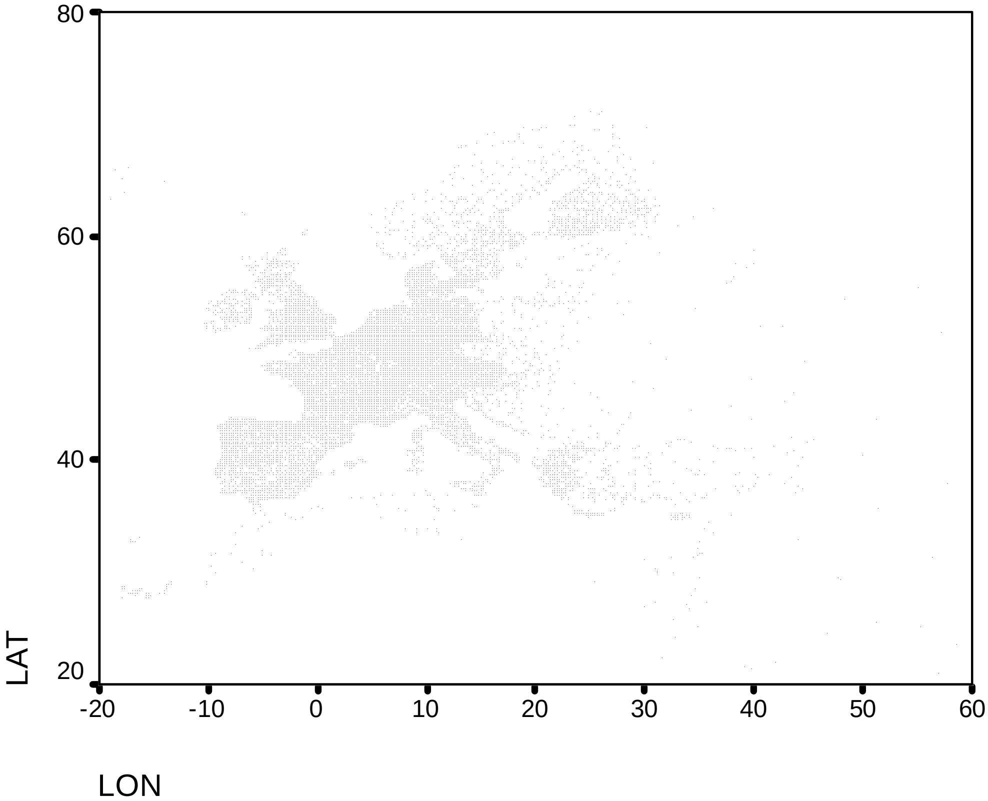


Discrete Choice Models for Long-Distance Travel based on the DATELINE Survey

Jeremy Hackney, IVT

Conference paper STRC 2004

STRC | 4th Swiss Transport Research Conference
Monte Verità / Ascona, March 25-26, 2004

Discrete Choice Models for Long-Distance Travel based on the DATELINE Survey

Jeremy Hackney
IVT
ETH
CH-8093 Zürich

Phone: +41-1-633 32 25
Fax: +41-1-633 10 57
eMail: hackney@ivt.baug.ethz

August 2005

Abstract

The influences on the route and mode choices of long-distance travellers on trans-European routes are modelled based on DATELINE, the most recent, comprehensive, and coordinated survey of long-distance travel in Europe. The results will be reported at an aggregated level to support decision making in the European Union. For the mode choice model, DATELINE's revealed preferences will be merged with the non-chosen routes/modes that were available to the travellers. These are represented in network models for rail, air, and road. The lowest-cost routes for the transportation alternatives will be calculated on these modal networks. Using this choice-set of modes and routes, along with information about trip, traveller, and mode characteristics, a discrete choice model will simulate the route/mode choices. The models will be weighted with socioeconomic data for extrapolation to reflect annual behavior at the scale of the respective NUTS populations. DATELINE is a survey of Europeans about their long-distance travel. It was carried out in the 15 EU countries and Switzerland from October, 2001 through October, 2002. DATELINE yields 55,000 households, 132,000 trips, and 101,000 unique OD pairs. The available network models do not have the resolution to calculate the flows on links connecting these points. The shortest routes will be calculated on the network models through a combination of regional aggregation to NUTS 5 zones, and refinement of the network models' connections. The benefits of combining transportation modes from separate networks into one network model will be weighed against increased model size and complexity and difficulty intrepreting the results of the shortest path calculations. Policy-relevant results will require a higher level of spatial aggregation using the NUTS coding scheme. The SABE dataset will be used to relocate and weight origins and destinations in the appropriate NUTS region. Cost and schedule data for commercial air, ferry, and road tolls are being obtained through ETIS-Base. Tour buses are a major long-distance mode for leisure travel in the DATELINE dataset. Characterization of the attributes of this mode will be more difficult. Though all train stations are present in the rail network model, their schedules are not all represented. The nearest major and minor train stations to each DATELINE response point will be calculated and their schedules completed. There are two ways to proceed in modeling mode choice. A classical model would establish a choice alternative for each major mode. Recent research into trip chaining performed in the INVERMO long-distance travel project (Last and Manz 2003) suggests better results could be achieved in a hierarchical approach, in which choice sets are available to each decision maker according to that person's trip and personal information. The presentation at STRC describes progress in the shortest path calculations and the formulation of the discrete choice model.

Keywords

Discrete Choice Models, Long-Distance Travel, DATELINE Survey

Table of Contents

Overview

A NUTS2 OD matrix for long-distance travel in Europe, using a destination and mode choice model based on the results of the DATELINE survey (Socialdata 2001) will be derived to correct the synthetically-derived OD matrix of the ETIS Base project of the European Union. The resulting corrected OD matrix in ETIS Base will be reported by mode and by trip purpose (work, non-work). This paper describes the method of this correction, focusing on the mode choice modelling that is necessary for extrapolating the DATELINE survey results to trip rates.

The paper describes the method in the following steps:

1. The description of ETIS Base and the context of the work in the ETIS Base project
2. The DATELINE survey and results
3. The modes in the model and their generalized cost functions
4. Destination and mode choice model
5. Assignment of DATELINE constrained by mode choice parameters
6. Calibration of DATELINE assignment matrix with counts
7. Fusion of the synthetic ETIS matrix with the DATELINE-derived matrix
8. Critical evaluation of the method

1 The description of ETIS Base and the context of the work in the ETIS Base project

The full title of ETIS Base is ‘Core Database Development for the European Transport policy Information System (ETIS)’. As the name implies, the project provides a dataset to support Pan-European transportation decision making. Users of the system will be able to download transportation policy indicators via a browser. The indicators will be calculated in the ETIS system from the core database of transportation , geographic, and socioeconomic parameters.

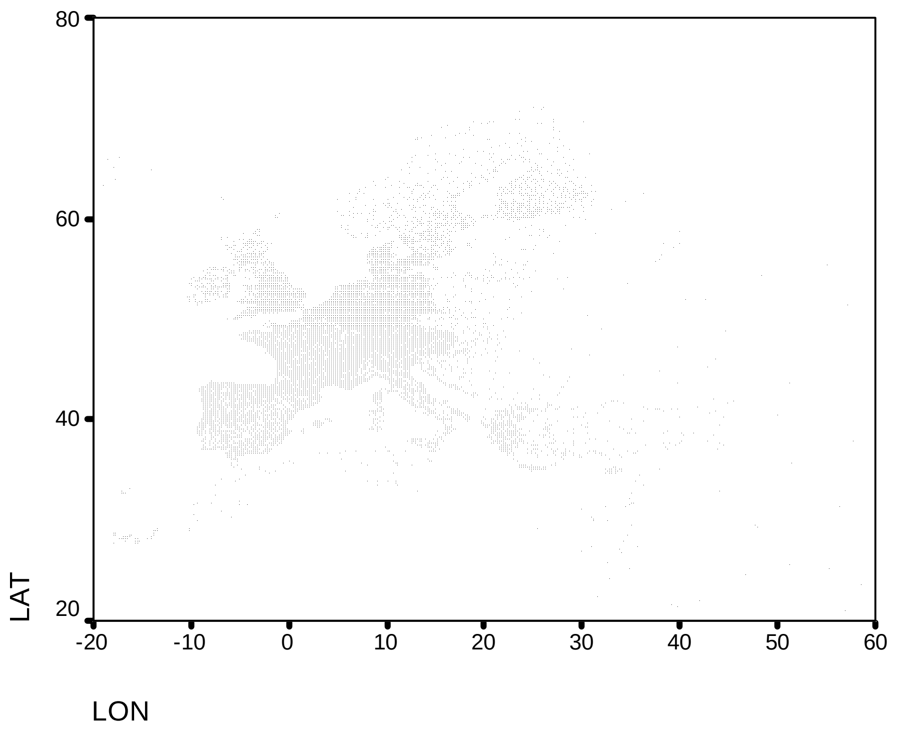
This system consists of three components: ETIS Base, which compiles the internally consistent dataset and calculates some complex policy indicators; ETIS Agent, which provides a web-based data access tool that also calculates indices to support political decision making, and ETIS Link, which promotes and integrates the project in EU decision making. At the moment, the focus in the project is to derive and document reproducible data procurement methods and index calculations, such that the database and data access system can be sustained indefinitely. Part of this process involves the identification of weak and lacking data. A perfect dataset or perfect data models are therefore not the goal of the project, rather one seeks a certain understanding of the limitations of the data and methods for making informed decisions despite these limits.

The geographical scope of the project includes the extended European Union (EU25) and the EFTA countries, Switzerland and Norway. To anticipate border interactions of the EU in the near future, a Pan European scope will establish the boundary conditions of the matrix and include data and models for Rumania, Bulgaria, Turkey, Bosnia & Herzegovina, Croatia, Serbia & Montenegro, Albania, FYRO Macedonia as well. For air transport, additional information and models will be provided or estimated on the continental level for Africa, America, Asia and Oceania. The resolution of the data and any models that are necessary for the policy indices are based on the Nomenclature of Territorial Units for Statistics (NUTS) of the European Union. The passenger OD matrix for the EU 25 and EFTA countries will be delivered at NUTS2 level (211 zones in the EU 15, corresponds to Swiss Kantonen) but calculated at NUTS3 level (1093 zones in the EU 15, corresponds to Swiss Bezirke). The OD matrix for other areas will be at NUTS0 (country) level or higher aggregation.

A critical element of the undertaking is the internal consistency of data, whose quality varies widely across the countries in the scope of the project. Characteristics of data such as resolution, reference year, sampling, modelling, etc. must be documented and corrected, such that the derived policy indices use compatible measures and are meaningful. A reference year of 2000 was chosen because a majority of the existing datasets and models had already been constructed for this year. If necessary, data from a year other than 2000 which is suspected to have significantly different values than a sample from 2000 may have had, will have to be adjusted to match other data from 2000.

The most important element of the policy indices for passenger transportation is the OD Matrix of the number of long-distance trips made per person per year between NUTS2 regional centroids. Policy decisions regarding links, such as investments on the TEN (Trans European Network) depend on assignment results based on a correct OD matrix. Long-distance trips (>100km) make up fewer than 1.5% of all trips (Last, Manz and Zumkeller 2003, Germany is the reference), and few countries have this data for their residents. As a result, a partner in the ETIS project will derive a synthetic OD matrix based on the attractiveness of destinations for business and non-business long-distance travel. In parallel, the DATELINE survey results will be enriched and used to extrapolate a calibrated OD matrix based on the revealed preferences of EU residents. The calibration will be done on a network model with a constraint given by actual link traffic counts. Finally, the synthetic and the DATELINE-based OD matrices will complement each other via a data fusion technique.

Figure 1 Distribution of Dateline origins and destinations



2 Description of the DATELINE dataset

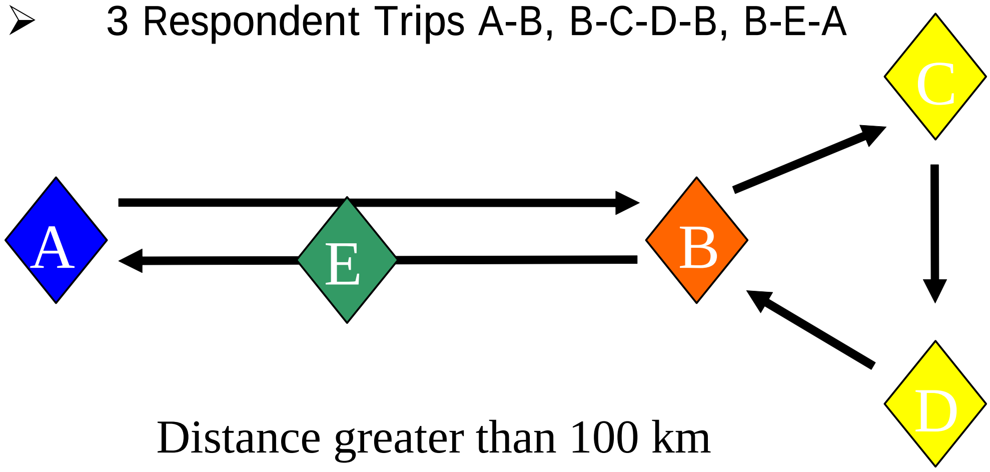
DATELINE is a household-level survey of 86,000 residents of the EU 15 and Switzerland about their long-distance travel. Individuals over 15 years of age reported travel of over 100km crow-fly distance for the purposes of “holiday” in the previous 12 months, as well as “other private” and “business” in the previous 3 months, and “commuting” for the previous 4 weeks. The survey was carried out from October, 2001 through October, 2002. Among other variables, the dataset contains travel date, destination, duration, and mode. Specific travel decisions like mode changes were recorded for travel between significant stops en route. Thus, a coarse record of route choice, in the form of stops en route, is also available, in particular for public transport based trips. The data is available to browse or to download (with permission) on the ETH Travel Data Archive (2004): http://129.132.96.89/nesstarlight/index.jsp

2.1 Trip-Chain Structure of DATELINE

DATELINE is based on trip chains (Figure 2) which form round-trip journeys or commutes, in which the final destination was over 100km distant from home. Excursions are a further form of trip chain which embark from and return to the journey destination. The end or beginning of a trip is determined by a stop of over 2 hours with a distinct purpose. To make the DATELINE data useful for corridor studies and OD Matrices, the trips have to be extracted from the journeys, excursions, and commutes to provide a list of all movements. This file of point-to-point trips provides a set of observed OD matrices by mode and journey/excursion/commute purpose. A true “trip purpose” is not recorded in DATELINE (Davidson 2003), but it can be assumed to be the same as the purpose of the trip chain. The origins and destinations of trips (also outside the EU and Switzerland) are geocoded to a point in the area of the true origin or destination, not necessarily corresponding to a NUTS centroid, but at a resolution on the order of NUTS5. The SABE database will enable the correction of this geocoding to the EU-standard NUTS centroid location.

Figure 2 Journey Concepts

- 1 Journey A-B-A
- 1 Excursion BCDB
- 6 Modelling Trips A-B & B-C, C-D, D-B, B-E, E-A
- 1 Trip Chain A-B-C-D-B-E-A
- 3 Respondent Trips A-B, B-C-D-B, B-E-A



Source: Peter Davidson Consultancy (2003)

2.2 Macroanalysis of DATELINE

A summary of the data is provided in Table 1

Table 1 Summary statistics

Item or Variable	N
Nationalities surveyed	EU 15 + CH = 16
Usable surveys	86,251
Trips from Journeys, Excursions, Commutes	144,011
Unique points (origin or destination)	20,549
Unique OD Pairs	101,591

An observed OD Matrix of the observed DATELINE trips at NUTS0 level for all modes has been reconstructed (not included in this paper). The observed mode share of the trips is in Table 2.

Table 2 Dateline: Observed modal shares

Mode	Share (%)
Private car	59.0
Tour bus/coach	8.5
Air	19.5
Ferries	1.9

3 Overview of Method for Deriving an OD Matrix from Observed DATELINE data

The mode share and the distributions of trip duration and distance compare with other surveys of long distance travel (sampling is discussed in Peter Davidson Consultancy 2003).When weighted and extrapolated to the population of each country, the dataset exhibits much lower trip rates (trips/person/year) than national long-distance travel surveys have shown (compare with national surveys in Hubert and Potier 2003, Swiss Einkommen und Verbrauchserhebung SFSO 1998, INVERMO results for Germany summarized in Axhausen and Frick 2004). The observed OD matrix therefore will not be used directly as a target for the calibration of the synthetic matrix flows. Instead, the observed DATELINE matrix will seed the estimation of a DATELINE-based flow matrix, which will form a basis for comparison with the ETIS Base OD matrix.

In a first step, the DATELINE observations will be used together with socioeconomic datasets, geographical datasets, and level of service (LOS) datasets for chosen and non-chosen transportation modes to estimate a destination/mode choice model. The model will be estimated with 4 modes (car, train, air + car, air + train/bus) for 2 trip purposes (work/non-work). After the estimation, the parameters (generalized costs) of the model will be compared with those of the synthetic matrices. This is a first check of the OD matrices' consistency.

In parallel with the enrichment with level of service variables, existing network models for the transportation modes will be improved with denser service infrastructure and higher-frequency service for the public modes to make a usable network at NUTS3 resolution.

The observed DATELINE OD matrix will then be assigned to the updated networks with the discrete choice parameters as constraints. Traffic counts on primary routes will be used to calibrate, actually to extrapolate, the OD matrix, until a threshold is reached for the improvement in the iteration. At this point, the DATELINE-derived OD matrix is ready for further use in analysis and fusion with the synthetic OD matrix from ETIS.

4 Enrichment of DATELINE

DATELINE will have to be enriched with socioeconomic data, geographic data, and level of service data for the chosen and non-chosen transportation modes. Previous work on decision variables to use in the discrete choice destination/mode model is important for the consideration of the data to use to enrich DATELINE.

4.1 Decision Variables and Form of Discrete Choice Models for Destination/Mode in Long-Distance Travel

DATELINE allows the estimation of destination/mode choice for three travel purposes at a municipal (not building or address) level. Clearly, “work” is a prime long-distance trip generator. The other trip purpose, “non-work”, will consist of the DATELINE purposes “holiday” and “other private”. The decision variables will clearly have different values for the two cases, and it could also be that certain variables are critical to one case but not relevant to another.

Primarily because of a lack of data of sufficient quality and detail, disaggregate models have not been used as much in long-distance travel as in everyday travel, and modelling progress trails correspondingly. It is well-known that long-distance journeys and the people making them are much more difficult to describe in models than everyday travel behavior due to their heterogeneity. This section describes the form of the logit destination/mode choice model and the possible relevant decision variables to include.

The choice of mode in long distance travel is not only a function of modal attributes and personal characteristics of the traveler, but is also closely tied to the characteristics of the destination and to trip purpose. A multinomial (independent destination/mode attributes) or nested logit (shared destination/mode attributes) model must be estimated on destination and mode attributes for the different trip purposes (Ben Akiva and Lermann 1985). The form of the model must take into account statistical corrections for the necessary aggregation of destinations into zones. Here, the data will be aggregated from municipalities to NUTS3, for estimation of the logit model. Ben Akiva and Lermann recommend estimating the model in one or two choice dimensions while holding the other dimensions constant (for example, choice of mode and frequency keeping the destination constant) and observing the stability of the estimated parameters. Stable parameters indicate that the form of the model is likely to be a good model when the values in all the choice dimensions are unconstrained.

In a study of the INVERMO longitudinal survey results in Germany (Last, Manz, Zumkeller 2003), it is clear that long-distance travel is more heterogeneously distributed among the population than everyday travel. Half of the population in the produced over 90% of the long-distance trips. The most mobile one percent of the population made long-distance trips ten times more frequently than the national average. This shows the importance of properly identifying the socioeconomic variables describing these groups.

Frequent trips are strongly tied to characteristics that are associated in our society with success: middle age, high education, a job with a high income. The respondents need to be divided into groups of trip purpose however, in order to describe behavior such as mode or destination choice, or to identify which socioeconomic characteristics and which mode attributes are important. 90% of the highly mobile people either travel extensively on business and little privately, or vice-versa. The latter group (those who travel a lot, but much more for leisure than for business) is four times as large as the former. Meanwhile, only 10% of highly mobile people divide their trips equally between work and non-work. This finding accents the importance of segmenting the respondents by the proportion of trip purposes that the person makes because sensitivities to supply characteristics like cost, time, and comfort will be correspondingly differentiated. Indeed, for travelers who make more work trips than non-work trips, the mode choice, which is distance-dependent, tends toward higher cost public modes with longer distance quicker than for people who travel more frequently for non-work purposes than for work purposes. The former group is averse to using more time than necessary in travel, and the latter prefers to save money.

The number of modes used by an individual in the year can also be used to classify the person as a “multi-“ or “mono-“modal personality, to help quantify the person’s awareness of other modes. In INVERMO, 60% of the people used only a car in the last year, 40% used a car and something else, and only a handful of these used train, car, and airplane. The most mobile people used the most diverse modes of transport.

So there is information in the longitudinal data itself that can let us filter the respondents into homogeneous groups that will give more meaningful parameter estimates in destination/mode choice models. This will be taken into consideration while planning the form of the logit model to use.

4.2 Socioeconomic and geographic data

Limtanakool, Dijst, and Schwanen (2004) present a summary of long-distance travel modelling and discuss the influence of travel time and destination characteristics on mode choice with respect to the socioeconomic characteristics of the traveller. The density and diversity of land use types, as calculated in an index defined by the authors, tends to discourage car use. Good access from the urban center to train stations increases the use of trains, particularly for business trips. The most important characteristic for train use by vacationers is the presence/absence of a train station. Education and car availability, as in everyday traffic, are consistent influences on mode choice for all trip purposes. Including the variable travel time is crucial in estimating the model, provided the socioeconomic corrections are right.

The authors use 7 household types, grouped by measures of household size, number of adults in the household who work outside the home, and the presence of children under 12. At the person level, the authors use age, gender, education level, household income, and a car availability index.

The land use variables used for both origin and destination are population density, type of municipality, availability of a train station, local specialization index of land use, national specialization index of land use, and land use balance.

Hubert and Potier (2003) concur with the assertion that long-distance trips are more heterogeneously distributed in the population, and that explanatory variables are difficult to identify. For example, the ownership of a vacation home might be the most important determinant of a household’s vacation travel, but this information would rarely appear in a transportation survey and is most likely independent of the level of service of transportation modes to the destination. They find in their study of national long-distance travel surveys that the number of long-distance trips per year increases with income, with a saturation effect for high incomes. The effect of the size of the origin and destination towns is found to be a relevant factor in determining trip rates, for towns of certain size, as is the regional urbanization of area around the towns. In the latter case, the trip purpose confounds the explanatory power of urbanization on trip rate. People who make international trips are described by other characteristics than people who make generic long-distance trips. Trips across borders might therefore be best treated by a geographic dummy (0,1) linked to destination so that the parameters of these individuals are estimated correctly. Heterogeneity is also observed in macroeconomic trends, such as the productivity of industry and services, reduced working time, growth in international business relations, or people going into retirement, all of which have large influence on travel rates on a macro scale.

Much of the household information mentioned, save income, is present in DATELINE. Socioeconomic data at NUTS3 or more aggregate level will characterize the survey respondents where data about the household is lacking. They will also provide a basis for normalizing parameters to extrapolate them to the 11 countries that are not included in the DATELINE survey, so that the mode choice model can be used to fill the OD matrix for these countries, as well.

Socioeconomic and geographic data will be drawn primarily from the GETIS and GISCO datasets, which have been obtained from ETIS.

4.3 Generalized costs of modes (supply models) to DATELINE

The level of service data is needed for both chosen and non-chosen modes for the destination/mode choice model. Impedance (utility) measures or functions for three (five) modes are needed:

- Road (car)
- Rail
- Ferry (same as the mode it carries: train or car)
- Air
    - Air + road
    - Air + rail

The modes are car, rail, air + car, air + rail/bus, and ferry. Tour buses are left out of the supply for now due to a lack of summarized data about the bus schedules, routes, restrictions in countries of operation (such as laws restricting competition with national trains), price and special package offers, and other service variables. In this simplified set of transportation alternatives, ferry transport is associated with the mode that the ferry carries (train = rail; motor vehicle = road). The characteristics of the intermodal access and egress to airports for the air mode are included as generalized costs from each NUTS centroid to each airport in the air mode. At least for the first tests of the method, air travel will be represented as composite modes “air + road” and “air + rail”.

The obvious impedance measures are travel time and out of pocket cost. Other contributions to the level of service that are relevant for long-distance travel include waiting time, check-in time, and service frequency (for business travel above all). The revealed preference dataset (DATELINE) does not contain precise information about which train was taken, so modeling the traveller's sensitivity to frequency, for example, is not possible. Another modal attribute will have to substitute as a proxy for frequency.

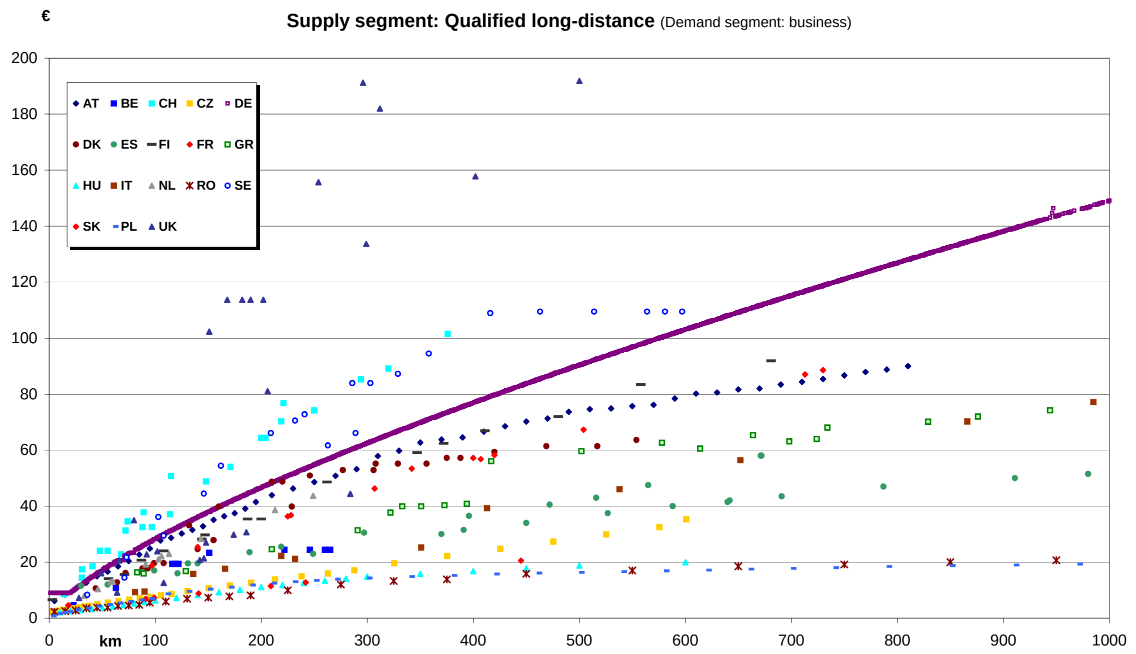
4.3.1 Road

Travel time is calculated for each DATELINE origin and destination based on the PTV basis network (PTV, 2004). Tolls on the TEN routes are available from IWW Karlsruhe. IWW has also estimated the driving cost per km based on fuel consumption and average driving speeds. A question remains as to how to estimate the price per kilometer of highway stickers (vignette, Switzerland and Austria) that are paid for once a year, even though some drivers only use them once (e.g. to pass through the country) and others use them daily.

4.3.2 Rail

Travel time from main train stations between the NUTS2 regions can be collected from the HAFAS server, which IVT will receive in the ETIS project. Alternatively, this travel time is available in the BAK rail model (Bleisch and Fröhlich 2003), for September, 2002. A network travel speed can be calculated from the DATELINE NUTS3 locations to these train stations. The number of train changes is also available in the respective datasets and can be substituted with a time penalty (to be found in the literature) so that all units of the time cost are in minutes, or the true time to switch can simply be used. There is no European database of rail prices, and the price in some countries can be a complicated function of advance booking discounts, discount cards, train type, route, and traveling class. IWW has analyzed rail prices, based on a web search of railway sites, and suggested a set of models to use for price per kilometer. This price can be adapted to trip purpose with some assumptions (for example, that business travellers use first class. IWW 2003).

Figure 1 Train Cost per Distance and Country



Source Eckhard Szimba, IWW 2004

The plot shows a regression line and data of price per kilometer for first class on high-speed trains, by distance, in the different countries.

4.3.3 Air

Check-in time, travel time and number of changes is included in the IVT flight model for September 2002 (Bleisch and Fröhlich 2003). The ticket price is a very variable element in the calculation of air impedances and is the subject of research by ETIS partner Mkmetrik in Germany. The difference in fares paid for work/non-work trips is large and having it modeled well will be very important for the estimation of the purpose-specific discrete choice model.

5 Enrichment of IVT Rail Network Model

In parallel to the enrichment of the DATELINE dataset with the service attributes of the different modes, the network models used to calculate travel times and mode/vehicle changes (public transportation) will also be appropriately updated to the task of assignment to NUTS3 resolution. The road and air models should have sufficient resolution and service quality, but they will be checked relative to the distribution of locations and chosen modes in DATELINE. The rail model will need refinement for the NUTS3 assignment.

IVT uses the IRPUD European railway network (Schürmann, 2001), with improvements added by IVT (Bleisch and Fröhlich 2003). The rail network geometry consists of 36,000 nodes and 78,000 links. There are 11,000 rail lines implemented, serving 2000 stops. To this level, the geometry is consistent with the GISCO network. The workday and weekend schedules for high-speed intercontinental and interregional trains were added from the Thomas Cook Timetables of September, 2002. Trains to the regional express level were added in the Alps for workdays. Only the schedules of the high-value connections were included for the rest of Europe. The model is fine enough for the correct representation of the access to 400 destination zones. The distribution of the lines in the European countries is presented in the Table 3:

Table 3 Number of Trains to each Country

Country	Lines	Country	Lines	Country	Lines
Albania	8	Greece	37	Slovenia	6
Austria	161	Holland	58	Spain	204
Belarus	4	Hungary	103	Sweden	103
Belgium	50	Ireland	77	Switzerland	664
Bulgaria	32	Italy	680	Turkey	18
Croatia	8	Macedonia	2	Ukraine	8
Czech Republic	53	Norway	13	Slovenia	6
Denmark	174	Poland	107		
Finland	83	Portugal	40		
France	592	Romania	40	International*	698
Germany	653	Russia	9	Airport Connections	166
Great Britain	414	Slovakia	93		

* International lines are those which cross areas of several countries, especially Eurocity trains

The real time tables permit the calculation of shortest-path matrices, including switching trains. For a valid connection, a maximum of 6 train changes is allowed, and the waiting time for train changes my also not exceed six hours. This treatment sets a minimum measure for level of comfort.

The airports from the air model are linked in this model to the rail network, so that the accessibility of any point on the network to any airport can be calculated.

6 Calibration of DATELINE OD Matrix: Assignment Constrained with Link Counts

A flow matrix must be derived based on DATELINE, to correct for the low estimates of trip rate. The DATELINE matrix is first assigned to the mono-modal network models. The route is determined by the shortest (time) paths on the loaded network, including waiting and vehicle changing time. Some correction will have to be made for the average everyday traffic (< 100 km) that occupies the networks and which is not included in DATELINE. This data will be solicited from national governments in ETIS Base. The proportion of the mode share in the long-distance OD matrix is corrected to match the mode share estimated by the parameters of the discrete choice model. Then, the size of the flows are adjusted according to UN/ECE (road) and ICAO (air) link flows using VISUM VStrom Fuzzy or, for comparison, another stochastic OD estimation method such as Path Flow Estimators (PFE, Vrtic, Axhausen, Bell, Grosso and Matthews 2004 or Ortuzar and Willumsen 1995). The stochastic assignment recognizes the presence of equally low-cost routes that could be assigned, and the travellers' imperfect knowledge of the available routes. The fuzzy algorithm allows for errors between the link flows and the total number of people traveling to account for the fact that the links flows and the OD survey were not performed on precisely the same people at the same time, but that there is a variation (one could even consider it a lag) in both flows and link counts that have to be kept independent of one another in this calibration, so an envelope of freeplay is built into the algorithm. This OD matrix will be taken to represent average long-distance travel however. It is therefore likely that the best assignment model has a steady-state, uncongested network solved by a simple maximum likelihood estimator and not a PFE.

6.1 Contents of ICAO and UNECE Datasets

The UNECE (United Nations Economic Commission for Europe) European Road Census (E-Road) contains data on road traffic flows (average annual daily traffic, AADT) by vehicle category, as well as breakdowns of vehicle flows by night, peak-hour and holiday traffic. In addition, the census provides information on infrastructure parameters of E-roads such as length and width of roads, number of carriageways and lanes, the latter of which are required for (for example) a Path Flow Estimator calculation. The 2000 road census is available for use in this project.

The Traffic by Flight Stage Digest (TF) of the International Civil Aviation Organization (ICAO) is an annual publication that provides traffic on-board aircraft on flight stages of international scheduled services. The data classified by international flight stage show by air carrier the aircraft type used, the number of flights operated by each aircraft type, the aircraft capacity offered and the traffic (passengers, freight and mail) carried.

The UN/ECE will also be carrying out a link count survey for TEN rail links in the 25 EU countries in 2005. The results of this survey could eventually be used to update the calibrations of the ETIS matrix.

6.2 Basis for Extrapolating OD Flows

The link flow-corrected OD matrix is iteratively re-assigned, re-distributed, and re-calibrated with the link flows until a sufficiently small change is observed between iterations. One baseline measure could be the MYSTIC dataset (http://www.cordis.lu/transport/src/mystic.htm). The resulting DATELINE-based flow matrix can then be used as a basis for adjustments to the synthetic OD flow matrix using a suitable data fusion technique (See for example Peter Davidson Consultancy, 2003).

7 Fusion of the DATELINE-derived OD matrix with the synthetic matrix

Variance-weighting was used by the Peter Davidson Consultancy to blend the MYSTIC long-distance survey with DATELINE. In this procedure, the number of trips per person per year is grossed up by weighting. The variance of the extrapolated number of trips on an OD pair is taken to indicate less certainty about the true number of trips taken between that origin and destination, and a normalized factor proportional to the inverse of the variance is used as a weighting factor in summing the value with a similarly weighted value for the same OD pair from the other dataset. This and other simple fusion techniques will be explored and compared for blending the DATELINE-derived OD matrix with the ETIS synthetic demand matrix.

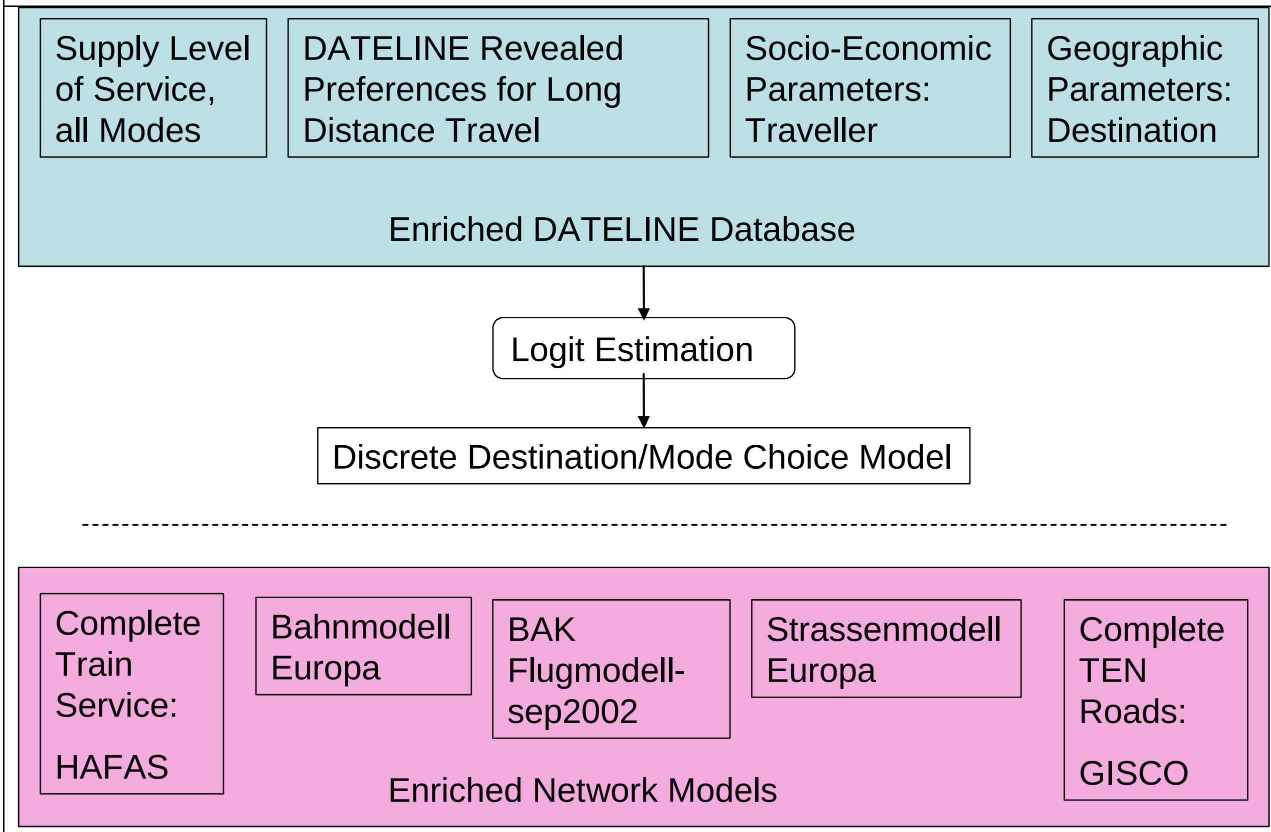
8 Summary of the Calibration Method

The DATELINE-based OD matrix will be extrapolated based on destination/mode choice parameters from the observed DATELINE matrix, and calibrating its flows based on link flow data. The establishment of the DATELINE-based OD matrix requires an assignment and a calibration based on link flows:

8.1 Enrichment Phase

- Generation of mode choice sets for each trip, based on least impedance
    - Road: individual motorized mode
    - Rail
    - Air + transport from NUTS centroid to/from airport
- Enrichment of DATELINE with the non-chosen modes
- Enrichment of DATELINE with socio-economic parameters at the appropriate NUTS level, including measures of attractiveness for destinations
- Enrichment of network models to appropriate spatial resolution and appropriate level of service for observed trip purposes and modes chosen

Figure 3 Enrichment of DATELINE and Network Models



8.2 Calibration Phase

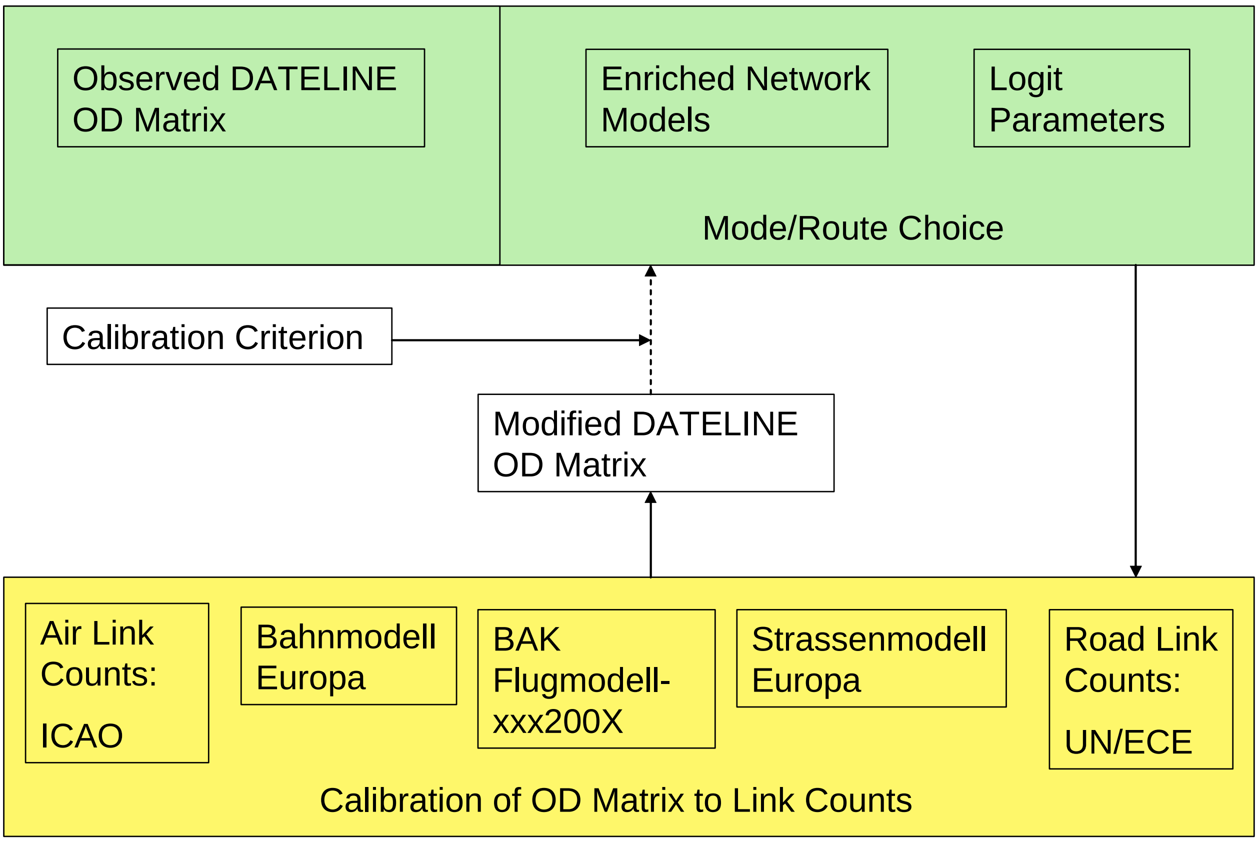
- Estimation of Destination/Mode choice parameters for trip purposes “work” and “non-work”, based on revealed choices, mode/route alternatives, geographic and socio-economic variables. Here, a route and a mode have the same meaning since the route (least cost mode) is mode-dependent.

- Assignment of the observed DATELINE OD matrix on the modal network models with the use of the Logit parameters from DATELINE.

- Comparison of the results with the available ICAO and UN/ECE or other flows, rail flows if available.

- Calibration of the DATELINE-based OD matrices to the link flows with VISUM VStrom Fuzzy or by the approach of path flow estimators (Vrtic, Axhausen, Bell, Grosso, and Matthews 2004)

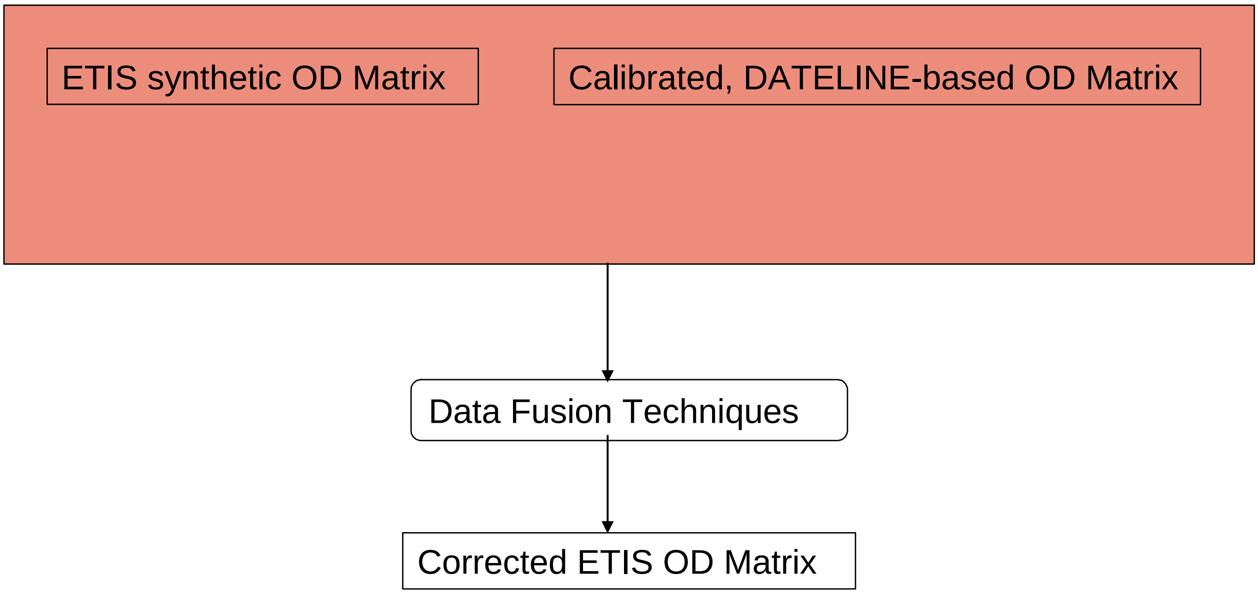
Figure 4 Enrichment of DATELINE and Network Models



8.3 Fusion of ETIS Base OD Matrices with the DATELINE-derived matrix

- Analysis of the changes in the matrices, and, respectively, the mode changes

Figure 5 Enrichment of DATELINE and Network Models



8.4 Feasibility and Reproducibility of Method

The OD matrix and destination/mode choice modelling can be repeated if the DATELINE survey is renewed or expanded for other countries, if the socioeconomic or link count datasets change, or if the quality of the supply networks (e.g. development of the TEN corridors and thus the link flows) change. This enables a means of renewing the OD matrices in ETIS.

The method relies on obtaining link traffic counts or their estimates, and national estimates of daily travel on the TEN routes which will have to be subtracted in order to yield the long-distance counts. If long-distance traffic counts cannot be obtained, for passenger rail for example, they could be modelled based on boarding/alighting rates, provided the latter can be obtained, though this would greatly complicate the process. It is more likely however, that the scaling up of the observed DATELINE OD matrix will have to be based entirely on the parameters calculated for the air and road modes.

While the trip distribution in DATELINE extends to destinations in all regions of the world, very few journeys originating outside the EU15 and Switzerland were captured. These countries (and corridors) are represented in the data above all in proportion to their popularity as destinations for work, private, and holiday travel from origins within the EU and

Switzerland. Until parameter estimates are available from residents of these countries, the Dateline parameters will have to be used as a first approximation. To estimate OD passenger flows outside the EU + CH, the DATELINE-based parameters will be used in the first instance with corrections based on the usual socio-demographic data about these countries. The mode choice and therefore flow amounts will depend on adjustments to the destination/mode choice parameters according to the different values of the socioeconomic parameters in each CEC and EFTA zone. This assumes the same model structure for these cultures, which is a reasonable assumption for a first model estimation but which may be questioned after more study of the issue. Nothing else can be done about it however, until long-distance travel data is obtained from these countries.

9 Literature

Axhausen, K. W.and M. Frick (2004) Nutzungen, Strukturen, Verkehr, *Arbeitsberichte Verkehr- und Raumplanung*, **205**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Ben-Akiva, M.E. und S.R. Lerman (1985*) Discrete Choice Analysis*, MIT Press, Cambridge.

Bleisch, A. and Ph. Fröhlich (2003) Die Erreichbarkeit von Regionen, IBC Modul Erreichbarkeit Phase 1, Schlussbericht, BAK Basel Economics, Basel.

BOKU-ITS (2003) DATELINE Deliverable 10b Weighting and Grossing up Report, Socialdata, Munich.

Cook's Timetable September, 2002.

Hubert, J. P. and F. Potier (2003) What is known? In K. W. Axhausen, J.-L. Madre, J. W. Polak and Ph. L. Toint (eds.) *Capturing Long Distance Travel*, 45-70, Research Studies Press, Baldock.

Last, J, W. Manz and D. Zumkeller (2003) Heterogenität im Fernverkehr – wie wenige Reisen wie viel? *Internationales Verkehrswesen*; **55,** Heft 6, Deutscher Verkehrsverlag, Hamburg, Germany.

Limtanakool, N., M. Dijst, and T. Schwanen (2004) The Influence of Socio-Economic Characteristics, Land Use and Travel Time Considerations on Mode Choice for Long-Distance Trips, 83rd Annual Meeting of the Transportation Research Board, January, 2004, Washington, D.C.

Mott, P. und K. Nöckel (1999) Fortschreibung von Nachfragematrizen: Ein neues Verfahren und seine praktische Erprobung, *Heureka ´99*, 287-294, FGSV, Köln.

Ortuzar, J. de D. and L. G. Willumsen (1995) *Modelling Transport, Second Edition*. John Wiley & Sons Ltd., Chichester, England.

Peter Davidson Consultancy (2000), Methodology for Statistical Analyses, Modelling and Data Collection (MYSTIC) http://www.cordis.lu/transport/src/mystic.htm.

Peter Davidson Consultancy (2003) DATELINE Deliverable 11 O-D Matrices, Socialdata, Munich.

PTV (2000): Benutzerhandbuch VISUM 7.5. Planung Transport Verkehr AG, Karlsruhe.

PTV (2004) Europäisches Basismodell Strasse, Planung Transport Verkehr AG, Karlsruhe.

Schürmann, Carsten (2001): Networks.txt. Institut für Raumplanung, Universität Dortumund.

Swiss Federal Statistical Office (2002) Methodologies and principal results, *Travel Behaviour of People Living in Switzerland in 1998*, SFSO, Neuchatel.

TIS.pt + UNEW (2003) DATELINE Deliverable 7 Data Analysis and Macro Results, Socialdata, Munich.

Vrtic, M., K.W. Axhausen, M.G.H. Bell, S. Grosso and W. Matthews (2004) Methoden zum Erstellen und Aktualisieren von Wunschlinienmatrizen im motorisierten Individualverkehr, Forschungsauftrag SVI 2000/379, IVT, ETH Zürich, Zürich.