BENEFITS AND LIMITS OF RECENT FLOATING CAR DATA TECHNOLOGY – AN EVALUATION STUDY

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

Stefan Lorkowski
TomTom Mobility Solutions Berlin
Rudower Chaussee 29, 12489 Berlin, Germany
phone: +49 30 6392 6293, fax: +49 30 6392 6291
email: stefan.lorkowski@tomtom.de

Peter Mieth
TomTom Mobility Solutions Berlin
Rudower Chaussee 29, 12489 Berlin, Germany
phone: +49 30 6392 6295, fax: +49 30 6392 6291
email: peter.mieth@tomtom.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 at the
11th WCTR Conference
June 24th- 28th, 2007
Berkeley, USA

Submission date: December 31st 2007

Keywords:
Floating Car Data, evaluation studies, traffic monitoring

**Corresponding author*

Abstract. Probe Vehicle Data or Floating Car Data (FCD) collected from vehicle fleets are an excellent technology for the traffic surveillance needed to support various applications of Intelligent Transport Systems. Data from an FCD-fleet result in travel time maps of the area under surveillance. Despite the simplicity of such a system, traffic surveillance based on such data has some disadvantages. The main shortcoming is caused by the insufficient penetration rate which has been achieved so far. Operating systems of Floating Car Data use at maximum several hundred up to few thousand vehicles in urban street networks composed of some ten or even hundred thousands of road kilometers. Questions arise about the reliability of travel time information on routes derived by very few probe vehicles. Unfortunately, most systems lack of a systematic performance evaluation. In this contribution, a taxi-based FCD system with about 500 taxis, operational running in Nuremberg, Germany, is tested with data from a measurement campaign. This campaign of four days duration had been conducted along a main street in Nuremberg using license plate recognition to estimate the travel times along the street. These data are compared to the travel time calculations obtained from the taxi-FCD system. The main result is that the FCD system is particularly able to detect jammed situations and the travel times calculated by the system deliver valuable data for mobility and traffic information systems.

1 INTRODUCTION

Traffic monitoring using Floating Car Data (FCD) 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 floating car data technology 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 data, especially taxi-fleets have been used in several applications as floating cars.

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 ([1]). Recently Jang et al. presented analyses concerning the optimum number of probe vehicles to get reliable travel times in Seoul ([2]). As in the latter study, this paper describes a measurement campaign which tries to clarify about the quality which can be reached using such an approach. A short description of the basic taxi-FCD system is described in section 2. The measurement campaign was conducted in Nuremberg, Germany, and is described in section 3. The main objective of the validation was the suitability of taxi-FCD systems for real time jam detection and is described in the results section 4. Finally, the conclusions are presented in section 5.

2 THE TAXI FCD SYSTEM

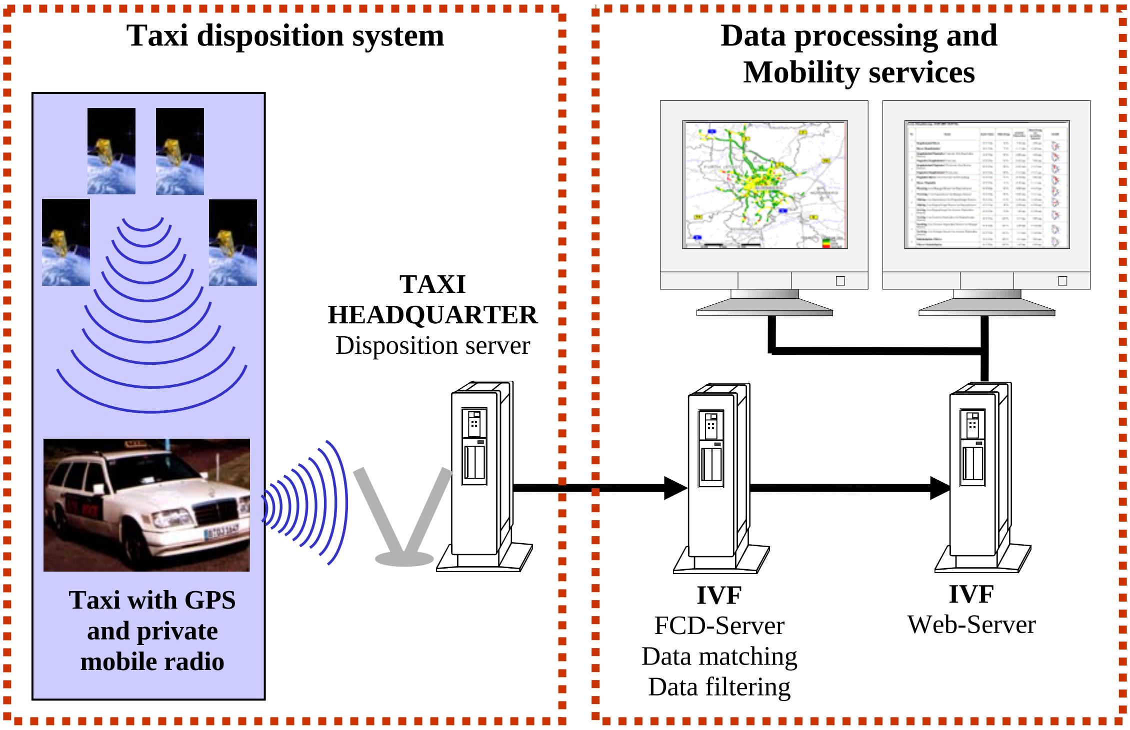


Figure 1: Basic structure of the taxi-FCD system.

The taxi-FCD system was developed by the institute (IVF) for 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 ([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 FCD-server of the institute. There the data are processed and used for mobility services like a traffic state map in the internet.

In Nuremberg about 500 taxis are collecting traffic data for the system. The data are sent in real-time and are used primarily for the disposition 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 10 and 120 seconds, 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 with cycle times of about 10 minutes to the server of the institute, which calculates the traffic states 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, which are matched then 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 singular 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 in cases of non-representative data. These occur when taxis are allowed to use special bus lanes or the taxis are picking up / deboard 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 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, usually with the additional information about the recent coverage by FCD on the links of the route.

3 DATA COLLECTION OF REFERENCE TRAVEL TIMES

3.1 Area of observation

The area of observation was a 2 km piece of road in Nuremberg which was relatively straight along the Regensburger Straße, Hainstraße and Münchener Straße. The area is shown in figure 2 together with the location of the two with license plate recognition systems equipped vehicles which are located nearby the streets.

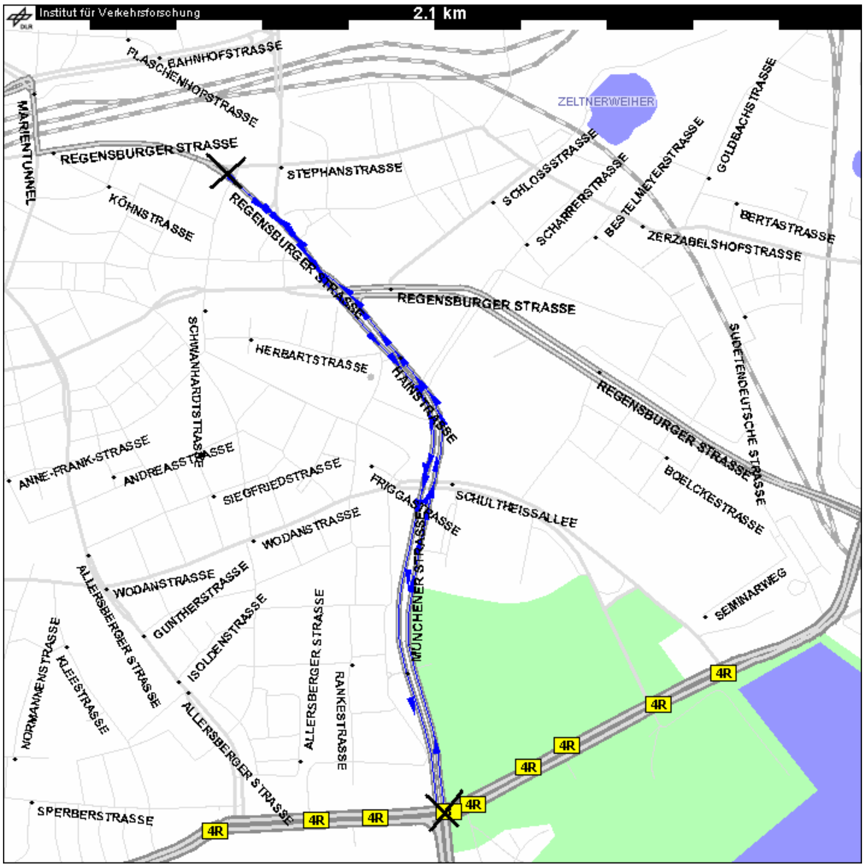


Figure 2: Area of observation (Regensburger Straße, Hainstraße, Münchener Straße)

It is an arterial towards/from the city centre (located in the north-west part of the map in figure 2). This road was taken especially because some of the signalized intersections cause strong congestion during the morning peak as well as during the evening peak, respectively. As can be seen in figure 3 which shows the average daily variation of speed on this road segment (three month average calculated from taxi-FCD), especially for weekdays the variation is significant.

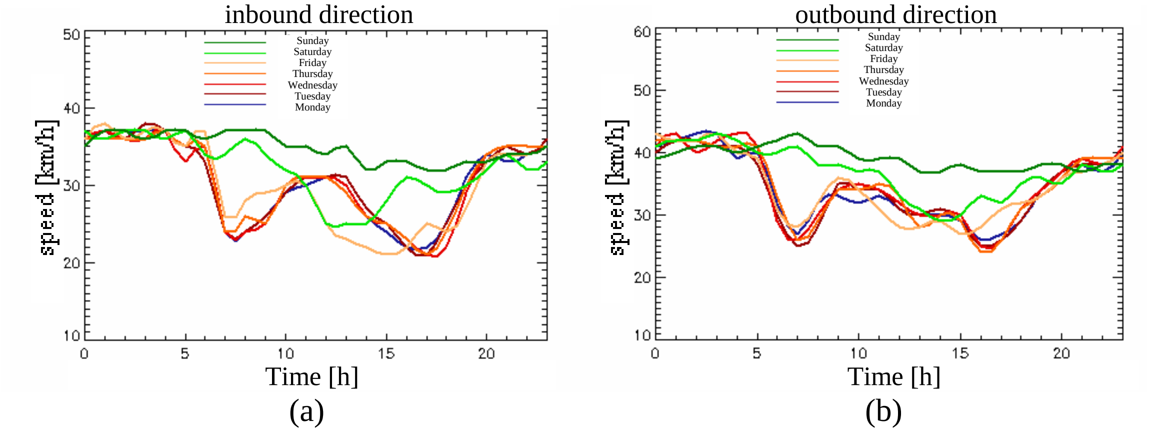


Figure 3: Daily variation of speed towards the city center (a) and away from the city center (b) (averages for each weekday in a total period of three month, gained by historical taxi-FCD).

Because of the congestion, this arterial seems to be very suitable to test the quality of the FCD system to detect traffic jams. In addition, the segment is relatively often used by taxis as can be seen in figure 4. Here, the coverage with taxi data on the main road network of Nuremberg is shown for a typical Friday, 8 AM (left) and 4 PM (right), respectively. The average values per hour are displayed resulting from the processing of single vehicles, which are matched on the network. On average, eight vehicles per hour drove on the chosen road segment, which seems to be sufficient for the evaluation.

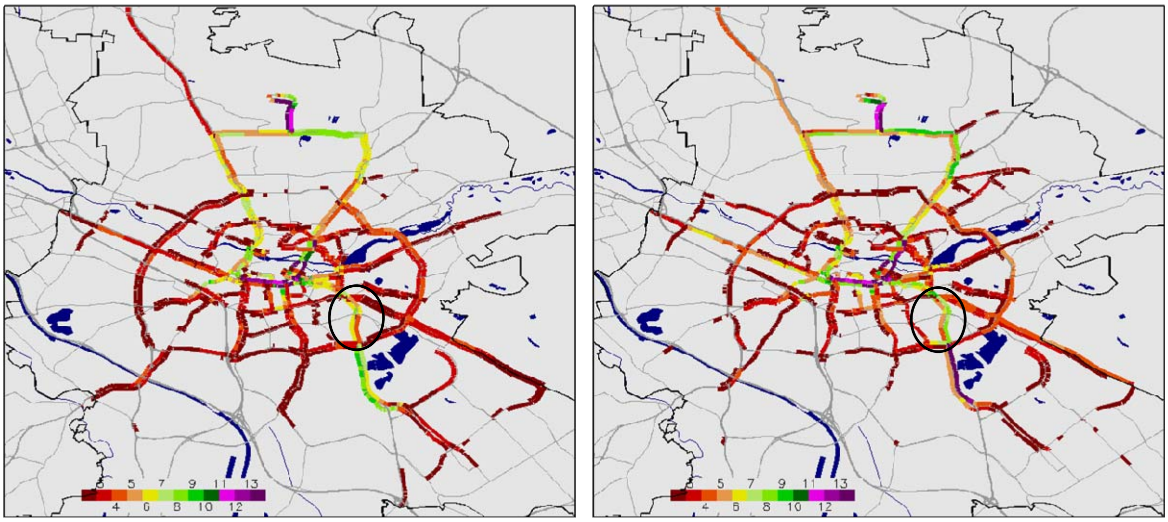


Figure 4: FCD coverage for a typical Friday at 8 AM (left) and 4 PM (right). The area of observation is marked with an oval.

3.2 Measurement procedure (License plate recognition (LPR))

The measurement campaign was conducted from Tuesday, 2005-09-13 to Friday, 2005-09-16. On each of the four days data were collected from 7 AM to 11 AM and 3 PM to 7 PM. Data were collected for the outbound direction on the first two days and for the inbound direction on the last two days. Each day was a workday and not in holiday times, thus a high traffic volume was observed. As all measurement systems worked properly all the time, a data set of the complete time period could be received.

For the license plate recognition (LPR) the hardware of the institute was used. These are two systems which are identical in construction, composed of a high resolution and a low resolution CCD-camera and a personal computer. Each camera system was installed on an equipped vehicle and positioned besides the streets at the locations marked in figure 1 and they recorded video sequences all the time. These video sequences were then analysed using the software “vidanPLATE” of the company “recognitec”. The software detects the license plates and writes them together with the correct time stamp to a file. Of course, not all license plates are detected correctly, but the number of detections received was significant. Having the list of both systems finally, they are matched to receive the travel times between the two locations for individual vehicles (same license plate in both lists). At the end, the license plates were coded to numbers and all data concerning the original license plates were deleted because of privacy concerns.

3.3 Results

At each of the two locations more than 10.000 license plates were detected, but caused by false detections not all of them were correct or could be found in both data-sets. Altogether 5351 matches had been extracted from the lists. For these the travel times could be calculated which are about 2.8 travel times per minute or about 170 travel times per hour on average. As will be seen in detail in the validation section 4, the data show characteristic daily variations and some other structural properties can be identified. Thus, this amount of data seems to form a valid basis and is adequate to validate the travel time predictions delivered by the taxi-FCD system.

A number of statistics had been computed from the data. Surprisingly, the recorded travel time data do not show a simple Gaussian structure. Instead, the distribution of travel times is a multi-modal one. This is shown exemplarily in figure 5 for the afternoon data on 2005-09-14.

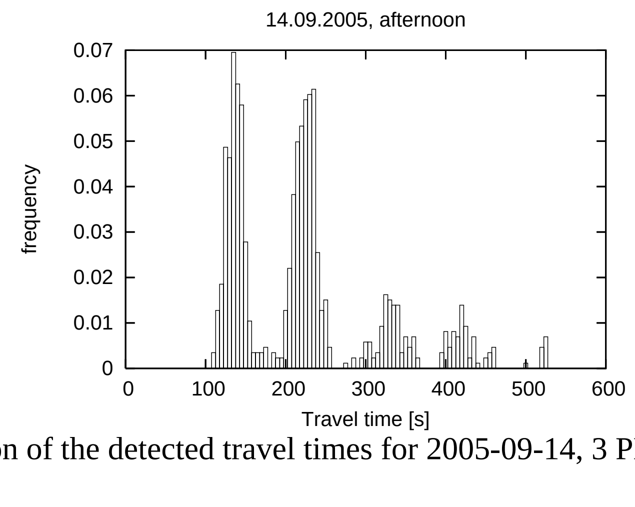


Figure 5: Distribution of the detected travel times for 2005-09-14, 3 PM to 7 PM.

It is strongly conjectured that this effect is caused by traffic signals between the two locations of measurement and also by signals upstream of the first location in driving direction. The vehicles pass already the first location in platoons, caused by a signal they passed before. Two large signalized intersections let the platoons split and the vehicles have to wait different amounts of cycles at the traffic signals. As can be seen in figure 5, most vehicles are able to drive through the 2km piece of road without any disturbance (first peak in frequency at about 140 seconds, which means an average speed of about 51 kph) or stop for one cycle at one intersection (second peak, 240 s, 30 kph). During the typical daily peak hours in this case some vehicles have to wait up to 4 cycles passing this 2km section. This can be seen, too, when looking at the development of the travel time during this period in figure 6. The diagram shows stripes at some travel times which represent the different amount of cycles the vehicles have to wait at intersections. Additionally, a typical afternoon peak from about 16:40 to 17:50 can be recognized.

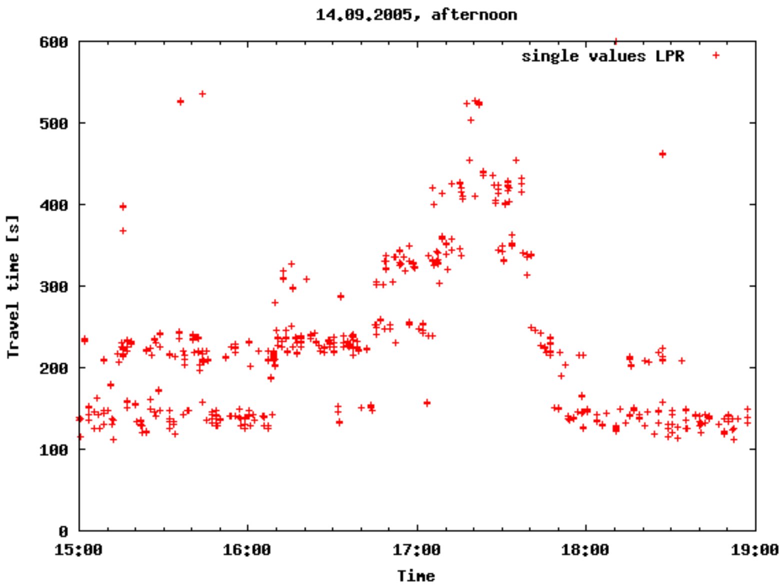


Figure 6: Travel times calculated from the LPR in the afternoon of 2005-09-14 showing stripes caused by signalized intersections and an afternoon peak.

The complete results for all four days of measurement will be shown in the following section together with the travel times calculated from the taxi-FCD system. Note that the multi-modal travel time distribution is not always present in travel time data, but depends on the details of the study area. In a very recent campaign (not reported here), no indices of a multi-modal travel time distribution have been found. So, this effect is so far not very well understood. In the near future, simulation studies will be carried out to further investigate this effect.

4 VALIDATION OF THE TAXI-FCD SYSTEM

To validate the taxi-FCD system with the data from the license plate recognition (LPR), the taxi data had to be prepared. At first, the time-corresponding taxi data for the four days were taken for the corridor. As a time stamp for the taxi data the time was taken when they were completely processed by the system. So, not the data of the single taxis themselves were validated, but the output of the taxi-FCD system which takes some minutes to provide mobility services with the data. To adequately compare the two different data sources, the taxi data as well as the LPR data were aggregated to 15 minute intervals.

Figures 7 to 10 show the complete results for all four days. Displayed are all individual vehicles detected by the LPR, the 15-min-averages and the standard deviation of the values to these averages. The 15-min-averages of the taxi-FCD system are displayed for direct comparison to the LPR averages. Additionally, the times when the complete route was covered with taxi data in the last 20 minutes is marked. This indicates that a good coverage with taxi data exists for these times. As already described in section 2, for the street segments of a route where no current data were present, historical values were taken to calculate the complete travel time on a route.

As can be seen in figures 7 to 10, the visual comparison between taxi-FCD and LPR shows that:

- significant jams and especially jams with a long duration are detected by the taxi-FCD system (see figure 7, 16:30), but sometimes only with a time delay (see figure 8, 17:40),
- short-term jams are probably not detected (e.g. figure 7, 09:45, figure 8, 18:30), if no taxi drives through the jam,
- for short time periods taxis may detect “phantom jams” (e.g. figure 8, 8:20 or figure 10, 16:00), if only a single taxi had driven on the route.

Taking anything together, the taxi-FCD-estimated travel times typically fall into the variance displayed by the LPR-measured travel times.

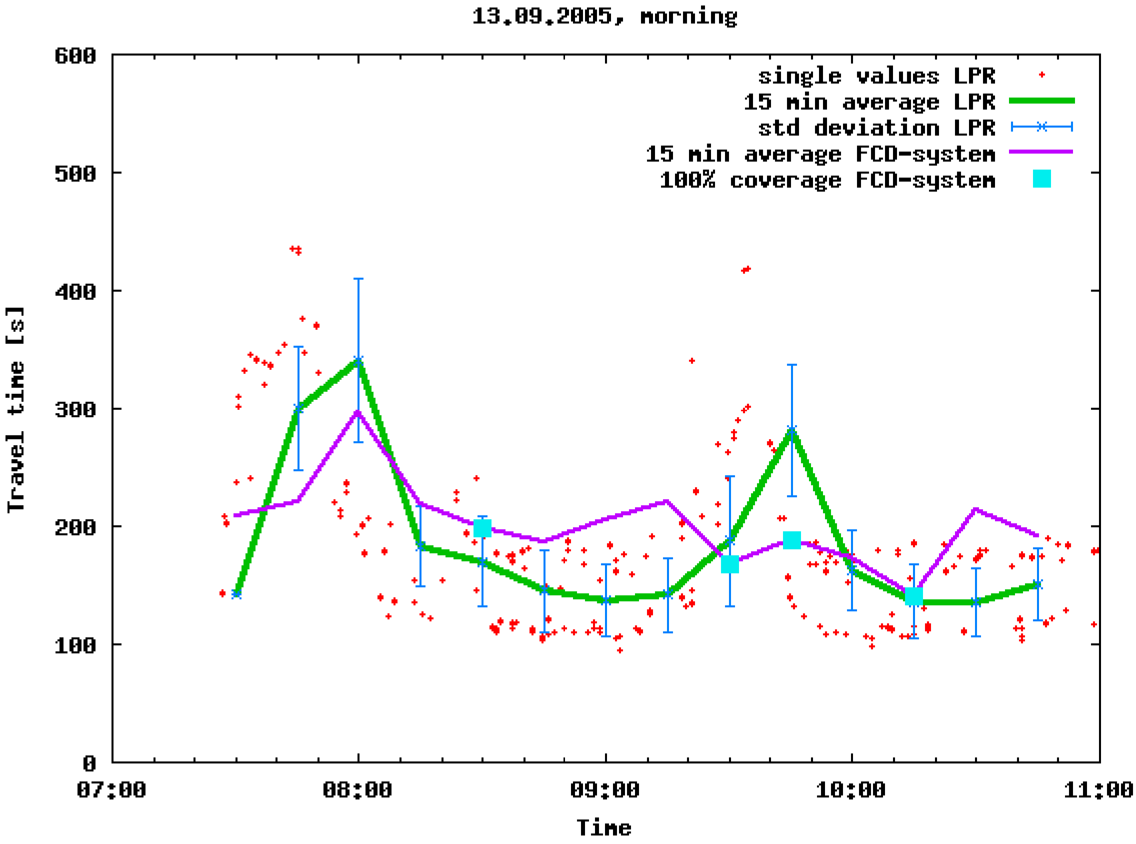
A more detailed validation was done by comparing the 15-min-averages of the two data sources statistically by calculating the standard deviations and variation coefficients. The standard deviation σ_x of the individual travel times to the 15-minute-intervals is defined as:

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=0}^{N}(x_i - \bar{x})^2} \, .$$

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

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

For the LPR data a standard deviation of about 45 seconds is obtained and a variation coefficient of about 19.4 %. This latter value may be understood as the system inherent variation coefficient. Calculating the variation coefficient of the taxi-FCD estimation to the 15-min-averages of the measured values gives a value of about 17.7 %. This is only slightly lower than the variation coefficient of the LPR data of 19.4 % and thus a satisfying result. It means that the few taxi data may be able to represent the characteristics of the whole traffic stream.



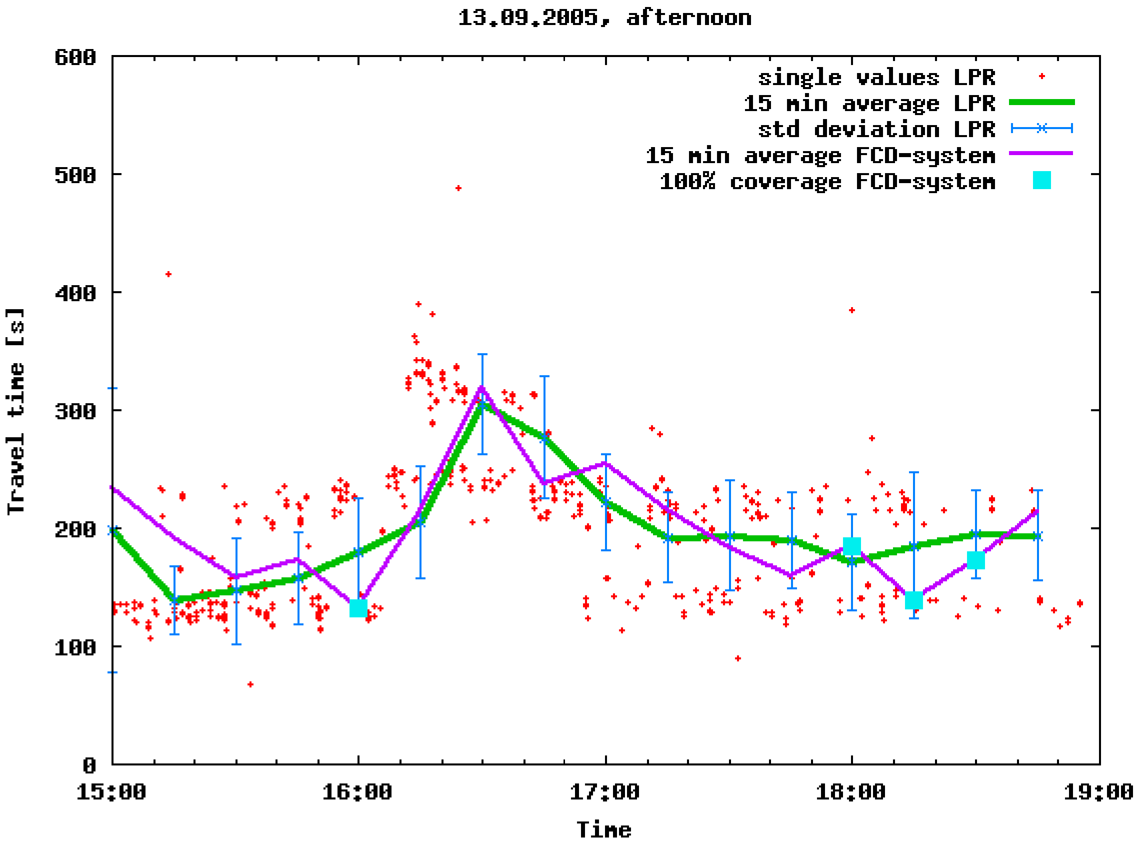


Figure 7: Travel times in the area of observation on 2005-09-13 in the morning (top) and the afternoon (bottom).

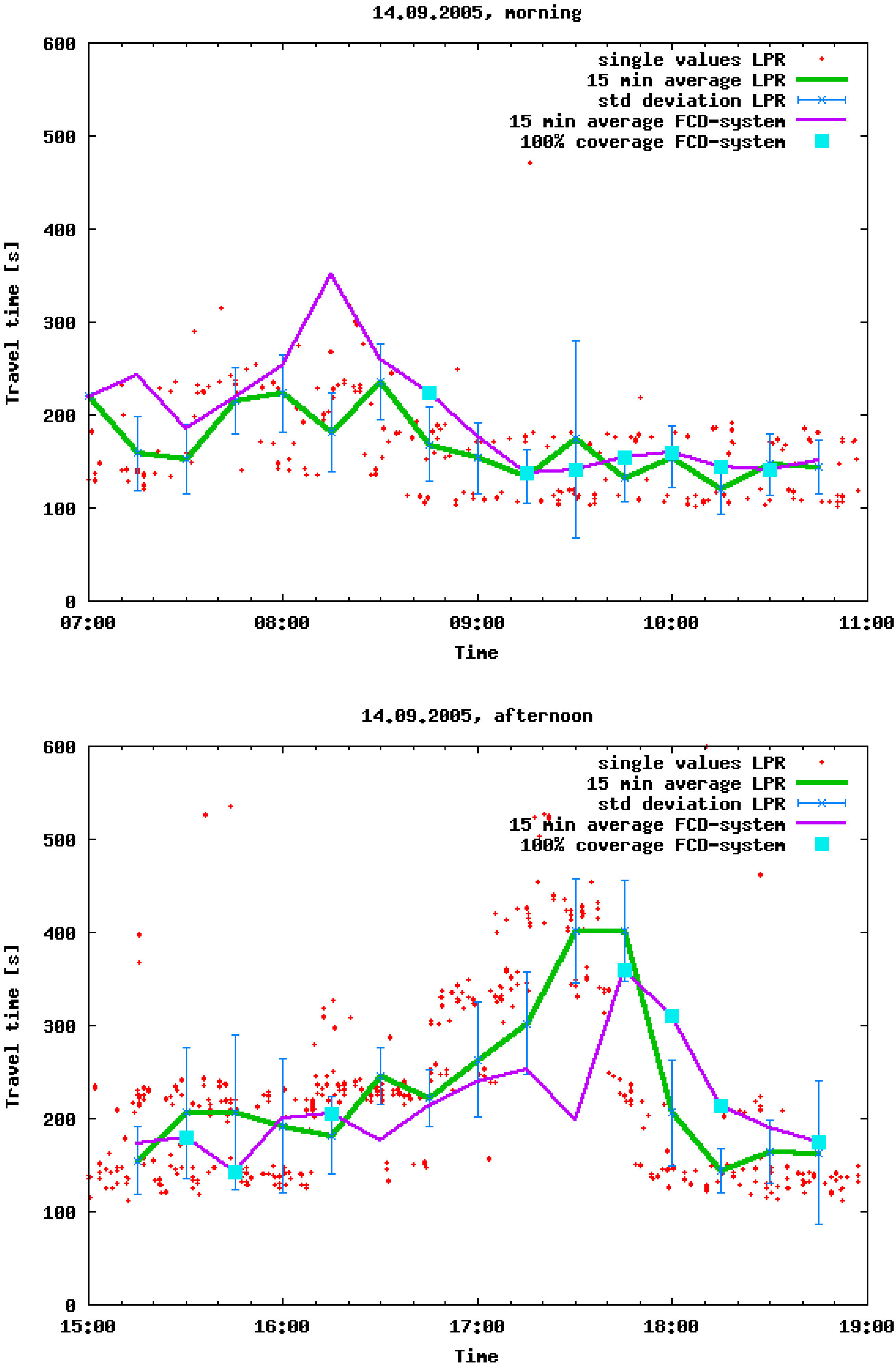
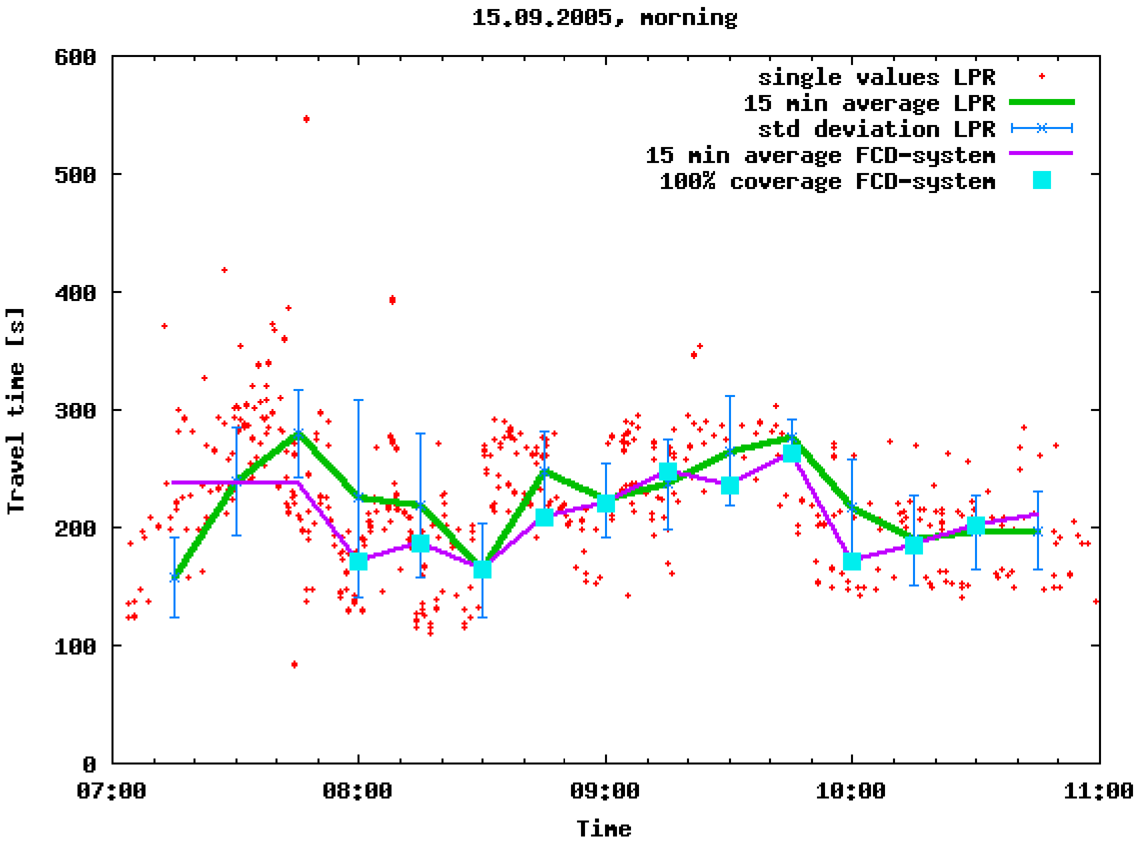


Figure 8: Travel times in the area of observation on 2005-09-14 in the morning (top) and the afternoon (bottom).



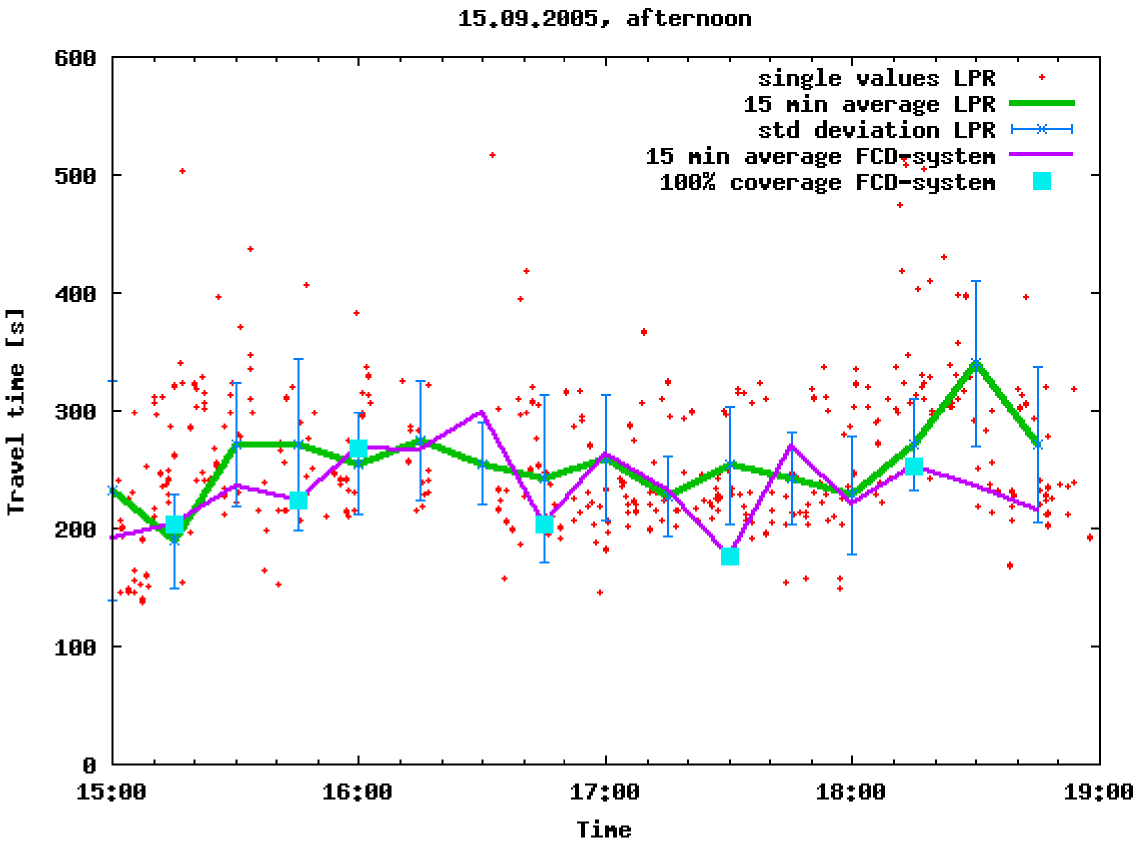
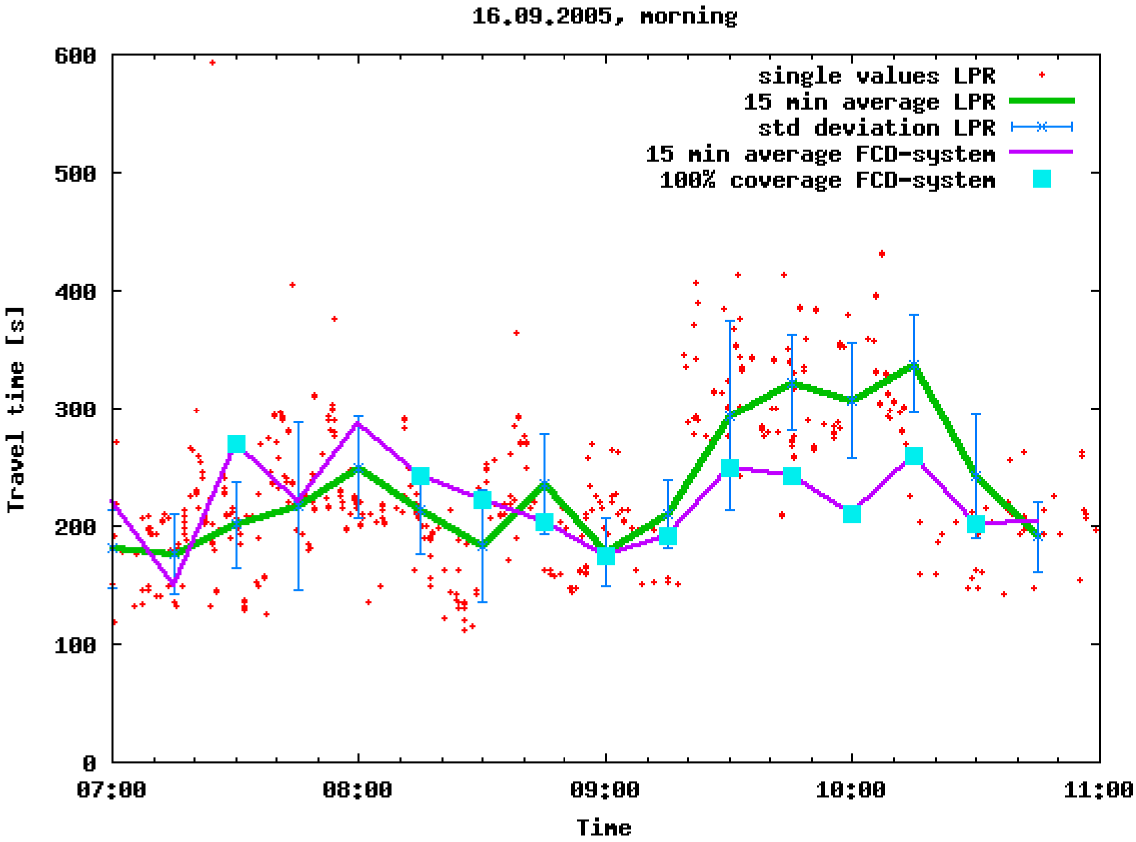


Figure 9: Travel times in the area of observation on 2005-09-15 in the morning (top) and the afternoon (bottom).



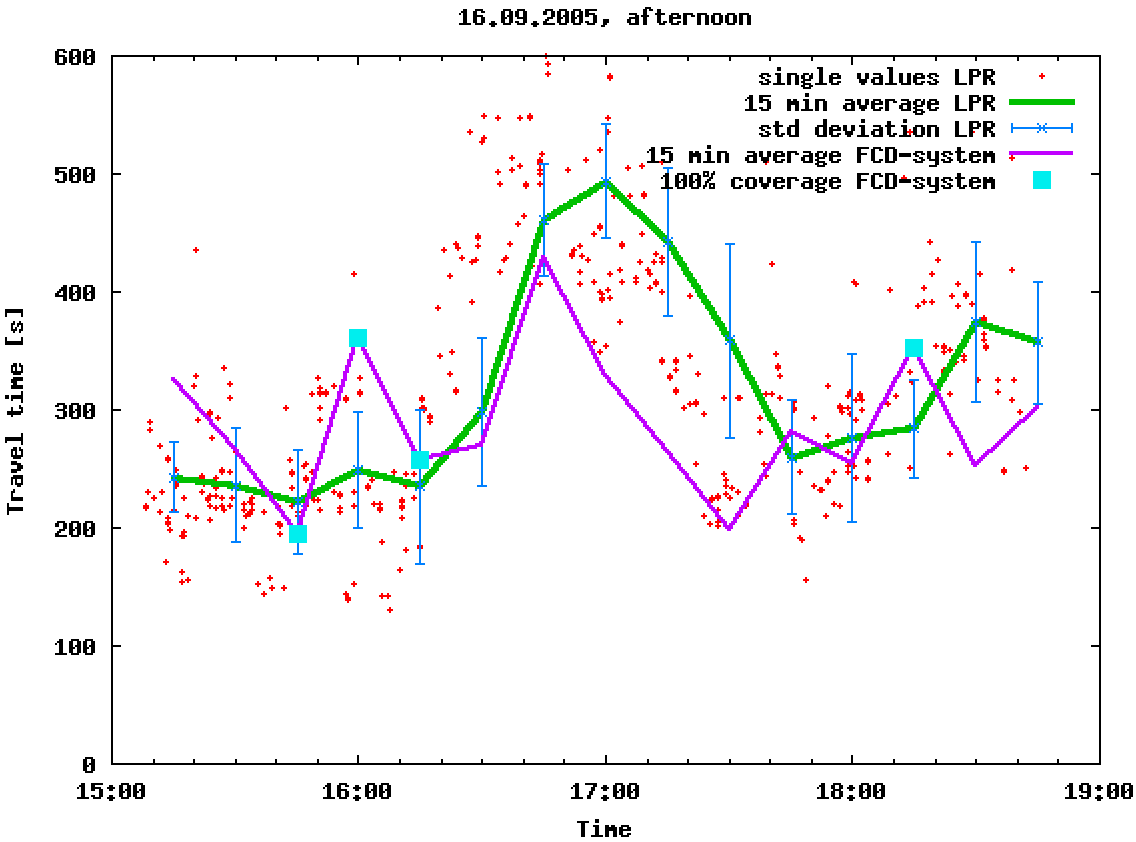


Figure 10: Travel times in the area of observation on 2005-09-16 in the morning (top) and the afternoon (bottom).

5 CONCLUSIONS

The measurement campaign to evaluate the reference travel times could be successfully conducted and it delivered a good reference for the validation of the taxi-FCD system under investigation.

Concerning the quality of the taxi-FCD system und for the area of observation the following conclusions can be drawn:

- The Taxi-FCD system is able to deliver valuable travel time information for mobility services. The average travel times are detected and calculated reliably. Because of the stochastic of the data coverage using FCD, 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.
- Long-term jams are detected with a very high trustiness, but typically with a time delay of some minutes.
- Because of the relatively low quantity and high variation of the data coverage, the taxi-FCD system as a self-sufficient sensor seems not to be suitable for special applications like controlling traffic lights in real-time.
- It might however be a very valuable tool for monitoring the quality of service of the traffic signals in a city, since on average it allows computing average waiting times upstream of a signal and therefore makes it possible to detect the slow changes caused by the ageing of the signal plans of pre-timed traffic signals.

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, 4^{th} - 9^{th} 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, 9^{th} - 12^{th} 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, 11^{th}-14^{th} 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, 31^{st} Jul - 1^{st} 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, 8^{th} - 10^{th} 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.