The Origin–Destination Matrix Estimation Problem — Analysis and Computations

Anders Peterson

Linköping University
INSTITUTE OF TECHNOLOGY

Department of Science and Technology

Linköpings universitet, SE-601 74 Norrköping, Sweden

Norrköping 2007

The Origin–Destination Matrix Estimation Problem

— Analysis and Computations

Anders Peterson   andpe@itn.liu.se   http://www.itn.liu.se

The cover illustrates the equipment for collecting link flow observations. Here, the traffic entering and leaving the car park at the commercial Ekholmen centrum in Linköping, Sweden, is being counted. Photo: April 8, 2007, by Inger Munkhammar.

Acknowledgements

Most important for the performance of a post-graduate study are the supervisors, and I sincerely acknowledge Jan Lundgren and Torbjörn Larsson for advice and support over the years. Jan has introduced me to traffic modeling and especially the Origin–Destination matrix estimation problem. He has been a good advisor by all occasions and I am very grateful for his help in defining and structuring the research work.

Torbjörn has guided me through the process of scientific writing and re-writing. He has patiently offered a lot of time for the details and the nuances in the thesis.

I thank all colleagues at the division of Communications and Transport systems, for discussions, inspiration and companionship. Henrik Andersson kindly read and commented my work. Clas Rydergren has been a good fellow in research, teaching — and train commuting. Carl Henrik Häll, with whom I used to share my office room, has always had time for a word or two whenever the research process has troubled me.

The benevolent people at the division of Optimization in Linköping, where my post-graduate studies once started, are acknowledged for hosting many research meetings.

Finally, I would like to express my greatest gratitude to my family and my friends. I have appreciated their encouragement and understanding, especially in situations where I did not manage to fully explain the problems troubling me.

Norrköping, May 2007
Anders Peterson

Abstract

For most kind of analyses in the field of traffic planning, there is a need for origin–destination (OD) matrices, which specify the travel demands between the origin and destination nodes in the network. This thesis concerns the *OD-matrix estimation problem*, that is, the calculation of OD-matrices using observed link flows. Both time-independent and time-dependent models are considered, and we also study the placement of link flow detectors.

Many methods have been suggested for OD-matrix estimation in time-independent models, which describe an average traffic situation. We assume a user equilibrium to hold for the link flows in the network and recognize a bilevel structure of the estimation problem. A descent heuristic is proposed, in which special attention is given to the issue of calculating the change of a link flow with respect to a change of the travel demand in a certain pair of origin and destination nodes.

When a time-dimension is considered, the estimation problem becomes more complex. Besides the problem of distributing the travel demand onto routes, the flow propagation in time and space must also be handled. The time-dependent OD-matrix estimation problem is the subject for two studies. The first is a case study, where the conventional estimation technique is improved through introducing pre-adjustment schemes, which exploit the structure of the information contained in the OD-matrix and the link flow observations. In the second study, an algorithm for time-independent estimation is extended to the time-dependent case and tested for a network from Stockholm, Sweden.

Finally, we study the underlying problem of finding those links where traffic flow observations are to be performed, in order to ensure the best possible quality of the estimated OD-matrix. There are different ways of quantifying a common goal to cover as much traffic as possible, and we create an experimental framework in which they can be evaluated. Presupposing that consistent flow observations from all the links in the network yields the best estimate of the OD-matrix, the lack of observations from some links results in a relaxation of the estimation problem, and a poorer estimate. We formulate the problem to place link flow detectors as to achieve the least relaxation with a limited number of detectors.

Populärvetenskaplig sammanfattning

Matematiska modeller används av väghållare runt om i världen för att beskriva, planera och styra trafiken. Det kan handla om att beräkna restider, miljöpåverkan och tillgänglighet, eller att prognostisera effekterna av nybyggnationer, ändrad framkomlighet och vägavgifter. Ofta räcker det med att beskriva trafiken med genomsnittsförhållanden, men för att exempelvis producera realtidsinformation, kontrollera kövarningssystem och skapa handlingsplaner för olika olycksscenarier behövs tidsberoende modeller.

Noggranna uppgifter om antal resande från varje startpunkt (eng. origin) till varje målpunkt (eng. destination) i ett trafiknät är viktiga indata till de flesta av dessa modeller, och de brukar organiseras i s.k. OD-matriser (eng. Origin–Destination matrices). Att utarbeta OD-matriser utifrån resvaneundersökningar och olika statistiska sammanställningar är dyrbart och görs sällan. Observerade trafikflöden, som samlas in genom slangmätningar, vägkameror, tullportaler, etc., utgör däremot billig och lättillgänglig information om den rådande trafiksituationen. Denna avhandling behandlar problemet att skatta OD-matriser utifrån trafikflödesobservationer.

Den skattade OD-matrisen ska tillsammans med den ruttvalsprincip som antagits för trafiken återskapa de observerade trafikflödena så bra som möjligt. Det finns olika principer för att beskriva trafikanters ruttval. Vanligtvis antar man att varje resenär väljer den väg som tar kortast tid utifrån den trängsel och de körtider som råder i den givna trafiksituationen. När OD-matrisen ändras kommer trafiksituationen också att ändras, vilket påverkar körtider och ruttval. Ju hårdare trafikerat gatunätet är, desto större blir förändringarna. I en tidsberoende modell måste dessutom flödet av trafik återges korrekt i både tid och rum.

Att förutse hur ruttval och flödesfortplantning beror av OD-matrisen är svårt och kräver omfattande analyser av hur känslig en given trafiksituation är för förändringar i resandet. I den här avhandlingen utvecklas och utvärderas olika sätt att modellera detta beroende.

En intressant fråga är var i gatunätet flödesmätningar ska utföras för att ge så heltäckande information som möjligt av trafiken. I avhandlingen jämförs olika placeringar med avseende på hur väl man lyckas skatta OD-matrisen och vi kan dra generella slutsatser om vilka placeringsstrategier som bör användas.

Avhandlingen innehåller både generella beräkningsrutiner, som i framtiden kan integreras i programvara, och mer specifika metoder, som redan kommit till nytta, bl.a. för att skatta tidsuppdelade OD-matriser för Göteborgs-området. De metoder som utvecklats är intressanta både för dem som tar fram den programvara som används inom branschen och för specifika frågeställningar hos enskilda väghållare, dvs. i Sverige främst Vägverket och de större kommunerna.

Introduction and Overview

1   Background

The number of cars increased heavily in the decades following World War II. During the 1950s the number of registered passenger cars in Sweden was more than quadrupled and ended with 1.1 Million (Statistics Sweden, 2007). The situation was similar in all western countries and opened for the use of mathematical models in traffic planning: How can the available infrastructure be used in the most efficient way and what investments would give the best effects? These questions are still fundamental in traffic planning, though the valuation of accessibility, congestion, emission, travel time, etc. have changed.

To be able to evaluate different engineering alternatives, there is a need for mathematical models. The outcome of such models is important support for making decisions on how links in the traffic network should be built or rebuilt, and, in a broader sense, how the city should be arranged with new commercial, industrial and living areas. Except for estimating the utility of building new roads, traffic models can be useful for evaluating the effects of for example changing speed limit and capacity (number of lanes), introducing road tolls and re-designing the intersections (turning lanes, signals, roundabouts, etc.). Another application is the development of plans of actions for taking care of traffic interruptions that are caused by incidents.

A fundamental issue in this thesis is the mathematical modeling of route choices for vehicles in a congested road network, which we recognize as *the traffic assignment problem*. Models for uncogested road networks, as well as models including alternative travel modes, are not covered in this thesis.

Mathematical modeling of traffic requires a lot of data and other information about the road network and the travel demand. The intended user of the models and methods, that are discussed in the thesis, is the road administrator for a medium or large sized city. For Swedish conditions we identify the Swedish National Road Administration (Vägverket), which is responsible for the major national road network, and the municipal traffic administrators in the larger citites as important users.

The accuracy of the modeled traffic situation depends on the quality of the available information, and how this data is combined and weighted from different sources. The travel demand is a key component and nearly every traffic model requires a tableau

specifying the travel demand between different places in the network. Such a tableau is called an *Origin–Destination matrix*, or *OD-matrix* for short; synonymously used terms are *trip table* or *(origin–destination) trip matrix*. This thesis is devoted to the problem of estimating reliable OD-matrices, in a reliable way.

A major distinction between the different types of traffic models is drawn with respect to their level of detail. A *macroscopic model* uses fluid variables, such as flow and density, and does not model individual vehicles — these are aggregated into continuous variables. A *microscopic model* describes vehicles (and often even drivers) individually. A *mesoscopic model* is in between and combines the ideas from macroscopic and microscopic models. Typically it uses macroscopic speed–flow relations to depict the motions of individual vehicles.

Macroscopic models are traditionally used for planning in larger networks over a longer time period: How should the city network be developed in the coming 5–10 years, with respect to some assumption of population growth in different sub-areas of the city? A microscopic model on the other hand is used for a smaller network and the result is used for more specific measures: How can the available space be allocated to lanes in the best way, to handle a troublesome intersection or sequence of intersections? A microscopic model often uses data from a macroscopic model as input, for example by stating an average situation. A mesoscopic model can be used to catch the overall changes in the traffic induced by some detailed changes of the infrastructure.

Another distinction is made between *time-dependent* (dynamic) and *time-independent* (static) models. For recovering how a traffic scenario is developed over time, we need a dynamic model, which can reproduce the reaction of the traffic to a current situation. A time-independent model can be described as a steady state in a dynamic model, i.e. a situation where reactions and contra-reactions are balanced. Time-independent models give average descriptions of the traffic situations. They require less input data, and well-established analytical models exist. For time-dependent models, the input data must provide more details on the traffic situation and we must make assumptions on how the traffic flow propagates in both time and space. Until today most time-dependent models are based on simulation. Traditionally, macroscopic models are time-independent, whereas we need a smaller and more detailed network of microscopic type to analyze time-dependent effects.

In this thesis both time-dependent and time-independent traffic models are discussed. We start in the next Section with an overview of macroscopic models for the time-independent case, by introducing the traditional *four-stage model*. We also briefly describe the problems which arise when a time-dimension is introduced. Thereafter, we introduce the OD-matrix estimation problem, which is the subject of this thesis. This problem is considered for the time-independent case and the time-dependent case in Section 3 and Section 4, respectively. Finally, in Section 5, the contribution of this thesis is presented. We discuss the purpose and motivation for the work presented and give a summary of the five annotated papers.

2   The four-stage model

The traffic planning process traditionally follows four sequential stages: trip generation, trip distribution, modal spilt and traffic assignment. *The four-stage model* was originally developed during the 1950s and 1960s for the planning of major highway facilities (Papacostas and Prevedouros, 2001). Soon, however, the model was applied also in other traffic planning situations and recognized as a standard for macroscopic modeling. Many software tools for traffic planning are still based on the ideas from the four-stage model.
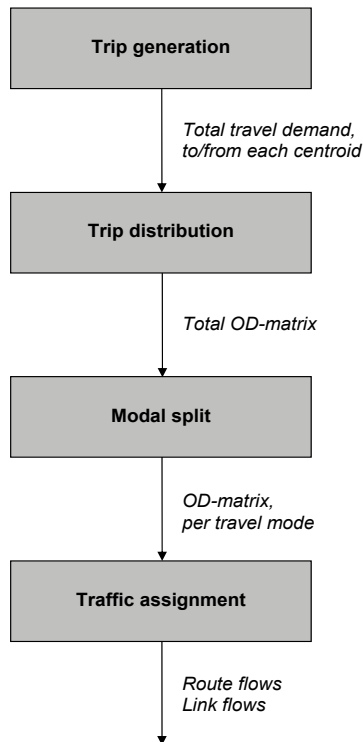
Trip generation

Total travel demand, to/from each centroid

Trip distribution

Total OD-matrix

Modal split

OD-matrix, per travel mode

Traffic assignment

Route flows
Link flows

Figure 1:   The four-stage model in its basic form.

Depending on the situation, some stages might not be applicable; for example, if no alternative travel modes are available. Over the years several alternative and/or extended planning schemes have been presented and some of them are implemented as options in the available software tools. Typical alternatives are to combine two or more stages, solve stages in reversed order, or to compute them iteratively for higher accuracy. Especially the second stage, deciding the distribution of travel

demand between origins and destinations, and the third stage, where the split onto different travel modes are computed, are often performed together as one stage. In the following the basic concept of the four-stage model, which is illustrated in Figure 1, will be presented to give a background to the OD-matrix estimation problem and its model setting. At the end of the section, we will briefly time-dependence as an extension to the model concept.

Many authors have already described the four-stage model. The following presentation has been inspired by the books of Wilson (1974), Patriksson (1994), Khisty and Lall (1998), Wright and Ashford (1998), Garber and Hoel (1999), Ortúzar and Willumsen (2001), and Papacostas and Prevedouros (2001), as well as the history by Bates (2000) and the thesis by Lundgren (1989).

2.1   Trip generation

The first of the four stages is the *trip generation*. The aim of this procedure is to determine how many trips there will be originating (trip production) or terminating (trip attraction) at each zone in the network. The size of these zones must be defined with an appropriate accuracy with respect to the purpose of the traffic model, and could range from some blocks in a city center up to a complete conurbation in a model for intercity travel demand forecasts. Each zone is represnted by a single node in the model, which we will refer to as a *centroid*. For performing this stage a broad variety of *survey data* is collected, concerning characteristics of the trip makers for each centroid, such as age, sex, income, auto ownership, trip-rate, land-use, and travel mode. In general it is a major project to collect the required data and it is therefore advantageous to coordinate the investigations to a specific base year, or, synonymously, a *target year*. In Sweden the database "Folk- och Bostadsräkningen", carried out by the Swedish authority for statistics, has been a valuable source. The latest "Folk- och Bostadsräkningen" was performed in 1990 (Statistics Sweden, 2007).

Depending on the puropose, it is possible to use different descriptions of the travel demand for different categories, trip purposes, travel modes, and time periods (time of day, day in week, week in season, etc.). Methodologically the two most frequently used techniques for trip generation analysis are cross-classification analysis (or category analysis) and multiple regression analysis. In the latter method values are aggregated for each centroid, whereas the central assumption in the cross-classification analysis is to disaggregate the demand to each household and calculate its generated trips by some pre-defined quota concerning the above given parameter types. It seems that most researchers today have agreed on the cross-classification analysis, essentially because it is based on census data (a priori given quota parameters). As a consequence, it is easy to compare and transfer models between cities, and the target year matrix can be used for validation, which is not the case in a regression analysis based model where the quota parameters depend on the data.

To conclude the description of the trip generation stage: Let P and Q denote the sets

of origin and destination centroids respectively. The output of the trip generation procedure is the sum of trips starting at all origins, o_p, $p \in P$, and the sum of trips ending at all destinations, $d_q, q \in Q$. (In most cases all centroids both produce and attract trips, which means that $P = Q$.)

2.2 Trip distribution

The second stage is the *trip distribution*. In this stage the generated sums of trips starting and ending at the centroids are connected to each other to form *travel demands* (OD-demands, OD-flows, or trip interchanges), for the OD-pairs. The aim of this stage is to find a trip distribution $g = \{g_{pq}\}$, $(p, q) \in P \times Q$ such that the aggregated information from the trip generation stage holds, i.e. such that $o_p = \sum_{q \in Q} g_{pq}$, $p \in P$ and $d_q = \sum_{p \in P} g_{pq}$, $q \in Q$. The trip distribution procedure is traditionally performed with some sort of *gravity model*. Alternatively, growth factor models and logit models can be used.

The name gravity model comes from Newton's law of gravitation, which states that the force of attraction between two bodies is directly proportional to the product of the masses of the two bodies and inversely proportional to the square of the distance between them. In the traffic distribution case it is assumed that the travel demand in OD-pair (p, q) is directly proportional to the trip ends production at the origin node, o_p, times the trip ends attraction at the destination node, d_q, weighted with a *deterrence function* (friction function), $f(\pi_{pq})$, of the travel impedance (time, distance, cost, etc.) between the two centroids. The travel demand from p to q can thus be expressed as

$$g_{pq} = k o_p d_q f(\pi_{pq}), \tag{1}$$

where k is some suitable weighting parameter. The deterrence $f(\pi_{pq})$ should be monotonically decreasing function of the travel impedance, π_{pq}, between p and q. The set $\Pi = \{\pi_{pq}\}$, $(p, q) \in P \times Q$ is known as the *skim table* (Papacostas and Prevedouros, 2001). Commonly a polynomial expression of the form $f(\pi_{pq}) = \pi_{pq}^{-\beta}$ is used. Here $\beta \geq 0$ is a parameter to be calibrated. (By setting $\beta = 2$, equation (1) states Newton's law of gravitation.) Alternatively, an exponential expression of the form $f(\pi_{pq}) = e^{-\beta \pi_{pq}}$ can be used.

In some models, it is also accounted for the relative attractiveness (accessibility) for each destination and the individual socioeconomic deterrence, k_{pq}, for each OD-pair. This leads to the following reformulation of equation (1):

$$g_{pq} = o_p \left(\frac{d_q f(\pi_{pq}) k_{pq}}{\sum_{q \in Q} d_q f(\pi_{pq}) k_{pq}} \right).$$

The bracketed expression can be interpreted as the probability that a trip originating at p will terminate at destination q.

There are many ways to calibrate the parameters required in the trip distribution procedure. Typically this is performed in an iterative process until a matrix close enough to the target year matrix is reproduced from the survey data. Most of these methods are of a heuristical nature; a common simplifacation is to estimate the value of each $f_{pq} = f(\pi_{pq})$ directly, instead of defining and calibrating the deterrence function.

For further reading on the use of gravity models in connection with transportation analysis, the reader is referred to the survey by Erlander and Stewart (1990).

The growth factor models are rather rough and cannot capture time-of-the-day variations. Instead, it is typically assumed that the travel demand is equal in both direction for all OD-pairs, i.e. that $g_{pq} = g_{qp}$, $p \in P$, $q \in Q$, and $P = Q$ (each centroid both produces and attracts trips).

To conclude the description of the trip distribution stage: Given the generated and attracted number of trips at each centroid, o_p and d_q, respectively, from the trip generation stage, the trip distribution procedure distributes the generated and attracted sums of trips onto travel demand g_{pq}, $(p, q) \in P \times Q$, such that $o_p = \sum_{q \in Q} g_{pq}$, $p \in P$ and $d_q = \sum_{p \in P} g_{pq}$, $q \in Q$. This distribution is typically performed with a gravity or a growth factor model.

2.3 Modal split

The third stage is the *modal split*. In the mode split (or mode choice) procedure the travel demand for each OD-pair is partitioned into different travel modes. In the simplest case there are only two travel modes available: private car and transit, but the travel modes can also specified in more detail, for example "sharing car with another person" might be a separate mode. The transit travelers can be classified into different sub-modes according to how they get to the bus stop or railway station (walk, bicycle, car, etc.). In some situations also the purpose of the trip ("home-to-work", "work-to-kindergarten", etc.) is considered, since, for example, it seems more likely that a person would choose to travel to his work with the transit system, but prefer a private car, if available, for social trips. Further, the purpose of the trip can be important since it might affect the acceptance for a delay or route guidance information. Besides the consideration of available modes and trip purpose, some models also consider the socioeconomic status of the trip-maker. For some persons, the travel time is more important than the travel cost, whereas the situation is the opposite for some other.

To determine how to disaggregate the OD-matrix into different travel modes, the *utility* of each mode must be calculated. The utility is a weighted sum of different attributes, like for example travel time, cost and comfort. When analyzing public

transport systems, parameters as walking time to transit line stop, waiting time, in-vehicle time, transit line frequency, transfer, and/or transfer waiting time can be included (Sjöstrand, 2001).

Let x_{hk} be the measured value of attribute $h \in H$ for travel mode $k \in K$ and let α_h, $h \in H$ be the corresponding weighting parameter. The weighted sum of all measured values, denoted by v_k, together with a random error term, ε_k, states the utility for travel mode $k \in K$:

$$u_k = v_k + \varepsilon_k = \sum_{h \in H} \alpha_h x_{hk} + \varepsilon_k. \tag{2}$$

The random error term expresses other attributes of travel mode k than those captured in H, the overall uncertainty of the measured values and also the variability in preferences among the individuals. Equation (2) is sometimes referred to as the *utility function*, which states the utility of travel mode k. We should mention that it is possible to specify a utility function separately for each centroid or OD-pair. Such a disaggregation can be used to capture effects that a trip maker might be more likely to use the transit system when traveling into the city center, with a shortage of parking places, than when traveling to another destination.

The utility for a travel mode is used to calculate the probability, p_k, that a certain traveler chooses travel mode $k \in K$, or, more precisely, the probability that a certain traveler perceives the highest utility by choosing that travel mode. In most applications of utility functions, the error term in equation (2) is assumed to be Gumbel (or Weibull) distributed, whereby the probability can be calculated through a *logit model*. The simplest case, where the error terms are independent and have equal variance, is the *multinomial logit model*, in which

$$p_k = \frac{e^{\mu v_k}}{\sum_{k \in K} e^{\mu v_k}}, \tag{3}$$

where v_k denotes the weighted sum measured values for travel mode $k \in K$, as defined in equation (2), and $\mu > 0$ is a scale parameter. A derivation of (3) was first proposed by McFadden (1973) and a complete derivation is presented by Domencich and McFadden (1975).

In the simple form of the logit model, i.e. in equation (3), the similarity between the travel modes, and thus the definition of the travel mode classes, is very important. Since the utility will be used in relation to the alternatives, the alternatives must be significantly different. Suppose, for example, that car, bus and tram are the three available travel modes in a system and suppose all of them have the same utility. By letting K consist of these three modes, the probability that a traveler chooses car will be 1/3. On the other hand, if bus and tram together are considered as "transit", this probability increases to 1/2.

To overcome this unwanted property the travel modes are often ordered in a hier-

archy, where the split of the travel demand on each level, has its own probability distribution. The probability for "tram" in the example above would be calculated as the product of the probability for "transit" and the probability for "tram, given transit". This, so-called *nested logit model* can of course have more than two levels, for example when, in addition to the type of "transit", also the way to access the transit network separates the travel modes. For further reading on nested logit models, see for example Ben-Akiva and Lerman (1985) and Oppenheim (1995).

Given the probability for a certain mode, it is easy to split the OD-matrix. The probability that a certain traveler chooses a certain travel mode can be interpreted as the *proportion* of all travelers choosing that mode. Thus the OD-demand for travel mode $k \in K$ in OD-pair $(p, q) \in P \times Q$ can simply be evaluated as $g_{pq}^k = p_k g_{pq}$.

To conclude the description of the modal split stage: Given the travel demand for each OD-pair g_{pq}, the modal split procedure determines how this is disaggregated into different travel modes, $k \in K$, such that $g_{pq} = \sum_{k \in K} g_{pq}^k$, $(p, q) \in P \times Q$. Normally this is done by finding a split proportion p_k for each mode $k \in K$.

2.4   Traffic assignment

The fourth and final stage is the *traffic assignment*. In the traffic (or trip) assignment the OD-matrix for each mode is assigned onto the traffic network, according to some principle. The aim of this procedure is to calculate the link flow volumes.

A traffic network can be represented by (N, A), which are the sets of its *nodes* and *links* respectively. Each node $n \in N$ is either a centroid or an intermediate node, modeling a network intersection. Thus the relation $P, Q \subseteq N$ holds. Each link $a \in A$ is either an actual street section, a transit connection or a generalized relation, for example an artificial "street" connecting a living area to the main network or symbolizing the walking path to the bus stop. All links are *directed*, and two-way streets are simply modeled as two separate links.

As indicated, each link must be a bearer of one or more travel modes and in the assignment procedure, this must be taken into account. A bus line, for example, is a link sequence itself, but each vehicle must also be assigned onto the street network, together with the private cars. In the rest of this thesis, we will only consider one travel mode, private cars. For notational convenience, the mode subscript k is therefore left out of the following presentation.

There are in general many possibal *routes* (or synonymously, paths) from one node to another. Let R_{pq} denote the set of acyclic routes from $p \in P$ to $q \in Q$. The set R_{pq} is finite, but typically very large. Therefore the assignment procedure must follow some assumption on how the routes are chosen.

The most common assumption is that each traveler will choose a route with the least instantaneous travel impedance. The travel impedance, π_{pq}, was introduced in

the trip distribution stage as a generalized measurement of time, distance, cost, etc., between the two centroids $p \in P$ and $q \in Q$. The term *generalized cost* is a more accurate description of the travel impedance, which instantaneously is experienced on a certain link, and is denoted by $c_a, a \in A$.

The generalized cost for a route is simply assumed to be the sum of the travel times on the included links, that is, travel impedance for passing an intersection must be expressed by the generalized cost on the links belonging to this intersection. The generalized link cost is a function of the *free flow travel time*, c_a^0, representing the constant link characteristics and, for a road link, the congestion level in the network, which is expressed by the link flow volumes, $v = \{v_a\}, a \in A$.

If each *link cost function* (link performance function) c_a is assumed to be independent of the flows on all other links, i.e. if $c_a(v) = c_a(v_a), a \in A$, the link cost functions are said to be *separable*. In most models, the link performance function $c_a(v_a)$ is an exponential or higher order polynomial function, which heavily grows rapidly as the link flow approaches the maximum capacity (typically around 1,800 vehicles per lane and hour). In many models, it is explicitly required that $c_a(v_a)$ is a monotonically increasing function.

The most widely spread method to distribute travel demand over alternative routes is the criterion "Equal Times", stated by Wardrop (1952):

> "The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route."

An assignment fulfilling this criterion, will be referred to as a *user equilibrium*. Introducing the non-negative route flow variables $h = \{h_r\}, r \in R_{pq}, (p, q) \in P \times Q$ and the link-route incidence variables $\delta_{ar}, a \in A, r \in R_{pq}, (p, q) \in P \times Q$, being 1 if link a is included in route r and 0 otherwise, the user equilibrium assignment is equivalent to the optimal solution to the following mathematical program:

$$\min_v \; f(v) \;\; = \;\; \sum_{a \in A} \int_0^{v_a} c_a(s)ds, \tag{4}$$

$$\text{s.t.} \quad \sum_{r \in R_{pq}} h_r \;\; = \;\; g_{pq}, \; \forall (p, q) \in P \times Q,$$

$$\sum_{(p,q) \in P \times Q} \sum_{r \in R_{pq}} \delta_{ar} h_r \;\; = \;\; v_a, \; \forall a \in A,$$

$$h_r \;\; \geq \;\; 0, \; \forall r \in R_{pq}, \forall (p, q) \in P \times Q.$$

This program was first stated by Beckmann et al. (1956) and is referred to as the *traffic assignment problem*. A complete derivation of the Wardrop's equilibrium principle from the this program is given in, for example, Patriksson (1994).

If the link cost is a monotonically increasing function of the link flow, the optimum solution to program (4) is uniquely determined in terms of link flows. The Frank–Wolfe method (Frank and Wolfe, 1956) is a well-known technique for linearly constrained convex problems and it has been successfully implemented for the assignment problem (4); for example the software tool Emme/2 (1999) is based on this method. For an overview of other techniques, proposed for the basic problem and its extensions, see Patriksson (1994).

Though the assignment problem has a unique link flow solution (as long as the link cost is a monotonically increasing function of the link flow) there are in general many corresponding route flows. This is one of the major obstacles when adjusting OD-matrices from observed link flows, assuming a user equilibrium assignment.

The assignment model discussed so far is *deterministic* in the sense that all travelers are assumed to drive the shortest path, i.e. one of the paths having the least generalized cost. This assumption takes no account for the variation in the travelers' different perception of travel cost. Daganzo and Sheffi (1977) extended the Wardropian user equilibrium condition to the principle of *stochastic user equilibrium* by stating that:

> "In a stochastic user equilibrium network no user believes he can improve his travel time by unilaterally changing routes."

Mathematically this is modelled by adding a error term to the generalized cost for each route. Depending on which distributions and which mutual dependencies that are assumed for these terms, different mathematical programs can be formulated. If the random terms are assumed to be independent and identically distributed Gumbel variates, a multinomial logit model can be used. An advantage of the stochastic assignment model compared to the deterministic model is that the route flows of an assignment are uniquely determined.

A consequence of the stochasticits assumption is that each route has a strictly positive flow. For a network including a cycle there is no upper limit for the number of routes, and thus an infinite number of route flows has to be considered, at least theoretically. In practice of course, the number of routes has to be restricted somehow. A remaining problem therefore is how to generate a restricted set of routes, which is representative for all possible routes. When just a few routes are generated for an OD-pair, these tend to be rather similar, i.e. to have a number of links in common. The overlap between the routes then gives an incorrect value for the proportion of a certain OD-demand passing each link. For further details on stochastic assignment models, see Sheffi (1985).

To conclude the description of the traffic assignment stage: Given the travel demand for each mode in each OD-pair, g_{pq}^k, $k \in K$, $(p, q) \in P \times Q$, the traffic assignment procedure assigns the travelers to routes in the network and predicts the traffic situation in terms of the link flows v_a for all links in the network, $a \in A$.

2.5 Time-dependent traffic models

Although the four-stage model is almost fifty years old, it is still fundamental for strategic traffic planning, such as, for example, identifying bottlenecks in the network, or dimensioning the infrastructure to a new living area. The outcome is, however, always a static description of the situation, which does not provide any information on how the traffic fluctuates over time. A time-dependent (dynamic) model must consider the influence from traffic conditions in a certain time period to any succeeding time period.

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, and controlling traffic-lights and other traffic facilites.

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 a dynamic network loading procedure.

Several attempts to extend Wardrop's "Equal time" criterion for a time-dependent model have been proposed. A major distinction is drawn between models relying on the instantaneous and actual travel time, respectively. The instantaneous travel time for a route is computed as the sum of the link travel times on the links included in the route at that time, when the journey is started, whereas the actual travel time relies on a forecast on what the travel time on a link will be at that time, when the traveler reaches the link. The instantaneous travel time might be combined with a model where re-routing during the trip is allowed.

In comparison to the static assignment models, where most of the researchers have agreed on the four-stage model and the basic ideas in each of its stages, the area of dynamic models is more unexplored. Though the first models for dynamic traffic assignment were proposed some thirty years ago (Merchant and Nemhauser, 1978a, 1978b), there is still no standard model framework. The time-dependent models are also surprisingly poor described in books and survey articles. Two exceptions are the literature overview by Peeta and Ziliaskopoulos (2001) and the foundations on dynamic modeling given by Han (2003).

3 Estimation of time-independent OD-matrices

The OD-matrix estimation problem amounts to finding an OD-matrix which, when it is assigned onto the network, induces link flows close to those which have been

observed in traffic counts. Most models also require some kind of information about the magnitude of the prior travel demand for each OD-pair, i.e. a *target* OD-matrix. The target OD-matrix is typically an old (out-dated) OD-matrix, possibly updated by some growth factor, or a matrix generated through a trip distribution procedure. When a target OD-matrix is used as input, the OD-matrix estimation problem is often referred to as a *calibration* or *adjustment* problem.

The problem of finding an OD-matrix which corresponds to some given link flow observations, can be seen as an inverse of the traffic assignment problem. However, there are some factors, which make the estimation problem much more complicated to handle. First of all, there is in general not possible to observe the flows on all links in the network. And even if all link flows could indeed be measured, the data would most likely be neither error-free nor consistent. Secondly, even though correct and error-free link flow data would be available for all links in the network, there are still many OD-matrices which assigned onto the network would induce the observed link flows. This is motivates the use of a target matrix.

The OD-matrix estimation problem for the time-independent case has been well studied in the past decades. The following overview is inspired by the literature surveys by Willumsen (1981), Barceló (1997), Bell and Iida (1997), Abrahamsson (1998), and Ortúzar and Willumsen (2001).

3.1 Problem description for the time-independent case

In the OD-matrix estimation problem we are interested in finding a feasible OD-matrix $g \in \Omega$, where $g = \{g_i\}$, $i \in I$, consists of the demands for all OD pairs. (We use the well-established term OD-matrix for g, although, for a convenient description, the elements are organized in a vector.) The assignment of the OD-matrix onto the links in the network follows an assignment proportion matrix $P = \{p_{ai}\}$, $i \in I$, $a \in A$, with p_{ai} being the proportion of the OD demand g_i that uses link a. We will use the notation $P = P(g)$ to emphasize that, in general, because of congestion, these proportions depend on the the traffic volumes, i.e. on the OD-matrix.

When assigned to the network, the OD-matrix induces a flow $v = \{v_a\}$, $a \in A$, on the links in the network. We assume that observed flows, $\tilde{v} = \{\tilde{v}_a\}$, are available for a subset of the links, $a \in \tilde{A} \subseteq A$, and that a target matrix $\hat{g} \in \Omega$ also is available.

The OD-matrix estimation problem can now be formulated as

$$\begin{aligned} \min_{g,v} \ F(g,v) &= \gamma_1 F_1(g,\hat{g}) + \gamma_2 F_2(v,\tilde{v}), \\ \text{s.t.} \ \sum_{i \in I} p_{ai}(g)g_i &= v_a, \quad \forall a \in \tilde{A}, \\ g &\in \Omega. \end{aligned} \tag{5}$$

The functions $F_1(g, \hat{g})$ and $F_2(v, \tilde{v})$ are generalized distance measures between the estimated OD-matrix g and the given target matrix $\hat{g}$, and between the estimated link flows v and the observed link flows $\tilde{v}$, respectively. These functions are assumed to be convex and they can be designed to account for variations in the quality of the given data.

The parameters $\gamma_1 > 0$ and $\gamma_2 > 0$ reflect the relative belief (or uncertainty) in the information provided by $\hat{g}$ and $\tilde{v}$, and the problem thus can be interpreted as a two-objective problem, where the two objectives are expressed in F_1 and F_2, and γ_1 and γ_2 are the corresponding weighting factors. In one extreme case, using $\gamma_1 = 0$, the target matrix will have no influence, and in the other, using $\gamma_2 = 0$, the target matrix will be reproduced and the observed link flows will have no influence.

The set of feasible OD-matrices, Ω, usually consists of all non-negative OD-matrices, but it can also be restricted to the matrices that are within a certain deviation from the target matrix, i.e. $\Omega = \{g \geq 0 | (1 - \alpha)\hat{g} \leq g \leq (1 + \alpha)\hat{g}\}$, for some parameter $\alpha > 0$ stating the tolerance level. An analogous restriction could be used to instead state a maximum allowed deviation from the link flow observations $\Omega = \{g \geq 0 | (1 - \beta)\tilde{v}_a \leq v_a \leq (1 + \beta)\tilde{v}_a, a \in \tilde{A}\}$, where $\beta > 0$.

The set Ω can also be constrained by relations between flows on different links, for example turn proportions at some intersections. Often intersections are designed with different lanes for different turning movements, which gives an opportunity to measure the turn proportions directly. This information can be added to strengthen the estimation problem. However, these proportions must be used carefully to avoid inconsistency with the observed link flows.

Another possibility is to restrict the total travel demand in all OD-pairs originating or terminating at a certain centroid, which in the four-stage model would represent an adjustment of the trip distribution with respect to the trip generation. A weaker requirement would be a total number of trips for the entire OD-matrix. In any case, the set of feasible OD-matrices Ω will remain convex, and be easily handed.

3.2   Different choices of objective

In the objective of the estimation problem (5) the deviations from the target matrix and the link observations are minimized. Obviously the resulting OD-matrix depends on the choice of distance measure, and we will therefore discuss this choice in more detail.

One type of distance measure is the *maximum entropy* function, which can be formulated as

$$F_1(g, \hat{g}) = \sum_{i \in I} g_i (\log g_i - 1). \tag{6}$$

With this choice, the target matrix is of no importance A special entropy formulation is derived from the principle of *minimum information* and was originally proposed by van Zuylen (1978) as

$$F_1(g, \hat{g}) = \sum_{i \in I} g_i (\log \frac{g_i}{\hat{g}_i} - 1). \tag{7}$$

If no target OD-matrix is available, it is sensible to assume that all values are equally likely, and by replacing $\hat{g}_i$ by unit weights for all $i \in I$, the expression in (7) reduces to that in (6). In this case, the OD-matrix estimation, however, turns into a trip distribution procedure.

Maximum entropy functions are used in the estimation models presented by Jörnsten and Nguyen (1979), van Zuylen and Willumsen (1980), Bell (1983), Bell (1984), Fisk (1988), Brenninger-Göthe et al. (1989), and Tamin and Willumsen (1989).

A second type of distance measure is the *maximum likelihood*, which states the likelihood of observing the target OD-matrix by the estimate. It is here assumed that the elements of the target OD-matrix are obtained as observations of a set of random variables. For a Poisson distributed system with a sampling factor ρ the distance measure can be formulated as

$$F_1(g, \hat{g}) = \sum_{i \in I} (\rho_i g_i - \hat{g}_i \log g_i). \tag{8}$$

Various types of maximum likelihood measures are used in the models proposed by Ben-Akiva (1987), Spiess (1987), and Tamin and Willumsen (1989). The maximum likelihood formulation proposed by van Zuylen and Branston (1982) is based on an assumption of normal distributed deviations.

Though the presentation here has focused on the deviation from the target OD-matrix only, both the maximum likelihood and the maximum entropy measures, of course, can be used for the deviation from the link flow observations as well, by formulating F_2 analogously to (8) and (6), respectively. Some of the refereed models do this, but there are also other possibilities.

The type of objective which is most common in the models proposed in the last decade is the *least-square* formulation. The least-square is a well-known deviation measure used in many types of estimation problems and is given by

$$F_2(v, \tilde{v}) = \frac{1}{2} \sum_{a \in \tilde{A}} (v_a - \tilde{v}_a)^2. \tag{9}$$

It is of course possible to give individual weights to the single deviations. One way of choosing the weights is to utilize information on the reliability of each observation. The measurements contained in $\tilde{v}$ are normally computed as means from a set of

observations for each link. In this case we can use the variance σ_a^2 among the measurements to account for how important each link observation is, and reformulate (9) as

$$F_2(v, \tilde{v}) = \frac{1}{2} \sum_{a \in \tilde{A}} \frac{1}{\sigma_a^2} (v_a - \tilde{v}_a)^2. \tag{10}$$

It is of course also possible to take the covariance between the observations on different links into account. Such measures are referred to as *generalized least-squares*. An analogous formulation can capture the deviation from the target OD-matrix, i.e. $F_1(g, \hat{g})$.

Various types of generalized least-square formulations have been used in the models proposed by Carey et al. (1981), Cascetta (1984), McNeil and Hendrickson (1985), Spiess (1990), Bierlaire and Toint (1995), Florian and Chen (1995), Maher and Zhang (1999), Bianco et al. (2001), Cascetta and Postorino (2001), Maher et al. (2001), Codina and Barceló (2004), Doblas and Benitez (2005), and Nie et al. (2005) among others.

To summarize this discussion, the functions $F_1(g, \hat{g})$ and $F_2(v, \tilde{v})$ represent distance measures, between the estimated OD-matrix g and the given target matrix $\hat{g}$, and between the assigned link flow v and the observed link flow $\tilde{v}$, respectively. Typically some combination of maximum entropy, maximum likelihood and least-square expressions are used, and they can be designed to take account varying data quality. We conclude that, in any case, the functions to be minimized are continuous, convex and at least two times differentiable with respect to their respective arguments.

3.3 Characteristics of the constraints

We have already mentioned that there are in general many OD-matrices which, when they are assigned to the network, induce the same link flows. In this section we will further explore the set of constraints defining the relationship between the travel demand (the OD-matrix) and the link flows.

Consider again the equations from the problem description (5), which connect the OD-matrix to the link flows:

$$\sum_{i \in I} p_{ai}(g) g_i = v_a, \forall a \in \tilde{A},$$

There is one equation for every link flow observation. This equation system is underdetermined as long as the number of OD-pairs, $|I|$, is greater than the number of link flow observations, $|\tilde{A}|$.

In general the number of OD-pairs is much greater than the number of link flow observations, especially for real-world networks. The number of OD-pairs is a subset of all possible node pairs, i.e. $I \subseteq N \times N$. Typically, the number of centroids are proportional to the number of nodes, and each pair of centroids defines an OD-pair. Therefore $|I| \propto |N|^2$ holds, which means that the number of unknown travel demands grows quadratically with the network size.

However, the mean number of links connected at each intersection, i.e. node, is independent of the network size. If we assume that some portion of the link flows are observed, the relation $|\tilde{A}| \propto |A| \propto |N|$ holds. Thus, the number of equations is proportional to the network size. We therefore conclude that the OD-matrix estimation problem has a greater freedom of choice, the greater the network is.

Topological dependencies in the network further delimits how well the OD-matrix can be determined by the equation system. Kirchoff's law, well-known from physics, states that the sum of incoming and outgoing flows at any intermediate node must coincide. This means that, for each intersection, at least one link flow is directly given from the others, which results in a row-wise dependency for the equation system.

The non-zero elements $p_{ai}(g)$ arise from one or more paths generated for OD-pair $i \in I$. However, since every subpath of a path is also a path, every pair of nodes along a certain path is also connected through parts of this path. This results in a column-wise dependency for the equation system.

We can conclude that the equation system most likely is not of full rank, which further increases the freedom of choice in the OD-matrix estimation problem.

3.4   Detector allocation

When OD-matrices are to be estimated from the information contained in link flow observations, the choice of links where detectors are placed is of course very crucial for the result. Therefore, some attention should be given to the problem of how to choose the set of detector links, aiming for a good and reliable estimate of the OD-matrix. The main reference for the overview in this section is the survey part in Paper IV of this thesis.

When link flow detectors are allocated to the traffic network with the aim to estimate a reliable OD-matrix, we want to maximize the coverage of the traffic (link flows, route flows, OD-pairs, OD-demands, etc.) in one way or the other. Clearly, there are different ways to specify how this coverage should be accomplished.

First of all, we must define under which conditions we consider a certain link to cover the travel demand in an OD-pair. There are basically two different definitions in the literature. In the first, and most commonly used approach, we consider an OD-pair to be covered, as soon as a certain portion of the travel demand passes at least one

link with a detector. Clearly, this approach requires some assumption about how the travel demand is distributed onto routes through the network. This route choice information is typically supplied from an assignment of the target OD-matrix.

In the other definition of coverage, we ignore the route choice and consider an OD-pair to be covered if and only if every possible route from the origin node to the destination node passes at least one detector. In practice this approach tends to allocate detectors to bridges and tunnels along natural boarders, such as rivers or railways.

Another aspect to take into account is whether we want to cover many OD-pairs, or many travelers. In the first case, all OD-pairs are equally important, whereas the second case counts every traveler, and thus, give more importance to the OD-pairs with a greater travel demand. Alternatively, we can formulate the detector allocation problem as to maximize the coverage of routes or route flows. The most commonly used strategy seems to be OD-pair coverage; see the survey part in Paper IV of this thesis.

The simplest approach for detector allocation is to choose those links where we expect the maximum flow to be observed, without taking account for any travelers being counted twice. Detectors are then typically allocated along one major road and most of the routes passing one detector also pass some other.

In practice, of course, link flow observations are not being performed only for providing information to the OD-matrix estimation problem. Detectors are also being placed for other purposes, such as the management of traffic signals, different real-time information systems, or road tolls. Therefore, when modeling the detector allocation problem, it is natural to consider the case where the placement of some link detectors are already at hand.

3.5   Constant assignment

From a modelling point of view, the most distinguishing difference between the approaches for the OD-matrix estimation problem, is how the assignment proportions in P are calculated and re-calculated throughout the estimation procedure. Especially it is crucial if the assignment matrix $P(g)$ is assumed to depend on g, i.e. if route choices are made with respect to congestion, or not. In the latter case, $P(g) = P$ is a constant assignment matrix, and the first set of constraints in the generic description (5), can be formulated as

$$\sum_{i \in I} p_{ai} g_i = v_a, \quad \forall a \in \tilde{A}. \tag{11}$$

The assumption that the assignment, i.e. the route choice, is independent of the load on the links, is realistic in a network with very low congestion rate, or in networks where in practice only one route can be used in each OD-pair. An example of this

is a corridor network, modeling a motorway through a city and its entrances and exits. In such a case the shortest path between origin and destination is uniquely determined, independent of the travel times.

A bit more sophisticated are the OD-matrix estimation methods where the used assignment matrix is taken from a carefully computed assignment of the target matrix, $P(\hat{g})$. If the OD-matrix to be estimated is close enough to the target matrix, this is a good approximation even for congested networks.

The models proposed by van Zuylen (1978), van Zuylen and Willumsen (1980), Carey et al. (1981), Willumsen (1981), Bell (1983), Cascetta (1984), McNeil and Hendrickson (1985), Spiess (1987), Brenninger-Göthe et al. (1989), Bierlaire and Toint (1995), Bianco et al. (2001) and Bierlaire (2002) all assume that the route proportions P are kept constant in the estimation problem. In the model by van Zuylen and Branston (1982) the route proportions P are replaced with flow conservation constraints for each node.

3.6   Equilibrium assignment

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 of the current traffic situation (travel times/link flows), which in turn depends of the OD-matrix. Thus, the relationship between the route proportions P and the OD-matrix g can only be implicitly defined. The set of feasible solutions to the estimation problem (5), is then defined as all the points (g, v) where v is the link flow solution satisfying an assignment of the corresponding OD-matrix $g \in \Omega$.

Nguyen (1977) presented the first model of this type and an extended version was proposed by Jörnsten and Nguyen (1979). Gur et al. (1980) suggested a way to obtain unique OD-matrices. Erlander et al. (1979) and Fisk and Boyce (1983) have proposed OD-matrix estimation methods based on a combined distribution and assignment model. In all these estimation procedures it is assumed that the assignment is made according to the deterministic user equilibrium assumption. This assumption will hold also in the following presentation.

The deterministic equilibrium assignment is an *inferior* problem to the *superior* problem of estimating the OD-matrix. A problem, which can be separated into one superior (or synonymously *outer* or *upper*) part, and one inferior (or *inner* or *lower*) part, is called a *bilevel problem*. The generic OD-matrix estimation problem given in (5), can be reformulated as a bilevel programming problem in the following way.

In the superior problem, the OD-matrix g defines the decision variables and we want to minimize $F(g, v)$ subject to $g \in \Omega$, that is

$$\min_g \ F(g, v) \ = \ \gamma_1 F_1(g, \hat{g}) + \gamma_2 F_2(v, \tilde{v}), \tag{12}$$
$$\text{s.t.} \qquad g \ \in \ \Omega.$$

The link flow v must satisfy the equilibrium assignment conditions, given the OD-matrix g. These conditions are fulfilled by solving the nonlinear inferior problem in which the link (and route) flows are decision variables:

$$\min_v \ f(v) \ = \ \sum_{a \in A} \int_0^{v_a} c_a(s)ds,$$

$$\text{s.t.} \qquad \sum_{k \in K_i} h_k \ = \ g_i, \ \forall \, i \in I, \tag{13}$$

$$\sum_{i \in I} \sum_{k \in K_i} \delta_{ak} h_k \ = \ v_a, \ \forall \, a \in A, \tag{14}$$

$$h_k \ \geq \ 0, \ \forall \, k \in K_i, \forall \, i \in I. \tag{15}$$

Fisk (1988) was first to give a bilevel formulation of the OD-matrix estimation problem. She used a variational inequality formulation to express the equilibrium conditions and in this way she allowed general link cost functions, which must not be separable (see Section 2.4).

It is well-known that bilevel programming problems are in general non-convex; see for example Bard (1998). By using methods known today, at the best a local optimum solution is obtained.

Spiess (1990) suggested a heuristic approach to solve problem (12). It is an iterative procedure, in which $\gamma_1 = 0$ and $\hat{g}$ is used as initial solution. In his approach, an approximate gradient of the objective function with respect to the OD-matrix is computed, under the assumption that the proportion matrix $P(g)$ is locally constant. Spiess' heuristic is efficient for large-scale applications and has been included in the software tool Emme/2 (1999). Doblas and Benitez (2005) have shown how the method could be made even more efficient by adding linear constraints bounding the possible changes of travel demands in the OD-pairs.

Florian and Chen (1995) reformulated the bilevel problem into a single level problem using the concept of marginal functions. They proposed to use a Gauss-Seidel type heuristic to solve the problem. 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 OD-pair are known beforehand, and it is thus appliciable only to very small problem instances.

Yang et al. (1992), Yang (1995) and Yang and Yagar (1995) all use the bilevel formulation and propose heuristics, which iteratively solve the upper and lower level

problems. Information from the lower level problem is transferred by so called influence factors, which are defined by route proportions or explicit derivatives. The derivatives are computed using the sensitivity analysis by Tobin and Friesz (1988). Assuming complementarity conditions to hold and disregarding any topological dependencies, they get approximate values of the derivatives, which are acceptable for small to medium sized networks. However, for larger networks the topological dependencies are significantly greater. Also, since all these methods include matrix inversions they are computationally very demanding for large problem instances.

Maher and Zhang (1999) have developed an iterative method where a first order Taylor approximation is used to express the changes of the assignment map $P(g)$ with respect to the OD-matrix g. In a first step the assignment map is kept constant and a tentative OD-matrix is estimated with some of the techiques discussed in Section 3.5. The tentative OD-matrix is, however, not taken as it is, but is used to give a search direction from the present OD-matrix. By assigning the tentative OD-matrix to the network, we get an approximation of how the assignment map will change along the search line direction. Maher et al. (2001) have further developed the method to the case where a stochastic assignment is assumed.

The method proposed in Paper I of this thesis is based on the more general sensitivity analysis presented by Patriksson (2004). It is a further development of the method by Spiess (1990), where a second order approximation is used for the partial derivatives. Further, a solution scheme is proposed, where the mutual influence between the OD-pairs, which is considered, can be restricted to keep a good balance between the computational time needed for yielding the search direction and finding an optimal step length.

The algorithm developed by Codina and Barceló (2004) is an application of the general method for non-differentiable convex minimization, proposed by Wolfe (1975). It is based on subgradients and does not need any sensitivity information.

Neither the method proposed in Paper I of this thesis nor the method by and Codina and Barceló (2004) is affected by topological dependencies, as the methods based on the sensitivity analysis by Tobin and Friesz (1988). Further, none of them involves matrix inversions and they therefore seem to be efficient also for larger networks.

The method proposed by Sherali et al. (1994), and further developed by Sherali et al. (2003), assumes a deterministic user equilibrium. It is based on error-free information on the travel times for all links in the network, and this information must be consistent with the link flow observations. In the method, it is assumed that an equilibrium link flow is observed, and thus that the set of used routes is known beforehand. This requirement makes a comparison to other methods unfair.

The methods proposed by Cascetta and Postorino (2001), Clegg et al. (2001), Maher et al. (2001), and Yang et al. (2001) differ from the other methods in the sense that they presume a stochastic user equilibrium.

4 Estimation of time-dependent OD-matrices

If the travel demand in the OD-matrix is assumed to vary over time, the OD-matrix is said to be time-dependent. 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 and the new techniques for supplying interactive information via internet, variable message signs and so on. An accurately estimated time-dependent OD-matrix is a basis for the decisions in many *Intelligent Transportation Systems* (ITS). Lind (1997) gives an overview of ITS applications with special attension to Swedish conditions.

Time-dependent OD-matrices are used both for strategic and operational purposes. In the strategic area the computations are made off-line and the aim is to model the normal traffic situation as well 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 incidents. This type of ITS applications are sometimes referred to as *Advanced Traffic Management Systems* (ATMS).

The pre-calculated scenarios and plans of actions are also used as a default description of the traffic conditions in the operational management. For an accurate real-time model of the traffic, these values, of course, must be combined with instantaneous estimates. This type of operational models are used to produce travel time forecasts, which in turn are essential for different kind of route guidance systems. *Advanced Traveler Information Systems* (ATIS) is a commonly used term for such applications.

In the following, first a generic formulation of the estimation problem for the time-dependent case is given, analogous to the formulation of the time-independent case in Section 3.1. We will then discuss some of the algorithms which have been proposed to solve the time-dependent OD-matrix estimation problem, and, finally, we will present some of the methods used for real-time estimation. It should be pointed out that the models where time-dependent OD-matrices are used, in general consider networks smaller than those considered in the time-independent case. Especially, the methods for real-time estimation are mostly designed for very small networks, typically with a corridor structure.

Estimating time-dependent OD-matrices is relatively new research area and unlike the time-independent case, survey articles are hardly found. For some introduction, we refer to the overviews by Lindveld (2003) and Balakrishna (2006).

4.1 Problem description for the time-dependent case

A general formulation of the time-dependent OD-matrix estimation problem can be derived from the time-independent formulation, given as problem (5) in Section 3.1. In the time-dependent estimation problem we aim at finding an OD-matrix

$g = \{g_{it}\} \in \Omega$, where element g_{it} expresses the travel demand in OD-pair $i \in I$ leaving the origin node in time period $t \in T$. The assignment of the OD-matrix onto the links in the network is made according to the assignment mapping $P = \{p_{it}^{ar}\}$, $a \in A$, $r \in T$, $i \in I$, $t \in T$, where each element in the matrix is defined as the proportion of the travel demand g_{it} passing link $a \in A$ during time period $r \in T$. As for the time-independent case these proportions may depend of the demand.

When assigned to the network, the OD-matrix induces a flow $v = \{v_{ar}\}$, $a \in A$, $r \in T$, on the links in the network. We assume that observed flows, $\tilde{v} = \{\tilde{v}_{ar}\}$, are available for a subset of the links, $a \in \tilde{A} \subseteq A$, in all time periods $r \in T$, although observations must not be performed in all time periods for all of the observed links. We could even consider a different discretization of time for the link flows than for the OD-matrix. In practice, however, the same set of time periods T is used for both the OD-matrix and the observed link flows.

Given a target OD-matrix $\hat{g} \in \Omega$ we can now formulate the generic time-dependent OD-matrix estimation problem as

$$
\begin{aligned}
\min_{g,v} \; F(g,v) &= \gamma_1 F_1(g,\hat{g}) + \gamma_2 F_2(v,\tilde{v}), \\
\text{s.t.} \; \sum_{i \in I}\sum_{t \in T} p_{it}^{ar}(g)g_{it} &= v_{ar}, \quad \forall a \in \hat{A}, r \in T, \\
g &\in \Omega.
\end{aligned}
\tag{16}
$$

Deviation measures $F_1(g,\hat{g})$ and $F_2(v,\tilde{v})$ for the time-dependent formulation are chosen as for the time-independent case, see Section 3.2.

The set of feasible OD-matrices, Ω, is bounded analogously to the time-independent case. A new possibility is to add constraints restricting the maximum deviation of the time-aggregated OD-demand from the target matrix, i.e. the distance of $\sum_{t \in T} g_{it}$ from $\sum_{t \in T} \hat{g}_{it}$. Also for the time-dependent case Ω will remain convex and be easily handed.

4.2   Methods for time-dependent estimation

One of the first models for time-dependent OD-matrix estimation was proposed by Willumsen (1984). This model simply makes the assumption that the route choice ratio is fixed and independent of time, i.e. that $p_{it}^{ar} = p_{it}^{ar}(g)$ holds for all $a \in A$, $r \in T$, $i \in I$, $t \in T$. This is an extension to the proportional assignment assumption in the time-independent case, see Section 3.5. As in the time-independent case, it is assumed that the congestion level is kept throughout the estimation, meaning that route choice and flow propagation are constant.

Willumsen's model is used in the OD-matrix estimation procedure in the software tool Contram; see Contram (2002). In this implementation the proportional as-

signment mapping P is taken from an assignment of the target matrix $\hat{g}$, which is performed by simulation. If the OD-matrix to be estimated, g, is close to the target matrix, $\hat{g}$, the assignment mapping for the target matrix, $P(\hat{g})$, is probably a good approximation to the actual assignment mapping, $P(g)$. If, however, the target matrix $\hat{g}$ is unreliable, then so is the assignment mapping $P(\hat{g})$. Especially if the network is congested and/or there are many alternative routes, the assignment mapping will be sensitive also to small changes in the travel demand.

Davis and Nihan (1991) uses a stochastic procedure for generating the assignment mapping, which is parameterized by the means and variances of the travel demand. They then develop a maximum likelihood estimator, which can be viewed as a development of the method proposed by Spiess (1987) for the static case. Davis (1993) extends 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) make assumptions on the travel time distribution and thereby they can account for different flow propagation in different time periods. This improvement is important for larger networks, where the assumption on equal travel times might be too rough. Hereby, the model becomes dynamic both in flow propagation and in route choice. However, none of these aspects is related directly to that congestion, which is actually given by the OD-matrix, but only to the time period. The relationship between the OD-matrix g and the assignment mapping P is based on statistics only.

Cascetta et al. (1993) develop a model for a general two-objective form of the problem. In their numerical tests a general least-square estimator has been used. The proportional assignment mapping,, P is expressed as a product of a time-dependent link–route incident mapping, $\Delta = \{\delta_{ar}^{kt}\}$, and a probability term, $\rho(k|t)$, expressing the probability that a traveler in OD-pair i, departing in time-period t, will choose route $k \in K_i$:

$$p_{it}^{ar} = \sum_{k \in K_i} \delta_{ar}^{kt} \rho(k|t). \tag{17}$$

This model is further developed by Tavana (2001), and Ashok and Ben-Akiva (2002). The latter authors also address a real-time formulation of the model (see Section 4.3).

Another statistical model is proposed by Hazelton (2000), who assumes P to be given by a Poisson distribution of the demand in the OD-matrix, in which the variation of route choice proportions is represented. The network is assumed to be uncongested in the sense that the demand has no influence neither to route choices nor to flow propagation. As objectives the maximum multivariate normal approximation of the likelihood is used. Some ideas for decreasing the complexity of the algorithm are given by Hazelton (2003).

In the model by Sherali and Park (2001) each time-dependent route flow, h_k^t, where $\sum_{k \in K_i} h_k^t = g_{it}$, is estimated individually. They develop a column generation procedure where the master problem is to find a non-negative route flow solution h, such that the objective in (16) is minimized. The column generation problem is to compute paths in a time-expanded network. Unfortunately, convergence cannot be guaranteed.

The model proposed by Lindveld (2003) is basically a time-dependent extension of the time-independent model proposed by Maher and Zhang (1999), for the case of deterministic user equilibrium, and further developed by Maher et al. (2001) for stochastic user equilibrium. The method has been successfully implemented for a small test network with a corridor structure, i.e. a network with no route choices.

Zhou and Mahmassani (2006) extend the generic formulation in (16) with a third objective capturing information from automatic vehicle identification data. This is used as a target for the proportion of the flow on a certain link in a certain time period, which passes another link in a succeeding time period. In the experiments, the software Dynasmart (2002) is used for the dynamic traffic assignment.

4.3 Methods for real-time estimation

A special case of the time-dependent OD-matrix estimation problem, introduced in Section 4.1, occurs when the estimation is performed instantaneously. We will refer to this as *real-time estimation* or *on-line estimation*. Such an estimate of the current travel demand in the network is a useful input for different kind of operational management, especially for computing travel time forecasts and producing route guidance information.

Most of the models proposed for real-time estimation can be used for off-line purposes as well. For a real-time estimator, however, the computing time available is very limited, and therefore the accuracy in the methods generally is lower. It should be noted, though, that some of the concepts presented for real-time estimation could be refined for more accurate computation, if the estimation is performed off-line.

For defining the real-time estimation problem, the set of time-periods must be specified as an ordered, possible infinite, set, $T = \{t_1, t_2, t_3, ...\}$. An estimation is performed at a time τ, when the link flow observations are available only for the present and preceding time periods, i.e. $\tilde{v} = \{\tilde{v}_{ar}\}$, $a \in \tilde{A}$, $r \in T^\tau$, where $T^\tau = \{t_1, t_2, t_3, ..., t_\tau\} \subseteq T$.

The model proposed by Cremer and Keller (1987) is one of the first of this type. In real-time they estimate the turning proportions at intersections utilizing the causal dependencies between the observed link flows. They allow for different route choices in different time periods. The turning proportions are expressed as functions time only, and the model contains no relations between the travel demand and the route choice. Further, it does not take the different flow propagations, and thus, the

24

differences in travel times, into account. Their work compares weighted least-square estimation, constrained optimization, simple recursive estimation, and a Kalman filtering approach. The results from Nihan and Davis (1987, 1989) and Sherali et al. (1997) are basically improvements of the computational techniques for the model.

The model concept is further considered by van der Zijpp and Hamerslag (1996), who use an improved Kalman filtering technique. Their method is developed for a corridor network, where only one possible route is available for every OD-pair. In the model proposed by van der Zijpp (1997), the estimation also incorporates automated vehicle identification data.

The models by Bell (1991), and Chang and Wu (1994) are more generalized in the sense that also varying travel times are taken into account. Bell uses statistical expressions for the travel times, comparable to those presented by Bell et al. (1991) for the off-line approach, which is discussed in Section 4.2 above. Chang and Wu compute the present travel demand recursively from the data collected from previous time intervals. Chang and Tao (1996) further extend the model to also incorporate signal settings. In the extension by Wu (1997) general network structures are considered.

Trying to restrict the computational effort required, Ashok and Ben-Akiva (1993, 2000) work with a state vector, only containing the (time varying) deviations from the target matrix $\hat{g}$. The authors also experiment with an aggregated formulation, estimating the deviations in departure rates per origin, rather than for a specific OD-pair. Though some accuracy is being lost, this second approach shows a promising performance. The reliability is improved by combining it with an off-line estimation, as is suggested in Ashok and Ben-Akiva (2002). (See also the discussion in Section 4.2.)

Camus et al. (1997) further develop the model by Ashok and Ben-Akiva (1993) to make forecasts of the future travel demand. Basically the idea is to propagate the flows from the origins and thereby predict future link flows, which replace the observed flows in the estimation problem.

In the algorithm proposed by Bierlaire and Crittin (2004) the state description by Ashok and Ben-Akiva is combined with a general technique for solving large-scale least-square problems. This iterative algorithm is shown to be more efficient than the Kalman filter approach suggested by Ashok and Ben-Akiva.

The approach proposed by Li and de Moor (2002) explores the effects of incomplete link flow observations. As objective they use a recursive generalized least-square estimator. The algorithm is comparable to that of Sherali et al. (1997).

Beside the well-defined models described above, different types of genetic algorithms and neural networks are widely implemented for handling various transportation problems in practice, for instance parameter estimation, traffic forecasting, traffic pattern analysis, and traffic control. Neural network estimators are given by, for example, Vythoulkas (1993), Dougherty (1995), Dougherty and Cobbett (1997), Kirby

et al. (1997), Barceló and Casas (1999), Mozolin et al. (2000), Jiao and Lu (2005), and Balakrishna et al. (2006). These methods are easy to implement and use. Only considered as black box tools, however, a fair comparison to the above-described methods, which rely on a traffic assignment model, cannot be performed. A drawback of the neural network approaches is their computational expense. Practical experiences from the Netherlands (Dougherty and Cobbett, 1997) have shown that though a neural network, which makes use of all available input information, can produce well-fitted predictions, such a strategy is not viable for real-time use.

5   The thesis

This thesis is devoted to the problem of estimating reliable OD-matrices from link flow information, provided by traffic counters. It basically covers three aspects of the OD-matrix estimation problem: the time-independent case, the time-dependent case, and the problem to collect the link flow observations such that the reliability of the estimated OD-matrix can be ensured.

5.1   Motivation

For most analyses in the field of planning and control of traffic there is a need of an accurate estimate of the underlying travel demand. Forecasts on queues, travel times, emissions, and accessibility, which in turn are used to predict the need of new-building, re-building, regulation, and information in the traffic, all rely on estimates of the travel demand. Especially the interest for time-dependent models grows for several reasons: traffic signal control, dynamic speed limits, incident detection and management of road tolls, the supply of information for different kind of route guidance systems, etc. Link flows, which are observed on some of the links in the modeled network, constitute an important data source for the estimation of OD-matrices. Compared to other available information sources, observed link flows supply cheap, reliable and easily updated information on the travel demand.

The time-independent OD-matrix estimation problem is well-known and has been studied in the literature for the last 25 years, at least. Due to congestion and the assumption on an equilibrium assignment, the estimation problem is in general hard to solve and until today only simplified solution procedures are available in commercial software. In particular, the latest advances in sensitivity analysis for traffic equilibria are not fully utilized by the currently available OD-matrix estimation procedures.

When a time-dimension is introduced into the OD-matrix estimation problem, not only the amount of travel demand, but also the departure time of every trip, has to be estimated. The development of accurate methods for OD-matrix estimation is limited by the absence of well-established and theoretically founded methods for time-dependent traffic assignment. Until today only assignment methods based on

simulation are applicable for large-scale networks, and it is a challange to adopt consistent OD-matrix estimation procedures.

In the OD-matrix estimation problem, the observed link flows are the only factors that induce a change of the target OD-matrix. It is therefore important to measure link flows which cover the travel demand as well as possible. Link flow detectors are traditionally allocated manually to the network and their placement is often motivated by other reasons than to ensure the best possible quality of the estimated OD-matrix. The problem of allocating link flow detectors to the network with the pronounced purpose to estimate OD-matrices accurately, is surprisingly little studied in spite of its prfound importance for the quality of the estimated OD-matrices.

5.2   Contribution

The thesis gives the following contributions to the research on OD-matrix estimation from link flow observations:

- An accurate algorithm for solving the bilevel formulation of the time-independent OD-matrix estimation problem. The algorithm can be easily implemented, and performs well for large-scale networks.

- An efficient algorithm for computing a second order approximation for how the link flows depend on the OD-matrix. The algorithm utilizes the latest advances in sensitivity analysis of time-independent traffic equilibria.

- Pre-adjustment schemes for improving the quality of a time-dependent OD-matrix, which is subsequently estimated with conventional techniques.

- A case study, in which these schemes are successfully applied to a network modeling parts of Gothenburg, Sweden.

- A heuristic for solving the time-dependent OD-matrix estimation problem, which is one of the most accurate algorithms being proposed for the generic problem formulation, without any special requirements on the network structure, assignment procedure, etc.

- Important insights in the complexity of the time-dependent OD-matrix estimation problem, with implications on the importance of reliable input data.

- A survey of models and methods proposed for the problem of allocating link flow detectors with the aim to estimate OD-matrices.

- An experimental environment where different strategies for link flow detector allocation can be analyzed and evaluated.

- An empirical evaluation of some of the established detector allocation strategies, with respect to the quality of the estimated OD-matrix.

- Outline of a new method for the detector placement problem, where impact to the estimated OD-matrix is emphasized.

5.3   Methodology

Methodologically, traffic modeling combines different academic areas. Traffic modeling methods utilize concepts from statistics, economics, automatic control, computer science, and civil engineering. The methods applied in this thesis mainly comes from the mathematical disciplines, and especially from the area of mathematical modeling and optimization.

5.4   Appended papers

Five papers are annoted to the thesis. The author of the thesis has contributed to the papers by a major involvement in the development and implementation of the solution methods, in the writing process, and in the analyses of the results.

Paper I: A Heuristic for the Bilevel Origin–Destination Matrix Estimation Problem

Co-authored with Jan T Lundgren.

This paper considers the time-independent case. The estimation problem is given a nonlinear bilevel formulation, where the lower level problem is to assign a given OD-matrix onto the network according to the user equilibrium principle. The problem is reformulated into a single-level problem, where the objective function contains link flows that are implicitly given by the assignment of the OD-matrix at hand. A descent heuristic, which is an adaptation of the well-known projected gradient method, is proposed as solution procedure.

When computing a search direction, the difficulty lies in the calculatation of the Jacobian matrix, which represents the derivatives of the link flows with respect to a change in the OD-flows. We do this by solving a set of quadratic programs with linear constraints. If the objective function is differentiable at the current point, the Jacobian is uniquely determined and we obtain a gradient direction. Also if differentiability does not hold, the returned direction can be used heuristically for computing a good search direction. Numerical experiments, with both some well-known test networks and a larger network from Stockholm are presented. The results indicate that the solution approach is efficient for medium to large sized networks.

The content of Paper I has been presented at:

- The International Conference on Operation Research, Duisburg, Germany,

September 3–5, 2001.

- The 9th Meeting of the EURO Working Group on Transportation, Bari, Italy, June 10–13, 2002.

An earlier version of Paper I is included in the licentiate thesis (Peterson, 2003).

Paper I has been revised for probable publication in *Transportation Research B*.

Paper II: Methods for Pre-Adjusting Time-Dependent Origin–Destination Matrices — an Application to Gothenburg

Co-authored with Jan T Lundgren and Stellan Tengroth.

We here consider an application of OD-matrix estimation from Gothenburg, in which the initial time-dependent OD-matrix is compounded from information about the daily travel demand and its time distribution for each trip purpose. With conventional estimation techniques this detailed information is aggregated, before estimation with respect to the link flow observations is performed. The idea presented in the paper is to use pre-adjustment schemes, that utilize the structure of the OD-matrix, and thereby obtain a better initial target matrix for the subsequent estimation procedure, which is performed with a traditional methodology. Three schemes are proposed for pre-adjusting: the overall travel demand for each trip purpose, the aggregated time-distribution for the total travel demand, and the time distribution for each trip purpose. The schemes are based upon assumptions about how the observed link flows are distributed over different trip purposes and how the average travel time from the origins to the observation links can be approximated.

Numerical tests are presented for both a small test network and for the Gothenburg network, where more than one million OD-matrix elements are estimated. The results show that a significantly better agreement to the observed link flows is obtained by using the pre-adjustment schemes. We believe that the developed schemes, possibly with small modifications, are applicable to many other traffic models with similar characteristics.

The content of Paper II has been presented at:

- Transportforum, Linköping, Sweden, January 14–15, 2004.

A previous version of Paper II is published as:

- "Methods for pre-adjusting time-dependent origin–destination matrices", in proceedings of the *10th World Congress and Exhibition on Intelligent Transport Systems and Services*, Madrid, Spain, November 16–20, 2003, Paper No. 2436.

Paper II is included in the licentiate thesis (Peterson, 2003).

Paper III: A Heuristic for the Estimation of Time-Dependent Origin–Destination Matrices from Traffic Counts

Co-authored with Jan T Lundgren and Clas Rydergren.

This paper considers a generic formulation of the time-dependent OD-matrix estimation problem. Special attention is given to the assignment map, which describes how the travel demand is transferred onto route and link flows in the network, and its relationship to the level of congestion. We decompose the assignment map in two parts, handling the route choice (traffic assignment) and the traffic flow propagation (network loading), respectively, and discuss how these components are affected by an adjustment of the travel demand, and why the effects are more difficult to predict when some links are oversaturated in one or more time-periods. We suggest an algorithm which is an extension to previously proposed methods for the time-independent OD-matrix estimation problem, where the changes of the assignment map are approximated by a difference quotient between two assignment maps in the search direction.

The algorithm is implemented together with a mesoscopic tool for dynamic traffic assignment and verified for a small test network, Brunnsviken, in the northern part of Stockholm. By numerical experiments the importance of the parameter settings are illustrated.

The content of Paper III has been presented at:

- Nordic MPS '04, Norrköping, Sweden, October 21–23, 2004.
- The 10th Jubilee Meeting of the EURO Working Group on Transportation, Poznan, Poland, September 13–16, 2005.
- Transportforum, Linköping, Sweden, January 11–12, 2006.

An earlier version of Paper III was published in:

- *Advanced OR and AI Methods in Transportation, Proceedings of the 10th Jubilee Meeting of the EURO Working Group on Transportation, Poznan, Poland, 13–16 September 2005*, A. Jaszkiewicz et al. (eds.), Publishing House of Poznan University of Technology, Poznan, Poland, pp 242–246.

Paper IV: Allocation of Link Flow Detectors for Origin–Destination Matrix Estimation — a Comparative Study

Co-authored with Torbjörn Larsson and Jan T Lundgren.

This paper is devoted to the problem of allocating link flow detectors, with the aim to ensure the best possible quality of the OD-matrix to be estimated. To meet this goal, the "coverage" of the traffic should be maxmized. Coverage of traffic can be defined in different ways, and especially the coverage of OD-pairs and travel demand (weighted OD-pairs) are frequently used. We develop an experimental framework, where the effect on the estimated OD-matrix induced by different allocation strategies can be evaluated. This framework is based on the assumption that a synthetic "true" OD-matrix is available, and hence, that the deviation from the estimated OD-matrix can be measured. The framework enables studies not only on how successful the allocation strategies are, but also on how differences in the implementations affect the result. As a side-effect we can also study the importance of how the assignment map is chosen.

Beside the literature survey and an experimental framework, which can be used for further similar studies, Paper IV also contains an empirical study with three traffic networks. The results indicate that it is crucial to consider the travel demand, and not only the number of OD-pairs, when the covering the traffic. The choice of assignment map seems to be of less importance for the quality of the estimated OD-matrix.

Paper IV is under revision for possible publication in *Transportation Research B*.

Paper V: A Novel Model for Placement of Detectors for Origin–Destination Matrix Estimation

Co-authored with Torbjörn Larsson and Jan T Lundgren.

Paper V is based on the assumption that the most correct OD-matrix is estimated if consistent and complete link flow observations are available. Non-complete information is interpreted as a relaxation of this requirement. From this assumption, the problem to place link flow detectors, as to maximize the quality of the estimated OD-matrix, is expressed in terms of discarding those links, with the least importance for the estimation problem. We state the OD-matrix estimation problem by using a Lagrange-dual formulation, and show that the non-availability of a link flow observation is equivalent to fixing a dual variable to zero. This observation leads to an interesting max–min formulation of the detector placement problem. Two related measures, expressing how well the OD-matrix to be estimated is determined by the choice of detector links, are also discussed.

The content of Paper V is theoretical and its purpose is to analyze the relationship between OD-matrix estimation and link flow detector allocation.

5.5 Future research

The algorithms which are presented in the appended papers are all implementable for real-life applications, and may be included in commercial software for OD-matrix estimation. Some adaptations might be required and the computational efficiency can be improved.

The rapid development of applications of time-dependent traffic models, all of which require accurately estimated OD-matrices, is a challenge for the future. A lot of work remains to establish and validate procedures for time-dependent traffic assignment, for which sensitivity analyses can be derived and applied to improve a time-dependent OD-matrix estimation procedures.

Throughout the work in this thesis, link flow observations are assumed to be the only information collected for describing the traffic situation, which should be expressed in the OD-matrix to be estimated. Today, however, also observations of travel times and/or speeds are available. For a majority of the time-independent models this extra information is redundant, since reliable analytical relations exist between all three quantities. For a time-dependent model, however, it might be interesting to analyze how this information can be used to improve the quality of the estimation.

The collection of information about the current traffic situation is an interesting research area in itself. A natural continuation of the research presented in Paper IV and V is the development of new methods for allocating link flow detectors to the network, with the purpose to ensure the best possible quality of the OD-matrix to be estimated.

References

Abrahamsson, T. (1998) "Estimation of origin–destination matrices using traffic counts — a literature survey", Interim Report IR-98-021, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Ashok, K., and M.E. Ben-Akiva (1993) 'Dynamic origin–destination matrix estimation and prediction for real-time traffic management systems", in: *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 465–484.

Ashok, K., and M.E. Ben-Akiva (2000) "Alternative approaches for real-time estimation and prediction of time-dependent origin–destination flows", *Transportation Science* 34, pp 21–36.

Ashok, K., and M.E. Ben-Akiva (2002) "Estimation and prediction of time-dependent origin–destination flows with a stochastic mapping to path flows and link flows",

Transportation Science 36, pp 184–198.

Balakrishna, R. (2006) "Off-line calibration of dynamic traffic assignment models", PhD thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Boston, Massachusetts.

Balakrishna, R., Ben-Akiva, M., and H.N. Koutsopoulos (2006) "Time-dependent origin–destination estimation without assignment matrices", in: *Methodological Issues in Transport Simulation*, Proceedings of the International Symposium of Transport Simulation (ISTS06), September 4th–6th, 2006, Lausanne, Switzerland, Paper 0059 .

Barceló, J. (1997) "A survey of some mathematical programming models in transportation", *Sociedad de Estadística e Investigación Operativa Top* 5, pp 1–40.

Barceló, J., and J. Casas (1999) "The use of neural networks for short-term prediction of traffic demand", in: *Transportation and Traffic Theory*, Proceedings of the 14th International Symposium on Transportation and Traffic Theory, Jerusalem, Israel, July 20–23, 1999, A. Ceder (ed.), Pergamon, pp 419–443.

Bard, J.F. (1998) *Practical Bilevel Optimization — Algorithms and Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Bates, J. (2000) "History of demand modelling", in: *Handbook of Transport Modelling*, D.A. Hensher and K.J. Button (eds.), Elsevier, Oxford, England, pp 11–33.

Beckmann, M., McGuire, C.B., and C.B. Winsten (1956) *Studies in the Economics of Transportation*, Yale University Press, New Haven, Connecticut.

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

Bell, M.G.H. (1984) "Log-linear models for the estimation of origin–destination matrices from traffic counts: an approximation", in: *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*, J. Volmüller and R. Hamerslag (eds.), VNU Science Press, Utrecht, The Netherlands, pp 451–470.

Bell, M.G.H. (1991) "The real time estimation of origin–destination flows in the presence of platoon despersion", *Transportation Research B* 25, pp 115–125.

Bell, M.G.H., and Y. Iida (1997) *Transportation Network Analysis*, Wiley, West Sussex, United Kingdom.

Bell, M., Inaudi, D., Lange, J., and M. Maher (1991) "Techniques for the dynamic estimation of O–D matrices in traffic networks", in: *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. (1987) "Methods to combine different data sources and estimate

origin–destination matrices", in: *Transportation and Traffic Theory*, N.H. Gartner and N.H.M. Wilson (eds.), Elsevier, pp 459–481.

Ben-Akiva, M., and S.R. Lerman (1985) *Discrete Choice Analysis — Theory and Application to Travel Demand*, MIT Press, Cambridge, Massachusetts.

Bianco, L., Confessore, G., and P. Reverberi (2001) "A network based model for traffic sensor location with implications on o/d matrix estimates", *Transportation Science* 35, pp 50–60.

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

Bierlaire, M., and F. Crittin (2004) "An efficient algorithm for real-time estimation and prediction of dynamic od tables", *Operations Research* 52, pp 116–127.

Bierlaire, M., and P.L. Toint (1995) "Meuse: an origin–destination matrix estimator that exploits structre", *Transportation Research B* 29, pp 47–60.

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

Camus, R., Cantarella, G.E., and D. Inaudi (1997) "Real-time estimation and prediction of origin–destination matrices per time slice", *International Journal of Forecasting* 13, pp 13–19.

Carey, M., Hendrickson, C., and K. Siddharthan (1981) "A method for direct estimation of origin/destination trip matrices", *Transportation Science* 15, pp 32–49.

Cascetta, E. (1984) "Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator", *Transportation Research* 18, pp 289–299.

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

Cascetta, E., and M. Postorino (2001) "Fixed point approaches to the estimation of o/d matrices using traffic counts on congested networks", *Transportation Science* 35, pp 134–147.

Chang, G.L., and X. Tao (1996) "Estimations of dynamic O-D distributions for urban networks", in: *Transportation and Traffic Theory*, Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon, France, 1996, J.B. Lesort (ed.), Pergamon, pp 1–20.

Chang, G.L., and J. Wu (1994) "Recursive estimation of time-varying origin–destination flows from traffic counts in freeway corridors", *Transportation Research B* 28, pp 141–160.

Chen, Y. (1994) "Bilevel programming problems: Analysis, algorithms and applications", PhD thesis, Publication No. 984, Centre de Recherche sur les Transports, Université de Montréal, Montreal, Canada.

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

Codina, E., and J. Barceló (2004) "Adjustment of O–D trip matrices from observed volumes: an algorithmic approach based on conjugate directions", *European Journal of Operational Research* 155, pp 535–557.

Contram (2002) "CONTRAM8: User Manual for Contram 8 Version 1", Mott MacDonald Limited and TRL Limited, Doc Ref 26661/T&P/WIN/02A, Winchester, United Kingdom.

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

Daganzo, C.F., and Y. Sheffi (1977) "On stochastic models of traffic assignment", *Transportation Science* 11, pp 253–274.

Davis, G.A. (1993) "A statistical theory for estimation of origin–destination parameters from time-series of traffic counts", in: *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., and N.L. Nihan (1991) "Stochastic process approach to the estimation of origin–destination parameters from time series of traffic counts", *Transportation Research Record* 1328, pp 36–42.

Doblas, J., and F.G. Benitez (2005) "An approach to estimating and updating origin–destination matrices based upon traffic counts preserving the prior structure of a survey matrix", *Transportation Research B* 39, pp 565–591.

Domencich, T., and D. McFadden (1975) *Urband Travel Demand: A Behavioral Analysis*, North Holland, Amsterdam, The Netherlands.

Dougherty, M. (1995) "A review of neural netowrks applied to transport", *Transportation Research C* 3, pp 247–260.

Dougherty, M.S., and M.R. Cobbett (1997) "Short-time inter-urban traffic forecasting using neural networks", *International Journal of Forecasting* 13, pp 21–31.

Dynasmart (2002) "DYNASMART-P User's guide", Version 0.930.0, Maryland Transportation Initiative, University of Maryland, College Park, Maryland.

Emme/2 (1999) "EMME/2 User's manual", Software Release 9, Inro Consultants Inc., Montreal, Canada.

Erlander, S., Nguyen, S., and N. Stewart (1979) "On the calibration of the combined distribution-assignment model", *Transportation Research B* 13, pp 259–267.

Erlander, S., and N.F. Stewart (1990) *The Gravity Model in Transportation Analysis — Theory and Extentions*, VSP, Utrecht, The Netherlands.

McFadden, D. (1973) "Conditional logit analysis of qualitative choice behavior", in: *Frontiers in econometrics*, P. Zarembka (ed.), Academic Press, pp 105–142.

Fisk, C.S. (1988) "On combining maximum entropy trip matrix estimation with user optimal assignment", *International Transactions in Operational Research* 2, pp 165–179.

Fisk, C.S., and D.E. Boyce (1983) "A note on trip matrix estimation from link traffic count data", *Transportation Research B* 17, pp 245–250.

Florian, M., and Y. Chen (1995) "A coordinate descent method for the bi-level O–D matrix adjustment problem", *Transportation Research B* 22, pp 69–73.

Frank, M., and P. Wolfe (1956) "An algorithm for quadratic programming", *Naval Research Logistics Quarterly* 3, pp 95–110.

Garber, N.J., and L.A. Hoel (1999) *Traffic and Highway Engineering*, 2nd edition, Brooks/Cole Publishing Company, Pacific Grove, California:

Gur, Y., Turnquist, M., Schneiderm, M., and L. Leblanc (1980) "Estimation of an origin–destination trip table based on observed link volumes and turning movements", Vol 1. Report RD-80/034, FHWA, U.S. Department of Transportation, Washington, District of Colombia.

Han, S. (2003) "Dynamic traffic modelling and dynamic stochastic user equilibrium assignment for general road networks", *Transportation Research B* 37, pp 225–249.

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

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

Jiao, P., and H. Lu (2005) "Dynamic origin–destination flows estimation for freeway corridors using genetic algorithm", *Journal of the Eastern Asia Society for Transportation Studies* 6, pp 2702–2717.

Jörnsten, K., and S. Nguyen (1979) "On the estimation of a trip matrix from network data", Technical report LiTH-MAT-R-79-36, Department of Mathematics, Linköping University, Linköping, Sweden.

Khisty, C.J, and B.K. Lall (1998) *Transportation Engineering — an Introduction*, 2nd edition, Prentice Hall, Upper Saddle River, New Jersey.

Kirby, H.R., Watson, S.M., and M.S. Dougherty (1997) "Should we use neural networks or statistical models for short-term motorway traffic forecasting?", *International Journal of Forecasting* 13, pp 43–50.

Li, J., and B. de Moor (2002) "Dynamic identification of origin–destination matrices in the presence of incomplete observations", *Transportation Research B* 36, pp 37–57.

Lind, G. (1997) "Strategic assessment of intelligent transport systems", PhD thesis, TRITA-IP FR 97-29, Department of Infrastructure and Planning, Royal Institute of Technology, Stockholm, Sweden.

Lindveld, K. (2003) "Dynamic O–D matrix estimation: a behavioural apporach", PhD thesis, Trail Thesis Series, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands.

Lundgren J.T. (1989) *Optimization approaches to travel demand modelling*, PhD thesis, Linköping Studies in Science and Technology, Dissertations, No. 207, Department of Mathematics, Linköping University, Linköping, Sweden.

Maher, M.J., and X. Zhang (1999) "Algorithms for the solution of the congested trip matrix estimation problem", in: *Transportation and Traffic Theory*, Proceedings of the 14th International Symposium on Transportation and Traffic Theory, Jerusalem, Israel, July 20–23, 1999, A. Ceder (ed.), Pergamon, pp 445–469.

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

Merchant, D.K., and G.L. Nemhauser (1978a) "A model and an algorithm for the dynamic traffic assignment problems", *Transportation Science* 12, pp 183–199.

Merchant, D.K. and G.L. Nemhauser (1978b) "Optimality conditions for a dynamic traffic assignment model", *Transportation Science* 12, pp 200–207.

Mozolin, M., Thill, J.C., and E.L. Usery (2000) "Trip distribution forecasting with multilayer perceptron neural networks: a critical evaluation", *Transportation Research B* 34, pp 53–73.

McNeil, S., and C. Hendrickson (1985) "A regression formulation of the matrix estimation problem", *Transportation Science* 19, pp 278–292.

Nguyen, S. (1977) "Estimating an OD matrix from network data: A network equilibrium apporach", Publication No. 60, Centre de Recherche sur les Transports, Université de Montréal, Montreal, Canada.

Nie, Y., Zhang, H.M., and W.W. Recker (2005) "Inferring origin–destination trip matrices with a decoupled GLS path flow estimator", *Transportation Research B* 39, pp 497–518.

Nihan, N.L, and G.A. Davis (1987) "Recursive estimation of origin–destination matrices from input/output counts", *Transportation Research B* 21, pp 149–163.

Nihan, N.L, and G.A. Davis (1989) "Application of prediction-error minimization and maximum likelihood to estimate intersection O–D matrices from traffic counts", *Transportation Science* 22, pp 77–90.

Oppenheim, N. (1995) *Urban Travel Modelling — From Individual Choices to General Equilibrium*, John Wiley, New York, New York.

Ortúzar, J. de D., and L.G. Willumsen (2001) *Modelling Transport*, 3rd Edition, Wiley, West Sussex, United Kingdom.

Papacostas, C.S., and P.D. Prevedouros (2001) *Transportation Engineering and Planning*, 3rd edition, Prentice Hall, Upper Saddle River, New Jersey.

Patriksson, M. (1994) *The Traffic Assignment Problem — Models and Methods*, VSP, Utrecht, The Netherlands.

Patriksson, M. (2004) "Sensitivity analysis of traffic equilibria", *Transportation Science* 38, pp 258–281.

Peeta, S., and A.K. Ziliaskopoulos (2001) "Foundations of dynamic traffic assignment: the past, the present and the future", *Networks and Spatial Economics* 1, pp 233–265.

Peterson, A. (2003) "Origin–destination matrix estimation from traffic counts", licentiate thesis, Linköpings Studies in Science and Technology. Theses, No. 1057, Department of Science and Technology, Linköpings universitet, Norrköping, Sweden.

Sheffi, Y. (1985) *Urban Transportation Networks — Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs, New Jersey.

Sherali, H.D., Arora, N., and A.G. Hobeika (1997) "Parameter optimization methods for estimating dynamic origin–destination trip-tables", *Transportation Research B* 31B, pp 141–157.

Sherali, H.D. and T. Park (2001) "Estimation of dynamic origin–destination trip tables for a general network", *Transportation Research B* 35, pp 217–235.

Sherali, H.D., Narayanan, A., and R. Sivanandan (2003) "Estimation of origin–destination trip-tables based on a partial set of traffic link volumes", *Transportation Research B* 37, pp 815–836.

Sherali, H.D., Sivanandan, R., and A.G. Hobeika (1994) "A linear programming apporach for synthesizing origin–destination trip tables from link traffic volumes", *Transportation Research B* 28, pp 213–233.

Sjöstrand, H. (2001) *Passenger assessments of quality in local public transport —*

measurement, variability and planning implications, PhD Thesis, Bulletin 202, Department of Technology and Society, Lund Institute of Technology, Lund, Sweden.

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

Spiess, H. (1990) "A gradient approach for the O–D matrix adjustment problem", CRT Pub. No 693, Centre de Recherche sur les Transports, Universit de Montral, Montreal, Canada.

Statistics Sweden (2007), `http://www.scb.se/`, download 2007-02-05

Tamin, O.Z., and L.G. Willumsen (1989) "Transport demand model estimation from traffic counts", *Transportation* 16, pp 3–26.

Tavana, H. (2001) "Internally-consistent estimation of dynamic network origin–destination flows from intelligent transportation systems data using bi-level optimization", PhD thesis, Department of Civil Engineering, University of Texas at Austin, Austin, Texas.

Tobin, R.L., and T.L. Friesz (1988) "Sensitivity analysis for equilibrium network flow", *Transportation Science* 22, pp 242–250.

Vythoulkas, P.C. (1993) "Alternative approaches to short term traffic forecasting for use in driver information systems", in: *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 485–506.

Wardrop, J.G. (1952) "Some theoretical aspects of road traffic research", *Proceedings of the Institute of Civil Engineers*, Part II, Road Paper No. 36, pp 325–378.

Willumsen, L.G. (1981) "Simplified transport models based on traffic counts", *Transportation* 10, pp 257–278.

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

Wilson, A.G. (1974) *Urban and Regional Models in Geography and Planning*, John Wiley and Sons, Bristol, United Kingdom.

Wolfe, P. (1975) "A method of conjugate subgradients for minimizing nondifferentiable optimization", *Mathematical Programming Study* 3, pp 145–173.

Wright, P.H., and N.J. Ashford (1998) *Transportation Engineering — Planning and Design*, 4th edition, John Wiley and Sons, New York, New York.

Wu, J. (1997) "A real-time origin–destination matrix updating algorithm for on-line

applications", *Transportation Research B* 31, pp 381–396.

Yang, H. (1995) "Heuristic algorithms for the bilevel origin–destination matrix estimation problem", *Transportation Research B* 29, pp 231–242.

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

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

Yang, H., and S. Yagar (1995) "Traffic assignment and signal control in saturated road networks", *Transportation Research A* 29, pp 125–139.

Zhou, X., and H.S. Mahmassani (2006) "Dynamic origin–destination demand estimation using automatic vehicle identification data", *IEE Transactions on Intelligent Transportation Systems* 7, pp 105–114.

van der Zijpp, N.J. (1997) "Dynamic origin–destination matrix estimation from traffic counts and automated vehicle identification data", *Transportation Research Record* 1607, pp 87–94.

van der Zijpp, N.J., and R. Hamerslag (1996) "Improved Kalman filtering approach for estimating origin–destination matrices for freeway corridors", *Transportation Research Record* 1443, pp 54–64.

van Zuylen, H. (1978) "The information minimising method: validity and applicability to transportation planning", in: *New Developments in Modelling Travel Demand and Urban Systems*, G.R.H. Jansen et al. (eds.), Saxon, Farnborough, United Kingdom.

van Zuylen, H.J., and D.M. Branston (1982) "Consistent link flow estimation from counts", *Transportation Research B* 16, pp 473–476.

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