

# Self Evaluation of Floating Car Data Based on Travel Times from Actual Vehicle Trajectories

Günter Kuhns, Rüdiger Ebdndt, Peter Wagner, Alexander Sohr, Elmar Brockfeld

German Aerospace Center, Institute of Transportation Systems (DLR-TS)

Rutherfordstr. 2, 12489 Berlin, Germany

Email: Guenter.Kuhns@dlr.de

**Abstract— Floating Car Data (FCD) fleets are a valuable data source to obtain travel times as basis for traffic information or route guidance systems. To deliver reliable traffic information and to improve algorithms and systems for generating FCD from GPS positions their current quality has to be evaluated first. In this contribution the travel times from actual vehicle trips are compared with travel times for each edge on those trips as they result from the FCD algorithm. About 540,000 trajectories generated by more than 4,000 taxis at the four Wednesdays in October 2010 are the basis for this comparison.**

## I. INTRODUCTION

SINCE no installation of dedicated sensors is required, the effort to collect Floating Car Data (FCD) is relatively low compared to other conventional sources of traffic information and so FCD has gained increased popularity during the last decade. In contrast to other data sources FCD systems deliver travel times instead of local traffic states. Travel times are known to be more compatible when working together with popular mobility services or navigation devices.

During the last years the Institute of Transportation Systems at the German Aerospace Center (DLR-TS) has developed algorithms and technologies to exploit GPS-data from probe vehicles (“floating cars”). Especially taxi fleets have been used in several applications as probe vehicles realizing implementations for different cities in Europe and Asia.

In the past, several approaches to evaluate the quality of traffic data generated from fleets of probe vehicles have been described together with the conducted experiments. One object of investigation is the representativeness of this data which is collected by only a small fraction of vehicles within the traffic stream and the number of vehicles required for an area to generate reliable traffic data.

Thiessenhusen et al. did some first analysis of travel times on urban roads on the basis of several hundred probe

vehicles in major german cities (4). Jang et al. presented analyses concerning the optimum number of probe vehicles to get reliable travel times in Seoul (5). Reinthaler et al. analyzed travel times estimation based on FCD especially for transports of hazardous good with a field test in Vienna (6) and conducted analyses on average path speeds in Dusseldorf (7). For validation of FCD systems several measurement campaigns were conducted with test vehicles to obtain traffic data based on GPS positions with higher quality for comparison with FCD data. (3)

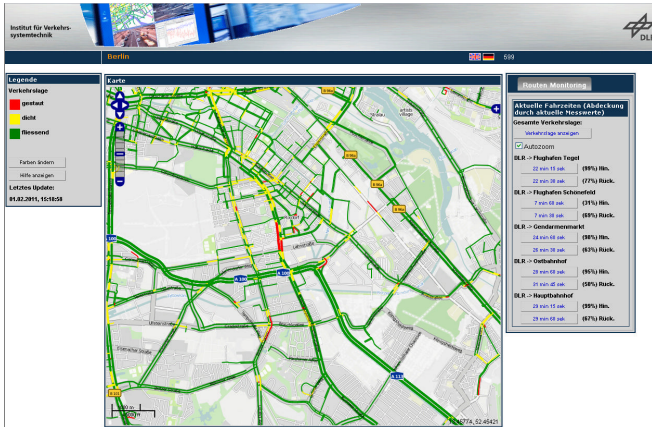
Compared to most approaches to testing the quality of FCD travel-times, this contribution uses the data itself to test their quality: any “completed” trip by a taxi is regarded as the actual measurement (it is known for sure, that the taxi has driven from GPS position 1 to GPS position 2 and that it took exactly this travel time) and compared with the travel time that the FCD system would have predicted for this trip. This opens up a convenient road to do really meaningful statistics with large numbers of data, and compute a wide range of statistical indicators from such a system than mere mean values or data from a few costly measurement campaigns.

## II. THE TAXI-FCD SYSTEM BERLIN

The taxi-FCD system for Berlin is running for several years now. Recently the system was reimplemented to integrate state-of-the-art technologies like SOA to improve modularity and maintainability (2), to improve efficiency and scalability (8), but core algorithms and the basic architecture as described in (1) did not change.

More than 4.000 taxis equipped with GPS and GPRS deliver their current positions and states in intervals of about 30 to 60 seconds to the taxi control center. These data are then forwarded to the FCD system of DLR-TS, which matches current taxi positions on a digital road network, generates travel times for each road section where current data is available and finally a traffic state map for the whole network as well as travel time information for selected routes is generated. For applications that require values at every edge, like route monitoring, historic travel times are

used as a fallback solution to provide better values for speed or travel times than heuristics.



### III. VALIDATION OF TAXI-FCD SYSTEM BERLIN

#### A. Design of this analysis

The approach presented here is to compare different output or intermediate values for routes traveled by taxis. To simplify comparisons, they are always based on travel times, since they can be easily combined and they are equally weighted.

First the travel time of a taxi on a sequence of edges (trajectory) of the road network can be computed by distributing the interval from one GPS measurement to the next one on the edges that are traveled between both. Errors that can occur could be caused by wrong matching to the digital road network as caused by the limited frequency of position reports (i.e., the “sampling error”), due to inexact GPS-positions (i.e., the “measurement error”), or by inaccurate distribution of the travel times on the edges. Since the comparison is based on the sums of travel times on a route, the second cause of errors can be ignored here. In the sequel, these travel times are referred to as trajectory travel times (denoted  $traj\_tt$ ).

Based on the trajectory travel times as well as on historic data for the current hour and week day, the current speed is computed periodically for all edges with available data. The influence of historic vs. measured data on the result depends on the amount and timeliness of measured trajectory travel times. Most recent data are included first and with the highest weight – until the sum of weights reaches a threshold or there are no more recent measurements. Otherwise, historic data is included to fill the gap between the accumulated weight of measurements and the threshold.

This is the main output of the taxi-FCD system that is used for the traffic state map and for monitoring of predefined routes within the road network. For this comparison these

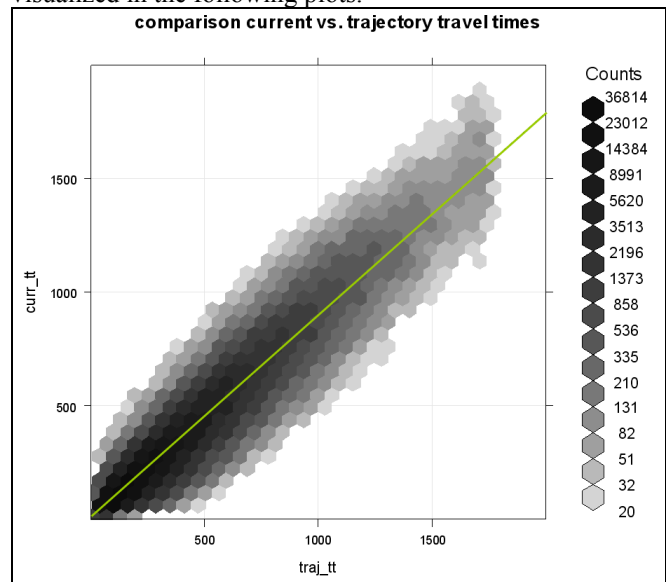
speed values are converted into current travel times (denoted  $curr\_tt$ ).

Finally historic speeds for edges in the road network are generated and updated based on current speeds. Since they are also the input for generating new current speeds this is a feedback loop that should adapt gradually to significant changes. Historic travel times (denoted  $hist\_tt$ ) for the analyzed trajectories are also compared to the measured trajectory travel times.

This analysis was conducted based on the data of all Wednesdays in October 2010 – which do not fall into the holidays in Berlin. About 540.000 trajectories with more than ten edges were used to generate trajectory, current and historic travel times.

#### B. Results

For each trajectory the travel times in seconds as calculated by the different methods introduced above are compared and visualized in the following plots.



**Fig. 1: Plot of trajectory travel times with corresponding current travel times**

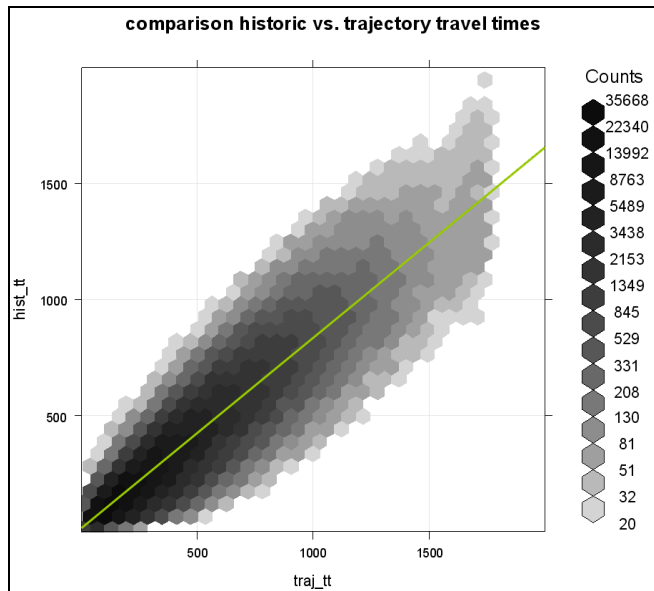
By comparing current with trajectory travel times the main output of the taxi-FCD system is compared with travel times that represent the real traffic situation. As you can see in the plot, real travel times are longer on average.

Since trajectory travel times have a strong influence on current travel times the significance of this comparison should be pointed out. With the current algorithm and for single edges the only criteria for the weight of one measurement is its timeliness – older values would be disregarded completely if there is enough newer data. If there are only older data, historic travel times are integrated with a high weight having a high influence on the resulting current travel time.

In contrast to single edges the effect of delay times or different speeds on subsequent edges, which can not be

measured directly, should even out for trajectories. Thus they can be used as ground truth to evaluate different algorithms to compute current travel times based not only on timeliness of single measurements.

The fact that current travel times generally appear to be shorter could be caused by over-optimistic historic speeds that are used to calculate the current speeds. Next, to examine the influence of historic data, travel times that are solely based on historic speeds are compared to taxi travel times.

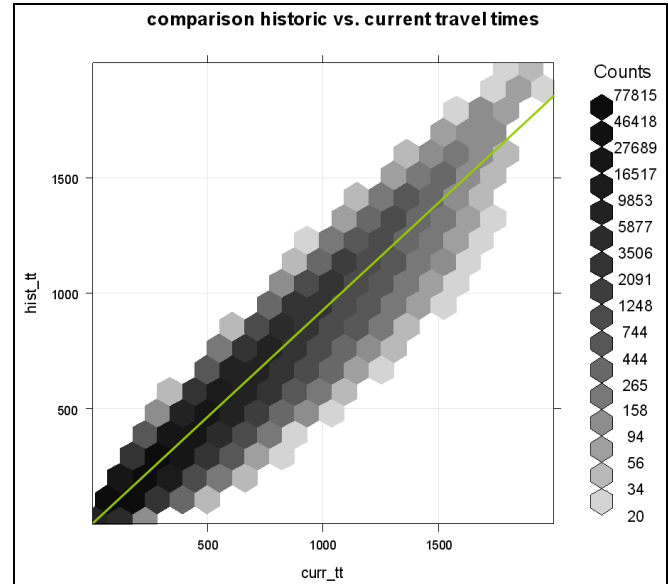


**Fig. 2: Plot of trajectory travel times with corresponding historic travel times**

The difference between historic and trajectory based data is even larger since historic travel times are smaller than those based on current travel times. It seems that historic travel times, for the course of this experiment, had an adulterant influence on the calculated current travel times (which turn out to be too small). A possible reason could be an inappropriate initialization of historic speeds: a short operating phase of the new taxi-FCD system may be responsible for historic speeds that are based on a time period that is simply too short. As a consequence, the historic speeds would have not been adapted adequately yet. Another possible source of error could be that the traffic situations that the historic travel times have been based on – namely those occurring in the holiday season during summer that ended just a few weeks before the day chosen for analysis – was one with generally shorter travel times. An additional small analysis for one Wednesday during the holidays in August displayed only a slightly smaller difference between trajectory and current travel times. This suggest that there may also be other sources of error besides the historical speeds.

Since historic travel times are often used as substitute for missing measured data, their quality should be as high as possible. One solution is a more granular clustering – not

only by day of week and hour, but also by separating holidays from other days. Also by including more data and gradual updates the quality of historic travel times should increase.



**Fig. 3: Plot of current travel times with corresponding historic travel times**

Finally, travel times based solely on historic data are compared with current travel times. The regression line forms an almost perfect diagonal, which shows the high influence of historic speeds in the calculation of current traffic times.

For longer trajectories with travel times higher than 1.200 seconds the peak in this plot is slightly above the regression line while particularly for shorter trajectories the peak is at or even below the regression line.

Since most of the trajectories are quite short (about 75% are shorter than 500 seconds, based on trajectory travel time) those have a very high influence on the overall result.

### C. Quality indices

Using travel times of individual trajectories as reference values, quality indices can be calculated to evaluate the quality of results and the algorithms they are based on. In the following, the targeted quality indices are briefly outlined and then the results of their calculation and a discussion of them are presented.

#### 1) Targeted Indices

One of the most significant indices is the *systematic error*. A first aim was to detect a possible systematic error, and, should it exist, to quantify it. A second aim was to further analyze the nature of the systematic error. A typical question would be whether it is constant or it is related to the value of a different, fluctuating quantity (such as e.g. the inherent variability of the observed traffic at a particular time of day).

Denoting

$X = \{t \mid t \text{ is travel time based on current speeds}\}$  and  
 $R = \{t \mid t \text{ is trajectory based travel time}\},$

the mean values of the two sets of measurements  $X$  and reference values  $R$  are

$$\bar{x} =_{\text{def}} \frac{\sum_{t \in X} t}{|X|} \text{ and } \bar{x}_{\text{ref}} =_{\text{def}} \frac{\sum_{t \in R} t}{|R|}.$$

Then, the systematic error is defined as

$$e_{\text{sys}} = \frac{\bar{x} - \bar{x}_{\text{ref}}}{\bar{x}_{\text{ref}}} \cdot 100(\%).$$

To ensure reliability and reproducibility of the experiments, also the *random error* has been determined as the *Standard Error of Mean (SEM)*. Where possible, the size of the sample has been adapted such that the resulting SEM can be considered to be sufficiently low. Let

$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

be the sample standard deviation of the series of measurements or observed reference values (i.e., in the formula,  $x_i$  denotes the  $i$ th element of the series of measurements constituting  $X$ , or it denotes  $i$ th element of the series of observations constituting  $R$ , and  $n = |X| = |R|$ ), an estimator for the standard deviation  $\sigma$  of the respective population. Then the random error can be quantified as

$$e_{\text{rand}} = \frac{s}{\sqrt{n}}$$

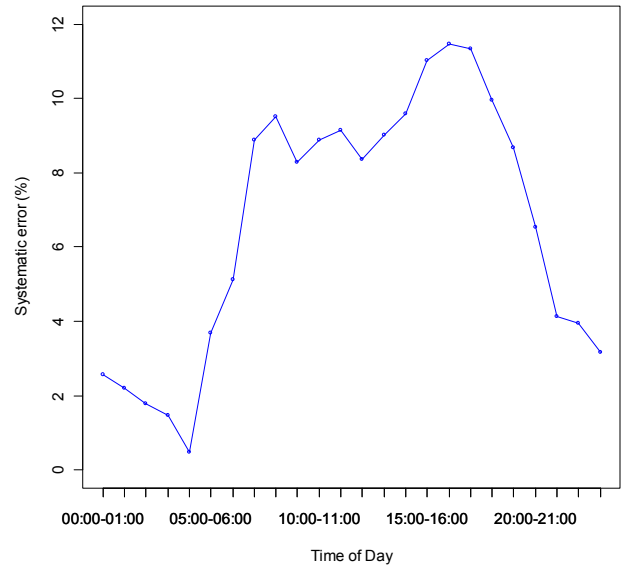
Calculation of the systematic error essentially conducts a comparison of two *mean values*, each referring to one of two distributions. The first distribution is that of the measured values (in the present case, the current FCD speeds) and the second one is that of the reference values (in our case, the trajectory based travel times). Besides the mean value, the second major parameter describing a distribution is the *standard deviation*. Our analysis of quality indices thus is extended by the comparison of the *Coefficients of Variations (CV)* of the two distributions: Since the CV is defined by

$$CV = \frac{\sigma}{\mu} \approx \frac{s}{\bar{x}},$$

this comparison also takes into account the standard deviation as the aforementioned second major parameter.

## 2) Results and Discussion

First, the trajectory data of only one Wednesday in October 2010 has been used as the basis for the calculation of the random error. As is well-known, the magnitude of the random error depends both on the size and the variance of the sample. For this smaller sample, and with the given degree of inherent variability of the traffic, it was still quite high (more than 27 percent). Therefore, the size of the sample has been increased to cover all Wednesdays in October 2010 (a total of about 540,000 trajectories), reducing the random error to 13.4%. All subsequent calculations then have been based on this second sample. The systematic error obtained was -7.46 %, that is, the travel times based on the current speeds of the FCD system (moderately) underestimate the reference values as obtained from the taxi trajectories (by around seven percent on average). To further investigate in the nature of this error, the daily course of this error has been determined. For this purpose, the systematic error has been calculated for each of the 24 hours of the day separately. The result of this calculation is depicted in Fig. 4.



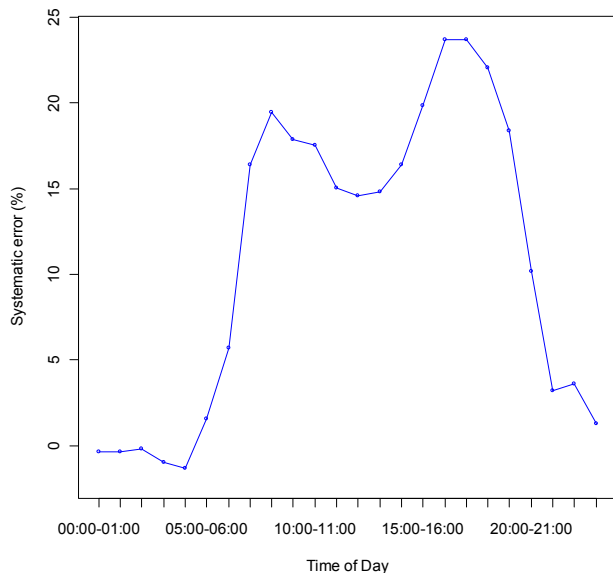
**Fig. 4: Daily course of the systematic error**

The largest error is observed in the time from 15:00 until 18:00 o'clock, i.e. during the afternoon rush hour. Of note is that, as an inevitable result of splitting the original data set into 24 smaller samples, the random error has been increased significantly (to about 67% on average). Nevertheless, the significance of the peak at rush hour time and the general shape of the curve in Fig. strongly suggest a non-constant systematic error that is dependent on a different quantity. Of note is also that there is only a small error during the nightly traffic where a high amount of free flow traffic can be expected.

Due to almost constant average travel times at night, both for one day and also between different days, those intervals can be described easier and more accurately by historic travel times. In contrast at times with more frequent changes between traffic states (e.g. morning or afternoon peaks) the difference between measured and historic travel times can be bigger since changes between traffic states do not occur at the same time on every days and also the magnitude of jams can differ.

Then, recalling that previous results indicated a certain adulterant influence of historic speeds on the quality of the current speeds (see Section III.B), together with the present observation again suggests, that historic speeds, having not (yet) been initialized adequately (due to a too short operation phase of the newly set up FCD system or perhaps also due to the influence of the holiday season), may be the source for the moderate systematic underestimation of the trajectory based travel times.

Aiming to verify this hypothesis, also the daily course of the systematic error has been calculated again, using the historic speeds instead of the current speeds obtained from the FCD system. The resulting course is shown in Fig. 5.



**Fig. 5: Daily error course for historic speeds**

The highest peak again falls into the time of the afternoon rush hour. Although the magnitude of the error for the historic speeds is much larger than that for the current speeds (about twice as large), the shape of the curve is very similar to the error course for the current speeds. Altogether, this experiment seems to confirm the presumed cause of the systematic error (i.e., the use of inadequate historic speeds). As the calculation of more adequate historic speeds requires a longer operation time of the FCD system, a mandatory direction of future research would be the repetition of the described experiments as soon as the newly established FCD

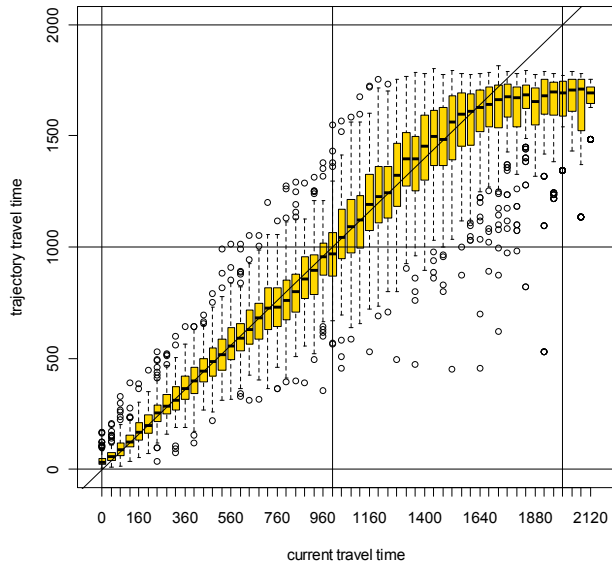
system has operated long enough to obtain a sufficiently large data base of historic speeds values.

To further pinpoint other sources of error, the sample was divided into two parts corresponding to short and long trajectories. The systematic error obtained for every part has been calculated. Thereby, a trajectory was considered to be “short” if and only if the corresponding travel time was below or equal a threshold of 500 seconds. The fraction of such short trajectories in the sample was about 75%. The systematic error obtained is 8.34% for the short trajectories, whereas it was only 6.61% for the long trajectories. That is, the FCD system reproduces the travel times on the longer trajectories more reliably. An explanation would be that the calculation of edge-based current speeds along a trajectory with more edges obviously has a better chance of arriving at a stable mean travel time for the trajectory (this should hold more or less regardless of the particular method of statistical aggregation that is used). Nevertheless, this problem also suggests that there is a certain potential for improvements in the algorithms currently used in the FCD system – a shift to more robust, e.g. median-based aggregation methods seem to be a promising direction for future research.

An analysis based on about 99,000 trajectories on a Wednesday during the holidays in August was conducted to test if the historic travel times used here are more compatible to holiday traffic. The systematic error of this sample was 6.33% with the trajectory travel times being only slightly less underestimated by current travel times as in the sample of the main analysis for October.

Next, the coefficients of variation of a) the current speeds and b) the trajectory based speeds have been calculated for the same data sample (including all Wednesdays in October 2010). The CV for the current speeds is appr. 0.806, and it was appr. 0.791 for the trajectory-based travel times. Therefore it can be concluded that the FCD system reproduces the inherent variability of the (trajectory based) reference travel times very well.





**Fig. 6: Median and quantiles of trajectory travel time for different ranges of current travel time**

In this plot, for all trajectory lengths based on current travel times the distribution of travel times based on trajectories driven by taxis are visualized. As one can see for most of trajectory lengths both travel times fit quite good with the current travel time being about the median of trajectory travel times. For shorter trajectories trajectory travel times are quite longer, due to many outliers being longer, which becomes more balanced for longer trajectories.

Results for longer trajectories (current travel time  $> 1500$  s) are influenced by a cap for trajectory travel times at 1800 seconds.

#### IV. Conclusion

Most previous approaches to evaluate the quality of Floating Car Data (FCD) rely on the data measured by only a few field tests. These approaches are costly to conduct and the collected data remain relatively scarce. In this paper, an interesting alternative called “self-evaluation” has been proposed: completed taxi trips with their known travel times are regarded as reference measurements, and the travel time that the FCD system would have predicted is compared to it. The much larger amount of test data that is made available by this approach (about half a million trajectories of more than 4,000 taxis) facilitated the analysis of data quality by sufficient statistics and a collection of credible quality indices. As a result of this analysis, a potential source of error of a newly established FCD system of German Aerospace Center has been identified and, during a subsequent analysis, the range of causes has been tracked down and further narrowed.

Apart from that, the overall validity and well-performance of this new FCD system has been confirmed.

#### Acknowledgment

The authors thank the taxi control center Taxi Berlin TZB GmbH for provision of GPS-data for the taxi-FCD system.

#### References

- [1] E. Brockfeld, P. Wagner and B. Passfeld (2007) Validating Travel Times calculated on the basis of taxi floating car data with test drives. ITS 2007 – 14th World Conference on Intelligent Transport Systems (Beijing) – Proceedings on CD-ROM (Paper-ID 4134)
- [2] L. Touko, S. Ruppe, E. Brockfeld and Y. Yahyaoui (2010) Traffic Data Platform based on the Service Oriented Architecture (SOA). WCTR 2010 – World Conference on Transport Research 2010 (Lisbon) – Proceedings on CD-ROM (Paper-ID 02539)
- [3] E. Brockfeld, A. Sohr, R. Ebendt (2010) Validation of a taxi-FCD system by GPS – testdrives. ITS 2010 – 17th World Conference on intelligent Transport Systems (Busan) – Proceedings on CD-ROM (Paper-ID TP014-4)
- [4] K.-U. Thiessenhusen, E. Brockfeld, R.-P. Schäfer, P. Wagner (2002) Analysis of travel times and routes on urban roads by means of floating-car data. European Transport Conference (Cambridge)
- [5] J. JeongAh, Y. Tae-Ho, Y. Yeong-Whon, L. Seung-Hwan (2006) Determining the optimum number of vehicle probes with accounting the reliability of link travel times. ITS 2006 – 13th World Conference on Intelligent Transport Systems and Services (London)
- [6] M. Reinthaler, J. Zajizek (2007) Real time route analysis based on floating car technology. 18th IASTED International conference: modelling and simulation (Montreal) – Proceedings pp. 609 – 612
- [7] M. Reinthaler, B. Novotny, R. Hildebrandt, F. Weichenmeier (2007) Evaluation of speed estimation by floating car data within the research project DMotion. ITS 2007 – 14th World Conference on Intelligent Transport Systems (Beijing) – Proceedings on CD-ROM (Paper-ID 2135)
- [8] R. Ebendt, P. Wagner (2010) An Integrative Approach to Light- and Heavy-Weighted Route Planning Problems. 5th IMA Conference on Mathematics in Transport, 12.-14. April 2010, London, Great Britain