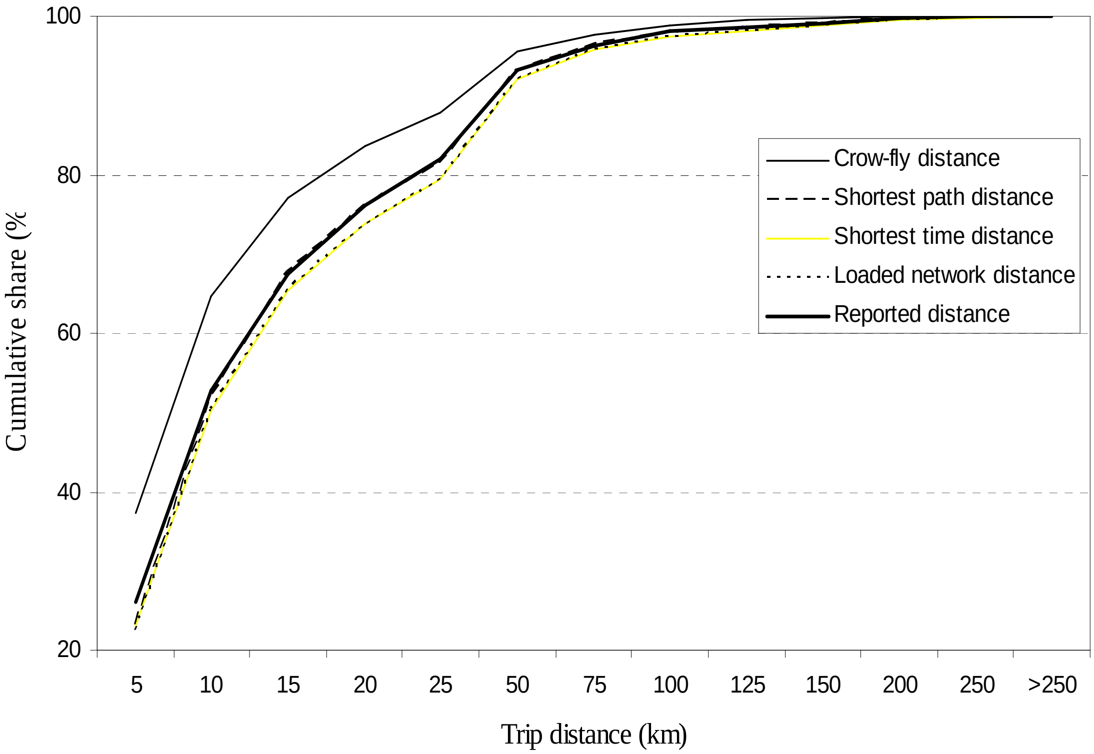


Travel distances estimations from household travel survey data: The case of Microcensus 2000

VS Chalasani

KW Axhausen

Arbeitsbericht
Verkehrs- und Raumplanung

224
June 2004

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Table of contents

Working paper

Travel distances estimations from the household travel survey data: The case of Microcensus 2000

VS Chalasani
IVT
ETH
CH-8093 Zürich

Telephone: +41-1-633 33 40
Telefax: +41-1-633 10 57
chalasani@ivt.baug.ethz.ch

KW Axhausen
IVT
ETH
CH-8093 Zürich

Telephone: +41-1-633 39 43
Telefax: +41-1-633 10 57
axhausen@ivt.baug.ethz.ch

June 2004

Abstract

Distance related information that has been collected from the travel surveys requires an in-depth analysis for the quality assessment of the directly reported distance data. This paper reports the issues involved in different distances calculations and comparing the same. The objective is to assess the accuracy of the reported travel distances. This report briefly introduced the loaded network distance and comparison of network distance, geo-code based crow-fly distance and the reported travel distance. Data that was collected in 2000 Microcensus is used for comparison. This sample data has been thoroughly screened and only within day car travel was considered for analysis. Variations in trip length distributions of reported distance and the calculated distances are observed. It is found that the reported distances are in between shortest path distance and loaded network distance. Using a descriptive statistical analysis, a series of detour factors are computed for different travel distance ranges. An average detour factor (operational) of 1.61 is obtained form the analysis. Accuracy of the reported information about origin and destination of the long-distance trips is assessed in the end.

Keywords

Travel distance; network distance; crow-fly distance, travel survey; Microcensus; n

Preferred citation style

Chalasani V.S. and K.W. Axhausen (2004) Travel distance computation from household travel survey data: The case of Microcensus 2000, *Arbeitsbericht Verkehrs- und Raumplanung*, **224**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

1 Introduction

Travel distance is an important factor that influences different travel parameters such as, mode of transport, destination location, time of departure, travel route, etc. But, the travel distance was not considered as one of the parameters in most of the transport models. Difficulty in collecting travel distance information without any additional burden from travel surveys and uncertainty in geographic location information are some of the prominent reasons for this ignorance. In general, travel surveys are used to collect data for a wide variety of purposes and are administered in a similarly variety of forms. However, there is one commonality among virtually all major travel surveys: the need to collect geographical information to describe the location of the respondent and/or the location of trips made (or to be made) by the respondent. The spatial dimension of the transportation system is an important context to travel surveys. Geographical location information provided by respondents can be used to determine the part of the transportation system that is available and how the respondent might use the system. To date, most of the travel surveys use the collected geographic data only after the survey is complete.

In general, depending on the targeted quality and quantity, a number of checks are carried out during the course of any travel survey. Travel surveys have certain operational limitations in observing both quantity and quality of travel parameters such as, travel distance, travel time, origin and destination. Reported travel distance observed through the travel surveys contain an error term that counts for the individual psychology in addition to the actual distance. This error is random and some of the key factors that influence the respondents to report erroneous distance data and geographic location information are:

- Familiarity with the local transportation system: When the respondent is travelling for the first time and no prior information is available, it is high possible for an erroneous distance data. In other situation, though the respondent reports the correct travel distance, which may not be the shortest.
- Existing traffic situations during the travel: Unusual delays create a psychological impression of travelling longer distance than that of actual.
- Purpose of the travel: Travel for the purposes such as leisure and shopping, may not follow the shortest path
- and other travel conditions (as driver or passenger).

Instead of real world distance, people usually carry cognitive, estimated or subjective distance, is very different from objective real world distance (Walmsley, 1988)

If we look at the problems associated in collection of geographic location information and distance data in travel surveys, reporting the actual distance travelled for a simple question depends on the above stated parameters. It is necessary to assess the accuracy of the reported distances with a standard distances such as network distances, crow-fly distances. It is equally difficult to report the geographical address of the origin and destination of each movement (trip, stage, journeys, etc.). Both the distance data and geographic location information are necessary and complement each other in assessing their accuracy. Various innovative data collection methods such as interactive geo-coding, GPS surveys, Internet-based surveys, were introduced to obtain reliable geographic location information from travel surveys (Resource system group, 1999; Wolf, 2003).

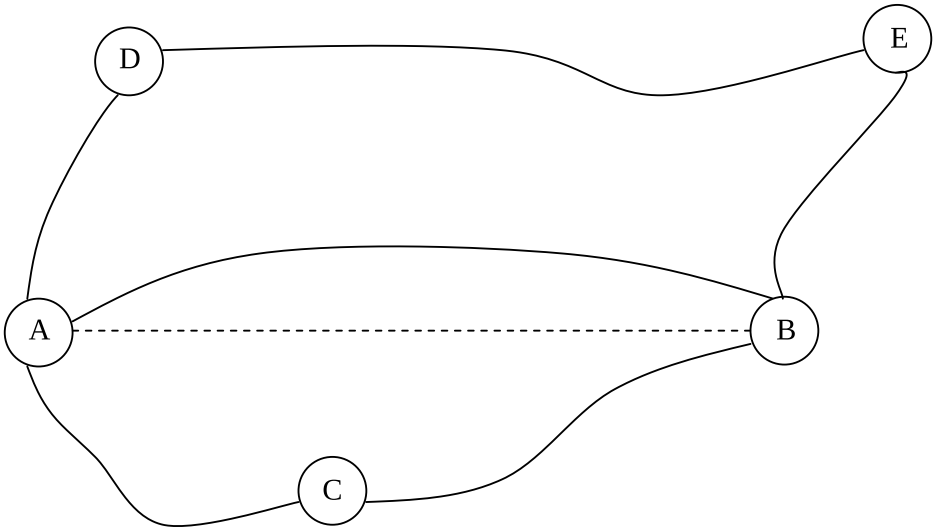
The geographic location information is later transformed into spatial coordinates through the geo-coding process. Geo-coding is generally defined as the referencing of address data in terms of coordinates. Geo-coding involves the assignment of spatial coordination (2D or 3D) from existing geo-data to the geographic location of each reported movement (stage in Microcensus 2000). Most importantly, the geo-coded information allows plausibility testing of the reported stage and trip distances and facilitates any necessary adjustment. Other survey data, e.g. speed and travelling time may also be validated. Three standard methods of co-ordinate logging have emerged to date: post survey geo-coding, online geo-coding (during survey) and the use of specialized equipment to capture co-ordinates. The post survey geo-coding is the most frequently employed to date. This is specifically implemented for the surveys which collection the geographical address of the locations. The 2000 Microcensus travel data was geo-coded using this method (Jermann, 2003).

This study aimed to calculate distances based on two different approaches, crow-fly and network based, and assess the accuracy of reported distances from Microcnesus travel 2000. Calculation of detour factors for the selected Swiss federal road network is also focussed in this study. Variations in the accuracy of reported distance for different purposes and various distance ranges are also scripted in this study.

2 Travel distance in travel surveys

Because of the difficulties involved in collecting travel distance information from travel surveys, most of the travel surveys inclined to calculate travel distance. The distance travelled is calculated using the geographic information collected from travel surveys. Traditionally, to understand the spatial aspects of the transportation system, transport researchers followed a standard set of distances. This set include, geo-based crow-fly distance, network based distances and direct observations. Basic definitions of these distances and their importance are discussed in this chapter.

Figure 1 A hypothetical network to explain different distances



A: Origin B:Destination C, D, and E: Intermediate stages

------: Crow-fly path _______: Network path

2.1 Crow-fly distance

By definition, crow-fly distance is straight-line distance between the origin and destination ignoring the earth’s curvature. Geographic location information in the form of longitude and latitude or the geo-codes is required to calculate the crow-fly distance. Irrespective of the structure and nature of the transportation system, crow-fly distance between two locations remains constant. From travel surveys, the geographical address (zip code, street name, and house number) of the origin and destination can be used to calculate the geo-codes of the point locations. Accuracy of the crow-fly distances depends on the collected geographical information. Crow-fly distance between A and B is the straight line distance i.e. crow-fly path in the figure 1.

2.2 Network distances

The two well known criteria in network distance calculations are shortest path distance and shortest time distance. Another distance used in this study is loaded network distance. In the hypothetical network there are three possible routes to reach B from A, namely ACB, AB, and ADEB.

Shortest path distance is the minimum distance to be covered on network to reach from one location (origin) to the other (destination). Shortest path distance depends only on the transportation network geometry, but not on its properties such as limiting speeds, number of lanes, etc. Shortest path distance is the minimum of the three routes.

Shortest time distance is the distance that can be covered in minimum time to reach the destination. Shortest time distance depends on both geometry and properties of the transportation network. Under free flow travel conditions in an unloaded network both the shortest path distance and shortest time distance need not be equal unless the speed limitations are identical for all links. Shortest time distance is the distance of the route which has the least travel time.

Loaded network distance is the average distance actually travelled to reach the destination under regular conditions. Vrtic et al (2003) calculated the travel distances between all municipalities in Switzerland by loading the network with an O-D matrix and calibrated with the average annual daily traffic (AADT). This loaded shortest network distance is the weighted average distance of all used paths between a particular origin and destination pair.

$$\text{Loaded shortest Network distance} = \frac{v_1 x_1 + v_2 x_2 + v_3 x_3 + \ldots\ldots + v_n x_n}{v_1 + v_2 + v_3 + \ldots\ldots + v_n}$$

Where v_1, v_2, ... v_n are the traffic volumes and x_1, x_2, ... x_n are length of the 'n' used paths between the origin and destination of interest.

The loaded network distances are naturally longer than the shortest distances. These are more realistic because they reflect the impedance of the regular traffic.

2.3 Reported distance

Reported travel distances are the distances as perceived by the respondents during their travel. Respondents reported distances need not necessarily be the actual distances travelled. As stated in the previous chapter, there are many factors that influence the respondent to report larger deviations. An attempt is made in this study to capture the impact of familiarity.

2.4 Detour factors

By definition, detour factor is the ratio between the network distance to the crow-fly distance. Conceptually, it is the relative distance needed to travel on the network to cover a unit of crow-fly distance. Different detour factors can be derived using different network distances:

$$\text{Detour factor (geometric)} = \frac{Shortest\ path\ dis\tan ce}{Crow-\ fly\ dis\tan ce}$$

$$\text{Detour factor (static)} = \frac{Shortest\ time\ dis\tan ce}{Crow-\ fly\ dis\tan ce}$$

$$\text{Detour factor (operational)} = \frac{Loaded\ network\ dis\tan ce}{Crow-\ fly\ dis\tan ce}$$

The significance and basic concept of each of these detour factors are explained below. Geometric based detour factor gives the additional distance to be traversed per unit crow-fly distance. It largely depends on the density (total length of roads per unit area) and geometry of the transport network and remains constant unless there is a change in the network geometry. Where as, static detour factor gives the additional distance to be traversed in minimum time per unit crow-fly distance. This factor depends on the operational constraint of the network and nothing to do with the flows. Under similar operational properties of the network, the static detour factor remains constant. The operational detour factor gives the units of distance to be traversed when the network is loaded with average traffic per unit crow-fly distance. It reflects the combined effect of the travellers behaviour, network geometry, and its properties. In addition to these network based detour factors, another detour factor, perceived detour factor, is introduced in this study. The perceived detour factor is the ratio of reported distance to that of corresponding crow-fly distance. Following conclusions can be drawn with the calculated detour factors:

- Differences in the Geometric and static detour factors represent the effect of operational constraints such as directions, capacity, speed limits, etc.
- Differences in static and operational detour factors are due to the induced traffic on the network under similar limitations such as directions, speed limits, etc.

Travel distances were analysed for all the above detour factors.

3 Data

For the calculation of travel distance from the travel surveys, Switzerland is considered as the study area. The data obtained in the recent national household travel survey, Microcensus 2000 is used in this study. The Microcensus 2000 is a daily mobility survey, which observed all the trips made on the specific day by each respondent. Travel information was collected at stage level. In this study travel distances were calculated at stage level and later aggregated to trip level. Only car trips were considered for distance comparisons. The data used in this study is from three approaches:

- Geographical data
- Network data, and
- Travel data.

However, data from above stated three categories has some limitations. Sections following discuss the data limitations in detail.

3.1 Geographical data

The Microcensus 2000 data was geo-coded in a separate study (Jermann, 2003). The following five geo-databases were used to calculate the geo-codes:

- Gebkoord BfS: All building entrances of Switzerland
- Gebkoord ZH: All building entrances of Canton Zurich
- Microspot: Swiss post office geo-codes
- Haltekoord: Public transport stops in Canton Zurich (Excluding Postbus stops)
- Bahnhoefe ARE: Swiss railway station geo-codes

Both the household and origin and destination of all reported stages were geo-coded. Geographical address (Zip code, place, street name, and house number) was used as the basis to calculate the household geo-codes. Mode of transport was additional considered in geo-coding the stages. The structure of the geo-coding is explained in the Table 1.

Table 1 Accuracy of geo-codes with respect to the trip origin and destination

Available geo-information	Geo-coded as
Zipcode + Name + Street Name + House number	Entrance geo-code
Zipcode + Name + Street Name	Geo-code of the first house in the street
Zipcode + Name	Post office geo-code
Public transport stops (stages only)	Geo-code of the public transport stop (except the air port)

Source: Jermann et al. (2003)

The quality of geo-codes is varying because of the inconsistencies prevailing in the reported address of either the household or origin or destination. In Microcensus 2000, respondents reported the address at various levels of geographic detail, such as Municipality, postal code, street and household number. All the geo-codes follow the as per the reported address. In the origin and destination geo-codes computation, in addition to the household geo-code database, public transportation geo-code database was also used. The consistency of the level of detail by origin and destination of the trips is analysed in Table 2.

Table 2 Accuracy of geo-codes with respect to the trip origin and destination

Reported information on the trip origin ↓	Reported information on the trip destination			
	Zip code, street name and house number	Zip code and Street name	Only zip code	Total
Zip code, street name and house number	17474(16.77)^{1}	0(0)^{2}	6528 (6.3)^{3}	24002 (23.0)
Zip code and Street name	0.7(0)^{2}	0 (0)^{4}	6528 (6.3)^{5}	6529 (6.3)
Only zip code	0(0)^{3}	0.7(0)^{5}	73684(70.7)^{6}	73685 (70.7)
Total	17475(16.77)	0.7(0)	86740(83.2)	104215

(x)^{a}: x->Percent share of the geo-codes a-> Reliability based rank of the geo-codes

When compared to the destination, more accurate information about the origin was reported. Overall, geographical information for 70% trips was reported at zip code level. As mentioned earlier, only car trips are considered for comparison. Table 3 shows the accuracy of geo-codes for the reported geographic location information of the car trips.

Table 3 Accuracy of geo-codes with respect to origin and destination of the car trips+

Description	Origin	Destination	Origin – Destination
Exact location	27.82	26.97	10.73
Nearest public transport	42.94	43.40	26.77
Post office	26.06	26.54	15.14
Missed	3.19	3.09	1.28
Total	100	100	53.92

+: Car trips (34195)

3.2 Network data

All the network distances were calculated using the IVT Swiss road network model. The IVT Swiss road network consists of 3,066 zones, 14,798 nodes, and 19,664 links (Vrtic et. al , 2003), is developed at IVT and it represents federal roads in Switzerland. The whole network model is developed at municipality level i.e. the lowest geographical unit in Switzerland. The network distances were calculated between the centroids of origin-destination municipalities. All the municipal centroids are connected to the network with an average distance of 141m. The origin-destination matrix (O-D Matrix) used to load the network. This O-D matrix was developed using the Microcensus 1994 and Census 1990 data. Average annual working day traffic (AADT, work day) for the year 2000 (ASTRA, 2001) was used in the study. To calculate the network distances, the O-D matrix was calibrated with the average annual working day traffic (cross-sectional counts).

Shortest path distances were calculated with no traffic flow and equal limiting speed on each link. While calculating the shortest time distances, traffic flows were set to zero, but the operational characteristics such as capacity, limiting speeds are assigned for each link. Loaded network distances were calculated by calibrating the O-D matrix with the average annual working day traffic. Loaded network distance is the weighted average of the all route lengths that contribute traffic between the origin-destination.

3.3 Travel data

The travel data collected in the Microcensus 2000, the Swiss national household travel survey, is used in this study. The following key variables were observed in Microcensus 2000 are relevant for this study:

- Address (Postal code, Street name, and House number) of the origin and destination
- Arrival time, departure time, and travel duration at trip level
- Arrival and departure time of all stages
- Purpose of the trip

A total of 104,715 same-day trips were reported by 29,407 respondents in Microcensus travel 2000. Mode of transport was observed at stage level. Only a part of the Microcensus travel 2000 data was selected for this study. The procedure in data selection is explained in the next chapter.

4 Calculation of travel distances

Using the rich and relevant data, the network distances and the crow-fly distance for all trips were calculated. Significant pre-analysis was performed on the Microcensus 2000 data in order to make the data ready for travel distance analysis, which is explained in the next section. In the subsequent sections, a detailed description of the network distance and crow-fly distance calculations are discussed.

4.1 Data preparation

The Microcensus 2000 dataset needed to be enriched and corrected in certain aspects for the planned analyses. The following assumptions were made in data preparation:

- The mode of transport was observed at the stage level. It is assumed, that the main mode of transport is the mode with maximum distance travelled during the trip.
- The geo-code of the post office was considered as the geo-code of the postal code area assuming that it is close to the centroid of the area.
- Network distance is independent of time of day, i.e. network travel does not vary between peak and off-peak periods.
- Only trips with car as the main mode of transport were considered for the trip length distribution analysis.
- Above this, the average annual daily traffic (AADT) on a working day was obtained using the automatic vehicle count on the streets, which counts mainly the cars.

In the Microcensus 2000, nearly 30% are walk trips and 50% are car trips (either as driver or passenger) - (main mode of transport).

The main mode of the transport was identified by:

- Main mode of transport by modal hierarchy, and
- Main mode of transport based on distance travelled.

For the first criterion, a modal hierarchy was developed. The main mode of transport is the mode with highest priority. In contrast, for the second criterion, the main mode of transport was determined based on the distance travelled. Table 4 compares the share of main modes of transport for the two criteria. Table 4 reveals that the assumptions made about the main mode of transport produce rather similar shares.

Table 4 Comparison of main mode of transport by hierarchy and travel distance

Mode of transport	Rank*	Percent share by the main mode of transport as per the		
		Modal hierarchy	Distance travelled	Difference
Plane	1	0.07	0.07	0.0
Train	2	3.70	3.60	0.1
Long-haul public transport	3	0.80	0.90	0.1
Short-haul public transport	4	6.20	6.10	0.1
Car	5	50.90	50.50	0.4
Motorcycle	6	1.30	1.30	0.0
Moped	7	0.80	0.80	0.0
Bicycle	8	7.30	7.30	0.0
Other	9	1.00	1.10	0.1
Walk	10	27.20	27.70	0.5

*: Rank of the transport mode as per assumed modal hierarchy

4.2 Network distance and crow-fly distance calculation

All the activity locations that were observed in Microcensus 2000 were geo-coded in the study by Jermann et al (2003). A two stage and three tier matching procedure was developed in the above study. In the first phase, addresses of the households, origin and destination of the stages, and trips at three levels (House number, Street name, and Postal code) were matched with the corresponding address at the equivalent level. First phase is exclusively for the addresses with private transport. Addresses involving public transport were dealt with in the second phase, where the addresses of the origin and destination of both the stages and trips were matched with the public transport geo-code database. The bias involving the geo-codes associated with Zurich main railway station will be different from that of a small railway station such as Zurich-Wipkingen. Overall, geo-codes carry a non-uniform error, which ultimately effects the crow-fly distances that are calculated using the biased geo-codes. One way to reduce this bias is to have addresses at the most precise level, i.e. house number.

At the same time, in travel surveys, too much emphasis on the addresses may lead to respondent fatigue and to further serious consequences. It is not unusual to travel to a place without knowing the precise address of the destination and vice versa. Overall, crow-fly distances for the Microcensus 2000 data are therefore imprecise. For the comparative purposes this is not an issue, as all calculations are based on the same-often erroneous-geocode. See Table 2 and Table 3 for the shares of different types of inaccuracy.

Network distance calculation: For the network distance calculations the locations (both origin and destination) were reassigned to the municipality centroids. The distances at stage level were calculated using the IVT's national road network (Vrtic et al., 2003). Stage distances within municipalities and zip codes are set to zero by default (about one-third of the trips). All the stages distances are summed to obtain trip distances.

4.3 Analysis of travel distances

Before analysing the travel distances, all calculated distances were analysed for their accuracy. It is found that only a part of the car trips have valid network distances, crow-fly distance and reported distance. If we look at the how the data has reduced, among the 104,215 reported same-day trips 53,500 (51%) are car trips. But, only 34,195 (33%) trips are eligible for distance comparison. In total, 17% trips are excluded because of inconsistency in distance calculations.

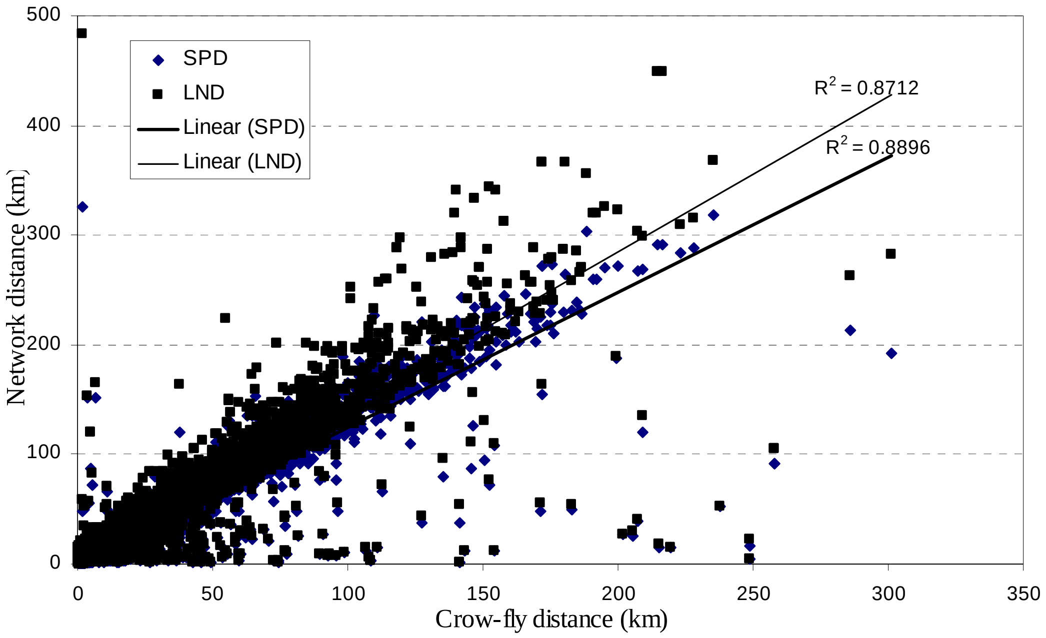
The different travel distances are analysed to understand the following issues:

- What relationships do the reported distances have with the network and crow-fly distances or above the network distances.
- to what extend can the crow-fly distances be considered as a good proxy for the distances on the actual road network
- variations in the trip length distributions and accuracy of reported distances
- What are the detour factors between different distances

4.3.1 Variations within Network distances

An attempt is made to understand the relationships between the IVT's national road network based distances and the crow-fly distances. Figure 2 shows the distributions of the two network distances (Loaded network distance and shortest path distance) against the crow-fly distances.

Figure 2 Distributions of network distances against the crow-fly distances



SPD: Shortest Path Distance LND: Loaded Network Distance

As the network distances are calculated between municipality centroids, there will be a large error in case of short distance trips. This is reflected in the Figure 2, a significant portion of short distance trips (below 50 km) are underestimated in the network distance calculations.

The trip length distributions of the same-day car trips (both as passenger and driver) based on crow-fly distances, network distances and reported distances are given in Table 5.

Table 5 Trip length distribution of same-day car trips (as passenger and driver) based on the distance data

Trip Length (km)	Crow-fly distances			Shortest path distance			Shortest time distance			Network distance			Reported distance		
	Fre-quency	Share [%]	Mean (km)	Fre-quency	Share [%]	Mean (km)	Fre-quency	Share [%]	Mean (km)	Fre-quency	Share %	Mean (km)	Fre-quency	Share [%]	Mean (km)
0 – 5	12769	37.34	3.05	7974	23.32	3.52	7843	22.93	3.50	7726	22.59	3.49	8924	26.10	3.58
5 – 10	9385	27.45	7.14	9779	28.60	7.34	9352	27.35	7.32	9494	27.76	7.31	9112	26.65	7.90
10 – 15	4214	12.32	12.29	5391	15.77	12.27	5192	15.18	12.34	5196	15.19	12.39	5017	14.67	13.06
15 – 20	2239	6.55	17.34	2865	8.38	17.27	2851	8.34	17.21	2858	8.36	17.22	3000	8.77	18.31
20 – 25	1458	4.26	22.38	1885	5.51	22.28	1893	5.54	22.17	1884	5.51	22.19	1955	5.72	23.34
25 – 50	2610	7.63	34.15	3996	11.69	34.05	4359	12.75	34.21	4330	12.66	34.23	3897	11.40	35.53
50 – 75	751	2.20	60.40	1086	3.17	60.97	1250	3.66	61.08	1244	3.64	61.04	1042	3.05	61.97
75 – 100	366	1.07	85.72	469	1.37	86.22	548	1.60	86.86	559	1.63	86.64	584	1.71	88.60
100 - 125	211	0.62	111.56	288	0.84	111.20	293	0.86	111.91	295	0.86	112.00	209	0.61	113.57
125 – 150	93	0.27	136.74	139	0.41	135.12	225	0.66	136.17	213	0.62	136.22	157	0.46	139.91
150 - 200	79	0.23	168.98	232	0.68	167.69	209	0.61	174.34	214	0.63	173.73	194	0.57	175.02
200 - 250	15	0.04	224.07	72	0.21	223.36	94	0.28	217.71	94	0.28	218.47	69	0.20	224.80
> 250	5	0.02	282.32	19	0.06	272.30	88	0.26	295.13	88	0.26	294.51	35	0.10	295.14
Total	34195	100.0	**13.14**	34195	100.0	**17.88**	34195	100.0	**19.60**	34195	100.0	**19.60**	34195	100.0	**18.44**

Average crow-fly distance is the least among all distances. Shortest time distance and loaded network distance follow closely with equal averages which are highest. The shortest path distances are in between crow-fly distances and other network distances. The average of reported distances is in between shortest path distance and loaded network distance.

In case of trips with distances less than 5 km, the mean reported travel distance is less than the mean network distance. Where as the mean reported distance is always greater than both the mean network distance and mean crow-fly distance in case of trips of above 5 km. Due to the inability in computing intra-communal trip distances and missing geo-codes, differences in percent share of small distance trips are significant. The graphical version of the variations in trip length distributions are shown in Figure 3. Significant differences were observed in trips of smaller distances and either intra-municipal trips or trips with missing geo-codes, whose travel distance could not be calculated. The cumulative trip length distributions of same-day domestic car trips shown in the Figure 4

Figure 3 Microcensus 2000: Variation in trip length distribution of same-day domestic car trips (passenger and driver)

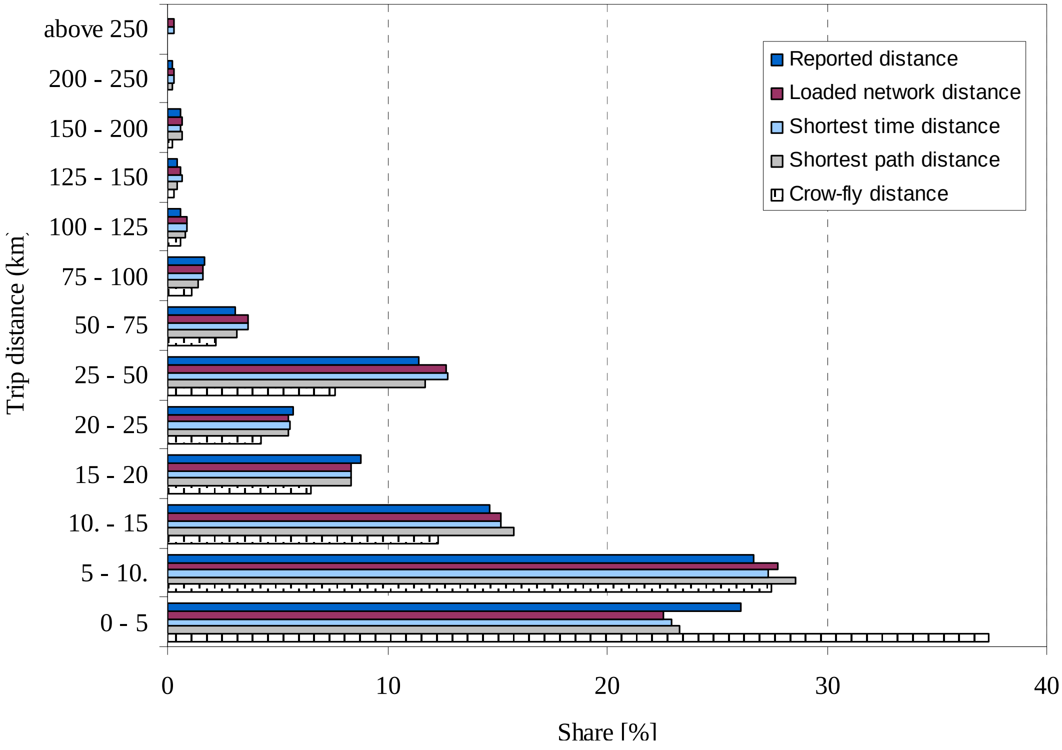
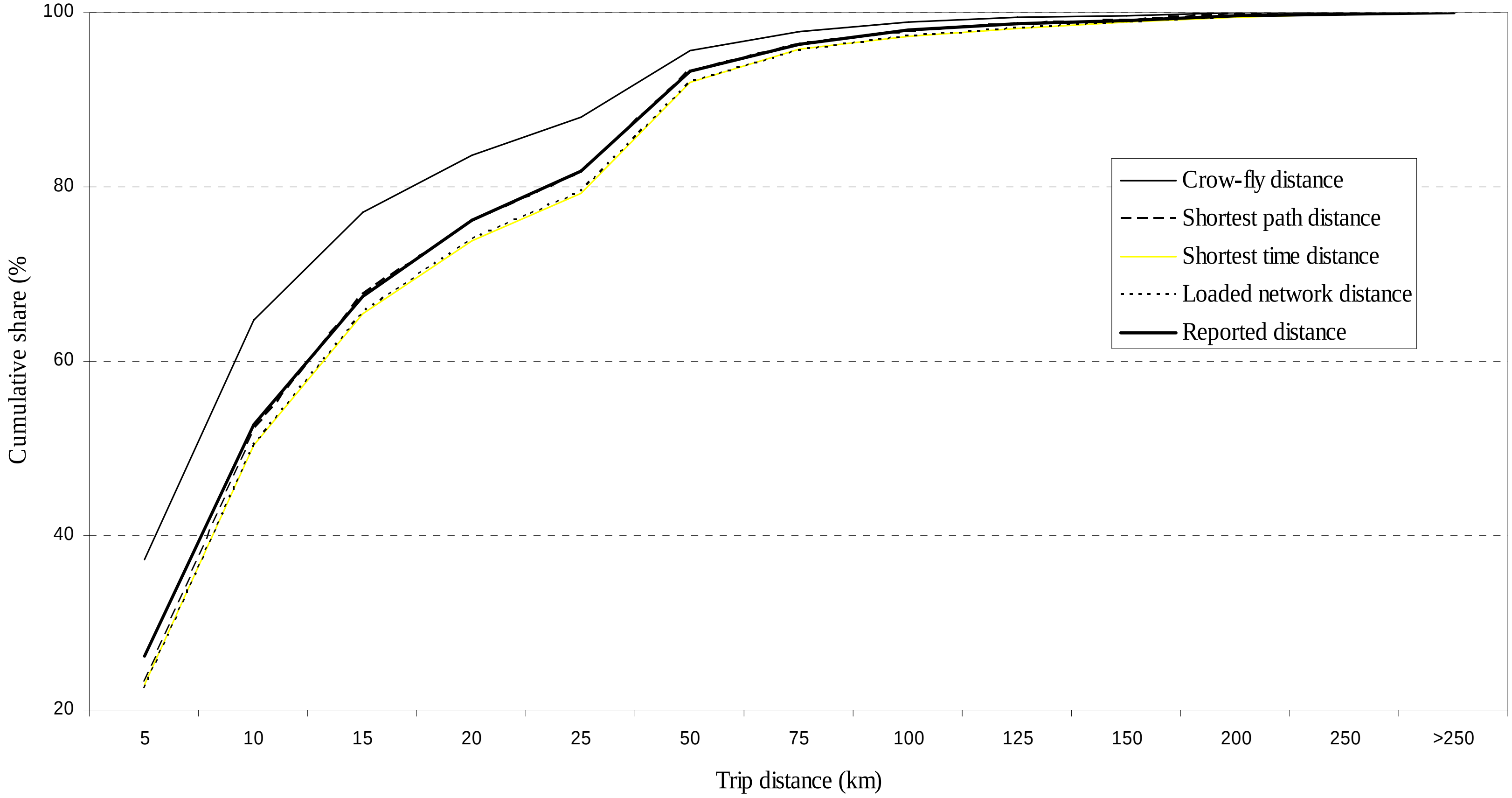


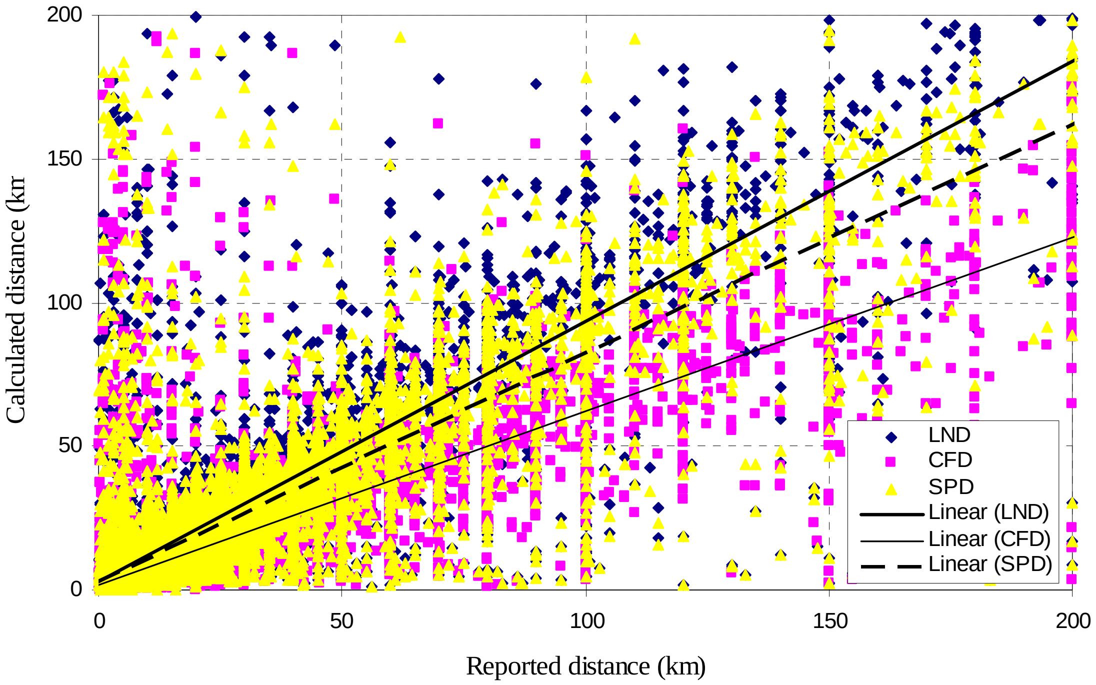
Figure 4 Cumulative trip length distribution for different distance measures



The reported distances cumulative distribution generally fall in between crow-fly distance and loaded network distance cumulative distributions. The Microcensus 2000 reported distances cumulative distribution is closer to the loaded network distance distribution.

Both the Table 5 and Figure 3 explained only the shares of pre-defined trip length categories, but not the mutual distribution. Mutual distributions of trip lengths allow understanding of correlations and the outliers. Figure 5 shows the distributions of reported distances with different network based distances. With a close range of moderate coefficient of correlations, all the three calculated distances are significantly scattered. These deviations are due to two prominent reasons, 1) Erroneous centroid based network distance calculations or post office based crow-fly distance calculations, and 2) outliers. First reason is most common in case of trips less than 50 km, where as second reason might be true for long-distance trips.

Figure 5 Distributions of reported distances against network based distances



LND: Loaded Network Distance $R^2 = 0.675$

CFD: Crow-fly Distance $R^2 = 0.692$

SPD: Shortest Path Distance $R^2 = 0.705$

To assess the accuracy of the reported distance, quantification of the deviations of the reported distances from the network distances is important. Figure 6 shows the distributions of reported distances deviations from the calculated distances. Following inferences are made by analysing the basic characteristics of these deviation distribution curves such as spread, peakness, skewness, etc.:

- With a low-peak and rightly skewed deviation curve against crow-fly distance demonstrates that most of the reported distances are greater than the crow-fly distances. However, around 13% trips are reported less than crow-fly distances, which is simply not correct.
- Deviation curve against shortest path distances has a high-peak and slightly right skewed. This reflects highly concentrated deviations within 20% from shortest path distances.
- Deviation curve against loaded network distances is similar to that of from shortest path distances, but skewed to the left. It can be concluded that the reported distances are mostly lower than the loaded network distances.

Figure 6 Distributions of reported distance deviations from calculated distances

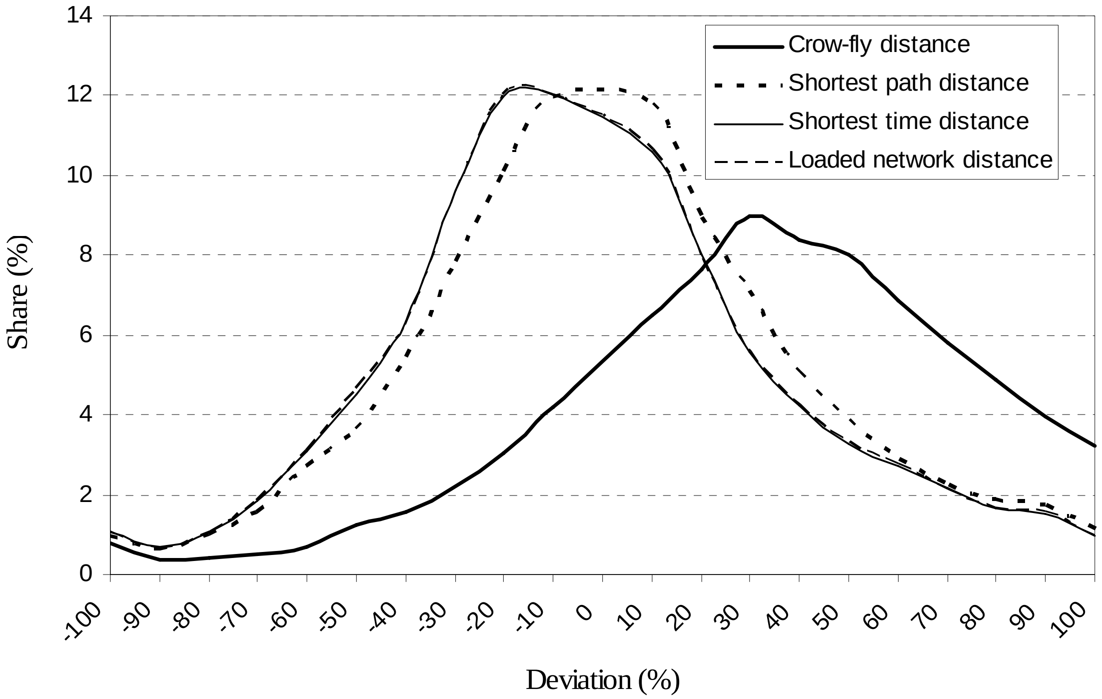


Table 6 Trip length distributions and accuracy of reported distances by trip purpose

	Trip purpose												
Trip distance (km)	Work		Education		Business		Shopping		Leisure		Service		Total
	Share*	Factor+	Share*	Factor+	Share*	Factor+	Share*	Factor+	Share*	Factor+	Share*	Factor+	
0 - 5	6.26	1.54	0.28	1.38	5.26	1.56	0.85	1.78	8.24	1.63	1.21	1.92	22.10
5 – 10	8.20	1.12	0.43	1.23	5.28	1.05	1.22	1.41	10.6	1.17	1.44	1.14	27.17
10 – 15	4.79	1.01	0.29	1.05	2.35	0.98	0.74	1.03	6.05	1.06	0.66	1.00	14.88
15 – 20	2.40	0.98	0.12	0.89	1.44	0.96	0.4	1.27	3.47	1.05	0.39	1.00	8.22
20 – 25	1.86	0.94	0.11	0.89	0.68	0.97	0.37	0.98	2.13	0.98	0.25	0.92	5.40
25 – 50	3.69	0.95	0.24	0.95	1.25	0.89	0.87	1.12	5.81	0.95	0.56	0.97	12.42
50 – 75	0.81	0.92	0.07	0.98	0.38	0.75	0.33	0.98	1.84	0.90	0.12	0.83	3.55
75 – 100	0.33	0.85	0.02	0.88	0.14	0.73	0.12	1.09	0.96	0.88	0.03	1.02	1.60
100 – 125	0.17	0.64	0	0.85	0.11	0.47	0.08	0.89	0.40	0.87	0.05	0.99	0.81
125 – 150	0.09	1.00	0.01	1.18	0.08	0.69	0.07	0.69	0.29	0.83	0.02	0.92	0.56
150 – 200	0.08	0.73	0	1.56	0.05	0.45	0.08	0.88	0.33	0.87	0.03	0.87	0.57
200 – 250	0.07	0.37	0	0.07	0.02	0.41	0.03	0.82	0.09	0.68	0.01	0.94	0.22
> 250	0.07	0.02	0.04	0.01	0.01	0.64	0.02	0.37	0.08	0.61	0.00	0.90	0.22
Total	28.82	1.13	1.62	1.09	17.05	1.16	5.16	1.26	40.28	1.17	4.79	1.26	97.72

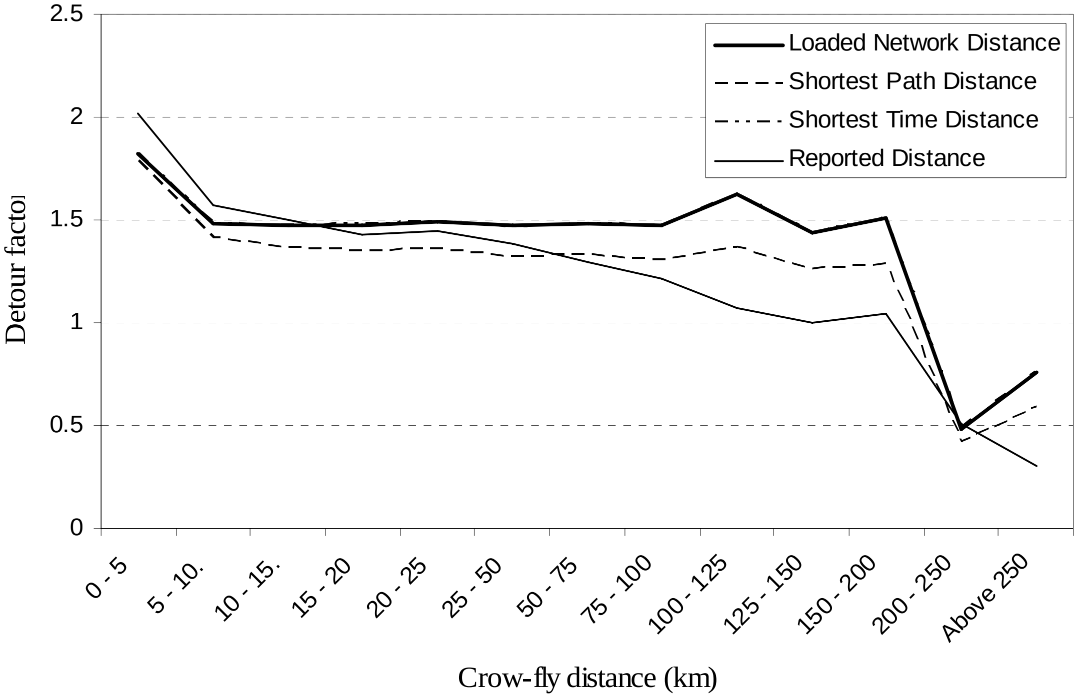
*: Percent share of valid car trips (34195) +: Factor = Average of (Reported distance / Loaded network distance)

As mentioned earlier, familiarity is one of important factors that influence the accuracy of reported distances. This study considered trip purpose to probe the effect of familiarity. Among the calculated distances, loaded network distances are consistent and represent the realistic scenario. The trip length distributions along with the average ratio (reported distance / loaded network distance) are tabulated in Table 6. Due to the centroid based network distance calculations, reported distances are greater than that of loaded network distances for short distances (up to 15 km). Where as a consistency in reporting is found for the trips with distance range of 20 – 75 km. Most of the long distance trips, which are small in number, were under reported. The trip length distributions for all trip purposes are more or less similar. Where as, the deviations in reported distances from loaded network distances are smaller in case of commuting trips (such as work, education, and business), when compared to other purposes.

4.4 Detour factors

Detour factor is an important parameter that exhibits not only the prevailing transportation network properties but also the user's knowledge. Detour factor normally consists of a network based static and an operational based dynamic part. Above to the network dependency, the detour factor is also distance dependent. Figure 7 shows the distributions of different detour factors. Perceived detour factor distribution has the largest variations. All the network based detour factor distributions are consistent in a distance range of 15 – 150 km. A drastic drop in detour factors is experienced with long distance trips. An average operational detour factor of 1.61 was calculated for the IVT Swiss federal road network.

Figure 7 Distributions of reported distance deviations from calculated distances



Tables 7, 8, and 9 show how the network distances (Loaded network distance, shortest path distance and shortest time distance, respectively) are distributed for a given range of crow-fly distances. Detour factors for different distance ranges are also calculated. It is found that the most of the shortest path distances are spread in immediate two lower ranges. The spread of travel distances into other distance ranges is more as travel distance increases At the same time, percent retained trips within a particular distance range is reduced as the distance increase

Table 7 Distribution of crow-fly distance based domestic trips with the domestic trips based on the network distance, and detour factor (passenger and driver)

Trip length (km)		Loaded network distance									
		0 - 25	25 - 50	50 - 75	75 - 100	100 - 125	125 - 150	150 - 200	200 - 250	above 250	Detour factor
Crow-fly distance	0 - 25	**99.93**	57.94	3.42	3.84	0.12	0.09	0.00	0.00	0.40	1.63
	25 - 50	0.07	**42.05**	77.02	10.34	1.32	0.09	0.00	0.00	0.31	1.47
	50 - 75	0.00	0.01	**19.55**	78.68	30.02	4.56	3.51	0.88	0.00	1.48
	75 - 100	0.00	0.00	0.00	**7.14**	67.84	54.77	14.15	0.00	3.83	1.48
	100 - 125	0.00	0.00	0.00	0.00	**0.70**	40.49	48.39	13.08	29.67	1.63
	125 - 150	0.00	0.00	0.00	0.00	0.00	**0.00**	32.97	32.88	4.26	1.43
	150 - 200	0.00	0.00	0.00	0.00	0.00	0.00	**0.97**	46.97	45.38	1.51
	200 - 250	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**6.18**	11.57	0.48
	above 250	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**4.58**	0.76
	Detour factor	1.60	1.57	1.76	1.75	1.69	1.60	1.63	1.68	3.23	**1.61**

Table 8 Distribution of crow-fly distance based domestic trips with the domestic trips based on the network distance, and detour factor (passenger and driver)

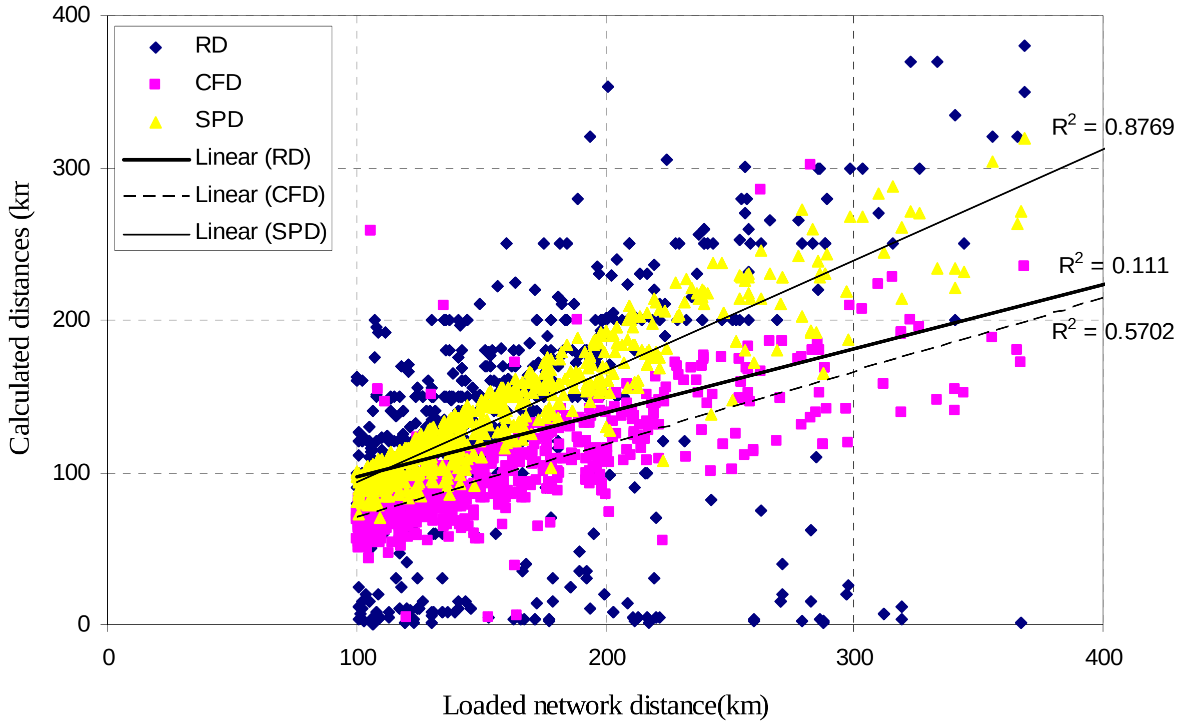
	Trip length (km)	Shortest path distance									
		0 - 25	25 - 50	50 - 75	75 - 100	100 - 125	125 - 150	150 - 200	200 - 250	above 250	Detour factor
Crow-fly distance	0 - 25	**99.31**	56.51	2.55	0.34	0.00	0.00	0.10	0.00	2.38	1.56
	25 - 50	0.29	**42.82**	72.54	6.63	0.07	0.00	0.00	0.00	0.00	1.32
	50 - 75	0.05	0.24	**24.32**	81.99	22.12	9.73	0.30	0.00	0.00	1.33
	75 - 100	0.02	0.25	0.21	**10.03**	73.05	55.32	5.55	0.00	0.00	1.30
	100 - 125	0.02	0.00	0.10	0.00	**3.64**	34.63	61.58	3.99	0.00	1.36
	125 - 150	0.01	0.04	0.00	0.53	0.00	**0.32**	28.74	25.25	0.00	1.25
	150 - 200	0.00	0.03	0.18	0.11	0.81	0.00	**2.94**	68.22	79.92	1.28
	200 - 250	0.02	0.10	0.10	0.00	0.31	0.00	0.00	**0.00**	17.70	0.42
	above 250	0.00	0.00	0.00	0.37	0.00	0.00	0.79	2.55	**0.00**	0.59
	Detour factor	1.54	1.43	1.58	1.43	1.38	1.49	1.43	1.42	5.41	**1.53**

Table 9 Distribution of crow-fly distance based domestic trips with the domestic trips based on the network distance, and detour factor (passenger and driver)

	Trip length (km)	Shortest time distance									
		0 – 25	25 - 50	50 - 75	75 - 100	100 - 125	125 - 150	150 - 200	200 - 250	above 250	Detour factor
Crow-fly distance	0 - 25	**99.61**	68.41	4.01	1.36	0.55	0.00	0.11	0.00	0.52	1.62
	25 - 50	0.26	**31.11**	83.06	25.56	1.49	0.00	0.09	0.00	0.00	1.47
	50 - 75	0.05	0.21	**12.20**	71.41	49.38	16.04	1.06	1.59	0.00	1.48
	75 - 100	0.02	0.16	0.34	**1.07**	47.04	68.31	24.34	0.72	0.00	1.48
	100 - 125	0.02	0.02	0.09	0.37	**0.00**	15.01	53.61	37.28	27.30	1.63
	125 - 150	0.01	0.00	0.06	0.00	0.15	**0.00**	19.36	33.28	15.42	1.44
	150 - 200	0.00	0.00	0.16	0.23	0.80	0.24	**1.42**	27.13	48.68	1.51
	200 - 250	0.02	0.02	0.09	0.00	0.00	0.39	0.00	**0.00**	3.91	0.48
	above 250	0.00	0.00	0.00	0.00	0.59	0.00	0.00	0.00	**4.17**	0.76
	Detour factor	1.59	1.57	1.76	1.76	1.70	1.60	1.62	1.68	3.24	**1.61**

In Figure 8, trips with a network distance of above 100 km are set out against the crow-fly distance, shortest path distance and reported distance. Reported distances has the most number of outliers and eventually has a lower R^2 (0.111). The shortest path distances R^2 value is consistent with both long and short distance ranges. Where as the crow-fly distances R^2 for long distances is lower than that of short distances.

Figure 8 Distributions of reported distance deviations from calculated distances



RD: Reported Distance CFD: Crow-fly Distance SPD: Shortest Path Distance

5 Conclusions

Four calculated distances namely, crow-fly distance, shortest path distance, shortest time distance and loaded network distance, were chosen to probe the accuracy of reported geographical information in the Microcensus travel 2000. The shares of aggregated mode of transport at trip level from two approaches are similar. The assumed modal hierarchy can be applied to aggregate the mode of transport. Only car trips were considered for comparison.

All the four distances were calculated at stage level. Geo-codes were used to calculate crow-fly distances. For network distance calculations, IVT's Swiss federal road network was loaded with an O-D matrix and calibrated with the average annual working day traffic. All the stage distances are summed up to obtain the trip distances. Both the shortest time distance and loaded network distance are equal. Only the trips with valid calculated distances, about one-third of total reported trips, are considered for comparison.

Accuracy of reported geographical information is probed by comparing the network distances and crow-fly distances. All the three network distance distributions are consistent and registered a high R^2 value. Trip length distributions of the selected trips are calculated based on the distance data. The following hierarchy was found from this analysis:

Crow-fly Distance<=Shortest Path Distance<=Reported Distance<= Loaded network distance

It is also found that, except for the short distances (up to 15 km), the reported distances are in between shortest path and loaded network distances. Reported distances are found higher than all the network distances for short distance trips (up to 15 km). The main reason behind this result is the calculation of network distances between centroids of the municipalities. All the calculated distances were plotted against the reported distances. Most of the deviations, which may include outliers, were observed in short distance trips.

Results obtained from this plot consolidate the above conclusions. Detour factors were calculated for different ranges of crow-fly distances. An average detour factor of 1.61 was found from the analysis. The network based distance distributions were analysed for different distance ranges. Analysis shows that the outliers are very few. Finally calculated distances and reported distances were plotted against loaded network distances for long distance trips. The network distances are consistent for both the long and short distance trips. Where as, more number of outliers / deviations were found in the crow-fly based long distance trips.

6 Literature

Jermann J. (2003) *Geokodierung Mikrozensus 2000*, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Resource Systems Group (1999) Computer-Based Intelligent Travel Survey System: CASI/Internet Travel Diaries with Interactive Geo-coding. U. S. Department of Transportation.

Rietveld P., B. Zwart, B. Van Wee, and T. van den Hoorn (1999) On the relationship between travel time and travel distance of commuters. Reported versus network travel data in the Netherlands, *The Annals of Regional Science,* Vol. 33, pp. 269-287

Tschopp, M. und P. Keller (2003) Raumstruktur-Datenbank: Gemeinde-Zuordnungstabelle, *Arbeitsberichte Verkehrs- und Raumplanung*, **170**, IVT, ETH Zürich, Zürich.

Vrtic M., P. Fröhlich und K.W. Axhausen (2002) Schweizerische Netzmodelle im Strassen- und Schienenverkehr, Jahrbuch der Schweizerischen Verkehrswirtschaft 2002/2003.

Welmsley D. J. (1987) *Urban living*, Longman scientific and technical, Essex.

Zmud, Johanna and Jean Wolf (2003). Identifying the Correlates of Trip Misreporting - Results from the California Statewide Household Travel Survey GPS Study, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

Die *Arbeitsberichte Verkehrs- und Raumplanung* dienen der schnellen Verbreitung der Ergebnisse der Arbeit der Mitarbeitenden und Gäste des Instituts. Die Verantwortung für Inhalt und Gestaltung liegt alleine bei den Autor/innen.

The *Working Papers Traffic and Spatial Planning* are intended for the quick dissemination of the results of the members and guests of the Institute. Their content is the sole responsibility of the authors.

Eine vollständige Liste der Berichte kann vom Institut angefordert werden:

A complete catalogue of the papers can be obtained from:

IVT ETHZ
ETH Hönggerberg (HIL)
CH - 8093 Zürich

Telephon: +41 1 633 31 05
Telefax: +41 1 633 10 57
E-Mail: sekretariat@ivt.baug.ethz.ch
WWW: www.ivt.baug.ethz.ch

Der Katalog kann auch abgerufen werden von:

The catalogue can also be obtained from:

http://www.ivt.baug.ethz.ch/veroeffent_arbeitsbericht.html