

# Generating Trajectories from Mobile Phone Data

## Autoren / Authors:

Johannes Schlaich

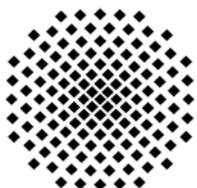
Thomas Otterstätter

Markus Friedrich

Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik, Universität Stuttgart  
vorname.nachname@isv.uni-stuttgart.de

## Veröffentlicht in / Published in:

Schlaich, J., Otterstätter, T., Friedrich, M. (2010), Generating Trajectories from Mobile Phone Data, *TRB 89th Annual Meeting Compendium of Papers*, Transportation Research Board of the National Academies, Washington, D.C., USA.



**Universität Stuttgart**

**Institut für Straßen- und Verkehrswesen**

**Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik**

**[www.isv.uni-stuttgart.de/isv/](http://www.isv.uni-stuttgart.de/isv/)**

# Generating Trajectories from Mobile Phone Data

## Corresponding Author:

Johannes Schlaich, Universitaet Stuttgart  
Seidenstr. 36, D-70174 Stuttgart, Germany  
Phone +49-711-68582485, fax +49-711-68582484, email [johannes.schlaich@isv.uni-stuttgart.de](mailto:johannes.schlaich@isv.uni-stuttgart.de)

## Co-Authors:

Thomas Otterstätter, Universitaet Stuttgart  
Seidenstr. 36, D-70174 Stuttgart, Germany  
Phone +49-711-68584414, fax +49-711-68582484, email [thomas.otterstaetter@isv.uni-stuttgart.de](mailto:thomas.otterstaetter@isv.uni-stuttgart.de)

Markus Friedrich, Universitaet Stuttgart  
Seidenstr. 36, D-70174 Stuttgart, Germany  
Phone +49-711-68582480, fax +49-711-68582484, email [markus.friedrich@isv.uni-stuttgart.de](mailto:markus.friedrich@isv.uni-stuttgart.de)

Submission date: July 30, 2009  
Submission of revised version: November 6, 2009

Word count: 5,089  
7,339 (including Figures and Tables)  
Figures: 7  
Tables: 2

## **Generating Trajectories from Mobile Phone Data**

### **Abstract:**

The paper reports on a method which generates time-space trajectories of travelers by analyzing data from the mobile phone network. The method utilizes mainly the data from location-area-updates, which are recorded from any mobile phone in the standby-mode and thus does not require phone calls or handset equipment. The main idea of the method is to derive location-area-sequences for single trips from the mobile phone data and to compare these with location-area-sequences, which are based on routes through the road and rail network. The method generates a vast number of trajectories on highways and can be used for several applications in the field of transportation planning.

### **Keywords:**

Floating Phone Data (FPD), time-space-trajectories

## INTRODUCTION

Currently, the primary data source for traffic state recognition are road side detectors (e.g. loop detector, overhead detector) which can measure occupancy rate, volumes and speeds for a specific position of a road. These detectors do not facilitate the transportation planner to derive time-space-trajectories, as it is not possible to recognize a specific car at different detectors. Only measurements with Automatic Number Plate Recognition Systems (ANPR, [1]) allow this recognition, but their usage for the purpose of transportation planning is often rather limited due to high installation costs and data protection issues.

Mobile detectors moving along with the traffic flow complement the stationary data by delivering time-space trajectories of single travelers. Usually mobile detectors are integrated in the navigation system of cars which broadcast floating car data (FCD) to a control centre. Main disadvantages of FCD are the relatively high communication costs and the still low equipment rates. As an alternative, observation of mobile phones can provide floating phone data (FPD) for a large number of travelers in the network. This paper reports on a study which used floating phone data recorded for several months in parts of the cellular phone network of T-Mobile, Germany.

At the beginning of this paper, the system architecture of a mobile phone network will be described. Then the paper will give a brief overview about existing projects using floating phone data and will emphasize on the difference between the existing methods and the method described in this paper. Thereafter, the method is described in four chapters:

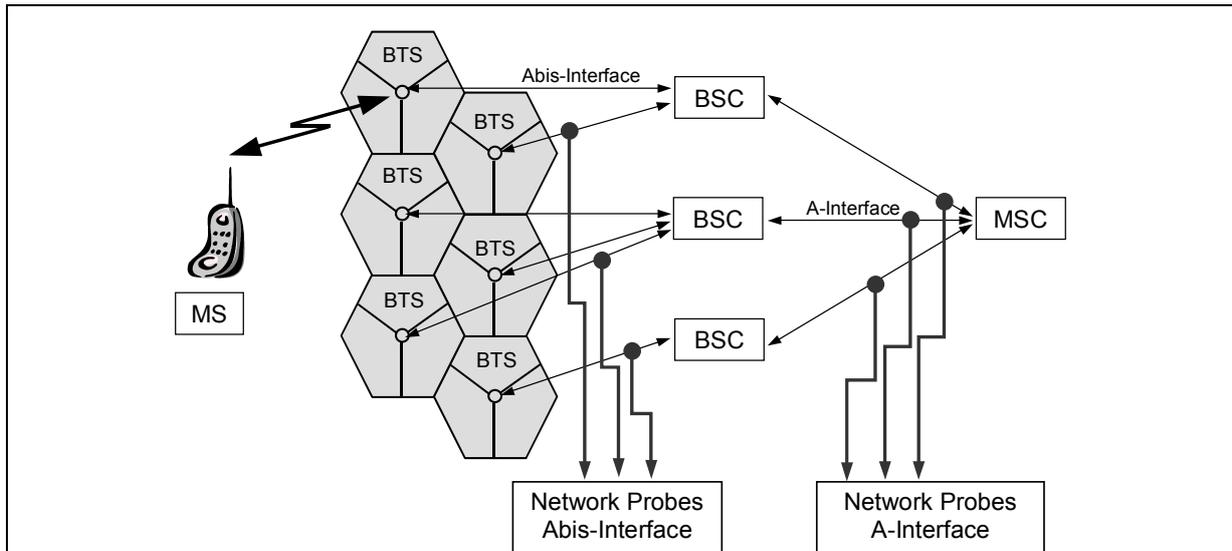
- Generating Location-Area-Code-sequences from floating phone data (“A-Data-LAC-sequence”)
- Generating Location-Area-Code-sequences from the road and rail network (“network-LAC-sequence”)
- Matching of the Location-Area-Code-sequences
- Results of the generation process

Finally, the conclusion will give a summary of the method and an outlook, what applications are enabled by the FPD-trajectories generated by the presented method.

## SYSTEM ARCHITECTURE OF A MOBILE NETWORK

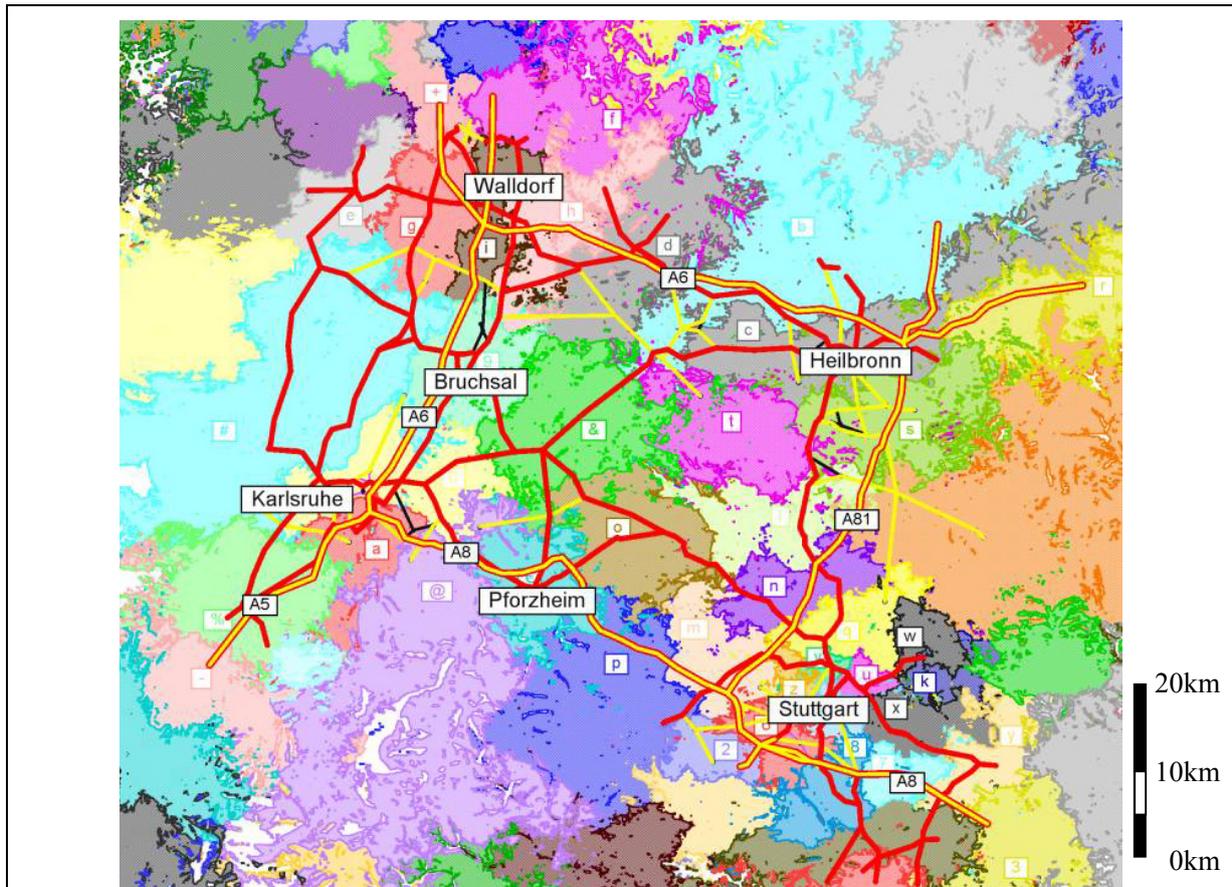
In Germany and many other countries, the main share of the mobile phone communication is processed on GSM networks today and in the near future. GSM stands for Global System for Mobile communications. FIGURE 1 displays a simplified illustration of the components of the GSM network required for mobile communication:

- Mobile Stations (MS): These are the mobile phones used by individual persons or embedded in the onboard units of the German toll collection system for trucks.
- Base Transceiver Station (BTS): They form the smallest unit of the GSM network. A BTS is commonly made up of three sector antennas with an angle of radiation of 120°. Each antenna corresponds to one radio cell. The mobile phone contacts the network via the BTS.
- Base Station Controller (BSC): A BSC manages a set of BTS. Data between the BTS and the BSC are transmitted via a special interface, the so-called Abis-interface.
- Mobile Switching Center (MSC): A MSC controls a set of BSC. The data is exchanged via the so-called A-interface.



**FIGURE 1: Simplified system architecture of a mobile phone network with the interface for the Network Probes [2].**

Roughly 20 radio cells are grouped together to form a location area (LA), which represents the spatial resolution of the A level (see FIGURE 2). The location of a mobile phone in standby mode is only known on the level of a location area. Only when a phone connection is established the serving cell is identified. This two-level interface minimizes the amount of management data in the network as well as the energy consumption of the mobile phone.



**FIGURE 2: Network, location areas and their codes (colored areas and characters) in the study area in South-West Germany.**

The method in this paper utilizes only data from the A-Interface. At this level, in case of a mobile phone in the standby-mode, data are only recorded when a mobile phone is changing from one location area to another. This process is called location area update. During a call, each change from one cell to another (handover) creates an entry to the A-Data. Each data set in the A-Data comprises amongst others the following entries:

- UsrID: A unique anonymous ID for each mobile phone
- timestamp: Timestamp of the data set (in seconds)
- LAC: Location area code of the current location
- CID: Cell-ID of the current cell

An important note is that all this data is necessary for the operation of a GSM-network according to the GSM-standard used in many countries. Thus, no additional data or equipment of mobile phones is required. Some network operators even store some of the described data for their own purposes or forced by governmental rules (e.g. for law enforcement), which further reduces the costs of this data source.

## STATE OF THE ART IN USING FLOATING PHONE DATA

There are several projects using floating phone data, which are summarized by [3] and [4]. The academic and industrial projects focus on traffic state detection, often based on the data from the Abis-Interface

using the detailed measurement reports. According to [3] some projects were not successful, other projects were successful with generating reliable travel times in particular for longer segments (>30 km). Current research ([5], [6]) and implementations by producers of navigation systems [7] show that FPD is a reliable source for traffic state detection on highways and some selected inner-urban roads.

For traffic state detection it is sufficient to create a vast number of short trajectories covering the road segments, which are to be detected. Thus, it is not necessary to create trajectories from the origin to the destination. Actually this is not even possible with methods utilizing the data from the Abis-Interface, as only very few mobile phone users make calls covering their entire trip.

Another project in Israel used floating phone data to obtain Inter-City person trip tables for nationwide transportation planning purposes [8]. In this project, however, no trajectories were created, only origin and destination of a trip were derived from the mobile phone data.

To the knowledge of the authors there is no successful project creating large numbers of time-space-trajectories from floating phone data except the method described in this paper.

## DEFINITION OF LAC-SEQUENCES

A LAC-sequence is a sequence of location areas traversed by a mobile phone during a single trip. In the study area with a limited number of location areas each location area can be identified by a unique character. This way it is possible to store a LAC-sequence as a string, e.g. "ABC". In this paper, we differentiate between two different LAC-sequences:

1. An *A-Data-LAC-sequence* is a LAC-sequence, which is recorded in the A-Data.
2. A *network-LAC-sequence* is a LAC-sequence, which is generated from routes in the road and rail network.

## GENERATING A-DATA-LAC-SEQUENCES FROM FLOATING PHONE DATA

### Data cleansing

First step of the data processing is the data cleansing, which consists mainly of four steps.

1. Few data with erroneous data such as missing timestamps are deleted.
2. Data without reference to the study area are deleted. Here, it has to be distinguished, whether a mobile phone only changes for a short period to a location area outside the study area (due to signal shadowing effects etc.) or whether the trip really left the study area. Different rules either delete the respective data entries or split the data of a user into different trips.
3. All users with less than three different location areas are deleted, as a minimum of three location areas is required to create a trajectory.
4. Consecutive data sets from a user with the same location area (handover during a call, short messages) are deleted, as they do not contain additional information for the presented method.

### Trip Identification

Mobile phone users can produce a series of trips in the road or rail network but they can also make no trips at all. This chapter deals with the description of three rules which are applied to identify potential trips:

- 60min-rule
- Extended 60min-rule
- Jumpiness-rule

### 60min-Rule

A mobile phone user remaining a considerably longer time in a location area than required for directly traversing the area, potentially starts or ends a trip in the respective location area. Considering the size of the location areas the time limit is set to 60 minutes.

The application of the 60min-rule is illustrated using a real example shown in FIGURE 3 with the original data of user 306394488 on the left side of the table. The duration of stay for this user in the location area 29453 amounts to approximately 1.5 hours. For this reason the data of the user is split at this point. The first trip keeps the old UsrID 306394488 going from line 1 to line 4. For the second trip a new UsrID is generated (in this case 309496459) and the location area 29453 is duplicated as a new line, which represents the origin of the new trip presuming a trip start time 1 minute prior to leaving location area 29453.

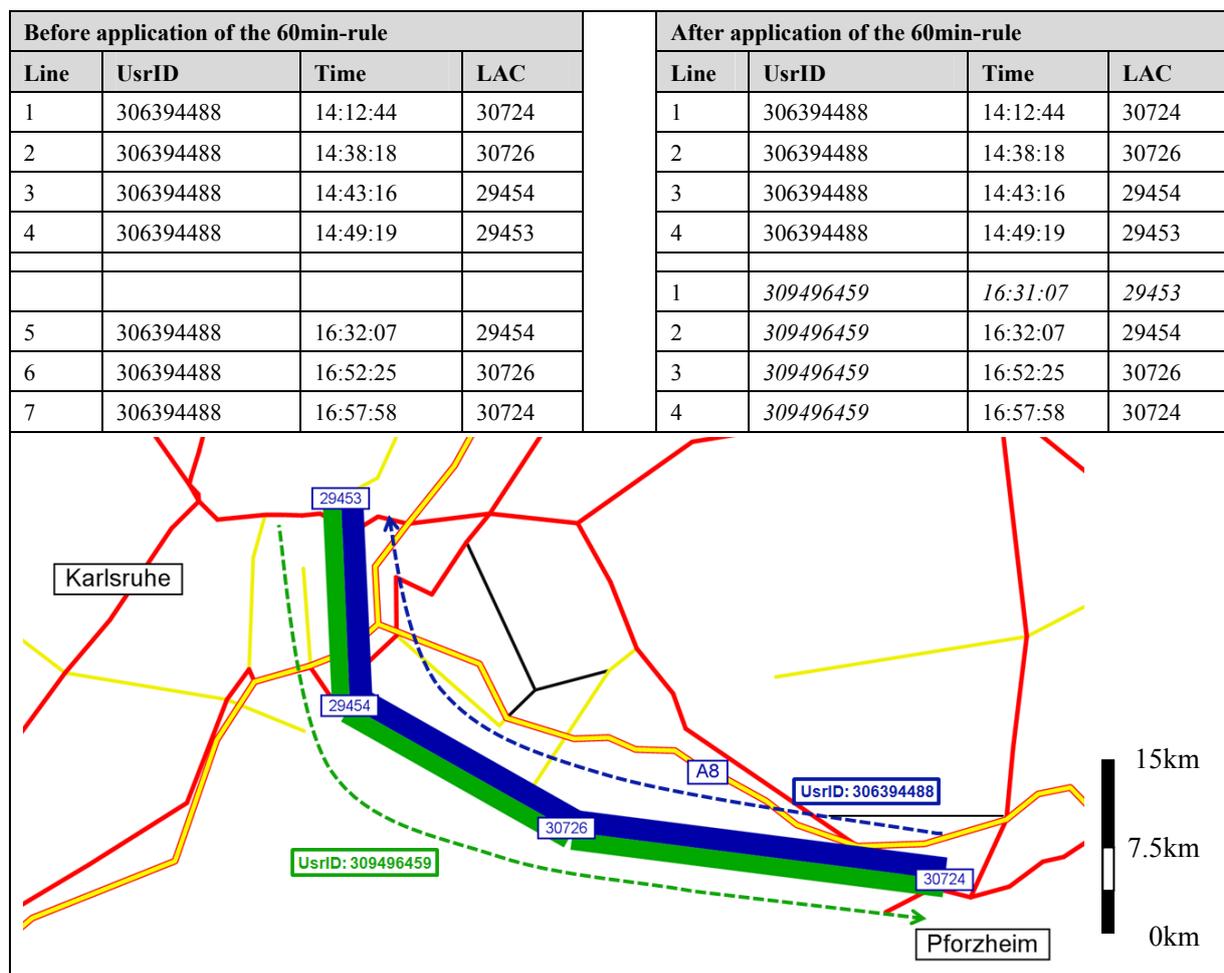


FIGURE 3: Splitting of the UsrID 306394488 based on the 60min-rule in line 5.

### Extended 60min-Rule

It is a particularity of the mobile phone network that some mobile phone users shift between different cells in short time intervals. If these cells belong to different location areas, such users may have several location area updates without actually moving, leading to a non-applicability of the 60min-rule. For this

reason the 60min-rule is extended by analyzing all location areas, which have multiple visits by a user. If both of the following conditions are fulfilled, the data sets of a user are split into two trips:

- Number of different LACs between the multiple visited location area is two or less (independent from the total number of location area updates between these LCAs).
- Time period between the first login and the last logout of the multiple visited location area is 60min or more.

### *Jumpiness-Rule*

As already mentioned in the previous chapter it regularly happens that mobile phone users change between different cells respectively location areas without any real movement taking place. In order to identify these users a jumpiness-factor for every trip is introduced:

$$\text{jumpiness\_factor} = \frac{\text{NumberOfLocationAreaUpdates}}{\text{NumberOfDifferentLACs}}$$

Mobile phone data of one user is deleted, if both of the following two conditions apply:

- $\text{jumpiness\_factor} > 2$
- $\text{NumberOfDifferentLACs} = 3$

Using these parameters, only users with high number of changes between three different LACs are deleted from the further data processing.

## **GENERATING NETWORK-LAC-SEQUENCES FROM THE ROAD AND RAIL NETWORK**

Each A-Data-LAC-sequence requires a matching route in the transport network. To create a set of potential routes for each origin-destination pair of location areas a multi-path route generation algorithm is applied. In contrast to shortest-path algorithms this multi-path algorithm builds for every origin a route tree which contains each destination not just once but multiple times. Using a branch-and-bound approach all nodes  $n$  are inserted in the route tree, if the following condition holds:

$$w(n) < w_{max}(n) = w_{min}(n) \cdot f + c$$

where

$w(n)$  impedance from the origin node to node  $n$  of the current path (e.g. travel time)

$w_{max}(n)$  maximum permitted impedance from the origin node to node  $n$

$w_{min}(n)$  impedance of the shortest path from the origin node to node  $n$

$f$  impedance dependent tolerance parameter  
(constant, e.g. 1.5, or as a function of the impedance)

$c$  constant tolerance parameter  
(constant or as a function of the impedance)

Main goal of the multi-path algorithm is to create a choice set of all relevant routes connecting each pair of location areas. However, the total number of routes should still be reasonable, especially only slightly different routes in inner-urban areas are dispensable. This can be achieved by the following steps:

- In the network model the road network is reduced by considering only links between TMC (Traffic Message Channel) locations. FIGURE 2 shows the resulting network model consisting of 1,128 directed links.

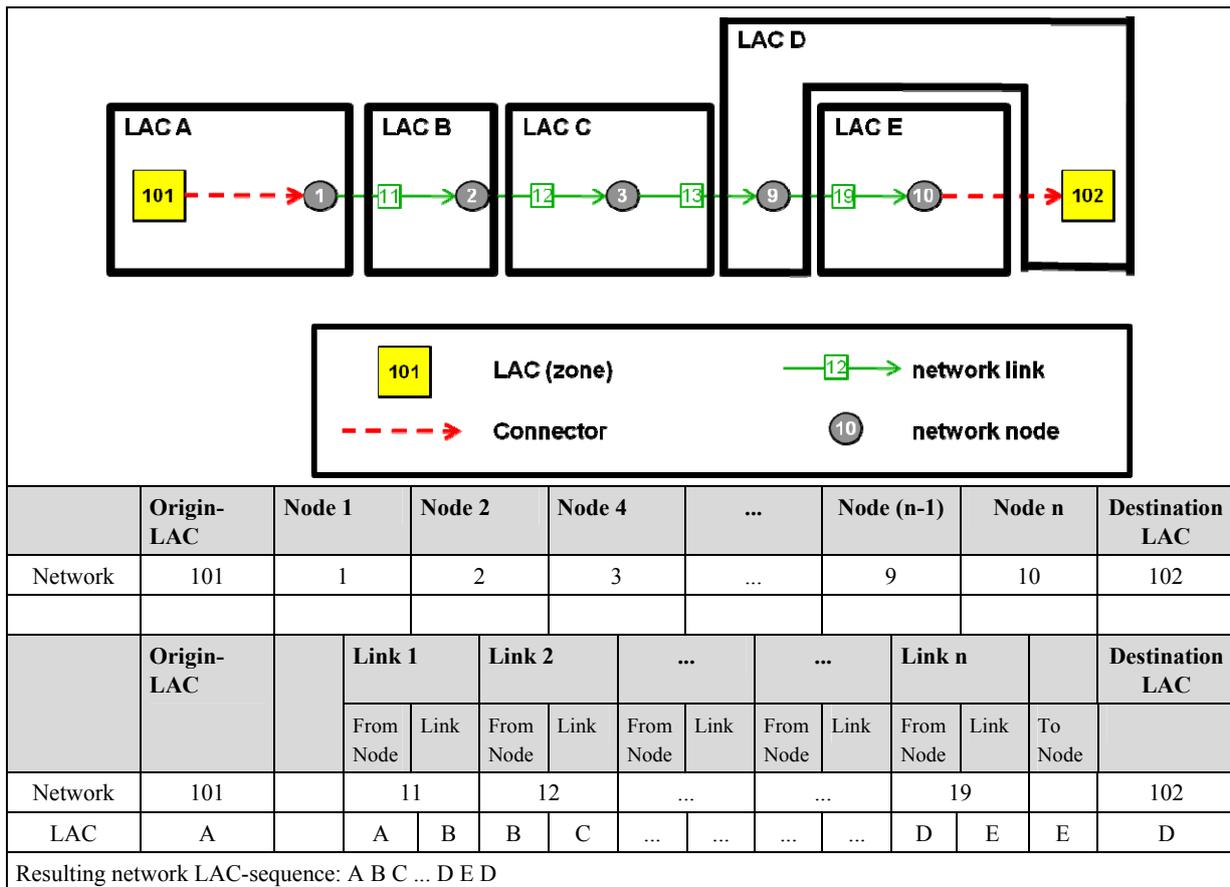
- The above condition is applied with different tolerance parameters  $c$  and  $f$  depending on the route length. For shorter routes (less than 15min), the tolerance parameters are rather strict in order to avoid too many routes, which differ only within the start-LAC, thus resulting in identical LAC-sequences.

- During the construction of the route tree, for each node added, the condition is not only checked from the beginning to the added node, but also for the last 15min of each route. This avoids that longer routes make senseless detours, which would not be allowed at the beginning of the route.

Applied to the network model shown in FIGURE 2 with 41 location areas about 260,000 routes were calculated. This makes an average of 155 routes for each origin-destination pair. In the rail network the number of routes is significantly lower. About 11,000 routes were calculated which makes an average of 6.5 routes for each origin-destination pair.

By intersecting the polygons of the location areas with the links and nodes of the transport network it is possible to assign a location area to each link and node in the network model.

FIGURE 4 shows one route connecting two location areas and the resulting string that encodes the network LAC-sequence. The string length is reduced by condensing sequences of the same character to one character. Different network routes can have the same LAC-sequences, leading to a total of 45,000 different network-LAC-sequences.



**FIGURE 4: Creating a network LAC-sequence from a network route.**

## MATCHING OF LAC-SEQUENCES

### Direct Matching of LAC-Sequences

At this point both the A-Data-LAC-sequences and the network-LAC-sequences are available. In a first step the complete congruence of the two strings is examined. All A-Data-LAC-sequences with complete congruence to a network-LAC-sequence are marked as processed and are not being taken into consideration for the following steps.

### “ABAB”-Rule

An analysis of the unidentified A-Data-LAC-sequences shows that many A-Data-LAC-sequences cannot be directly matched because of frequent shifts between two LACs. The LAC-sequence „CABABABD“ is an example for this, displaying several shifts between the LAC A and the LAC B without the extended 60min-rule being applicable because of the short total duration stay in the two LACs A and B. In order to obtain successful matches with the LAC-sequences from the multi-path route generation algorithm, the occurrence of „ABAB“ is reduced to „A B“. The following TABLE 1 shows the full simplification process for the example „CABABABD“.

**TABLE 1: Recurring application of the “ABAB”-rule.**

Current LAC-sequence	Description
C <u>A</u> B <u>A</u> B <u>A</u> B <u>D</u>	First occurrence of „ABAB“
C <u>A</u> <u>B</u> ABD	Deleting the middle LACs „BA“
C <u>A</u> <u>B</u> ABD	Second occurrence of „ABAB“
C <u>A</u> <u>B</u> <u>D</u>	Deleting the middle LACs „BA“

The reduced A-Data-LAC-sequences are subsequently matched once again directly with the network-LAC-sequences. All LAC-sequences from A-data that are still not identified during this step are considered in the next step.

### Sequence comparison

There are various reasons why A-Data-LAC-sequences from A-data cannot be matched to any network-LAC-sequence from the road or rail network in the first two steps. These reasons can be classified into two categories:

1. A-Data-LAC-sequence corresponds to a single trip, but there exists no suitable LAC-sequences from the route tree generator.
2. A-Data-LAC-sequence does not correspond to a trip, or corresponds to more than one trip.

For A-Data-LAC-sequences of the first category the best possible network-LAC-sequence should be determined whereas for A-Data-LAC-sequences of the second category no network LAC-sequences should be found. TABLE 2 lists a series of reasons for the occurrence in both categories.

**TABLE 2: Reasons for a deficient congruence in matching LAC-sequences from A-data with LAC-sequences from network routes.**

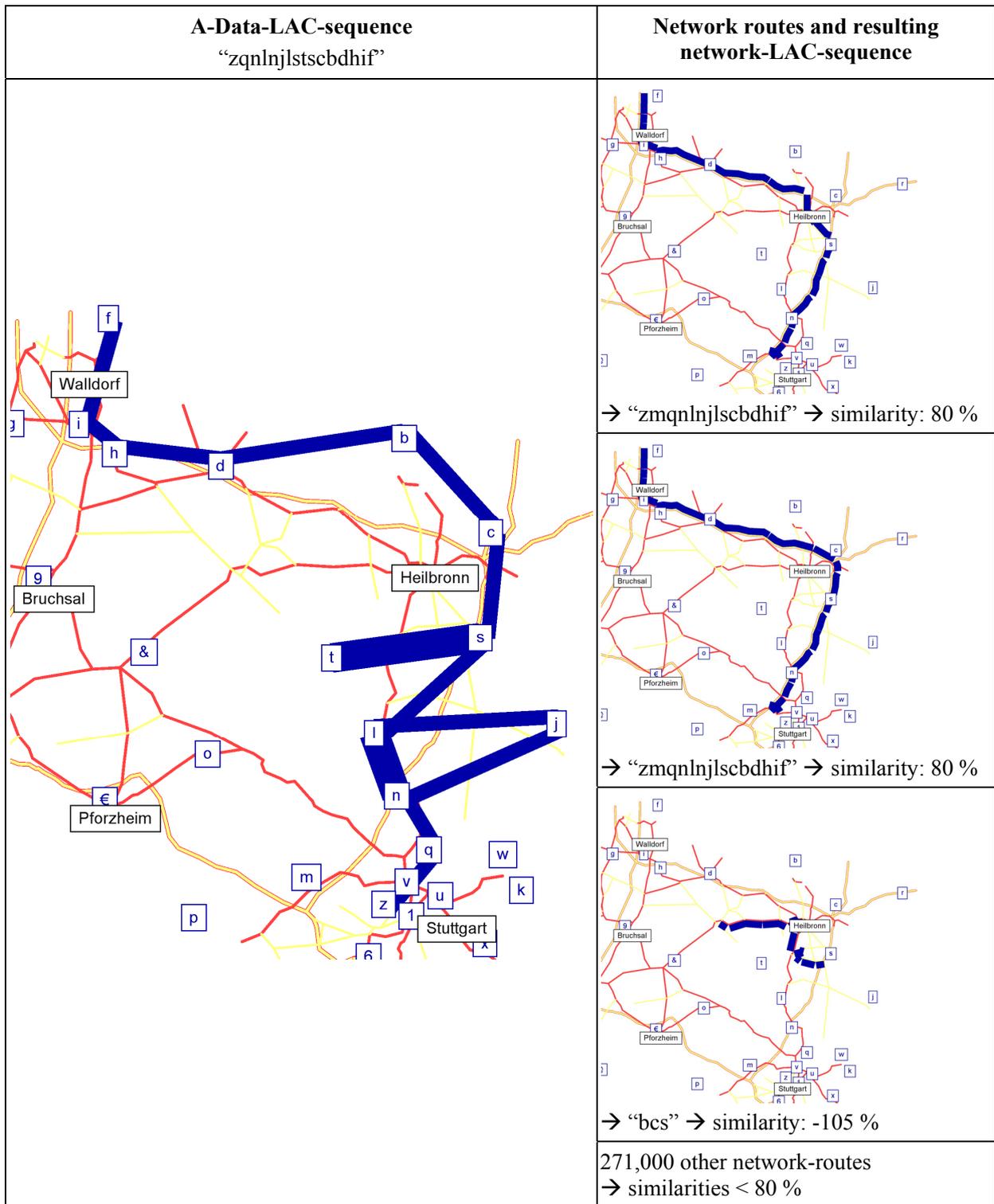
No.	Description	Example LAC-sequences from:	
<b>Category 1: A-Data-LAC-sequence corresponds to a trip in the road or rail network</b>			
		<b>... A-Data</b>	<b>... Network Route</b>
1	(Temporary) Login of a mobile phone user at the second next location area (due to shadowing effects etc.)	ABCDXDF	ABCD F Mobile phone user changes temporary to „X“, although this is not the next location area (“ABA-case”).
2	Fuzziness during the generation of the network LAC-sequences	ABCDEDF	ABCDE F Short return to „D“ is not registered in the network-LAC-sequence
3	Mobile phone user covers routes that are not part of the TMC-network (for instance routes using links in the secondary road network).	---	
<b>Category 2: A-Data-LAC-sequence is not a single trip</b>			
		<b>... A-Data</b>	<b>... Network Route</b>
4	Multimodal trip (road and rail).	ABCDEFGH	ABCD (road) & DEFGH (rail) Multimodal trips cannot be identified.
5	Insufficient identification of trip, for instance no split into two different trips.	ABCDEDCB	ABCDE & EDCB Destination in „E“ not identified. The total trip is not part of the route tree (detour factor is too high)

As illustrated in the examples of the category 1 the deviations follow certain rules (for instance „ABA“ instead of only „A“). For a successful matching of A-Data-LAC-sequences to the most probable network-LAC-sequence a sequence comparison based on the Needleman-Wunsch-algorithm [9] is applied. This algorithm was developed to determine a similarity of two sequences and is often used in bioinformatics in order to align protein or nucleotide sequences. The algorithm gives scores for matching characters and penalties for gaps.

The algorithm is adapted for the special purposes of LAC-sequences comparison and computes a similarity for each pair of LAC-sequences. In contrast to the normal Needleman-Wunsch-algorithm not only two sequences are compared, but also the structure of the A-Data-LAC-sequence is analyzed. As a result the “ABA-case” illustrated in TABLE 2 does not reduce the similarity value as much as another random mismatch with the same length.

For the parameters selected in this case a complete congruence of the two strings leads to a similarity value of 100%. Strings that differ strongly can have negative similarity values. The network LAC-sequence with the highest similarity value is assigned to the A-Data-LAC-sequence. However, only a sufficiently high similarity value will generate a trip from an A-Data-LAC-sequence. Analyses from many sequence comparison have shown that 60 % is an adequate threshold for this purpose.

FIGURE 5 visualizes the process for an A-Data-LAC-sequence. On the left side the A-Data-LAC-sequence is illustrated by connecting the centers of each LAC. On the right side three out of 271,000 networks routes and their resulting network-LAC-sequence are shown. In addition, the similarity between the A-Data-LAC-sequence and the network-LAC-sequence is given. The upper two network routes result in the same network-LAC-sequence, which has the highest similarity (80 %) of all network-LAC-sequences.



**FIGURE 5: Results of a sequence comparison with highest similarity of 80 %.**

The computation of the similarity for the example in FIGURE 5 is illustrated below. The optimal position of the gaps in the LAC-sequences is a result of the Needleman-Wunsch-algorithm, which requires linear programming with a matrix of the size 16x15 (lengths of the strings) and is not explained in this paper.

A-Data-LAC-sequence:	z qnl njl stscbdhif	(length: 16)	
Network-LAC-sequence:	zmqnl njls    cbdhif	(length: 15)	
match	+ ++++++    ++++++		→ 14.0 points
gap	-            -		→ -2.0 points
ABA-case	o		→ <u>.4 points</u>
			12.4 points

$\text{similarity} = \text{points} / (0.5 * \text{length of strings})$   
 $12.4 \text{ points} / (0.5 * (16 + 15)) = 80\%$

By taking into consideration the „ABA-case“ (here: „...sts...“) along with longer congruence of the LAC-sequences the string comparison generates a high similarity of 80%.

## ASSIGNING LAC-SEQUENCES TO NETWORK ROUTES

As shown in the previous FIGURE 5 some LAC-sequences correspond to two or more network routes. This results from the fact that the location areas are relatively large. In this case the demand of a LAC route is distributed to the set of network routes in three steps:

1. Mode choice: There are approaches (e.g. [10]) to distinguish trajectories from private and public transport. As only 457 (out of some 50,000) mostly short inner-urban LAC-sequences simultaneously represent a road and a rail route, a fixed distribution in accordance with the general results of mobility surveys (e.g. 80% road and 20% rail in [11]) is assumed in order to keep the computing time at an acceptable level.

2. Choice of origin and destination node: Each LAC is connected to one or more network nodes by connector links. Each network route starts and ends at these nodes. If different connector links result in identical LAC-sequences, information of the serving Cell-ID from the first or last location area update is used to determine the most likely network node.

3. Route choice: The final route choice is modeled applying a C-Logit model [12]. This model considers the travel time of each route in the uncongested network and the similarity of the routes. It also prefers routes which remain on one road category (e.g. motorway, rural road) and do not change road category at intersections.

These distributions have only very low impact on the traffic volumes on highways, as they affect mainly inner-urban roads.

At the end, for each trajectory the times at each node are computed based on the start time, end time, selected waypoints and the free-flow-speed of the links.

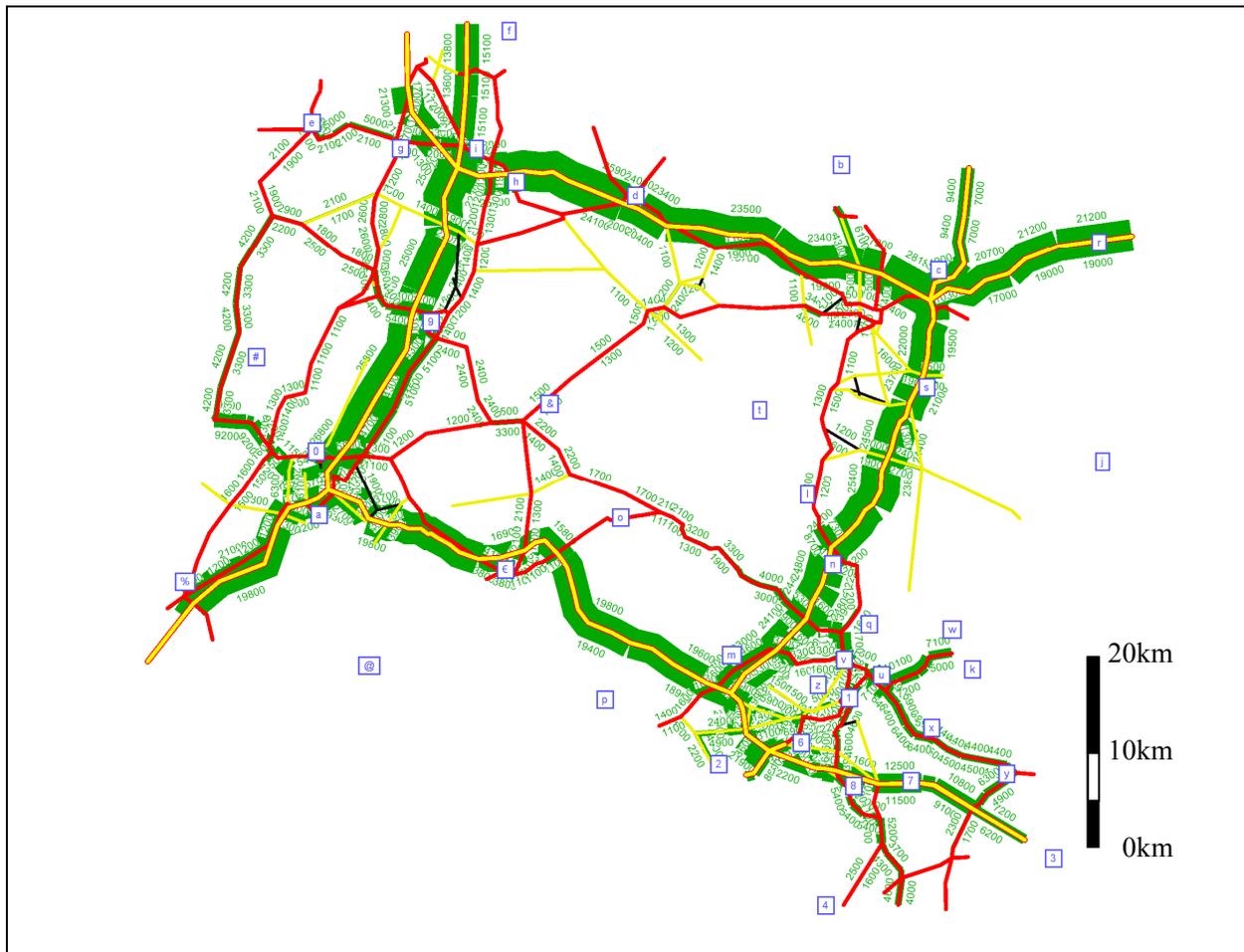
## RESULTS

The LAC-based generation of trajectories has been processed using data recorded in the cellular phone network of T-Mobile, Germany for roughly 90 days. In the following the results for an average working day are presented.

For one working day nearly 40 million data records are registered in the study area. From those, roughly 80 % are deleted by the data cleansing process. In particular deleting records of users with only

one or two different LAC reduces the data set significantly. These users make no trips at all or only short trips which cannot be detected with a LAC-based generation of trajectories.

The trip identification distinguishes roughly 600,000 trips per working day. From those, approximately 37 % can be directly matched to a LAC-sequence from the multi-path route generator, 13 % are assigned by applying the “ABAB”-rule and roughly 30 % are assigned using the sequence comparison. 20 % of the A-Data-LAC-sequences remain unassigned. Many of these have three different LAC-sequences and do not correspond to a trip on the roads of the study area. Others are trips in the study network, but are not detected properly during the process.



**FIGURE 6: FPD-volumes of a working day in the road network.**

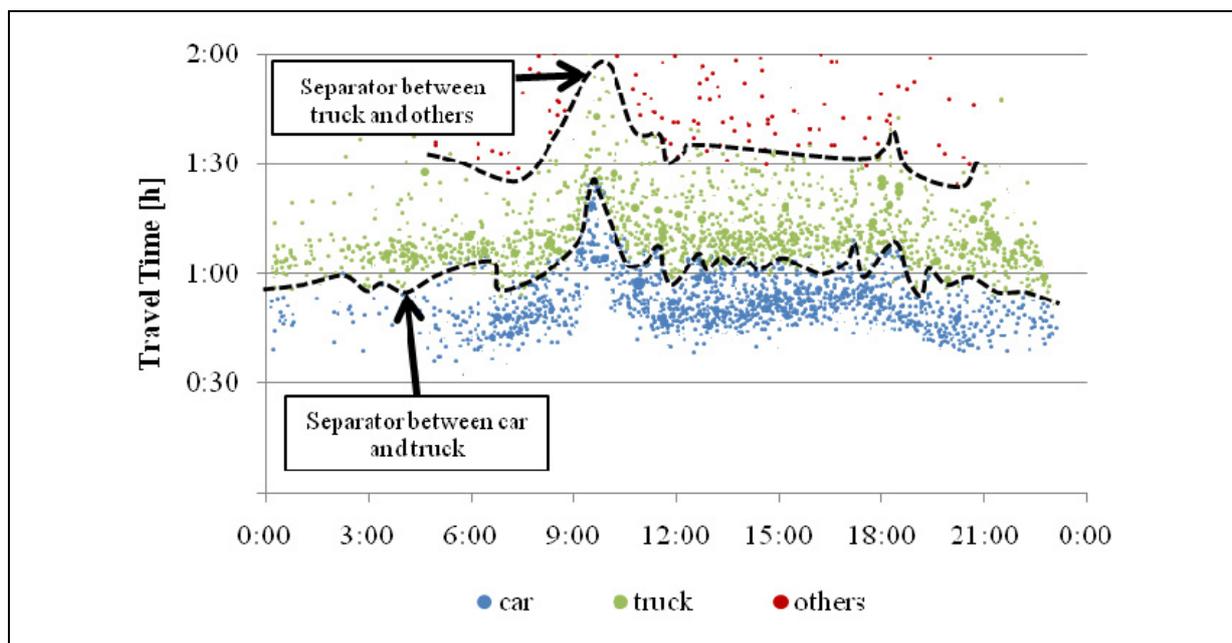
The above FIGURE 6 shows the traffic volumes in the study area for a working day. Plausibility controls show very good results for the highway network:

- The traffic volumes on the highways are constant at 40-50 % of the values from road-side detectors.
- The through-traffic volumes correspondent to ANPR-measurements [13].
- The trip length distribution shows longer trips than mobility studies, which is due to the non-consideration of short trips.
- Travel times show realistic patterns.

For the non-highway part of the network it has to be examined in each individual case and depending on the goal of the respective analysis, whether the trajectories have sufficient quality. In

certain cases it may be advantageous to include the Cell-ID-information into the analysis, which can be erroneous (as described by Gur et al. [8]), as the serving Cell-ID often does not correspond to the nearest Cell-ID.

The following FIGURE 7 shows the travel times between Stuttgart and Walldorf via Karlsruhe (ca. 90 km). Applying an algorithm based on the floating average of the travel time is possible to differentiate roughly between cars, trucks (respectively slow cars) and other vehicles (with stops along the route etc.). There is a traffic hold-up at around 10 a.m. leading to higher travel times for all vehicles. The free-flow travel times are about 45 minutes for cars (120 km/h) and 65 minutes for trucks (83 km/h). The high share of trucks results from the fact that most trucks are equipped with a GSM-module embedded in the on-board-unit of the German toll system.



**FIGURE 7: Division of through-traffic into the classes car, truck and others.**

## CONCLUSION

The presented LAC-based generation of trajectories has been proven to generate reliable and robust results for a period of approximately 90 days for the highway network in the study area. The main restriction of the generated trajectories is that observable trips need to contact at least three locations areas, so that only trips with a trip length longer than the diameter of a location area can be monitored. Considering this constraint, trajectories from floating phone data can close the existing gap in the continuous monitoring of the travel behavior of trips exceeding a length of approximately 20 kilometers.

Further limitation of the trajectories is the fact that using only the data from only one mobile network operator the sample might be skew. E.g. specific operators are preferred by young user or particular ethnic groups. In addition, the generated trajectories are no vehicle-trajectories, but SIM-card-trajectories. Each vehicle can have none, one or e.g. in buses several SIM-cards onboard.

The vast number of trajectories permits the generation of trip tables of mobile phone users. These trip tables can be projected to trip tables of passenger cars and trucks by comparing the traffic volumes of the trajectories and the traffic volumes of road-side detectors [14]. The trajectories also enable the

analysis of route-choice-behavior in order to determine the impacts of traffic state information and construction sites on the route choice of travelers [15].

As floating phone data have proven to be a reliable source also for the purpose of traffic state estimation with a commercial potential in the navigation system market it should be a goal of academics and transportation planners to use the data not only for traffic state estimation but also for the projection of demand matrices, route choice analyses and monitoring applications.

## ACKNOWLEDGEMENT

This research is part of the project Do-iT [2], which was funded by the German Federal Ministry of Economy and Technology (Bundesministerium für Wirtschaft und Technologie). The mobile phone data for the study was provided by T-Mobile, Germany.

## REFERENCES

1. Friedrich, M., Jehlicka, P. and J. Schlaich, Automatic number plate recognition for the observance of travel behavior, CD-OM, *Proceedings of the 8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability*, Annecy, France, 2008.
2. Do-iT, Nutzung von Mobilfunkdaten für das Verkehrsmanagement - das Projekt Do-iT, *Beilage Nr. 5 zur Straßenverkehrstechnik*, herausgegeben vom Projektträger Mobilität und Verkehr, Bauen und Wohnen (TÜV Rheinland Group), 2008.
3. Qiu, Z., Jin, J., Cheng, P., and B. Ran, State of the Art and Practice: Cellular Probe Technology Applied in Advanced Traveler Information Systems, CD-ROM, *TRB 86th Annual Meeting Compendium of Papers*, 1853, Transportation Research Board of the National Academies, Washington, D.C., 2007.
4. Caceres, N., Wideberg, J.P., Benitez, F.G. (2008), Review of traffic data estimations extracted from cellular networks, *Intelligent Transport Systems*, vol. 2, issue 1, pp. 179-192.
5. Schwieger, V., Ramm, K., Czommer, R. and W. Möhlenbrink, Mobile Phone Positioning for Traffic State Acquisition, *Journal of Applied Geodesy*, Walter de Gruyter, Berlin - New York, vol. 1, issue 3, 2007.
6. Wörner, A., Verfahren und Vorrichtungen zur Generierung von Verkehrsinformation und ihrer Lokalisierung innerhalb eines räumlichen Bereichs, Offenlegungsschrift (Patent Application Publication), DE 10 2007 013 220 A1, *German Patent and Trade Mark Office*, 2008
7. Cohn, N., Real-time Traffic Information and Navigation: An Operational System, *TRB 88th Annual Meeting Compendium of Papers CD-ROM*, Transportation Research Board of the National Academies, Washington, D.C., 2009.
8. Gur, Y., Bekhor, S., Solomon, C. and L. Kheifits, Use of Massive Cell Phone Data to Obtain Inter-City Person Trip Tables for Nationwide Transportation Planning in Israel, *TRB 88th Annual Meeting Compendium of Papers DVD-ROM*, Transportation Research Board of the National Academies, Washington, D.C., 2009.
9. Needleman, S.P. and C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, *Journal of Molecular Biology*, vol. 48, pp. 443-453, 1970.

10. Schollmeyer, R. and T. Wiltschko, Classification of public transport vehicles using cellular mobile radio data, *Proceedings on 6th European Congress and Exhibition on Intelligent Transport Systems and Services*, Aalborg, Denmark, 2007.
11. Mobilität in Deutschland (MiD), *Ergebnisbericht im Auftrag des Bundesministeriums für Verkehr, Bau- und Wohnungswesen*, Projekt 70.0736/2003, 2004.
12. Cascetta, E., *Transportation Systems Engineering: Theory and Methods*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.
13. Schlaich, J. and M. Friedrich, Staumeldungen und Routenwahl in Autobahnnetzen – Teil 1: Analyse von Staumeldungen & Teil 2: Routenwahl in Autobahnnetzen, *Straßenverkehrstechnik*, issues 10 & 11, pp. 621-627 & 706-712, Kirschbaum Verlag, Cologne, Germany, 2008.
14. Friedrich, M., Immisch, K., Jehlicka, P., Otterstätter, T. and J. Schlaich, Generating OD matrices from Mobile Phone Trajectories, *TRB 89th Annual Meeting Compendium of Papers DVD-ROM*, Transportation Research Board of the National Academies, Washington, D.C., 2010 (accepted paper).
15. Schlaich, J., Analyses of the Route Choice Behavior using Mobile Phone Trajectories, *TRB 89th Annual Meeting Compendium of Papers DVD-ROM*, Transportation Research Board of the National Academies, Washington, D.C., 2010 (accepted paper).