

RESEARCH

Open Access



Trajectory estimation based on mobile network operator data for cellular network simulations

Gerald Ostermayer*, Christoph Kieslich and Manuel Lindorfer

Abstract

In this paper, we present a framework for estimating trajectories of cellular networks users based on mobile network operator data. We use handover and location area update events of both speech and packet data users captured in the core network of the Austrian MNO A1 to estimate the subscribers' mobility behavior. By utilizing publicly available data, i.e., environmental information, road infrastructure data, transmitter power ranges and antenna characteristics, our approach allows the estimation of subscriber trajectories for both urban and semi-rural environments with a good accordance to the actual trajectories. Additionally, we present a method to estimate a particular subscriber's movement velocity, on the basis of mentioned data. Furthermore, we propose a methodology to estimate when a particular user started or ended a speech or packet data session during his journey, based on mobility-related network events. With this, our framework enables the creation of reproducible mobility situations for cellular network simulations at system level.

Keywords: Mobility, Trajectory estimation, Modeling, Simulation, Cellular networks

1 Introduction

During the standardization and development process of cellular systems, it is necessary to evaluate the performance of new features that are to be tested. Since it is not feasible to implement an entire test system for every planned feature in the early development stages, simulations are the only method to get performance figures that help to assess the value of the new features. Some features and algorithms strongly depend on the mobility of the subscribers (e.g., power control, handover, scheduling); therefore, dynamic system level simulations are necessary to incorporate the mobility of the subscribers.

This applies not only for new systems; also already deployed systems are continuously improved over their lifetimes. And again, performing simulations is the proper method to rate features and algorithms under evaluation. In this case, the simulations should be based on the real network, i.e., the real cell deployment in the real environment (comprising the real street network) with the

actual offered traffic. Additionally, the subscriber mobility behavior should be incorporated as realistic as possible. In that case, it is very beneficial to use the mobile network operator's (MNO) information related to the subscribers' mobility, i.e., anonymous handover (HO) and location area updates (LAU).

In our work, we use HO and LAU events of speech (GSM) and packet data (GPRS, UMTS, LTE) users captured in the core network of the Austrian MNO A1. The location accuracy of these events is limited to the coverage area of the concerned cells. Additionally, we use freely available data sets about the environment, the base station (BS) configuration, antenna characteristics, transmitter power ranges, etc. The mentioned data sets provide us with a sequence of sample points of the subscriber's trajectory where each sample point has a location inaccuracy based on the cell coverage area. In this work, we estimate trajectories and the velocity of subscribers based on the available data sets and compare these trajectories and velocities with those the subscribers moved along in reality. Additionally, we use specific events captured in the core network to determine when a certain

*Correspondence: gerald.ostermayer@fh-hagenberg.at
University of Applied Sciences Upper Austria, Softwarepark 11, 4232
Hagenberg, Austria

speech or packet data session started or ended. This information provides a time and spatial frame for modeling network traffic, whereas the actual traffic model to be applied can be selected on demand at a later point in time. This generic solution allows the creation of reproducible mobility situations for real network simulations, whereby the traffic model becomes exchangeable.

Figure 1 outlines the basic concept of our system: a trajectory is generated based on a given set of LAU and HO events. These events are recorded asynchronously, i.e., each event is logged on occurrence. A HO event occurs whenever the mobile terminal changes its serving cell whereas a LAU event is issued whenever a mobile terminal changes its location area. With that, we have information about the trajectory as sampling points (time-location tuples) with a certain accuracy in space domain and exact in time domain. Whilst LAU events are issued only when the subscriber is in idle mode, HO events are generated during active speech or packet data sessions. Based on the corresponding events, we can distinguish between speech and packet data sessions, which constitutes useful information when it comes to modeling and simulating network traffic.

We found out that trajectories can be estimated for urban as well as semi-rural environments with good accordance to the real trajectories. Our approach also allows the estimation of the subscribers' movement velocity and is therefore well-suited to describe their mobility.

The rest of this paper is organized as follows. In Section 2, related work in the areas of mobility modeling, travel time estimation, and trajectory estimation is presented. In Section 3, the developed trajectory estimation framework is outlined. Important figures of merit and results of performed experiments and simulations are shown in Sections 4 and 5, respectively. Finally, the paper is concluded in Section 6.

2 Related work

Over the past years research has been conducted in the field of mobility simulations and investigation of mobility related events in mobile communication networks. Besides mobility simulations and mobility modeling, mobile subscription data is also used for mobility behavior estimation in ITS applications. We summarize fundamental concepts in all relevant areas, namely mobility modeling, trajectory estimations, and travel time estimations. The field of traffic modeling remains unregarded, since our work does not focus on actually modeling network traffic, but instead provides a generic solution for applying a traffic model of choice.

2.1 Mobility models

Stochastic processes are often used in order to model the human mobility behavior. The most common mobility models are random way point (RWP, [1]) or random walk models (sometimes referred to as Brownian Motion, used in, e.g., [2–4]), Levy walks and flights [5, 6], and also the Gauss-Markov mobility [7]. Rhee et al. [8] compared the Levy walk model, which is an extended random walk model, with the human mobility, based on GPS traces of 101 volunteers. Their findings indicate that outdoor Levy walks with less than 10 km contain statistically similar features as the human mobility. To understand human mobility patterns, Gonzalez et al. [9] analyzed 100.000 anonymized trajectories of subscribers. They found out that human mobility is characterized by a time-independent travel distance and that people travel to a few highly frequented locations. In [10], the authors propose a mobility model based on the flocking behavior of birds in order to model a realistic movement of groups of mobile entities in mobile ad-hoc networks (MANETs). Similar to that Morlot et al. [11] propose an interaction-based mobility model for hot spots, i.e., their

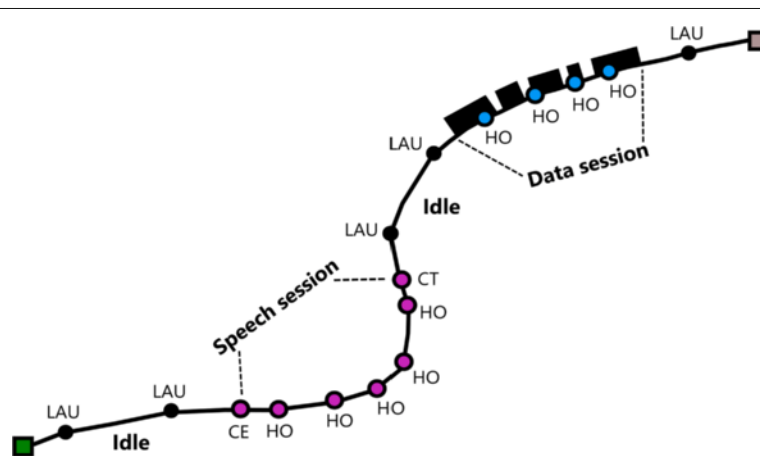


Fig. 1 Basic concept of an estimated subscriber trajectory based on a sequence of events captured in a cellular mobile network

formation and disaggregation as well as the time dynamics in between. However, these results were used to evaluate the mobility of a collectivity of users but not for the behavior of a single individual.

In [4], a mathematical model is proposed in order to track a mobile terminal in a cellular environment. Based on this model characteristic quantities determining the mobility behavior are derived, e.g., the distributions of cell residence and holding time as well as the average number of HOs. Misra et al. [12] present an approach towards modeling the mobility pattern of a target node in a wireless sensor network based on available tracking data that are collected by sensing nodes. Different methods are presented to determine and predict the trajectory of the moving target node. In contrast to this work, we consider cellular systems in real environments with underlying real street networks. In [13], the authors propose a random room mobility model (RMM) in order to describe the mobility behavior of patients inside a hospital that are monitored using wireless body area networks (WBAN). With the help of this model, they studied the performance of extra-WBAN communication.

Liu et al. [14] propose a hierarchical mobility model based on user profiles which consist of a set of user mobility patterns (UMPs). The model covers a global mobility model (GMM) and a local mobility model (LMM) in order to predict trajectories in wireless ATM networks. Whereas the GMM predicts the sequence of cells a user is moving through, the LMM determines the actual path within a cell. The sequence given by the UMP is randomly changed based on different operations (insert, delete, change). This model is somehow a random walk limited by deterministic components. Therefore, all positions within a cell are possible for the mobile. The proposed model is verified by simulations only.

2.2 Trajectory estimation

The estimation of trajectories based on mobile phone data is a topic which has been addressed by numerous researchers in recent years. Schlaich et al. [15] used sequences of LAUs in order to derive trajectories for mobile users. Using LAUs constitutes an eligible approach, since they are issued when the mobile terminal is in both idle and connected mode. The ascertained sequence of LAUs is compared to a set of pre-generated routes between an estimated start and end position. In the end, the route showing the highest similarity with respect to the sequence of LAU events is chosen as the trajectory describing the corresponding user's mobility. This approach works well for longer trajectories around 20 km, for short tracks however it does not work properly, since a minimum number of three LAUs is required.

Tettamanti et al. [16] used HO updates instead of LAUs, which makes their approach applicable not only for the

higher road network, but also for the minor one. HO updates, however, are only issued when the corresponding mobile terminal is in connected mode, i.e., during an active call, which constitutes a significant limitation for their approach. Using Voronoi tessellation, the coverage area for every cell which was traversed during an ongoing call was estimated. The start and end position of the trajectory correspond with the centroid of the cell where the call originated and terminated, respectively. Certain influences such as the population density or land use in the respective areas, which can be used to determine, e.g., where it is more likely that a certain trajectory starts or ends, are not taken into consideration. In order to generate different routes connecting the given start and end position, they used the traffic modeling simulation framework VISSIM. For each of the generated routes, the squared sum of all minimum distances between the route and the cell sites was calculated. In the end, the route with the minimal distance was chosen as the final trajectory.

Becker et al. [17] characterize the mobility pattern of hundreds of thousands of people using a huge number of anonymized call detail records (CDRs) from a cellular network. They characterize the human mobility in terms of daily travel distance and carbon emissions with a resolution of ZIP code areas. Additionally they determine traffic volumes carried by different roads. For that purpose, test users drive along predefined routes and record the corresponding sequences of cells their mobile phones are connected to. Afterwards they fit the actual cell sequence to the best predefined in order to estimate the path the user was moving along. The major difference to our work is that they distinguish only between a few predefined routes where for all of those the cell sequence has to be recorded in a training phase.

2.3 Travel time estimation

Travel time estimation methods can be categorized in such using active tracking and such using passive tracking methods. Active tracking requires the participation of the subscriber and/or the mobile terminal. Active tracking technologies are, e.g., Global Positioning System (GPS) and Time Difference of Arrival (TDOA). Methods like TDOA need BSs that support that feature and therefore require special equipment. These systems have a higher location accuracy than passive tracking systems, but they increase either the power consumption of the mobile terminal (e.g., GPS) or the network load (e.g., TDOA) or both. On the other hand, passive tracking uses events captured in the operator's core network to locate the subscribers. Therefore, passive tracking does not require any special equipment in the mobile terminal or involvement of the subscriber. The related work outlined in the following is based on passive tracking exclusively.

Alger et al. [18] from Vodafone Germany were one of the first who investigated travel time estimation on the German higher roads network. Double handover (DHO) events of speech users were used to compute the cell passage time, i.e., the time elapsed between entering and leaving the cell coverage area. By estimating the route a subscriber moved within the cell coverage and dividing it's length by the time difference, the velocity can be computed. An average filter was used to combine the travel times of other subscribers in the same cell. The system was used to investigate the traffic condition on the higher roads network in real time. This research showed that travel times can be estimated for the higher roads network without the need of expensive inductive loop detectors. However, their approach focused only on the highway networks and does not support the minor roads network. Especially, the route estimation is very simplistic on highways. Since the German road network¹ consists of 12.917 km highway and 230.377 km other roads their approach is only applicable for 5.6 % of the entire road network.

In order to estimate the traffic speed and travel time, Bar-Gera [19] used a proprietary system developed by Estimotion Ltd. This system uses information about HO events of speech users to estimate the traveled route and velocity. A sequence of locations derived from the HO footprints is matched to road segments. Unfortunately, no information is provided about the determination of the HO footprints which are the basis for further investigations.

3 Mobility estimation and traffic modeling

In this section, we give a detailed description of the methodology used in our framework. We introduce all the data that are required as input to our system as well as the algorithms to transform this data into a set of user trajectories. Figure 2 gives an overview of the particular steps which are performed in order to derive a users' trajectory from a given input. The individual steps are explained in the following.

3.1 Mobile network operator data

The data described in this section are provided by the Austrian MNO A1. The MNO captures a lot of events at the interfaces of its core network by installing network probes. This process is described by Valerio [20] in detail. Each event that is captured in the core network features the following attributes:

- Timestamp: specifies the exact time when the event was captured
- ID: identifies the subscriber
- Cell ID: identifies the cell the subscriber is currently connected to

- LAC: identifies the location area the subscriber is currently assigned to
- Event: defines the type of the captured event

Some of these events are strongly related to the mobility of the subscribers. The most important events in this respect are call establishment and termination, HO, and LAU. Each of these events can be related to a certain subscriber ID and the cell where the event occurred. In case of a HO event this is the cell the subscriber is handed over to (target cell), in case of a LAU event this is the target cell in the newly assigned location area. An example for a sequence of events related to a certain subscriber that was captured in the core network is depicted in Table 1. The first event describes a call establishment, followed by three HO events and a call termination event. Using the position of the involved BSs in combination with their antenna configuration (i.e., the angles of the electrical bore sight), a coarse position of the subscriber at the time the event occurred can be determined.

3.2 Road network

Based on the assumption that individuals typically move along roads, our framework maps the estimated subscriber trajectories on ways and streets existing in a real environment. The road network we use for that reason is obtained from OpenStreetMap. These freely available data cover a majority of the real road network and provide additional information such as the maximum allowed speed on particular road segments. This information is required for adapting and validating the estimated user's speed, whereas the road network defines the actual route of the final trajectory.

3.3 Cell area estimation

At this point, it is difficult to estimate the actual location of a particular subscriber at the time the specific event is captured as only the position of the BS the subscriber is currently connected to is known. In order to improve the estimated position, knowledge about the cell's coverage area (relevant for call establishment and termination events) and its borders (relevant for HO and LAU events) is inevitable. We therefore investigate two methods for determining mentioned characteristics, namely Voronoi diagrams and coverage prediction.

3.3.1 Voronoi diagrams

A simple approximation of a cell's coverage area and borders can be made using Voronoi diagrams as described by Baert and Seme [21]. Hereby, Voronoi tessellation is used to partition a plane with n points into n convex polygons. Each of these convex polygons contains one generator point, i.e., the location of the cell site. However, many cells are part of a sectorized deployment, which implies

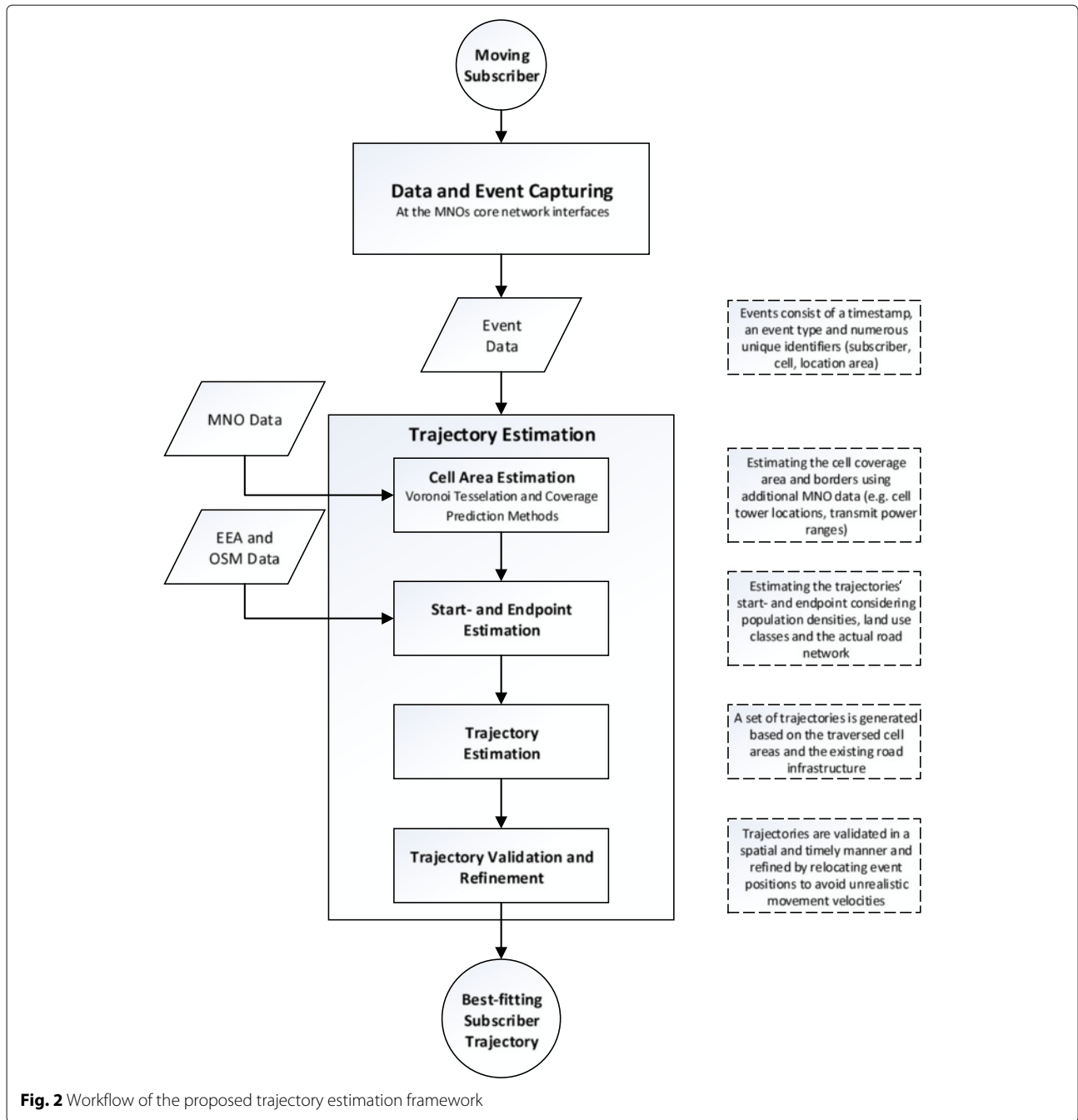


Table 1 Events captured in the core network of the mobile network operator A1

Timestamp	ID	Cell ID	LAC	Event
1327678833	4	51520	5501	Call establishment
1327678854	4	31510	5501	Handover update
1327678935	4	31210	5502	Handover update
1327678949	4	40510	5502	Handover update
1327679027	4	40510	5502	Call termination

that the cell towers of these sector cells have the same location but their antenna patterns are aligned differently. We incorporate this circumstance by moving the location of each cell tower in the direction of the main beam of the antenna pattern. The longitude's and latitude's location are shifted according to Eqs. (1) and (2), respectively. The constant factor $\frac{1}{50000}^\circ$ in both equations constitutes an approximate maximum movement of 1.6 m in each direction, resulting from the earth's polar and equatorial perimeter.

$$x = x + \cos(\alpha) \frac{1}{50000} \circ \quad (1)$$

$$y = y + \sin(\alpha) \frac{1}{50000} \circ \quad (2)$$

After performing the mentioned transformations, each cell's coverage area is described by a unique polygon, resulting from Voronoi tessellation. A subscriber located within such a polygon has the closest distance to the associated BS. Using Voronoi diagrams for cell area estimation has the advantage that every location has a well-defined affiliation to one of the BSs; moreover, the cells' borders are defined exactly. One drawback is that eventually different power classes of particular cells are not considered. Additionally, the influence of the BSs' environment is also completely neglected.

3.3.2 Coverage prediction

Certain limitations of Voronoi tessellation in the scope of cell area estimation can be overcome by using prediction calculations through the application of network planning tools. These tools allow the incorporation of additional data, which comprises, among other things, antenna characteristics, transmitter power, and landscape characteristics.

We estimated the coverage area for all cell sites of the Austrian MNO A1 in two areas of interest. Since A1 only provided us with the location and the beam direction of its antennas, we used publicly available data sets to enrich coverage estimation. A representation of the landscape is defined by a digital elevation model². This model offers a resolution of 25 m, which means that each pixel covers an area of 25 by 25 m. Additionally, we integrated a building block model for the two areas of interest. For the city of Linz, the footprint of every building is known from OpenStreetMap³, though, their heights is unknown. For these buildings we assumed an average height of 15 m. On the contrary, the city of Vienna⁴ provides a building model including height information. This model consists of the footprint of buildings and a category that specifies their heights in steps of approximately three meters.

In our building model we enriched the building layout from OpenStreetMap for the city of Vienna with the average height of the specified category. In total, we computed two coverage predictions for each area of interest. The first uses the transmitter's location, transmitter power, and the digital elevation model. The second prediction enhances the first one by additionally including the building model on top of the elevation model.

3.3.3 Transmitter power

A cell's coverage area directly depends on the transmitter power of the corresponding BS. We use transmitter power specifications from the Austrian Forum

Mobilkommunikation (FMK) as basis. The FMK provides a service where participating MNOs can upload information about their network infrastructure. For our work the cell sites' location and information about their transmitter power is relevant. However, there are a few limitations while using these data. First, the service provided by the FMK is voluntary and therefore the data can either be out of date or even absent. Second, only the highest transmitter power of all sectors in case of a sectorized cell deployment is provided. Finally, it is not indicated which network operator a particular cell site belongs to.

In the following, our approach to retrieve transmitter power estimations for each cell site in the areas of interest is outlined. At first, we define two sets of locations, one for the cell sites of the provider of relevance and one for the FMK transmitters. Let $A = \{pos_0, pos_1, \dots, pos_i\}$ be a set of BS locations of the network operator and $B = \{pos_0, pos_1, \dots, pos_j\}$ be a set of BS locations provided by the FMK. Set $C = A \cap B$ then contains the positions of cell sites of the MNO of interest for which the information regarding transmitter power is available. For BSs of the same MNO where this information is not at hand (set $D = A \setminus B$), an estimation is performed. Given the information (positions and transmitter powers) of the BSs in set C , the transmitter powers of BSs in set D are estimated by applying a scattered interpolation. An example therefore is depicted in Fig. 3. The triangles indicate the positions of BSs in set C , the transmitter power values outlined at all other positions are estimations of the transmitter power for a BS situated at the corresponding location. Different transmitter power values are depicted using different colors, according to the color scheme shown in Fig. 3.

3.4 Start- and endpoint estimation

At this point, the framework knows the road network on which subscribers can move and the coverage area of the cell sites in the areas of interest. In order to estimate trajectories for a particular subscriber, we need to know where the subscriber started and ended his journey. The call establishment and termination events are used to determine those cells where the trajectory starts and ends in, respectively.

It is obvious that the probability for the location of a trajectory's start- and endpoint is not homogeneously distributed over the entire cell area. Hence, a reasonable remedy is the inclusion of land use information. In our framework, we use the CORINE land cover⁵ (CLC) maps provided by the European Environment Agency (EEA). These maps define a number of land use classes and have a spatial resolution of 100 meters, i.e., each pixel covers an area of 100 by 100 m, which is smaller than the coverage area of many cells. We assign every CLC class a weight that

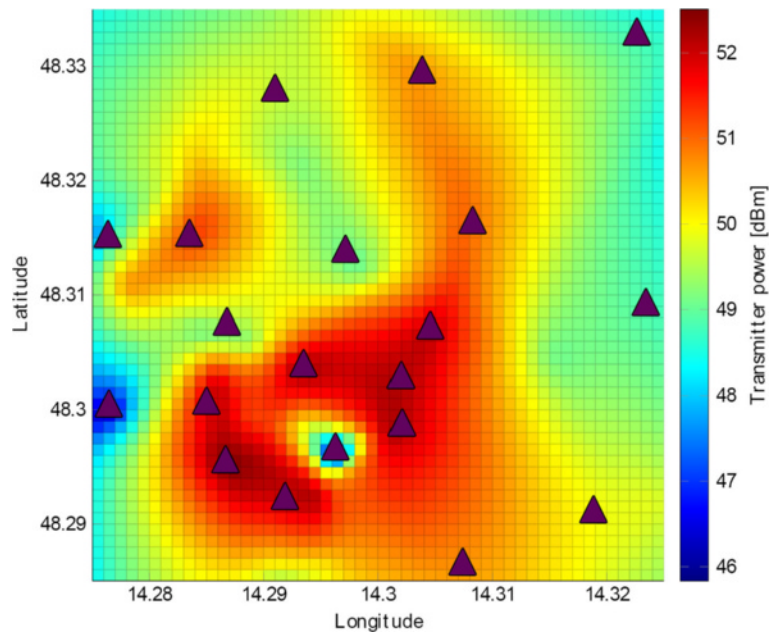


Fig. 3 Transmitter power interpolation. Sample power interpolation for 20 cell sites with a random transmitter power in the range of 45–52 dBm

defines the probability that a trajectory starts or ends in an area of this kind. In this manner, the start and end position can be restricted to a few defined CLC classes such as artificial surfaces, urban fabric and industrial, commercial, and transport units.

The second improvement is the use of socio-statistical maps (i.e., population density maps⁶). These maps are also provided by the EEA and have the same resolution as the CORINE land cover maps. The basic assumption is that it is more likely that a trajectory starts or ends in an area with higher population density compared to other areas of the same CLC class. These maps can be used to create trajectory sets that fit better to the morning as well as to the evening hours since it is possible to incorporate the commuter traffic.

An example is depicted in Fig. 4, where a region consisting of three different population density areas is illustrated. Based on our assumption the subscriber will more likely be located in the area with a population density of $240 \frac{\text{people}}{\text{km}^2}$ rather than in those with $45 \frac{\text{people}}{\text{km}^2}$ and $30 \frac{\text{people}}{\text{km}^2}$. Hence, our framework considers all different population density areas within the region’s boundaries and computes a spatial probability density function. After a population area has been selected by a random process, the exact start and end position will be selected randomly based on a uniform distribution within the bounds of the population area. In order to generate a realistic trajectory based on the underlying road network, the estimated start and end position are mapped to the closest road segment in their vicinity.

3.5 Trajectory estimation

Apart from the trajectory’s estimated start and end position, we know a series of HO and LAU events related to the locations of the corresponding BSs. Considering the estimated coverage area of every cell, a series of cells the subscriber traversed during his journey can be derived. We restrict our generated trajectories to be on top of the road network of the area of interest. In this manner, the subscriber’s route will be computed based on the OpenStreetMap road network between the estimated start and

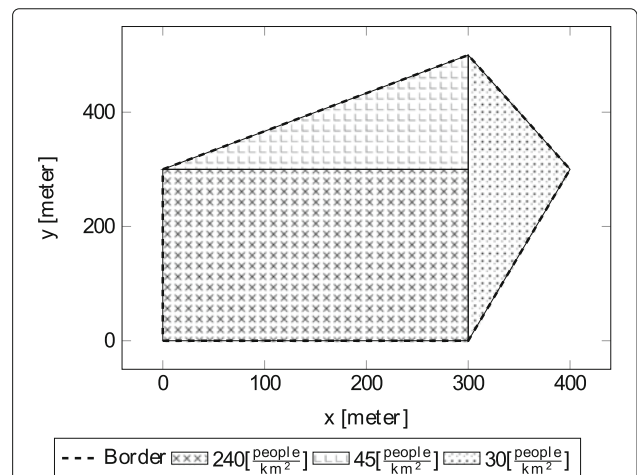


Fig. 4 Start- and endpoint estimation. Area with three different population densities. Our assumption is that people will more likely start or end their journey in a more densely populated area

end position. By default, 20 routes with different start and end positions will be computed. Afterwards, two plausibility ratings for these routes will be calculated, the first considering their geometry and the second incorporating the time it takes to traverse them.

As for the trajectory's geometric validation, the squared sum of minimum distances between the route and the centroid of each particular cell coverage area is calculated. For each of the pre-calculated routes j , the squared sum D_j of all minimum distances $d_{i,j}$ between the route j and cell site i is computed.

$$D_j = \sum_{i=1}^m \min(d_{i,j})^2 \quad (3)$$

The second validation step verifies the trajectories with respect to the time required to traverse them. For this purpose, we introduce a duration ratio $route_{ratio}$ between the call or packet session duration $t_{session}$ and the time it takes to traverse the route t_{route} , as depicted in Eq. (4). This metric provides information whether the estimated route is either too short or too long.

$$route_{ratio} = \frac{t_{session}}{t_{route}} \quad (4)$$

The squared sum D_j in combination with the duration ratio $route_{ratio}$ is used to select the most likely route the subscriber has traveled with respect to the distance to involved cell sites and ascertained travel time. More precisely, the route which is showing a minimal D_j and a $route_{ratio}$ which is closest to or equal to 1 is selected as the subscriber's trajectory.

3.6 Velocity estimation

The average velocity of a moving subscriber between two successive HO or LAU events can be calculated by simply dividing the distance on the road network between the two event locations by the passage time (difference of the two timestamps). Very accurate timestamps are available from the MNO data, whereas the estimation of the actual HO and LAU event positions needs a closer look.

3.6.1 Handover position

HO events in cellular networks are usually initiated if the network decides that the serving cell of a mobile terminal is no longer the best available cell. The best serving cell is typically the one that guarantees the aspired quality-of-service (QoS) with the least necessary transmitter power. Therefore, the position where a HO is executed correlates with the coverage areas of the two involved cells. In reality, the coverage areas of neighboring cells overlap in order to guarantee a seamless HO. Two aforementioned methods to determine the coverage area of a cell are used in our framework.

When using Voronoi diagrams, every cell has a well-defined and unique coverage area, i.e., no overlapping or gap between the coverage areas of two neighboring cells can occur. Since the HO can only take place on the trajectory between the intercept between this trajectory and the border between the involved cells is the estimated position of the HO event.

Using the coverage prediction method, we have to deal with three different situations concerning the estimated coverage areas of the involved cells,

1. **Overlapping handover:** The estimated coverage areas of the two involved cells intersect with each other (see Fig. 5a).
2. **Unconnected handover:** The estimated coverage areas of the two involved cells do not intersect (see Fig. 5b). This means that a HO is made to a cell site whose coverage area the subscriber has not entered yet. Although such situations are very unlikely in reality we sometimes have to deal with them. This is because our input data for the coverage prediction are not as complete as they should be. It would require additional knowledge of used antenna configurations and transmission power to estimate the coverage area of the concerned cell sites more precisely. As we are using only freely available data sources, we lack this kind of information.
3. **Ping-pong handover:** A HO is performed from cell A to cell B and later back to A although the trajectory does not contain lines that are used twice. These HOs are an undesirable effect not only for the network itself but also for the timing estimation. Therefore, we filtered out ping-pong HOs by removing the last HO event, in this specific case the HO event from B to A.

Algorithm 1 depicts how the system estimates the HO position for the overlapping and the unconnected HO type. For an overlapping HO the centroid of the intersection $C = A \cap B$ of the involved cells is computed and mapped onto the calculated trajectory. For an unconnected HO, the line between the closest points of the two coverage areas is computed. If this line intersects the previously computed route then the point of intersection is set as the HO position. If this is not the case then the midpoint of the line is mapped onto the route.

3.6.2 Location area update position

In contrast to HO events, LAUs are issued only when the mobile terminal is in idle mode, i.e., when no speech or data session is active. Similar to HO events, LAUs are afflicted with information regarding the cell site where the update was issued and an exact timestamp when the event was triggered. Analogous to the HO position estimation, this information is used to derive a reasonable location

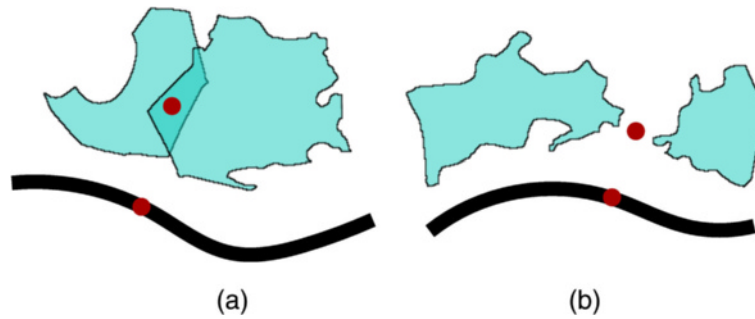


Fig. 5 Different types of handover. An overlapping handover (a) and an unconnected handover (b) are depicted together with the estimated handover position (red point)

Algorithm 1 Handover prediction

```

1:  $HP \leftarrow List()$ 
2: for  $i \leftarrow 1$  to  $length(handover) - 1$  do
3:    $cur = handover[i]$ 
4:    $next = handover[i + 1]$ 
5:   if  $Intersects(cur, next)$  then
6:      $centroid = Centroid(Intersects(cur, next))$ 
7:      $Add(HP, NearestPoint(route, centroid))$ 
8:   else
9:      $line = LineBetweenNearestPoint(cur, next)$ 
10:    if  $Intersects(line, route)$  then
11:       $Add(HP, Intersects(line, route))$ 
12:    else
13:       $midPoint = MidPoint(line)$ 
14:       $Add(HP, NearestPoint(route, midPoint))$ 
15:    end if
16:  end if
17: end for

```

for the particular LAU events. The derivation procedure however differs slightly and will be outlined hereinafter.

The first step is similar to what is done when dealing with an unconnected HO, which means that a line connecting the two closest points of both coverage areas is computed. In contrary, the end point of the line which is located closest to the cell site where the actual LAU event was issued is presumed to be the most reasonable position for the corresponding event (Fig. 6). This is due to the fact that LAUs occur in close proximity to the cell site in which the event was actually triggered (based on the available network data). The same applies if the two coverage areas are overlapping, in this case the closest point of the intersection area is chosen to be the location where the event was issued.

3.6.3 Average velocity

After the HO and LAU positions have been estimated, the average velocity of the subscriber between two consecutive event locations can be computed. Each event exhibits

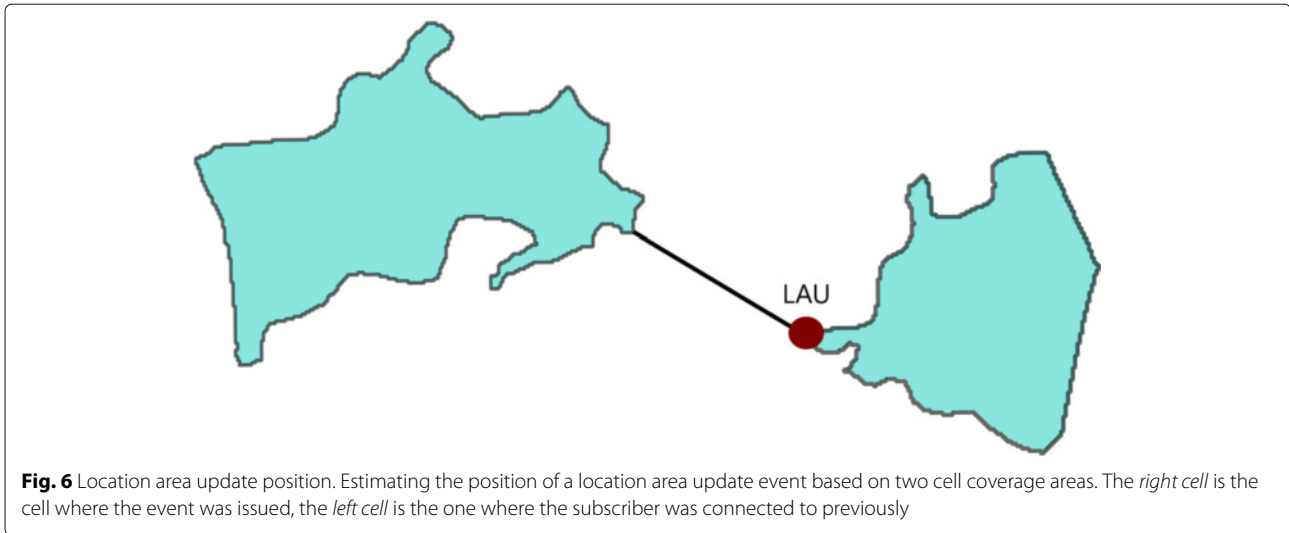
a timestamp which indicates when the event occurred. In order to derive the subscriber's velocity, the distance on the road network between two event positions is divided by the difference of their timestamps.

3.6.4 Velocity adaption

Whilst the timestamps of HO and LAU events are very precise, the estimated positions are not. This can yield in completely unrealistic average velocities between two successive estimated event positions. If the distance between two events is estimated as too long, the estimated average velocity can be much higher than the maximum allowed speed on that particular road. Since the speed limit is known for each street segment from OpenStreetMap data, this information can be used to adapt the estimated event positions in order to obtain realistic velocities (see Algorithm 2).

First of all, the trajectory is partitioned into so-called event segments, which represent the partial trajectories connecting two successive HO or LAU positions. Each of these segments is verified with respect to its conformance with the average velocity and the maximum allowed speed on the respective road segments. If the average speed exceeds the allowed limit by a certain factor (we decided to permit a maximum velocity of 1.7 times the allowed speed limit, this factor however can be chosen on demand), the afflicted HO or LAU event positions are repositioned. This is done by moving the respective positions back and forth along the trajectory. By increasing or decreasing the distance between two event positions, the afflicted segments' average velocity will be altered in the same way.

In order to determine to which extent the corresponding event positions have to be displaced, a reference distance is calculated. This is done by multiplying the maximum allowed speed for the concerned road segments with the time difference between the two events, which is given by their respective timestamps. By comparing this reference distance with the distance between the estimated event positions, the actual shift can be obtained. The



Algorithm 2 Handover adaption

```

1: for  $i \leftarrow 2$  to  $length(trajecory) - 1$  do
2:    $allowedSpeed = MaxSpeed(trajecory[i])$ 
3:    $cur = trajecory[i]$ 
4:    $prev = trajecory[i - 1]$ 
5:    $next = trajecory[i + 1]$ 
6:
7:   if  $Speed(cur) \geq 1.7 * allowedSpeed$  then
8:      $curDist = Distance(cur)$ 
9:      $prevDist = Distance(prev)$ 
10:     $nextDist = Distance(next)$ 
11:     $nominalDist = allowedSpeed * Duration(cur)$ 
12:     $SetSpeed(cur, allowedSpeed)$ ;
13:
14:     $nextTempDist = nextDist + (curDist -$ 
15:     $nominalDist)/2$ 
16:     $prevTempDist = prevDist + (curDist -$ 
17:     $nominalDist)/2$ 
18:    if  $nextTempDist/Duration(next) \gg$ 
19:     $nominalSpeed$  then
20:       $nextTempDist = nextDist$ 
21:       $prevTempDist = prevDist + (curDist -$ 
22:       $nominalDist)$ 
23:    else if  $prevTempDist/Duration(prev) \gg$ 
24:     $nominalSpeed$  then
25:       $nextTempDist = nextDist + (curDist -$ 
26:       $nominalDist)$ 
27:       $prevTempDist = prevDist$ 
28:    end if
29:     $SetSpeed(next, nextTempSpeed/Duration(next))$ 
30:     $SetSpeed(prev, prevTempSpeed/Duration(prev))$ 
31:  end if
32: end for

```

displacement of particular positions however also affects the previous and/or the subsequent event segments. This implies that for certain segments only the first, the last, or both event positions have to be displaced in order to guarantee realistic velocities in the adjacent segments. In this manner, the entire trajectory is traversed.

3.7 Traffic model framing

As mentioned in the beginning, MNOs capture various kinds of events in their core network that allow to deduce, e.g., if a particular mobile terminal is in idle or connected mode or his coarse position at cell coverage level. Furthermore, these events can be used to distinguish between speech and data users, as different events are issued for the respective connection types. These HO events (or LAU events in case the mobile terminal is in idle mode) can be used in order to infer the subscriber's mobility behavior in terms of a movement trajectory, as outlined in the previous sections. For the reason of reproducible mobility situations in cellular real-network simulations at system level, the modeling and simulation of network traffic is an important aspect. In our work we do not focus on particular traffic models (examples are given in [22–24]); instead, we offer a generic solution that allows to estimate when a speech or data session started and ended, both in a time, and spatial dimension. The actual traffic model to be applied at a later point in time is freely selectable, which adds additional flexibility to the simulations to be performed.

Based on particular events captured in the core network, it is possible to derive at which points in time and at which spatial locations a network subscriber was participating in an active packet data or speech session. In both cases, specific HO events are issued, whose locations are derived and mapped onto the road network in

order to describe the subscribers mobility behavior. From that point, it can be ascertained when a particular speech or data session started or ended. Whereas for a speech session the start and end position can be derived from the respective call establishment and termination event (analogous to the estimation of HO positions), the position when a session started or ended remains unclear for packet data users. This is due to the fact that in the core network only HO events are recorded which indicate that a data session is currently active, but no specific initiation or termination events are issued as it is the case for speech users. On that account, we assume that a packet data session starts somewhere between the last LAU and the first related HO event. The same applies for the termination of the session, just the other way round. For estimating the position when a particular data session starts or ends, we use a uniformly distributed random variable in order to obtain the definite estimation of the location of the particular event. Figure 7 outlines this concept based on two simple scenarios.

4 Figures of merit

In order to evaluate our findings, we make use of a number of metrics that allow us to assess the estimated trajectories with respect to the ground truth trajectory, which was recorded using a GPS tracker. The first part of the evaluation affects the estimated route of the trajectory, which is compared to the actual one in order to derive a similarity factor. Secondly, the estimated HO positions are confronted with the real ones. For that reason, a mobile terminal was used during test rides in order to record the exact position where the moving subscriber was handed over to a new cell site or location area. Finally, the estimated velocities are verified using the recorded GPS tracks. Hereinafter, the used metrics are described in more detail.

4.1 Route geometry

In order to evaluate the estimated trajectory with respect to the actual, recorded one, we use two different metrics,

namely the Hausdorff and the Fréchet distance. These metrics allow to ascertain the similarities between the real and the estimated route.

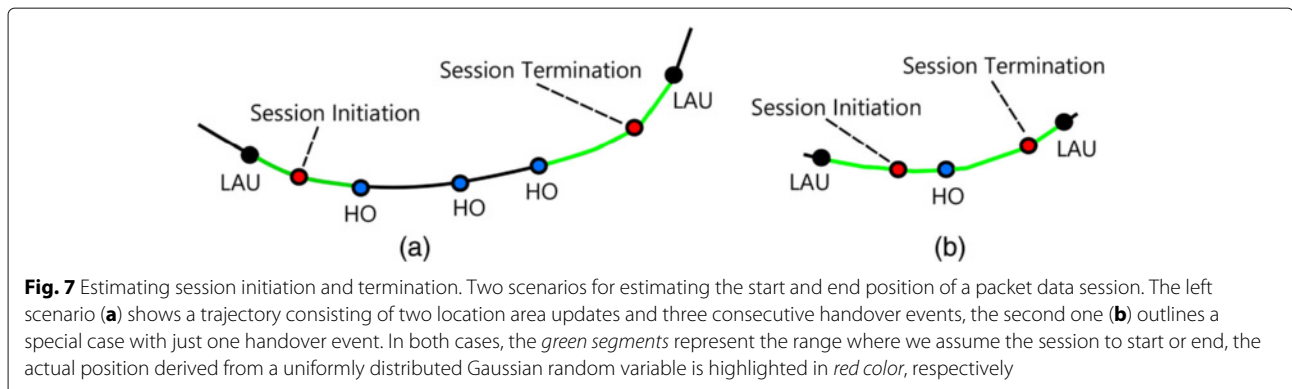
1. The *Hausdorff distance* $d_H(X, Y)$ (first introduced in [25]) measures how far two subsets X and Y of a metric space are apart from each other. Two subsets are close in the Hausdorff distance if every point x of one set is close to some point y of the other set. The Hausdorff distance then is the greatest of all the distances from a point x in subset X to the closest point y in subset Y . In Eq. (5), X and Y are the sets representing all points x of the estimated and all points y of the actual route, respectively, $d(x, y)$ constitutes the distance between a point x and a point y .

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\} \tag{5}$$

2. The *Fréchet distance* constitutes a measure of similarity between curves, which takes the actual position and order of points along the curves into consideration. We consider the algorithm developed by Alt and Godau [26] which computes the Fréchet distance of two polygonal curves in Euclidean space. Hence, the Fréchet distance for two curves $A, B : [0, 1] \rightarrow V$, in our case two routes, is defined as

$$\delta_F(f, g) = \inf_{\alpha : [0,1] \rightarrow [a,a']} \max_{\beta : [0,1] \rightarrow [b,b']} \|f(\alpha(t)) - g(\beta(t))\| \tag{6}$$

where α and β are continuous functions with $\alpha(0) = a, \alpha(1) = a', \beta(0) = b$ and $\beta(1) = b'$. Our evaluation was done using Christophe Genolinis' implementation of the Fréchet distance [27].



4.2 Handover deviation

Since we estimate the HO and LAU position based on the geometry of the coverage area of the involved cell sites, we need a metric that indicates how far the estimated positions deviate from the observed ground truth positions. The Euclidean distance $d(x, y)$ (in two-dimensional space, Eq. (7)) between the estimated position $x = (x_1, x_2)$ and the observed one $y = (y_1, y_2)$ is used as a metric. It is the basis to evaluate the position prediction as well as the coverage area estimation.

$$d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} \quad (7)$$

4.3 Velocity comparison

In order to verify our approach, we compare the estimated and adapted (refer to Algorithm 2) velocity with the ground truth velocity which can be obtained from recorded GPS information.

One of the metrics we use for this reason is the *mean absolute error* (MAE, Eq. (8)) between the observed average velocity v_i and the computed velocity \hat{v}_i . The observed average velocity v_i is the velocity obtained from GPS between the two successive HO positions. The computed velocity is derived by calculating the distance between the successive HO locations divided by the time difference between the corresponding events.

$$MAE = \frac{1}{n} \sum_{i=1}^n (|\hat{v}_i - v_i|) \quad (8)$$

The second metric is the *root mean square error* (RMSE, Eq. (9)), which is stronger influenced by large errors than by small ones. In both equations the variable n represents the number of event segments a trajectory consists of.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{v}_i - v_i)^2}{n}} \quad (9)$$

5 Experiments and results

In this section, we present results which have been achieved using the described methodology. For the reason of evaluation, we performed test drives in urban and semi-rural areas in Upper Austria and the city of Vienna. The recorded data are used as a reference in order to validate the estimated subscriber trajectories. First of all, we characterize the test setup and the test environment. Subsequently, we present various metrics and figures that allow to quantify the achieved results with respect to actual data. We show that our approach constitutes an eligible method to derive realistic trajectories for subscribers in a cellular network in both urban and semi-rural environments.

5.1 Environment

As mentioned beforehand, we performed several test drives in order to capture data which is required to verify our approach towards trajectory estimation. For each of these test drives, we were using the identical mobile terminal with the same subscriber identity module (SIM). The handset in use was a Samsung S3 running the Android 4.1.1 JellyBean operating system. During the test drives, the handset was used to initiate speech and data sessions. Throughout the entire drives, the mobile terminal's current position was recorded using a GPS tracker. Additional information regarding the actual serving cell sites and location areas was logged in order to derive the effective position of potential HO or LAU events. At a later time, the Austrian MNO A1 provided us with the subscriber information and events captured in the core network for the particular test drives. These data were used to derive the subscriber's trajectory, whereas the recorded data was required for validation and quantification.

Table 2 gives an overview of the start and end position and the number of HO events for each of the performed test drives. The number of HOs in brackets corresponds to the unfiltered raw events captured by the core network, the second number represents the same set of HO events after removing ping-pong HOs. The first test drive started in the city of Linz and ended in Treffling, a small town in the outskirts of Linz. This test drive was performed entirely on the higher road network and features a semi-rural scenario that consists of an urban and a rural section (trajectory depicted in Fig. 8). The second test drive took place within the borders of the city of Linz. This drive was performed on streets of the minor road network as well as on the city highway. The third and fourth test drive (refer to Fig. 9) took place in the city of Vienna. Both test drives were performed on the major and minor road network and constitute an urban mobility scenario.

5.2 Results

In order to be able to quantify and rate the developed mechanisms and procedures, evaluating the results achieved using our system with actual data is inevitable. For that reason, all in all, four test drives in different

Table 2 Overview of the start and end position and the number of HO events for the four test drive trajectories

#	Start [Lat,Lon]	End [Lat,Lon]	Handover
1	48.28105,14.30415	48.33499,14.3780	18 (22)
2	48.31616,14.29052	48.28169,14.30275	17 (19)
3	48.13993,16.32367	48.19880,16.25597	16 (20)
4	48.19283,16.27260	48.15278,16.30172	27 (37)

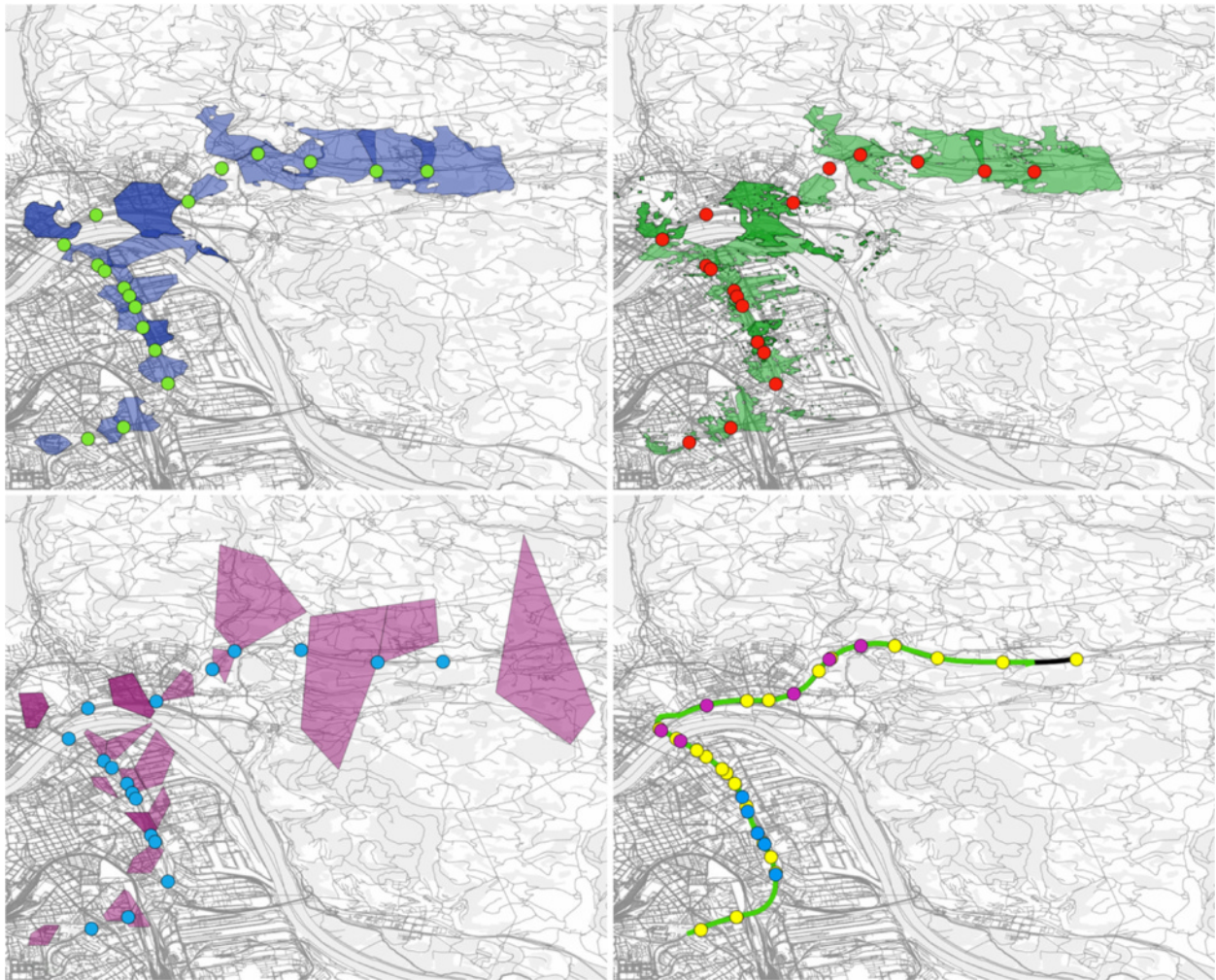


Fig. 8 Trajectory 1. Estimation of event positions based on different coverage prediction methods: Coverage prediction (*top left*), coverage prediction including height and building model (*top right*) and Voronoi tessellation (*bottom left*). The last picture shows the original trajectory recorded using a GPS tracker (*black*) and the estimated trajectory based on event positions derived from the top right image (*green*). The *purple points* indicate an active speech session, the blue ones a data session, respectively. The *yellow points* are the actual event positions which have been recorded using a handset

environments have been performed in order to collect reference data. This data basis allows to verify the estimated trajectories using real data under different aspects. First of all, the route geometries of the estimated trajectory and the actual one are confronted, and a similarity factor is derived. Furthermore, the deviation of the effective HO positions and the estimated ones is computed. The last evaluation aspect covers velocity estimation, which is done by comparing the derived subscriber velocity with the ground truth recorded by a GPS tracker. The findings presented on the following pages indicate that our approach constitutes an appropriate method for deriving subscriber trajectories based on mobile network operator data.

5.2.1 Route geometry

In order to quantify the similarity between the estimated trajectory and the actual route, two different metrics, namely the Hausdorff and the Fréchet distance, are used. Whereas the Fréchet distance is well-suited for curve matching, the Hausdorff distance represents the maximum distance between the estimated and the actual route. Hereby, the Hausdorff distance either indicates the maximum distance between the estimated start and end positions or the maximum distance between the estimated and the actual route. The results in Table 3 show a minimum and maximum Hausdorff distance of 0.0075°, which is equivalent to 590 m, and 0.0151° (1189 m), respectively. The minimal and maximal Fréchet distance of



Fig. 9 Trajectory 4. Estimation of event positions based on different coverage prediction methods: Coverage prediction (*top left*), coverage prediction including height and building model (*top right*) and Voronoi tessellation (*bottom left*). The last picture shows the original trajectory recorded using a GPS tracker (*black*) and the estimated trajectory based on event positions derived from the top right image (*green*). The *purple points* indicate an active speech session, the blue ones a data session, respectively. The *yellow points* are the actual event positions which have been recorded using a handset

0.0012°(94 m) and 0.0047°(370 m), respectively intimate a high similarity between the actual and the estimated trajectory for all four test cases.

5.2.2 Handover deviation

In order to verify the performance of our HO positioning algorithm, we compare three differently estimated HO

positions—each of them using a specific coverage estimation method—with the actual positions recorded by a mobile terminal.

Voronoi diagrams constitute the simplest, yet an appropriate method for deriving the position of a HO or LAU event. Though this mechanism only takes the location of the cell sites into consideration, the HO positions can be estimated reasonably well, compared to the actual positions. In the course of our investigations, we recognized that Voronoi diagrams work best in urban areas, which naturally feature a high density of cell sites. In semi-rural regions, where less cell sites usually cover larger areas, the emerging Voronoi diagrams become fairly large, since their extent is only limited by the surrounding cell sites, which can lead to certain HO positions being estimated very inaccurately.

Table 3 Route similarity computed with the discrete Hausdorff distance and the Fréchet distance

Trajectory	$d_H(X, Y)$	$\delta_F(f, g)$
1	0.0135°	0.0047°
2	0.0075°	0.0015°
3	0.0151°	0.0021°
4	0.0107°	0.0012°

Apart from Voronoi diagrams, we use cell coverage predictions performed using a network planning tool in order to incorporate additional information, e.g., transmitter power or antenna characteristics. For an even more realistic view on the cell sites' coverage we enhance this prediction with a building model. Figure 10 shows the deviations of HO positions for the first test trajectory. On the horizontal axis the particular HO events are listed whereas on the vertical axis the distance between the estimated and the real position is shown. The differently colored bars indicate the three different coverage estimation methods. It can be seen that the deviation increased in the second half of the plot. This was caused by leaving the urban area and driving into a semi-rural one. Here, Voronoi diagrams performed worse compared to the other coverage predictions. The reason for this is that in rural environments the cell coverage areas get larger and the differences between Voronoi polygons and real coverage areas increase since Voronoi tessellation does not consider any terrain information. On the other coverage prediction hand is able to take such environmental information into account. This in further consequence leads to smaller errors in estimating HO positions.

Table 4 shows the minimum, the first quartile, the median, the mean, the third quartile and the maximum position deviation in kilometers for all four test

Table 4 The minimum, first quartile, median, mean, third quartile and maximum handover position deviation in kilometers for each of the four test drive trajectories using three different coverage predictions M (coverage prediction P, coverage prediction with building PB and Voronoi diagrams V)

#	M	min	Q ₁	\bar{x}	\bar{x}	Q ₃	max
1	P	0.161	0.468	0.663	0.705	0.878	1.403
1	PB	0.101	0.547	0.716	0.711	0.928	1.315
1	V	0.142	0.420	0.837	0.715	1.023	1.155
2	P	0.219	0.286	0.321	0.401	0.483	0.900
2	PB	0.132	0.261	0.346	0.399	0.525	0.884
2	V	0.048	0.229	0.343	0.386	0.482	0.900
3	P	0.111	0.360	0.449	0.677	0.604	3.786
3	PB	0.148	0.338	0.451	0.679	0.536	4.062
3	V	0.013	0.281	0.409	0.669	0.639	4.085
4	P	0.020	0.223	0.309	0.355	0.504	0.863
4	PB	0.030	0.213	0.325	0.359	0.506	0.850
4	V	0.010	0.175	0.321	0.328	0.443	0.802

trajectories. All HO positions have been estimated with Algorithm 1 and were performed with the three mentioned coverage estimation methods.

By combining the HO deviations over all four test trajectories and computing the MAE and RMSE, coverage

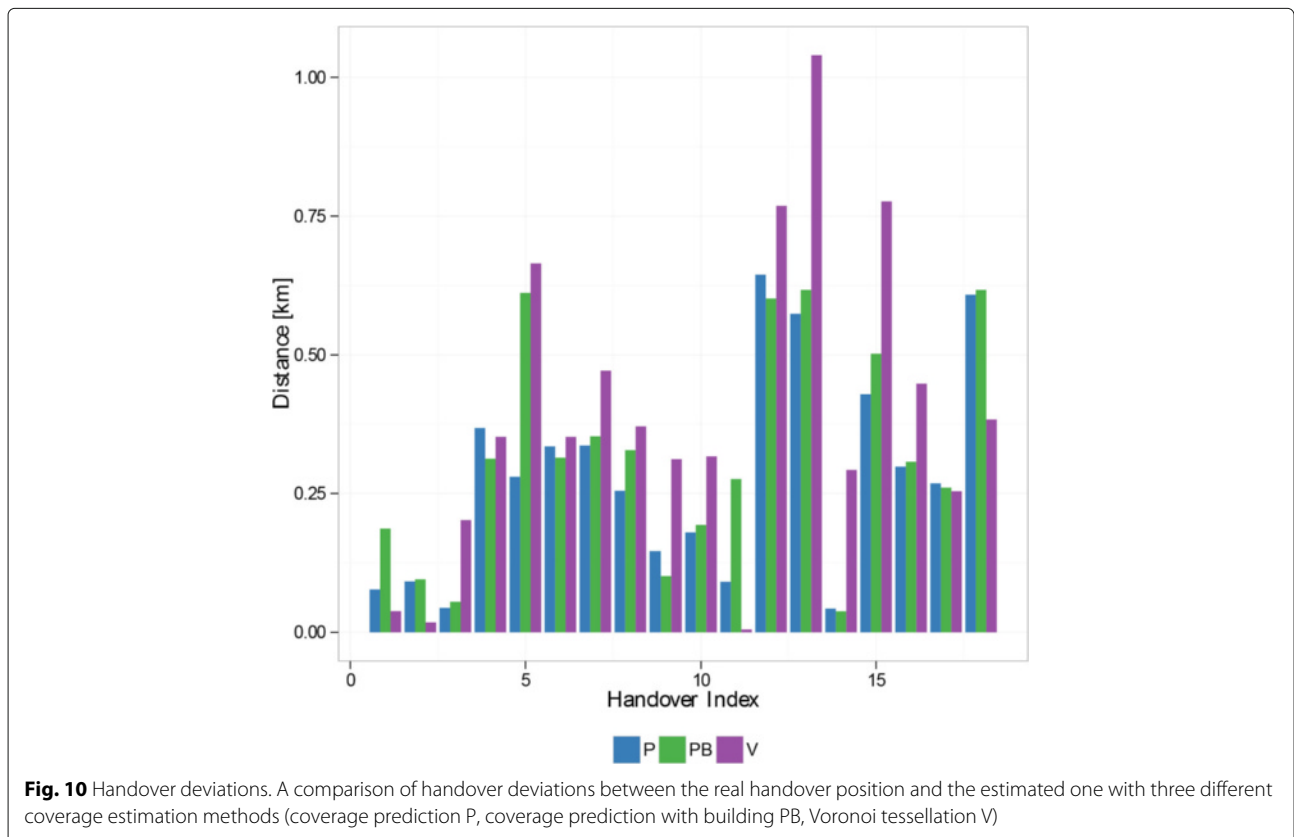


Fig. 10 Handover deviations. A comparison of handover deviations between the real handover position and the estimated one with three different coverage estimation methods (coverage prediction P, coverage prediction with building PB, Voronoi tessellation V)

prediction without building information performed best with a MAE of 0.176 km and a RMSE 0.231 km, followed by the extended coverage prediction with a MAE 0.189 km and a RMSE 0.253 km and Voronoi diagrams with a MAE 0.238 km and RMSE 0.311 km. The slightly higher errors when using coverage prediction including a building model originate from the inaccurate building heights, which are either completely unknown (e.g., city of Linz, where an average height of 15 m is assumed) or available in an inadequate resolution (e.g., city of Vienna).

For the fourth trajectory in particular the minimum, maximum and mean HO position deviations are 0.010, 0.802, and 0.328 km, respectively, which shows that the HO positions can be estimated with high accuracy.

The third trajectory shows a high maximum HO deviation which was caused by the network not performing a HO while driving for 3.55 km within the city of Vienna. Since the HO algorithm computed a HO position between the two cell site coverage, areas that were far-off a large HO deviation occurred.

5.2.3 Velocity estimation

In this section, we want to discuss the results of the raw velocity estimation and the adapted velocity estimation after HO repositioning. We compared each of the velocities against the ground truth velocity derived from the

recorded GPS data. Since the estimated velocity accords to the average velocity between two HO positions we computed the average GPS velocity between the respective HO positions. In Fig. 11, the raw estimated velocity for the first trajectory is illustrated. It can be seen that there occur many velocity overruns for all coverage estimation methods. By comparing this figure with the HO deviation (see Fig. 10) it becomes obvious that a velocity overrun is caused by a large deviation between the estimated and the actual HO position (≥ 0.5 km).

In contrast, Fig. 12 depicts the estimated velocity after applying the adaption and HO repositioning algorithm. Hence, we see that the bigger deflections could be reduced. To verify the accuracy of the estimated velocity, the MAE and RMSE for both velocity estimations and for each trajectory and coverage prediction method were computed. The results are shown in Table 5 and indicate that the adaption algorithm can reduce the error introduced by a wrong HO position estimation significantly.

For the coverage prediction without building model, the velocity MAE using the adaption algorithm was reduced on each trajectory by an average of 60.40 %. For the coverage prediction enhanced with a building model the reduction was 65.75 %, for Voronoi diagrams 73.01 %.

However, this indicates only how well the adaption algorithm works but not how accurate the velocity estimation

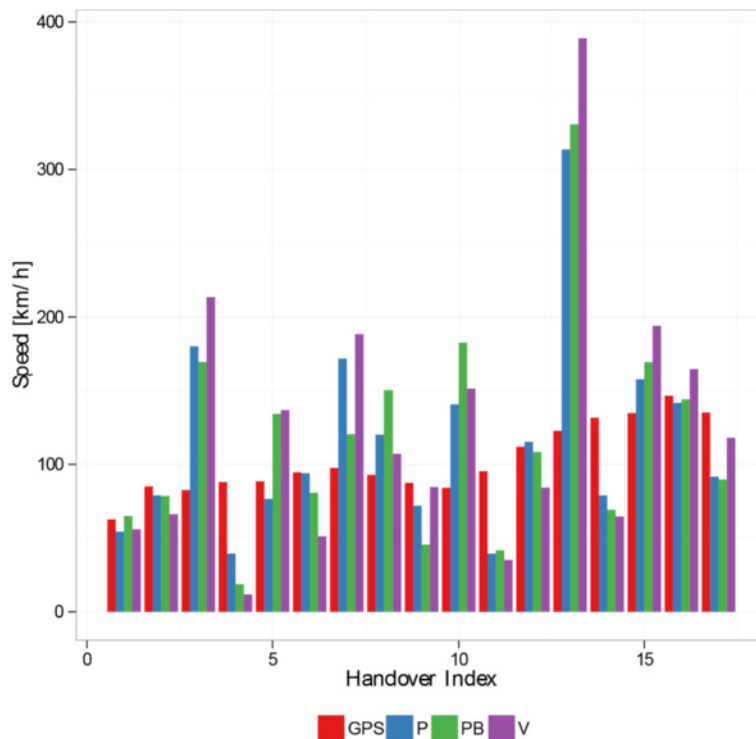
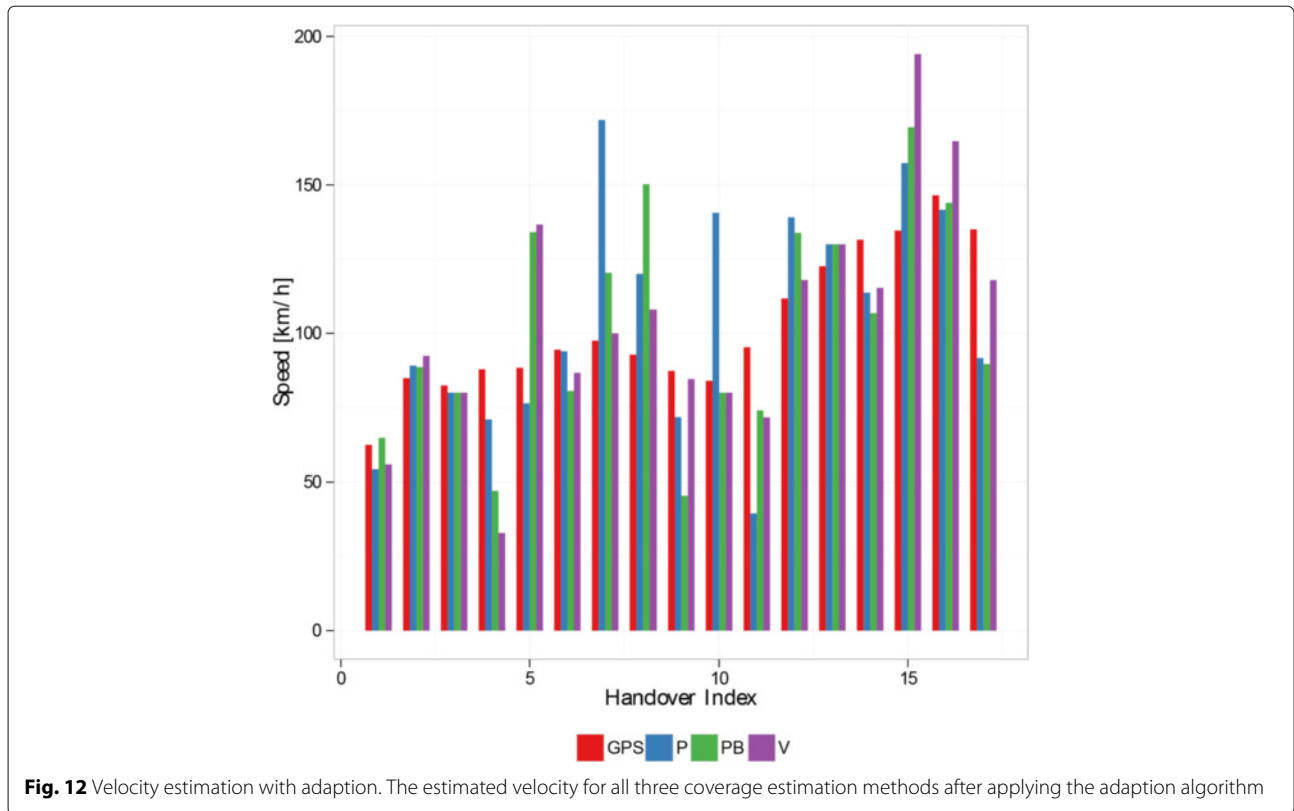


Fig. 11 Velocity estimation without adaption. The estimated velocity for all three coverage estimation methods before applying the adaption algorithm



is in general. To validate this for the three different coverage prediction methods, the absolute adapted velocity errors from the four trajectories were combined in order to compute the MAE. Here, the coverage prediction enhanced with a building model had the minimum absolute error for all trajectories with 14.887 km/h, followed by Voronoi diagrams with 14.921 km/h and coverage prediction without building model 16.004 km/h. This indicates

Table 5 MAE and RMSE for the four test drive trajectories using three different coverage predictions

#	M	MAE	RMSE	MAE _{adapted}	RMSE _{adapted}
1	P	42.365	62.593	23.360	31.514
1	PB	50.315	69.985	23.127	29.142
1	V	59.704	85.549	17.648	25.250
2	P	81.304	138.707	8.970	16.047
2	PB	60.964	99.090	9.436	12.086
2	V	113.999	181.854	13.029	16.932
3	P	21.125	26.808	13.305	17.799
3	PB	26.384	33.786	11.941	16.607
3	V	28.449	40.088	13.036	18.781
4	P	58.425	150.199	17.081	22.000
4	PB	48.061	125.747	14.553	17.795
4	V	72.715	191.037	15.388	21.599

that post-processing using the adaption algorithm is necessary in order to estimate an accurate movement velocity for a particular subscriber.

6 Conclusions

We have shown that trajectories can be estimated for cellular subscribers by using subscriber information captured in the core network of a MNO and the location of the respective BSs. We presented three methods for the cell site coverage estimation and compared them in terms of the accuracy of HO positions and velocity estimation. Our results show that the HO positions can be estimated with a mean deviation of 179 m by using coverage prediction without a building model. The average (recorded) distance between two consecutive HO events was 625 m in semi-rural environments and 283 m in urban regions, respectively. To minimize the effect of a deviated HO position, we developed an adaption algorithm that is able to reduce velocity overruns and decrease the mean absolute error. The evaluation has shown that the adaption algorithm reduces velocity overruns for all four test drive trajectories. The mean absolute velocity error over all four trajectories was 14.887 km/h (the actual speed of all four test drives was in the range between 50 and 145 km/h) when using coverage prediction enhanced with a building model. This indicates that the velocity can be estimated for urban and rural areas with good accuracy.

The presented results show that cellular network subscriber trajectories can be generated for the purpose of mobility simulations. These trajectories cover the subscriber's route and provide its velocity and are therefore well-suited to describe the subscriber's mobility behavior. Furthermore, based on the estimated position of the corresponding HO and LAU events, it is possible to derive a time and spatial frame for applying a traffic model for a single subscriber, which is useful, e.g., for real network traffic simulations.

Endnotes

¹Statistics about the German road network [28].

²The Digital Elevation Model over Europe from the GMES RDA project: <http://www.eea.europa.eu/data-and-maps/data/eu-dem>.

³More information about how building information can be retrieved from OpenStreetMap: <http://wiki.openstreetmap.org/w/index.php?title=Buildings&oldid=1050850>.

⁴The GIS service of the city of Vienna: <https://www.wien.gv.at/kultur/kulturgut/architektur/gebaeuedaten.html>.

⁵Corine Land Cover 2006 seamless vector data: <http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-3>.

⁶Population density disaggregated with Corine land cover 2000: <http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2>.

Acknowledgements

This project was supported by the program Regionale Wettbewerbsfähigkeit OÖ 2010-2013, which is financed by the European Regional Development Fund and the Government of Upper Austria.

Competing interests

The authors declare that they have no competing interests.

Received: 18 January 2015 Accepted: 27 June 2016

Published online: 06 October 2016

References

1. J Broch, DA Maltz, DB Johnson, Y-C Hu, J Jetcheva, in *Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking*. A performance comparison of multi-hop wireless ad hoc network routing protocols, *MobiCom '98* (ACM, New York, 1998), pp. 85–97
2. I Rubin, CW Choi, Impact of the location area structure on the performance of signaling channels in wireless cellular networks. *Comm. Mag.* **35**(2), 108–115 (1997)
3. JJ Garcia-Luna-Aceves, EL Madrga, in *Proceedings of the Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*. A multicast routing protocol for ad-hoc networks (IEEE, New York, 1999), pp. 784–792
4. MM Zonoozi, P Dassanayake, User mobility modeling and characterization of mobility patterns. *IEEE J. Sel. Areas Commun.* **15**(7), 1239–1252 (2006)
5. BB Mandelbrot, *The Fractal Geometry of Nature*. (W.H. Freeman and Company, New York, 1982)
6. MF Shlesinger, GM Zaslavsky, J Klafter, Strange kinetics. *Nature.* **163**, 31–37 (1993)
7. B Liang, ZJ Haas, in *Proceedings of IEEE Information Communications Conference (INFOCOM 1999)*. Predictive distance-based mobility management for pcs networks (IEEE, New York, 1999), pp. 1377–1384
8. I Rhee, M Shin, S Hong, K Lee, SJ Kim, S Chong, On the Levy-walk nature of human mobility. *IEEE/ACM Trans. Networking.* **19**(3), 630–643 (2011)
9. MC Gonzalez, CA Hidalgo, A-L Barabasi, Understanding individual human mobility patterns. *Nature.* **453**(7196), 779–782 (2008)
10. S Misra, P Agarwal, Bio-inspired group mobility model for mobile ad hoc networks based on bird-flocking behavior. *Soft. Comput.* **16**(3), 437–450 (2012)
11. F Morlot, I École, N Supérieure, F Baccelli, SE Elayoubi, in *Proceedings of IEEE INFOCOM*. An interaction-based mobility model for dynamic hot spot analysis (IEEE, New York, 2010), pp. 1–9
12. S Misra, S Singh, M Khatua, MS Obaidat, Extracting mobility pattern from target trajectory in wireless sensor networks. *Int. J. Commun. Syst.* **28**(2), 213–230 (2015)
13. S Misra, J Mahapatro, M Mahadevappa, N Islam, Random room mobility model and extra-wireless body area network communication in hospital buildings. *IET Networks.* **4**(1), 54–64 (2015)
14. T Liu, P Bahl, I Chlamtac, Mobility modeling, location tracking, and trajectory prediction in wireless atm networks. *IEEE J. Sel. Areas Commun.* **16**(6), 922–936 (2006)
15. J Schlaich, T Otterstätter, M Friedrich, in *Proceedings of the 89th Annual Meeting Compendium of Papers*. Generating Trajectories from Mobile Phone Data (Transportation Research Board of the National Academies, Washington D.C., 2010)
16. T Tettamanti, H Demeter, V István, Route choice estimation based on cellular signaling data. *Acta Polytechnica Hungarica.* **9**(4), 207–220 (2012)
17. R Becker, R Cáceres, K Hanson, S Isaacman, JM Loh, M Martonosi, J Rowland, S Urbaneck, A Varshavsky, C Volinsky, Human mobility characterization from cellular network data. *Commun. ACM.* **56**(1), 74–82 (2013)
18. M Alger, E Wilson, T Gould, R Whittaker, N Radulovic, Real-time Traffic Monitoring using Mobile Phone Data (2004). <http://www.maths-in-industry.org/miis/30/>. Accessed 12 Jan 2015
19. H Bar Gera, Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. *Transp. Res. C.* **15**(6), 380–391 (2007)
20. D Valerio, Road Traffic Information from cellular network signaling. Technical Report FTW-TR-2009-003 (2009)
21. A-E Baert, Semé, in *International Symposium on Algorithms, Models and Tools for Parallel Computing in Heterogenous Networks*. Voronoi Mobile Cellular Networks: Topological Properties (IEEE Computer Society, Los Alamitos, 2004), pp. 29–35
22. 3rd Generation Partnership Project 2, CDMA2000 Evaluation Methodology, Revision A (2009). http://www.3gpp2.org/public_html/specs/C.R1002-A_v1.0_Evaluation_Methodology.pdf. Accessed 12 Jan 2015
23. European Telecommunications Standards Institute, Universal Mobile Telecommunications System (UMTS): Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS (UMTS 30.03 version 3.2.0) (1998). http://www.etsi.org/deliver/etsi_tr/101100_101199/101112/03.02.00_60/tr_101112v030200p.pdf. Accessed 12 Jan 2015
24. ND Tripathi, in *Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th*. Simulation based Analysis of the Radio Interface Performance of an IS-2000 System for various Data Services, vol. 4 (IEEE, Atlantic City, 2001), pp. 2665–2669
25. F Hausdorff, *Grundzüge der Mengenlehre*. (Veit and Company, Leipzig, 1914)
26. H Alt, M Godau, Computing the Fréchet distance between two polygonal curves. *Int. J. Comput. Geom. Appl.* **5**, 75–91 (1995)
27. C Genolini, Computing the Fréchet Distance between two Trajectories (2014). <https://cran.r-project.org/web/packages/longitudinalData/longitudinalData.pdf>. Accessed 14 Sept 2016
28. Statistische Ämter des Bundes und der Länder, Verkehr - Strassen des überörtlichen Verkehrs (2014). http://www.statistik-portal.de/Statistik-Portal/de_jb16_jahrtab36.asp. Accessed 14 Sept 2016