# VALIDATING TRAVEL TIMES CALCULATED ON THE BASIS OF TAXI FLOATING CAR DATA WITH TEST DRIVES


**Elmar Brockfeld ***
**Institute of Transport Research, German Aerospace Center**
**Rutherfordstrasse 2, 12489 Berlin, Germany**
**phone: +49 30 67055 231, fax: +49 30 67055 202**
**email: elmar.brockfeld@dlr.de**

**Bert Passfeld**
**Institute of Transport Research, German Aerospace Center**
**Rutherfordstrasse 2, 12489 Berlin, Germany**
**phone: +49 30 67055 284, fax: +49 30 67055 202**
**email: bert.passfeld@dlr.de**

**Peter Wagner**
**Institute of Transport Research, German Aerospace Center**
**Rutherfordstrasse 2, 12489 Berlin, Germany**
**phone: +49 30 67055 237, fax: +49 30 67055 202**
**email: peter.wagner@dlr.de**



For Presentation as a scientific paper
at the
14th ITS Conference
October 9th- 13th, 2007
Beijing, China

Submission date: January 31st 2007

Keywords:
Probe Vehicle Data, evaluation study, traffic monitoring



**Corresponding author*

**Abstract.** In contrast to traditional traffic sensors like induction loops, the position data from fleets of Probe Vehicles are an excellent technology to determine travel times which are needed for information and route guidance systems. This contribution describes an application of a taxi-PVD system in Hamburg with about 700 taxis. Here, an information system was set up which delivers travel times accessible via internet. The information is validated with test drives of independent vehicles driving on the routes. It is demonstrated that a taxi-PVD system like this is able to deliver reliable information about travel times.

# INTRODUCTION

Traffic monitoring using “Probe Vehicle Data” (PVD) has become an important issue of transport research recently. There are three reasons for this: First, new popular mobility services require such data which are not available with already established data collection technologies like inductive loops, infrared sensors or video images. Especially the boom of navigation devices drives an increasing demand for true and up-to-date travel time information not observable with conventional measuring methods. Second, a lot of communities suffer from heavy traffic on one hand and very short resources for additional infrastructure investments on the other hand. And third, more and more fleet operators use positioning techniques to manage their vehicles producing some kind of Floating Car Data as a side effect with nearly zero additional costs.

The probe vehicle data technology (PVD) is a relatively new approach to collect traffic data. In contrast to usual approaches vehicles which float with the traffic stream are used as sensors to give information about the traffic states. In recent years, the Institute for Transportation Research (IVF) has developed algorithms and technologies to exploit these data, especially taxi-fleets have been used in several applications as probe vehicles.

The quality of this new sensor is - by now - not extensively analysed. The big challenge is that in contrast to usual traffic sensors only the data of a small amount of vehicles can be used to represent the traffic state of a complete traffic stream. Thiessenhusen et al did some first analysis of travel times on urban roads on the basis of some hundred probe vehicles in some cities (see (1)). Recently Jang et al presented analyses concerning the optimum number of probe vehicles to get reliable travel times in Seoul (see (2)).

This paper describes the set up of a taxi-PVD system with about 700 taxis in Hamburg, Germany, and a small measurement campaign with independent vehicles. This campaign was used to assess the reliability of the calculated travel times using such an approach. A description of the basic taxi-PVD system and the concrete set-up in Hamburg is described in section 1. The design of the measurement campaign with two test vehicles is described in section 2, followed by the main results in section 3. Finally, the conclusions are presented in section 4.

# THE TAXI-PVD SYSTEM

The taxi-PVD system was developed by the institute (IVF) and has been implemented in various cities in Europe and Asia (see (3), (4), (5) for some examples). The main output of the system, which is travel times on the links of the network under surveillance, is used in web applications where the traffic states of a whole city are being displayed in a map (see (6), (7)). Since detailed travel times are available, the data can in principle be used for other information services and traffic management purposes.

So far, the basic data sources for these systems are taxis which are equipped with GPS devices. The basic structure of the system is displayed in figure 1. The GPS data of the taxis are sent to the taxi headquarter, which sends them to the PVD-server of the institute. There the data are processed and used for mobility services like a traffic state map in the internet.

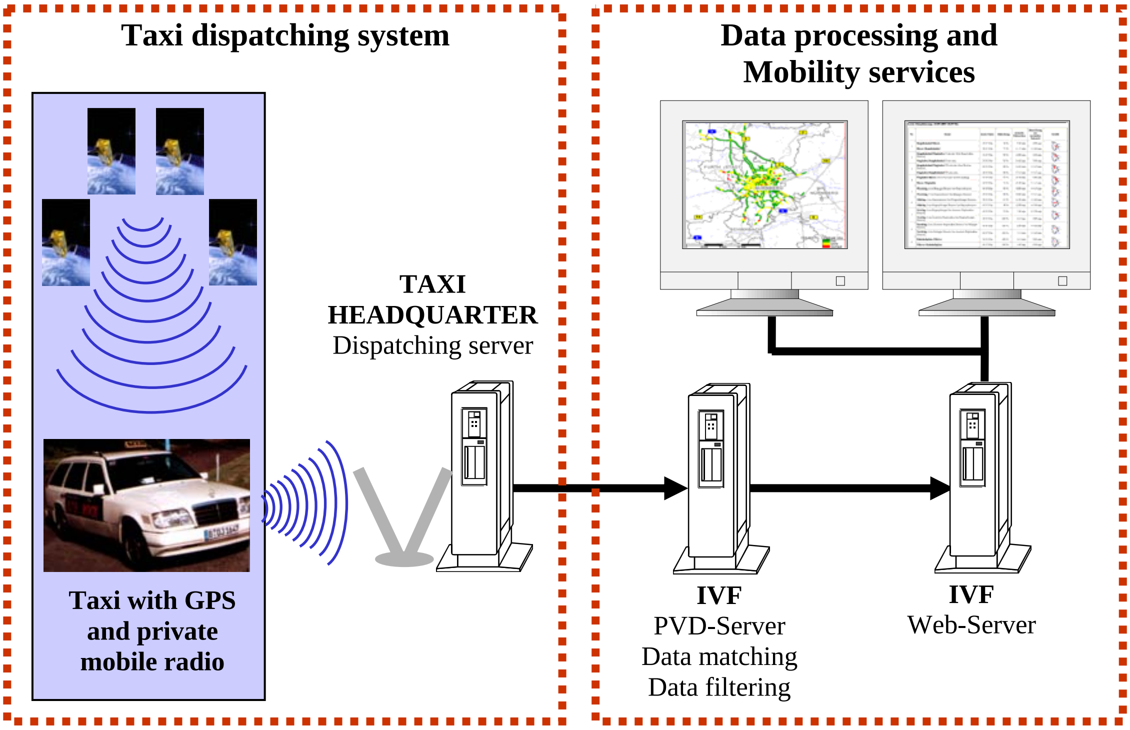


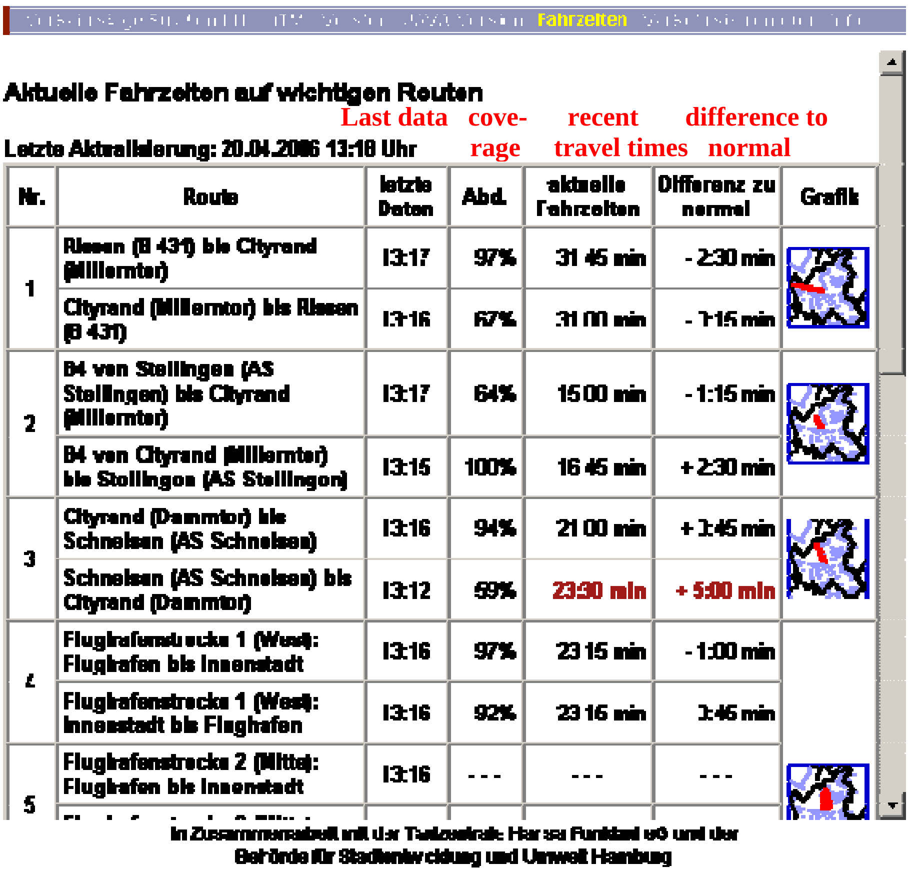
Figure 1: Basic structure of the taxi-PVD system.

In Hamburg about 700 taxis are collecting traffic data for the system. The data are sent in real-time and are used primarily for the dispatching system of the taxi headquarter, which uses the information to assign carriage orders to the taxis in a quick and fair way. The GPS positions are sent by any taxi periodically to the headquarters. The cycle times of the positioning are limited by the bandwidth of the communication channel and vary between about 20 seconds and 5 minutes - in most cases up to 3 minutes - dependent on the status of the individual taxi, which may be “searching for people to carry”, “carrying people to destination” and so on. The collected GPS positions are sent then every 10 minutes to the server of the institute, which calculates the traffic state on each street segment of the network and creates a traffic state map out of it. The GPS positions together with the identifiers of the taxis are put together to form individual vehicle trajectories and these trajectories are matched to a commercial digital map. In particular, if the positions of two subsequent GPS values of a taxi are far away, a routing algorithm (like Dijkstra’s shortest path algorithm (8)) tries to identify the most likely route the taxi drove and assigns the corresponding travel times to the connecting links. This way, travel times can be assigned to each single link in the street network where taxis recently drove. Of course, all data are filtered for bad GPS signals (caused, e.g., by clouding or multi path signals) and implausible values in the data sources and the calculated results. Furthermore, data are filtered to eliminate non-representative data. These occur when taxis are allowed to use special bus lanes or the taxis are picking up / dropping off a passenger. Altogether, roughly 10-20 % of the data have to be discarded.

The resulting travel times on each street segment are then used to display maps which show the recent traffic states or - more important here - typical routes in a city which can be monitored concerning their travel times. Of course, the taxis are not always driving a complete route to be monitored, but parts of it. Thus, because typically not for every time and place recent taxi data are present, a historical travel time database - which contains daily variations - is used to supplement those links which have no recent data available.

This way routes can be monitored at every time, typically with the additional information about the recent coverage by PVD on the links of the route.

The travel time monitoring solution for Hamburg is as shown in figure 2. The recent travel times on the routes are displayed together with the difference to the normally expected travel time, a coverage value and the information about the age of the most recent data obtained for a route. The website is updated every 10 minutes and in case of significant deviations from the normal traffic states, the corresponding values are highlighted by red colour (see figure 2, route 3 f. ex.).




| Nr. | Route | letzte Daten | Abd. | aktuelle Fahrzeiten | Differenz zu normal | Grafik |
|---|---|---|---|---|---|---|
| 1 | Rlssen (B 431) bis Cityrand (Millernter) | 13:17 | 97% | 31 45 min | - 2:30 min | |
| | Cityrand (Millerntor) bis Rissen (B 431) | 13:16 | 67% | 31 00 min | - 1:15 min | |
| 2 | B4 von Stellingen (AS Stellingen) bis Cityrand (Millernter) | 13:17 | 64% | 15 00 min | - 1:15 min | |
| | B4 von Cityrand (Millerntor) bis Stellingen (AS Stellingen) | 13:15 | 100% | 16 45 min | + 2:30 min | |
| 3 | Cityrand (Dammtor) bis Schnelsen (AS Schnelsen) | 13:16 | 94% | 21 00 min | + 3:45 min | |
| | Schnelsen (AS Schnelsen) bis Cityrand (Dammtor) | 13:12 | 59% | 23:30 min | + 5:00 mln | |
| 4 | Flughafenstrecke 1 (West): Flughafen bis Innenstadt | 13:16 | 97% | 23 15 min | - 1:00 min | |
| | Flughafenstrecke 1 (West): Innenstadt bis Flughafen | 13:16 | 92% | 23 16 min | 3:45 min | |
| 5 | Flughafenstrecke 2 (Mitte): Flughafen bis Innenstadt | 13:16 | --- | --- | --- | |
| | [illegible] | | | | | |

In Zusammenarbeit mit der Taxizentrale Hansa Funktaxi eG und der Behörde für Stadtentwicklung und Umwelt Hamburg

Figure 2: Display of recent travel times on important routes (www.verkehrsinfo.hamburg.de).

Looking into detail exemplarily for the results of a single route in figure 3, it can be seen that the calculated travel times are of course fluctuating, but seem to be reliable. The averages of the typical weekdays were calculated from a three month period. They show a daily variation on workdays with a morning peak from about 7 to 10 A.M. The following broad afternoon peak reaches its end at about 7 P.M.

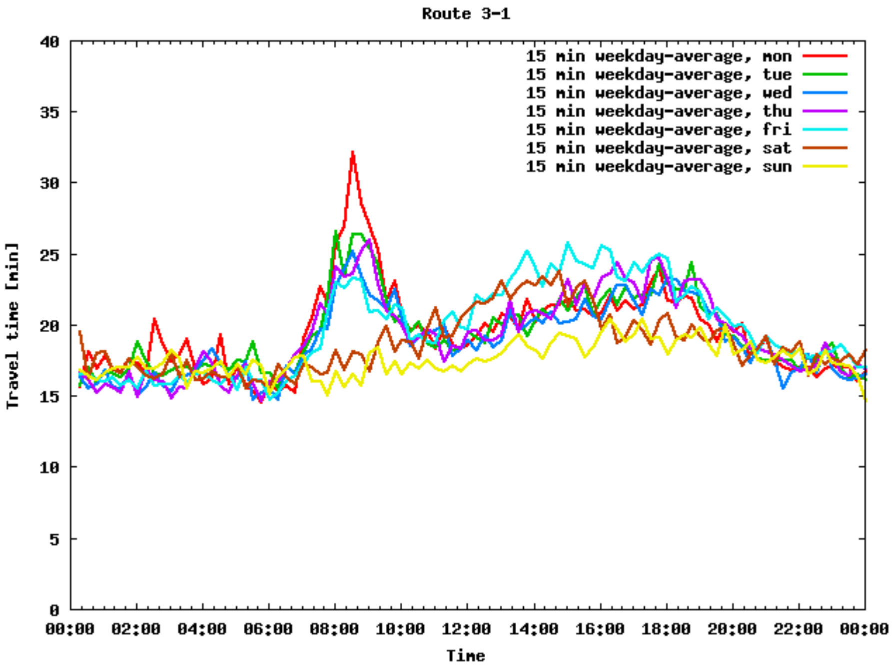



Figure 3: Typical daily variations on a route aggregated to weekdays.

By analysing the distribution around the averaged weekday values, the variations of the travel times becomes apparent. Figure 4 shows an example for the Wednesdays on a route. The average standard deviation to the 15-minute averages is about 1.8 minutes, which means a variation coefficient of about 10 %. Thus, a variation of 20 % is given in the travel times.

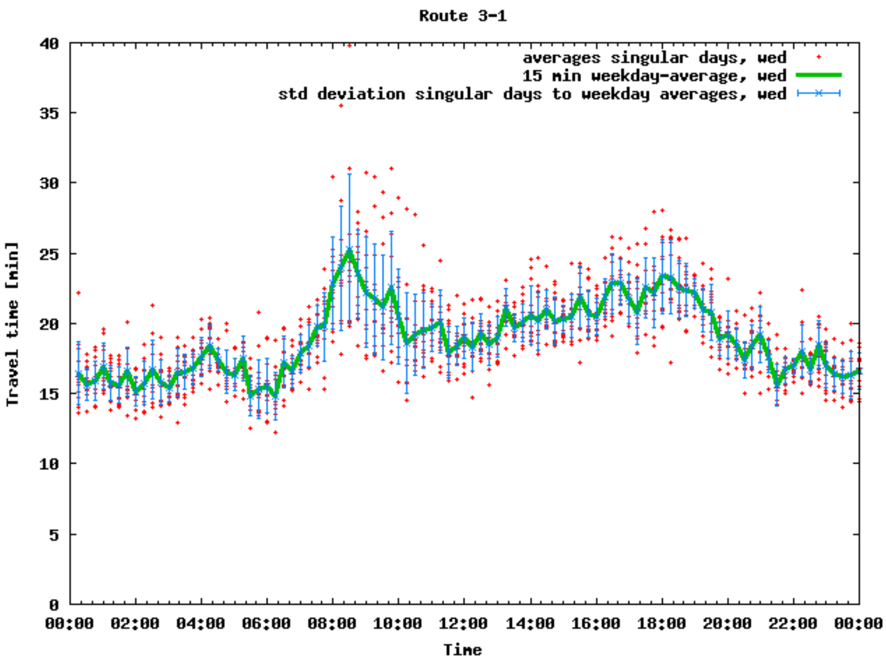



Figure 4: Daily variation curve for Wednesdays on Route 3-1 together with the 15-min averages on single Wednesdays in a three month period.

# MEASUREMENT CAMPAIGN

## MEASURING TRAVEL TIMES ON DEFINED ROUTES

The goal of the measurement campaign was to validate the information about the travel times on the routes which are presented in the internet at www.verkehrslage.hamburg.de. On the website twelve routes were already defined to be of big interest for the citizens. These are typical arterials towards/from the city center and ring roads. Out of these, seven routes were taken for the test drive campaign. The drivers were told to follow the given routes precisely.

The definition and numeration of the corridors is shown in table 1, following the notation on the website. Most routes are roughly 10 km in length, which will be long enough to provide and measure realistic travel times. The number of a corridor is composed here from the route number and a digit “1” or “2”, where “1” is typically the direction towards the city center (routes 3,4,5,6 and 7) or in case of ring roads (routes 11 and 12) the clockwise driving direction (“2” directions from city center or against clock). The exact route definitions are shown in table 1. Note, that the routes 4, 5 and 6 are a set of alternative routes from the city center to the airport. Especially in the area of the airport some short route segments are identical. Route 5 is at every day only opened for corridor 5-1 (towards center) from 4 A.M. to 12 A.M. and for corridor 5-2 from 12 A.M. to 4 A.M.

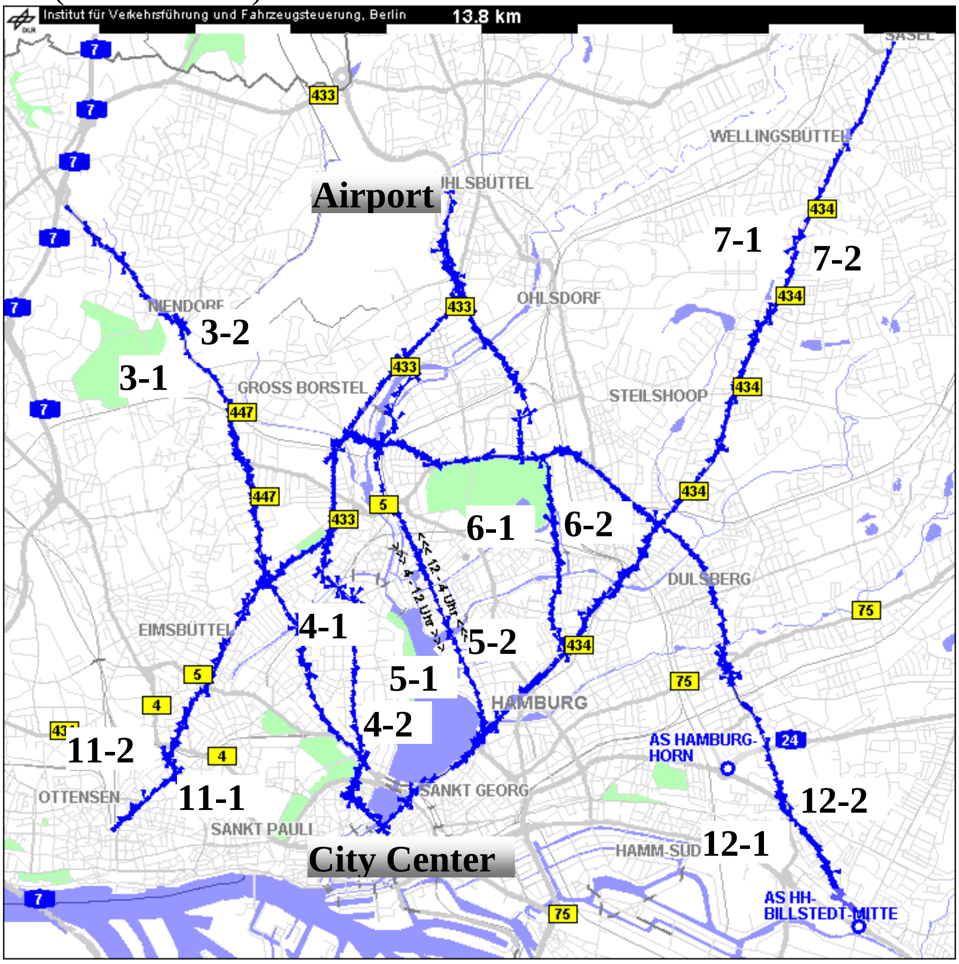



Figure 4: The seven routes which were to drive by the test drivers.

| Route-nr. | Corridor-nr. | Length (km) | Description |
|---|---|---|---|
| 3 | 3-1 | 9.8 | Schnelsen (AS Schnelsen) to border of city center (Dammtor / Stephansplatz) via B447 and Grindelallee |
| | 3-2 | 9.8 | Border of city center (Dammtor / Stephansplatz) to Schnelsen (AS Schnelsen) |
| 4 | 4-1 | 9.7 | Airport gateway to city center (Jungfernstieg / Ballindamm), west route via B433 and Rothenbaumchaussee |
| | 4-2 | 11.1 | City center (Jungfernstieg / Ballindamm) to airport access, west route |
| 5 | 5-1 | 10.3 | Airport gateway to city center (Ballindamm / Jungfernstieg), middle route via B433, Rathenaustraße, Bebelallee and Sierichstraße |
| | 5-2 | 11.2 | City center (Ballindamm / Jungfernstieg) to airport access, middle route |
| 6 | 6-1 | 10.5 | Airport gateway to city center (Ballindamm / Jungfernstieg), east route via Sengelmannstraße, Saarlandstraße and Adolph-Schönfelder-Straße |
| | 6-2 | 11.4 | City center (Ballindamm / Jungfernstieg) to airport access, east route |
| 7 | 7-1 | 13.1 | B 434 from Sasel (Ring 3) to border of city center (Ferdinandstor) |
| | 7-2 | 13.1 | B 434 from border of city center (Ferdinandstor) to Sasel (Ring 3) |
| 11 | 11-1 | 9.6 | Ring 2 (West) from Altona to city north (Railway station Altona to east intersection Überseering) |
| | 11-2 | 9.6 | Ring 2 (West) from city north to Altona (east intersection Überseering to railway station Altona) |
| 12 | 12-1 | 9.0 | Ring 2 (Ost) from city north to Billstedt (east intersection Überseering to Schiffbeker Weg) |
| | 12-2 | 9.0 | Ring 2 (Ost) from Billstedt to city north (Schiffbeker Weg to east intersection Überseering) |

Table 1: Description of the chosen corridors.

## MEASUREMENT PROCEDURE AND PREPARATION OF OBTAINED DATA

To evaluate the travel times calculated by the taxi-PVD system, two vehicles (denoted as “MFZ” in the following) were driving during two days on the routes as defined in 2.1. This happened on Tuesday, 14.3.2007 and Wednesday, 15.3.2007, respectively. On each day the vehicles drove from 9 A.M. to 6 P.M. along the defined routes. At each time the two vehicles drove different routes. The following table 2 gives an overview on the distribution of the corridor on the vehicles and days. As can be seen, the amount of test drives on each route is between 3 and 9, which will make it possible to compare some travel times obtained for each route.

| MFZ | Day | Weekday | Daytime | corridors | Amount of test drives |
|---|---|---|---|---|---|
| A | 14.03.2006 | Tuesday | morning | 51 und 62 | each 3 |
| | | | afternoon | 52 und 61 | each 5 |
| B | 14.03.2006 | Tuesday | whole day | 41 und 42 | 8 on corridor 41<br>9 on corridor 42 |
| A | 15.03.2006 | Wednesday | whole day | 111, 112, 121, 122 | each 5 |
| B | 15.03.2006 | Wednesday | morning | 71 und 72 | each 4 |
| | | | afternoon | 31 und 32 | each 6 |

Table 2: Cover of corridor drives (73 in total).

The vehicles were equipped with GPS-data loggers of type “Royaltek BlueGPS RBT 3000” which can store up to 32000 data sets in the NMEA file format. The frequency of

logging was set to five seconds, which gave finally about 4900 data sets (about 6.8 hours) from each vehicle for each day. Thus, in total 27 hours of data could be taken for the analyses.
The raw data obtained from the GPS-loggers were split into single trajectories according to the defined corridors. Thus, the complete data collected for a corridor (as displayed in figure 5, left) were split into trajectories which are defined as a sequence of GPS positions covering a complete route in one direction. Figure 5 (right) shows an example for a resulting trajectory in the segment of the route.

To calculate the travel times from this data set, the trajectories have been split manually with the help of a visualisation program (display similar to that in figure 5). Trajectories were split exactly at the positions were a route begins/ends and the according time stamps were taken for the direct calculation of a travel time.

In total, the collected GPS data were split to 73 single trajectories. One of them was not used because the driver left the prescribed route. As can be seen in figure 5 and the overview in figure 6, the GPS signals show a very good positioning and can be taken as reliable.

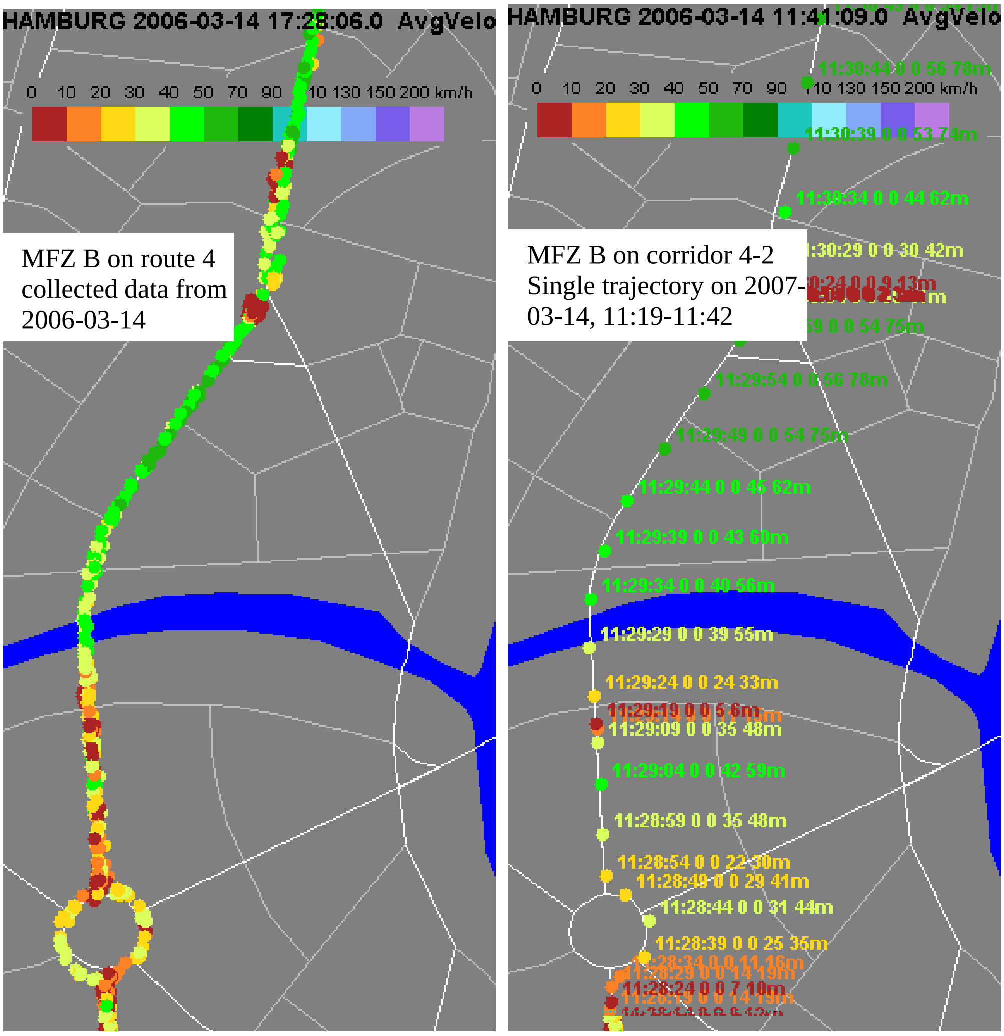



Figure 5: Collected GPS data on a segment of route 4. Complete GPS data collected (left) and example of a single trajectory (right) taken for the travel time calculation.

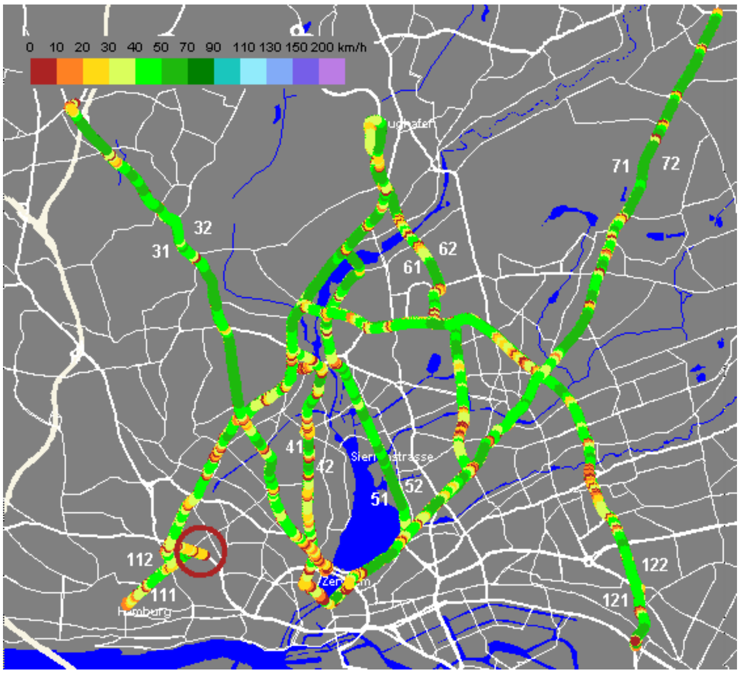



Figure 6: Collected GPS data on the seven defined routes (the red circle marks an inadvertent aberration during the campaign on corridor 11-1)

# VALIDATION OF THE TAXI-PVD SYSTEM

The validation of the taxi-PVD system with the test drive data is presented in two ways. First, a visual verification is presented in figures 8 to 14 which plot the measured travel times against the current and historic travel time values calculated from the taxi-PVD system. Second, the whole results are statistically analysed.

Figures 8 to 14 show visually the complete results for all days and all routes. In general, the typical situation (historical information) which should be expected on that day is shown and the real situation at the corresponding day as measured by the taxi-PVD system and the test drives are shown. In detail, the 15-minute averages on each single day are displayed together with the average values over all corresponding weekdays of a three month period. In addition to that, the standard deviation is shown to give an insight into the variation. Finally, the daily variation curve of the actual day calculated from the taxi-PVD system is shown and compared with the travel times resulted from the test drives. As already described in section 1, for the street segments of a route where no current data were present, historical values were taken to calculate the complete current travel time on a route.

As can be seen in figures 8 to 14, the visual comparison between taxi-PVD and the test drives shows that:

- In general, there is a good match between the travel time data from the test drives and the travel times as calculated by the taxi-PVD system.
- In most cases, the test drives match relatively good with the historical information.
- In some cases the current situation is not matched, but relatively good historical information (figure 8, route 5-1). That may occur if a jammed situation is detected by the taxis at segments the test driver already passed some minutes before.

- In single cases the test drives do not match the current nor the historical curve at all (figure 8, route 5-2). This may be because a current jammed situation detected by the test driver on a segment of the route is not detected by the taxi-PVD system.

Taking anything together, the measured travel time values typically fall within the range specified by the variance of the taxi-PVD system, concerning the historical values as well as the current values.

A more detailed validation was done by comparing the deviations between travel times obtained from the test drives and the current 15 min averages. This was done by first calculating the absolute deviation (TT_MFZ – TT_TAXI) and then the relative percentage deviation ((TT_MFZ – TT_TAXI) / TT_TAXI). Doing this for all 72 travel time deviations, gives the frequency distribution of the percentage differences as shown in figure 7.

58 out of 72 values lie between -10 % and 10 % (standard deviation is 9.4), which means that for 80 % of the compared travel times the difference is less than 10 %. Translated to travel times these 10 % mean for a typical route which can be driven in 20 minutes a deviation of about 2 Minutes.

These results may be compared to the system immanent standard deviation and variation coefficient. The standard deviation $\sigma_x$ of the current travel times to the historical 15 minute intervals is defined as:

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \; .$$

N is the number of current ravel times in a 15-min interval, $\bar{x}$ is the historical 15-min average. The variation coefficient VarK gives a percentage value and is defined by:

$$\mathrm{VarK} = \frac{\sigma_x}{\bar{x}} * 100\% \; .$$

Calculating these values over all routes and all days, an average standard deviation of about 2.0 minutes and a variation coefficient 9.7 % is obtained.

Thus, the differences between the taxi-PVD system and the results obtained from the test drives of about 10 % are in average in the range of the system immanent variation of 9.7%.

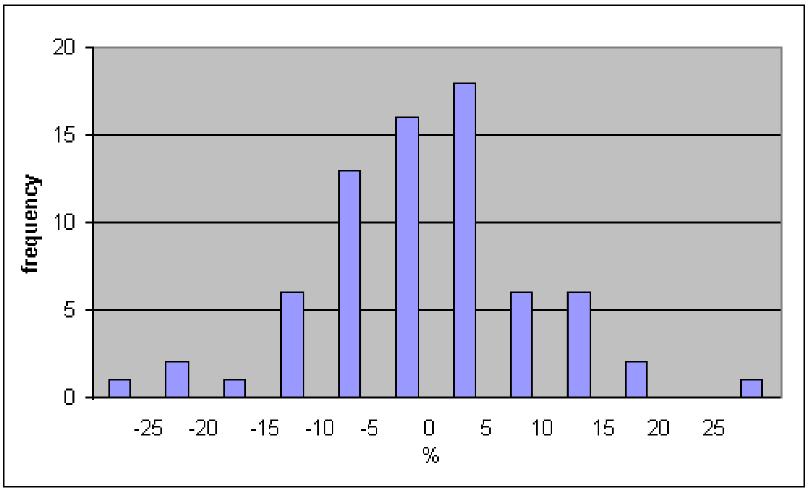



Figure 7: Absolute frequency of the percentage travel time deviations

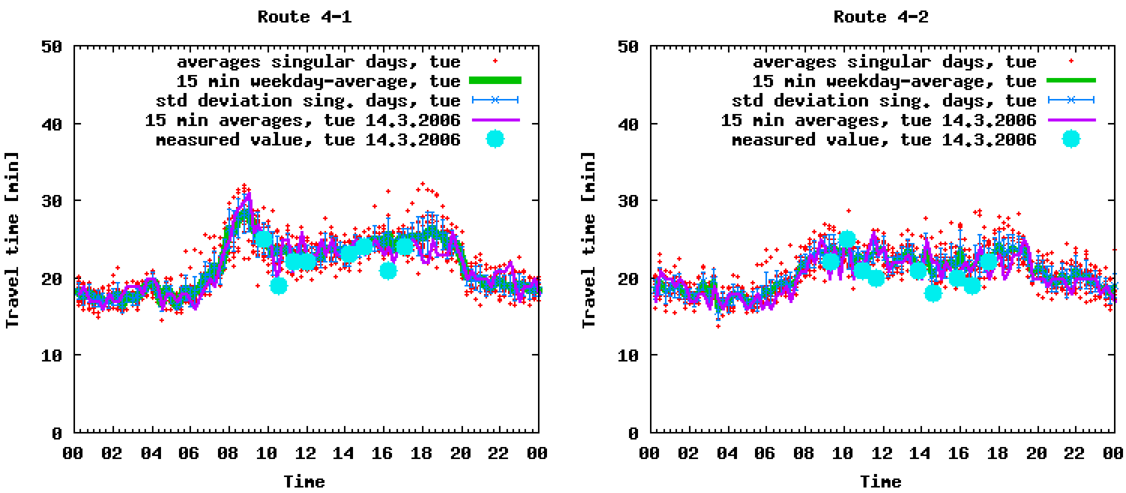



Figure 8: Travel times on Route 4 on 2006-03-14 in comparison to weekday averages, direction 1 (left) and 2 (right).

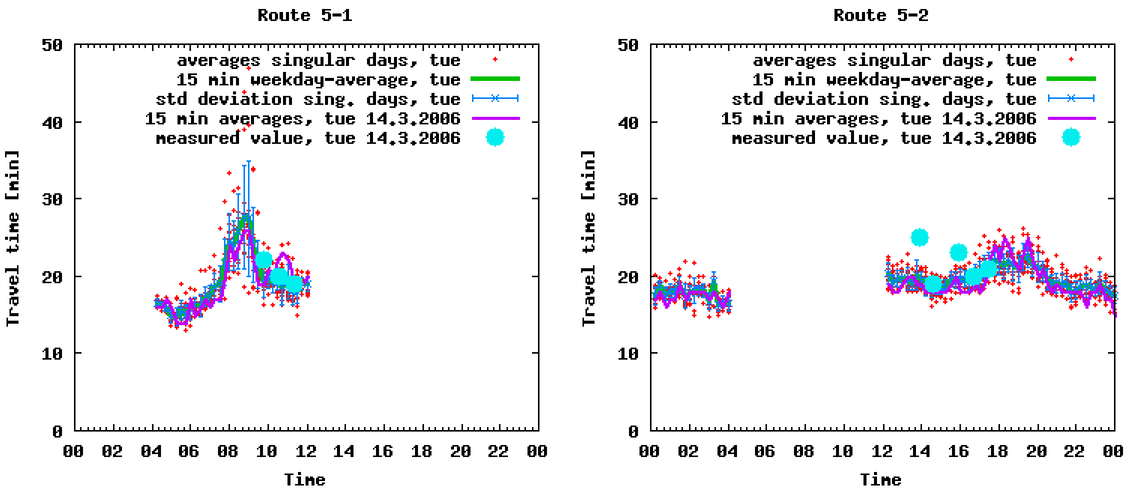



Figure 9: Travel times on Route 5 on 2006-03-14 in comparison to weekday averages, direction 1 (left) and 2 (right).

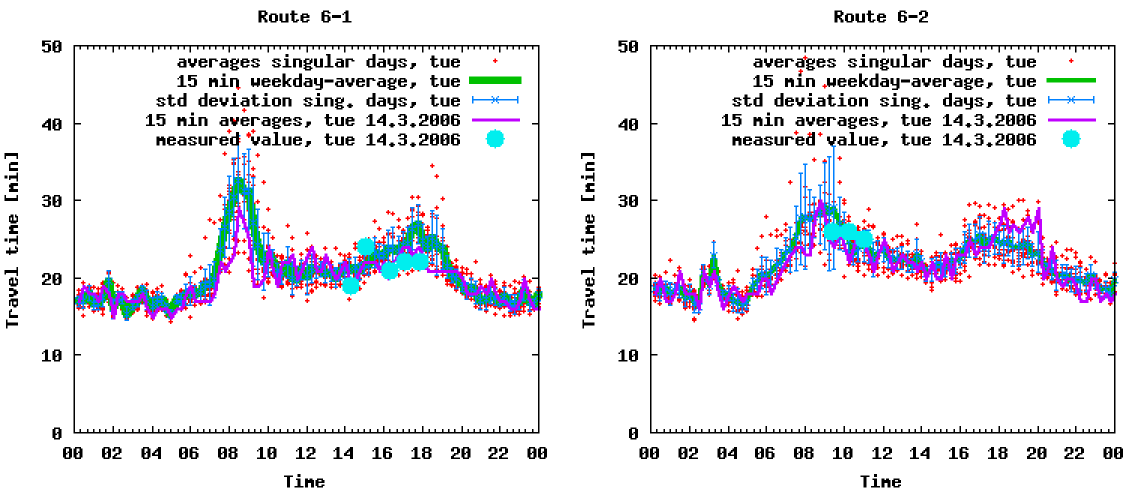



Figure 10: Travel times on Route 6 on 2006-03-14 in comparison to weekday averages, direction 1 (left) and 2 (right).

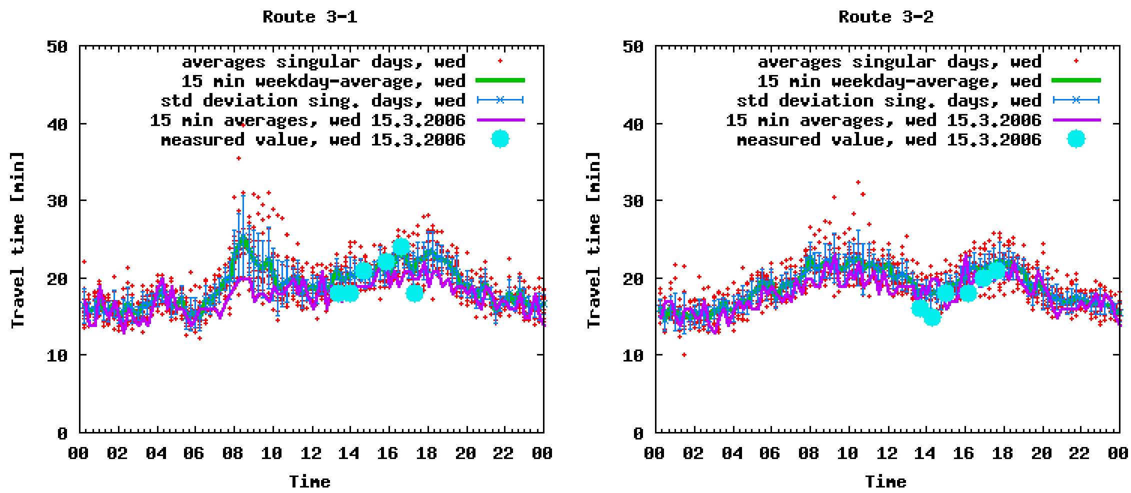



Figure 11: Travel times on Route 3 on 2006-03-15 in comparison to weekday averages, direction 1 (left) and 2 (right).

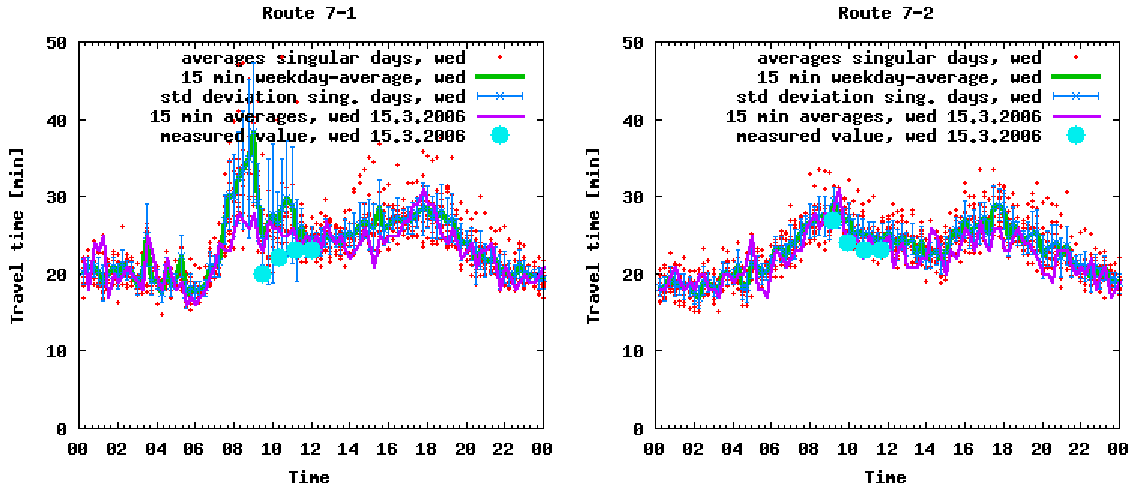



Figure 12: Travel times on Route 7 on 2006-03-15 in comparison to weekday averages, direction 1 (left) and 2 (right).

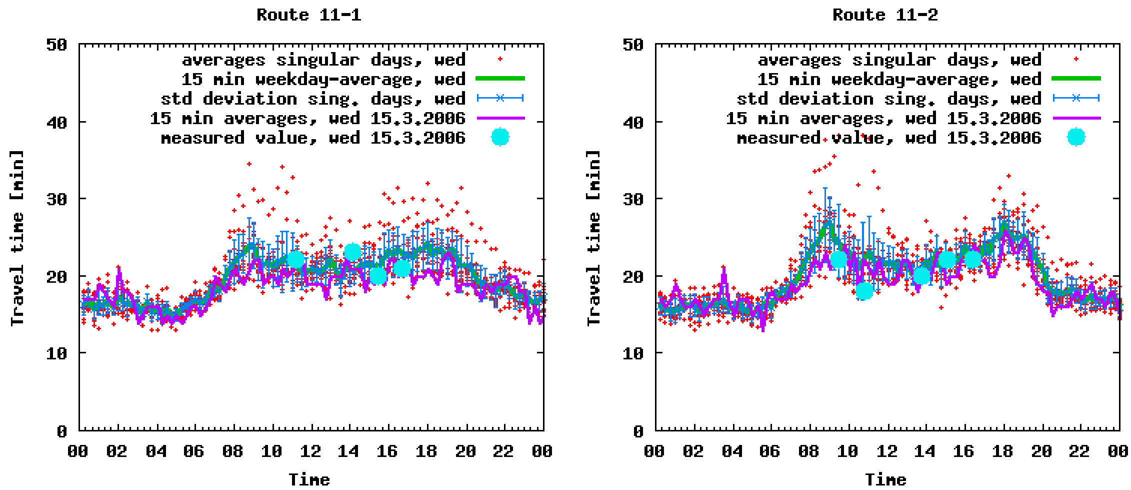



Figure 13: Travel times on Route 11 on 2006-03-15 in comparison to weekday averages, direction 1 (left) and 2 (right).

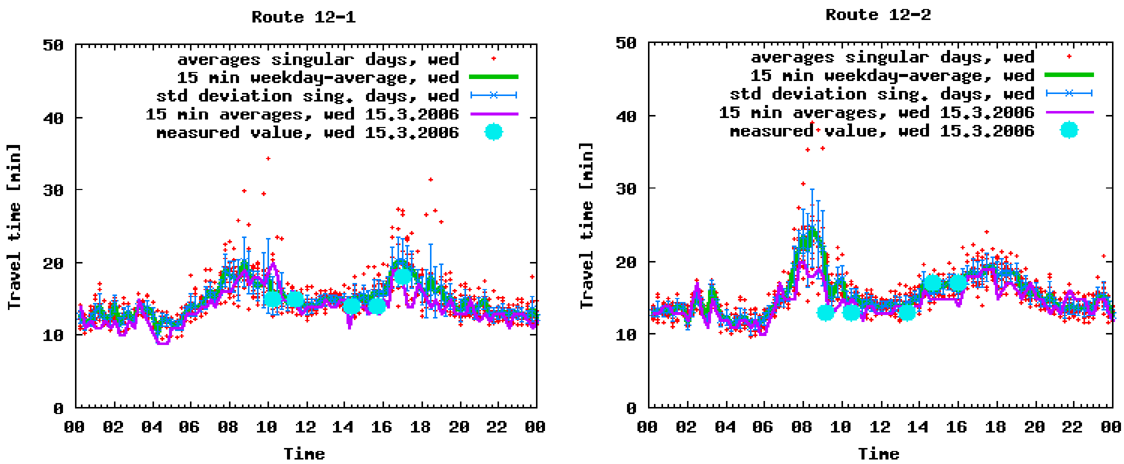



Figure 14: Travel times on Route 12 on 2006-03-15 in comparison to weekday averages, direction 1 (left) and 2 (right).

# CONCLUSIONS

The test drives could be conducted successfully and a valid data set was obtained. The current travel times generated by the taxi-PVD system could be successfully validated with the test drives.

The taxi-PVD system is able to deliver representative and valuable travel time information for mobility services. The average travel times are detected and calculated reliably. Because of the stochasticity of the data coverage using PVD, for the real time traffic state detection a data fusion with locally continuous detecting sensors like induction loops is very likely to improve the performance of the system. Especially local jammed situations may be detected better leading to more reliable travel time information on the routes.

Besides the fact that the taxi-PVD system delivers reliable information at the day, especially the analysis of the development of the most interesting morning rush hour could not be addressed appropriately with the test drives, because the test drives were only conducted from 9 A.M. to 6 P.M. For future tests this should be taken more into account.

# REFERENCES

(1) Thiessenhusen, K.-U.; Brockfeld, E.; Schäfer, R.-P.; Wagner, P. (2002): Analysis of travel times and routes on urban roads by means of floating-car data. European Transport Conference PTRC, 4th - 9th Sept 2002, Cambridge, UK.
(2) JeongAh Jang; Tae-Ho Yoo; Jeong-Whon You; Seung-Hwan Lee (2006): Determining the optimum number of vehicle probes with accounting the reliability of link travel times. 13th World Congress on Intelligent Transport Systems and Services, 9th - 12th Oct 2006, London, United Kingdom.
(3) Schäfer, R.-P.; Thiessenhusen, K.-U.; Brockfeld, E.; Wagner, P. (2002): A traffic information system by means of real-time floating-car data. ITS World Congress 2002, 11th-14th Oct 2002, Chicago, USA.
(4) Lorkowski, S.; Brockfeld, E.; Mieth, P.; Passfeld, B.; Thiessenhusen, K.-U.; Schäfer, R.-P. (2003): Erste Mobilitätsdienste auf Basis von "Floating Car Data". In: RWTH Aachen [Hrsg.]: Tagungsband zum 4. Aachener Kolloquium "Mobilität und Stadt", Stadt Region Land, 75, pp. 93 - 100, 4. Aachener Kolloquium "Mobilität und Stadt" AMUS, 31st Jul - 1st Aug 2003, Aachen (Deutschland), ISBN 3-88354-140-0.
(5) Kühne, R.D.; Schäfer, R.-P.; Mieth, P.; Lorkowski, S.; Bei, X. (2005): Vehicle Probes as Data Collectors for Asian Metropolitan Areas. 4th Asia Pacific Conference, 8th - 10th Nov 2005, Xi'an, China.
(6) www.nuernbergverkehr.de, accessed 2006-12-21.
(7) www.verkehrsinfo.hamburg.de, accessed 2006-12-21.
(8) Dijkstra, E. W. (1959): A Note on two Problems in Connections with Graphs, Numer. Math., 1, pp. 269-271.