelgium project, where weekly activity chains will be extrapolated for the population of the whole country, with the ambition to provide aggregate travel demand on a day by day basis.

4. DISSEMINATION AND VALORISATION

I. COMMUNICATIONS

CORNELIS E., 2011. Connaître la mobilité quotidienne, cela suffit-il pour mener des politiques de transport ? , Congrès ATEC 2011, Versailles (F) 2 et 3/2/2011

MA T.-Y. Travel and activity temporal rhythm over a week: results from the BMW survey in Belgium. Eurocities-DATTA 3rd workshop, October 15 & 16 2009, Lausanne, Switzerland.

MA T.-Y., 2010. Travel and activity temporal rhythm over a week: results from the BMW survey in Belgium. 12th WCTR Conference, June 2010, Lisbon, Portugal

VITI F., 2010. Analyzing weekly activity-travel behavior from behavioral survey and traffic data. 12th WCTR Conference, June 2010, Lisbon, Portugal.

WALLE F., 2009. Comparing weekly mobility patterns from mobility survey and from traffic counting. Eurocities-DATTA 3rd workshop, October 15 & 16 2009, Lausanne, Switzerland.

II. VALORISATION

One of the innovative aspects of BMW was to collect (both from a survey and from traffic counts) information about mobility over a week. More and more, transport researchers are aware that grounding mobility policies on daily mobility could lead to some bias and has undoubtedly limitations. Moreover activity based models try also to be closer to the households behaviours and therefore to be able to take into account weekly cycles in households' activities profile.

Many researchers are therefore express interests for BMW data and results. More precisely, two close research collaborations have already been built to go further into the exploration of the collected data.

The Eurocities Datta project (http://www.eurocities-datta.eu/) project funded by the ANR (French Agency for Research) is conducted by LET (Lyon, F), LASUR (Lausanne, CH) and GRT (Namur, B) aims to study the travel time budgets and compare them amongst different countries. One of the workpackages in this project is devoted to an analysis of temporal rhythms in mobility and activity patterns. This study is essentially based on the data from BMW. Several communications and publications were or will be produced from this research.

Tijs NEUTENS (UGENT) will also use BMW data for a Post-doctoral research on person-based accessibility. Up to now, his study of this accessibility was based on OVG data, therefore only on one-day mobility agenda (cf his PhD thesis). However, people could plan their needs to access public services over a larger period than a day; a week seems more realistic. Since BMW data seems very useful for a more realistic approach of these accessibility indicators.

Finally, the work achieved during the BMW project seems absolutely interesting in current trends for transport research trying to go beyond the traditional framework of daily mobility.

III. PRESS ARTICLES

De Standaard, June 2nd 2009. Studie komt op helft minder ritten dan De Lijn.

5. PUBLICATIONS

CORNELIS E., 2011. Connaître la mobilité quotidienne, cela suffit-il pour mener des politiques de transport ? Proceedings (CD) Congrès international ATEC-ITS France 2011.

MA T.-Y., WALLE F., CORNELIS E., RAUX C., 2010. Travel and activity temporal rhythm over a week: results from the BMW survey in Belgium, in Proceedings WCTR 2010.

VITI F., TAMPERE C.M.J., FREDERIX R., CASTAIGNE M., WALLE F., CORNELIS E., 2010. Analyzing weekly activity-travel behavior from behavioral survey and traffic data. Proceedings of the 12th WCTR Conference, June 2010, Lisbon, Portugal.

6. REFERENCES

Abrahamsson T. (1998). Estimation of Origin-Destination Matrices Using Traffic Counts – A Literature Survey. Internal report IR-98-021 /May on work of the International Institute for Applied Systems Analysis, Laxenburg, Austria.

Antoniou, C. (1997). Demand Simulation for Dynamic Traffic Assignment. MSc Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Armoogum J., Castaigne M., Hubert J.-P., Madre J.-L. (2005). *Immobilité et mobilité observées à travers les enquêtes ménages de transport ou d'emploi du temps*, Actes des Journées de Méthodologie Statistique, INSEE, 12 p.

Ashok, K. (1996.) Estimation and Prediction of Time-Dependent Origin-Destination Flows. Ph.D. dissertation, Dept. Civil and Environmental Eng., Massachusetts Institute of Technology.

Axhausen KW., Schönfelder S., Wolf ., Oliveira M., Samaga U. (2003), "80 weeks of GPS-traces: Approaches to enriching the trip information", Submitted to the 83rd Transportation Research Board meeting.

Axhausen, K.W., Zimmermann A., Schönfelder S., Rindsfüser G., Haupt T. (2002). Observing the rhythms of daily life: A six-week travel diary, Transportation, 29 (2) 95-124.

Balakrishna R. (2006) "Off-Line Calibration of Dynamic Traffic Assignment Models", PhD thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Balakrishna, R. and H.N. Koutsopoulos. (2008). Incorporating Within-Day Transitions in the Simultaneous Off-line Estimation of Dynamic Origin-Destination Flows without Assignment Matrices. 87nd annual meeting of the Transportation Research Board, Washington, USA.

Barcelo, J. (2002). Microscopic traffic simulation: A tool for the analysis and assessment of its systems. Proceedings of the half year meeting of the Highway Capacity Committee, Lake Tahoe.

Bell M.G.H. (1983). The estimation of an origin destination matrix from traffic counts. Transportation Science Vol. 17, No. 2, pp. 198-217.

Bell M.G.H. (1991). The estimation of origin-destination matrices by constrained generalized least squares. Transportation Research Part B, Vol. 25, No. 1, pp. 13-22.

Bell M.G.H., Inaudi D., Lange J., Maher M. (1991). Techniques for the dynamic estimation of O–D matrices in traffic network. Advanced Telematics in Road Transport, Proceedings of the DRIVE Conference, Brussels, Belgium, February 4–6, 1991, Vol. 2, Elsevier, pp 1040–1056.

Ben-Akiva M., Bierlaire M., Koutsopoulos H., Mishalani R. (1998). DynaMIT: a simulation-based system for traffic prediction. Paper presented at the DACCORD Short Term Forecasting Workshop February, 1998 Delft, The Netherlands

Bianco L., Confessore G., Reverberi P. (2001). A network based model for traffic sensor location with implications on OD matrix estimates. Transportation Science, Vol. 35, pp. 50-60.

Bielaire M. (2002). The total demand scale: a new measure of quality for static and dynamic origin–destination trip tables. Transportation Research Part B, Vol. 36, pp. 837-850.

Bierlaire M., Crittin F. (2003). An efficient algorithm for real-time estimation and prediction of dynamic OD tables. Operations Research 52 (1), 116–127.

Bierlaire M., Toint P.L. (1994), Meuse: An origin-destination matrix estimator that exploits structure. Transportation Research Part B, Vol. 29B No. 1, pp. 47-60.

Bliemer, M.C.J., Versteegt, H.H., and Castenmiller, R.J. (2004). INDY: A New Analytical Multiclass Dynamic Traffic Assignment Model. Proceedings of the TRISTAN V conference, Guadeloupe.

Cascetta E. (1984). Estimation of trip matrices from traffic counts and survey data: a generalized least squares approach. Transportation Research Part B, Vol. 18, pp. 289-299.

Cascetta E., Cantarella, G.E. (1990). A day-to-day and within-day dynamic stochastic assignment model. Transportation Research Part A Vol. 25, 277–291.

Cascetta E., Improta A.A. (2002). Estimation of travel demand using traffic countsand other data sources. Chapter 5 of *Transportation and Network Analysis: Current Trends : Miscellanea in Honor of Michael Florian*. Florian M., Gendreau M., Marcotte P. (Eds.)

Cascetta E., Inaudi D., Marquis G. (1993). Dynamic estimators of origin–destination matrices using traffic counts. Transportation Science 27, pp 363–373.

Cascetta E., Postorino M. (2001). Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks, Transportation Science Vol. 35, pp 134–147.

Castaigne M., Cornélis E., Hubert J.-P. (2004). Les contraintes du ménage dans le choix modal des déplacements, FUNDP, Dpt de Mathématique, GRT, Rapport 2004/05, Namur.

Chen Y. (1994). Bilevel programming problems: analysis, algorithms and applications. PhD thesis, Report CRT-984, Centre de recherche sur les transports (CRT), University of Montreal, Montreal, Quebec, Canada.

Chlond, B., Kuhnimhof, T. (2003). Rules of Non-Response and Selectivity – analysing the drop-out in the multistage recruitment process for the German Mobility Panel, Paper presented at the European Transport Conference, Strasbourg, France October 2003.

Chung I.-H. (2001). An optimum sampling framework for estimating trip matrices from day-to-day traffic counts. PhD thesis, University of Leeds.

Cirillo, C., K.W. Axhausen (2006). Evidence on the distribution of values of travel time savings from a six-week diary, Transportation Research A, 40A (5) 444-457.

Cirillo C., Toint Ph. (2001). An activity-based approach to the Belgian national travel survey, Rapport GRT 07/01, Groupe de Recherche sur les Transports, Département de Mathématique, Facultés Universitaire ND de la Paix, Namur, Belgique.

Clegg J., Smith, M., Xiang, Y., Yarrow R. (2001). Bilevel programming applied to optimising urban transportation. Transportation Research Part B Vol. 25, pp 41–70.

Cortese V., Hamzaoui M., Valkeneers A., Veinstein M. and Desmarez P., (2006). Working paper scolarisation et niveau d'instruction, SPP Politique Scientifique Fédérale.

Cremer M., Keller H. (1987). A new class of dynamic methods for the identification of origin–destination flows. Transportation Research Part B Vol. 21, pp 117–132.

Daganzo C.F. (1994). The cell transmission model: a dynamic representation of highway traffic consistent with hydrodynamic theory, *Transportation Research B* 28, 269-287.

Daganzo C.F. (2006). Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B: Methodological, Vol. 41, No. 3, pp. 49-62

Davis G.A. (1993). A statistical theory for estimation of origin–destination parameters from time-series of traffic counts. Transportation and Traffic Theory, Proceedings of the 12th International Symposium on Transportation and Traffic Theory, Berkeley, California, 21–23 July, 1993, C.F. Daganzo (ed.), Elsevier, pp 441–463.

Davis G.A., Nihan N.L., (1991). Stochastic process approach to the estimation of origin–destination parameters from time series of traffic counts. Transportation Research Record 1328, pp 36–42.

Drissi-Katouni O., Lundgren J. (1992). Bilevel origin-destination matrix estimation using a descent approach. Technical report LiTH-MAT-R-92-49, Department of Mathematics, Linkoping, Institute of Technology, Linkoping, Sweden.

Ehlert A., Bell M.G.H., Grosso, S. (2006). The optimization of traffic count locations in road networks. Transportation Research Part B, Vol. 40, pp. 460-479.

Fei X., Eisenman S.M., Mahmassani H.S. (2007). Sensor Coverage and Location for Real-Time Traffic Prediction in Large-Scale Networks. Transportation Research Record Vol. 2039

Florian M, Chen Y. (1993). A coordinate descent method for the bilevel OD matrix adjustment problem. Work presented at the IFORS Conference in Lisbon, Portugal.

Fisk C.S. (1988). On combining maximum entropy trip matrix estimation with user optimal assignment. Transportation Research Part B Vol. 22 No. 1, pp. 69-73.

FOD Economie, K.M.O. Middenstand en Energie. Algemene Socio-Economische Enquete 2001. http://statbel.fgov.be/census/home_nl.asp

Gentile G., Meschini L., Papola N. (2007) Spillback congestion in dynamic traffic assignment: a macroscopic flow model with time-varying bottlenecks, *Transportation Research B* 41, 1114-1138.

Gentili M. (2002). New Models and Algorithms for the Location of Sensors on Traffic Networks. PhD thesis. Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma "La Sapienza".

Hajnal, I., Miermans, W. (1996). Onderzoek Verplaatsingsgedrag Vlaanderen. Analyseopdracht. Eindverslag. Diepenbeek: Provinciale Hogeschool Limburg, Departement Architectuur.

Hazelton, M. (2000). Estimation of origin–destination matrices from link flows on uncongested networks. Transportation Research 34B, 549–566.

Hazelton, M. (2003). Some comments on origin–destination matrix estimation. Transportation Research 37A, 811–822.

Henderson, J., and Fu, L. (2004). Applications of Genetic Algorithms in Transportation Engineering. Presented at the 83rd annual meeting of the Transportation Research Board.

Janson, B.-N., Southworth, F. (1992) Estimating departure times from traffic counts using dynamic assignment. Transportation Research 26B, 3-16.

Jayakrishnan, R., Mahmassani H.S., and Hu T.-Y. (1994). An Evaluation Tool for Advanced Traffic Information and Management Systems in Urban Networks. Transportation Research C, Vol. 2C, No. 3, 1994, pp. 129-147.

Kang, Y. (1999). Estimation and Prediction of Dynamic Origin-Destination (O-D) Demand and System Consistency Control for Real-Time Dynamic Traffic Assignment Operation. Ph.D. Dissertation, The University of Texas at Austin.

Kattan, L. and B. Abdulhai. (2006). Non-Iterative Approach to Dynamic Origin/Destination Estimation using Parallel Evolutionary Algorithms. Presented at the 85th annual meeting of the Transportation Research Board.

Kim, H., Baek, S., Lim, Y. (2001). Origin-Destination Matrices Estimated with a Genetic Algorithm from Link Traffic Counts. Transp. Res. Record No.1771, 156–163.

Lam W.H.K., Lo H.P. (1990). Accuracy of O-D estimates from traffic counting stations. Traffic Engineering and Control, Vol. 7, No. 1, pp. 105-114.

LeBlanc L.J., Fahrangian K. (1982). Selection of a trip table which reproduces observed link flows. Transportation Research Part B, Vol. 16, pp. 83-88.

Lindveld K. (2003). Dynamic O-D Matrix estimation : a behavioral approach. PhD thesis, Delft University of Technology.

Liu S., Fricker, J.D. (1996). Estimation of a trip table and the θ parameter in a stochastic network. Transportation Research Part A Vol. 30, pp. 287–305.

Lo H.-P., Chan C.-P. (2003). Simultaneous estimation of an origin–destination matrix and link choice proportions using traffic counts. Transportation Research Part A Vol. 37, pp. 771–788

Maher M.J. (1983). Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. Transportation Research Part B, Vol. 17, No. 6, pp. 435-447.

Maher M.J., Zhang, X., van Vliet D. (2001). A bi-level programming approach for trip matrix estimation and traffic control problems with stochastic user equilibrium link flows. Transportation Research Part B, Vol. 35, pp 23–40.

Mahmassani, H.S., Abdelghany, A.F., Huynh, N., Zhou, X., Chiu, Y-C. & Abdelghany, K.F. (2001). DYNASMART-P User's Guide. Technical Report STO67-85-PIII, Centre for Transportation Research, University of Texas at Austin.

Marzano V., Papola A., Simonelli F. (2007). Investigating the reliability of the o-d matrix correction procedure using traffic counts. Presented at TRISTAN VI, Phuket, Thailand.

Messmer, A. and Papageorgiou, M. (1990). METANET: A macroscopic simulation program for motorway networks. Traffic Engineering and Control Vol. 31, pages 466-470.

Nguyen S. (1977). Estimating an OD Matrix from network data: a network equilibrium approach. Publication No. 60, Centre de Recherche sur les Transports, Universite de Montreal, Quebec.

Nuits E., Zwerts E. (2001). Onderzoek verplaatsingsgedrag, stadsgewest Hasselt – Genk (deel 1 : methodologische analyse), Diepenbeek.

Pas E., Kitamura R. and Fujii S. (1997). Time-use data, analysis and modeling: toward the next generation of transportation planning methodologies, Transport Policy, Vol. 4, No. 4, pp. 225-235.

Peeta, S. and A. K. Ziliaskopoulos. (2001). Foundations of dynamic traffic assignment: The past, the present and the future. Networks and Spatial Economics, 1:233–265. Kluwer Academic Publishers, The Netherlands.

Peterson A. (2007). The Origin–Destination Matrix Estimation Problem - Analysis and Computations. MSc Thesis, Department of Science and Technology Linkoping University, Norrkoping, Sweden

PTV (2005). PTV Vision software suite (VISUM/VISEM/VISEV A/VISSIM). PTV AG-Traffic Mobility Logistics.

Robillard P. (1975). Estimating the O-D Matrix from observed link volumes. Transportation Research Vol. 9, pp. 123-128.

Schlich, R., K.W. Axhausen (2003). Habitual travel behaviour: Evidence from a six-week diary, Transportation, 30 (1) 13-36.

Smith M. J. (1979). The existence, uniqueness and stability of traffic equilibria. Transportation Research Part B, Vol. 13, pp. 295-304.

Spiess H. (1987). A maximum-likelihood model for estimating origin-destination matrices, Transportation Research Part B, Vol. 21, pp. 395-412.

Spiess H. (1990). A descent based approach for the OD matrix adjustment problem. Publication no. 693 at Centre de recherche sur les transports, Universite de Montreal, Montreal, Canada.

Tampere, C.M.J., Stada J., Immers L.H., Peetermans E., and Organe K. (2007). Methodology for Identifying Vulnerable Sections in a National Road Network, in Transportation Research Record No. 2012, pp. 1-10.

Thierens, D. (1999). Scalability Problems of Simple Genetic Algorithms, Journal of Evolutionary Computation, MIT Press, 7(4), pp. 331-352.

Treiber M., Helbing D. (2002). Reconstructing the spatio-temporal traffic dynamics from stationary detector data, Cooper@tive Tr@nsport@tion Dyn@mics **1** 3.1-3.24

Wardrop G.J. (1952). Some theoretical aspects of road traffic research. Proceedings of the institution of Civil Engineers II(1), pp. 325-378.

Weijermars, W. (2007), “Analysis of urban traffic patterns using clustering”, TRAIL Thesis Series T2007/3, The Netherlands TRAIL Research School

Willumsen L.G. (1984). Estimating time-dependent trip matrices from traffic counts. Proceedings of the 9th International Symposium on Transportation and Traffic Theory. J. Volmuller and R. Hamerslag (eds.), VNU Science Press, Utrecht, The Netherlands, pp 397–411.

Van der Zijpp N.J. (1996). Dynamic origin destination matrix estimation on motorway networks. PhD Thesis, Delft University of Technology.

Van der Zijpp, N.J., Lindveld, C.D.R. (2001) Estimation of origin-destination demand for dynamic assignment with simultaneous route and departure time choice. Transportation Research Record 1771: Transportation Network Modeling, 75-82.

Van Zuylen H.J., Willumsen L.G. (1980). The most likely trip matrix estimated from traffic counts. Transportation Research Part B, Vol. 14, pp. 281-293.

Viti F., Tampere C.M.J., Verbeke W. (2009). Sensor locations for reliable travel time prediction and dynamic management of traffic networks. Transportation Research Record No. 2049, pp. 103-110.

Wardrop J. (1952). Some Theoretical Aspects of Road Traffic Research Communication Networks," in Proc. Inst. Civ. Eng., vol. 2, no. 1, 1952, pp. 325–378.

Wu Y., Davis G., Levinson D. (2004). Improving the Estimation of Travel Demand for Traffic Simulation: Part II. Report CTS-04-11, University of Minnesota, www.cts.umn.edu/pdf/CTS-04-11Part2.pdf

Yan, T., Jans, M., & Curtin, R. (2006). Changes in Nonresponse to Income Questions. American Association for Public Opinion Research Annual Conference, Montreal. Published in the American Statistical Association Proceedings.

Yang H., Meng, Q., Bell, M.G.H. (2001). Simultaneous estimation of the origin–destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium. Transportation Science, Vol. 35, pp. 107–123.

Yang H., Sasaki T., Iida Y., Asakura Y. (1992). Estimation of origin-destination matrices from link traffic counts on congested networks, Transportation Research Part B, Vol. 26, pp. 417-434.

Yang H., Zhou J. (1998). Optimal traffic counting locations for origin-destination matrix estimation. Transportation Research Part B, Vol. 32, No. 2, pp. 109-126.

Yperman I. (2007) The Link Transmission Model for dynamic network loading, PhD Thesis, Katholieke Universiteit Leuven, Belgium.

Zhang, H. M. and Y. Nie. (2005). Modeling network flow with and without link interactions: properties and implications. 84rd annual meeting of the Transportation Research Board, Washington, USA.

Zhou, X. and H. S. Mahmassani (2004). A Structural State-Space Model for Real-Time Origin-Destination Estimation and Prediction in a Day-to-Day Updating Framework. Proceedings of the 83rd TRB Annual Meeting, January 2004, Washington D.C.

Zhou X., Qin X., Mahmassani H.S. (2003). Dynamic Origin-Destination Demand Estimation Using Multi-Day Link Traffic Counts for Planning Applications. In Proceedings of the 82nd TRB Annual Meeting, January 2003, Washington D.C.

Zumkeller D., Chlond B., Lipps O. (1997). The German Mobility Panel: Options, Limitations and the complementary Use of secondary Data, Proceedings of the International Conference on Transport Survey Quality and Innovation: Transport Surveys: Raising the Standard, Grainau, Germany, Mai 1997.

Zumkeller, D., Chlond, B., Kuhnimhof, T (2002). Panelauswertung 2002– Schlussbericht zu FE 70.0702 / 2002 (Evaluation of the Mobility Panel 2002, Final Report, in German), Institute for Transport Studies, University of Karlsruhe November 2003.