Building a minimal traffic model from mobile phone data

Michael Zilske Kai Nagel

March 17, 2013

We investigate setting up a traffic simulation scenario only from a high spatial resolution set of mobile phone trajectories and an OpenStreetMap road network model. Mixed road traffic is modelled as the result of a choice for each user between a simulated congested car mode and a non-congested alternative mode parameterized by its speed.

1 Introduction

Transport planning is a multi-faceted exercise, necessitating many inputs from many different sources. One such source are simulations of the transport system. With such simulation system, one can insert a considered infrastructure or policy measure into the model, and observe the simulated reactions of the transport system. Downstream modules, such as the calculation of emissions (e.g.[1]), accessibilities (e.g.[2]) or inputs to a cost-benefit-analysis, can be attached (e.g. [3]). An important issue with such models is that it is rather time-consuming and expensive to put them together; for example, the model for the German national assessment exercise takes several years to put together, which may be prohibitive especially for a developing country.

In this situation, it is interesting to consider alternative, and possibly faster and cheaper, approaches. With the availability of OpenStreetMap (OSM) data, one major obstacle has been removed. We have consistently found that it is possible to base traffic simulation models based on that data in spite of some shortcomings [4], and while the data quality of OSM differs heavily between urban and rural areas, it approaches that of commercial network data for large cities.[5] The main issue is that OpenStreetMap data does not contain flow capacities, i.e. the maximum number of vehicles that can leave a link during an hour. Instead, that number is estimated from road category information.

The other missing item in order to bring such a model up and running is the demand. The demand contains, in some way, information about all trips that are made from one location to another during one day. Such demand data is typically obtained from surveys. Two typical sources are the census or similar information (which typically contains, besides the home location, the work or education location), or trip diaries, which are in many countries obtained from asking a sample of the population about how, and at which locations, they spent a certain day. However, sometimes such information is not available, or it is difficult to procure, for example because of privacy issues. In this situation, the generation of synthetic demand from electronic sources has become an increasingly active research field. Some of these investigations have a focus on route choice (e.g. [6]), while others derive origin/destination matrices [7][8]. Most of them employ some sort of intermediate model of behavior or trip generation before assigning traffic to the road. In consequence, the guiding focus for the present paper is the question in how far a meaningful traffic simulation can be constructed directly just out of OpenStreetMap network data, and anonymous cell phone traces as provided by the "Data for Development" (D4D) challenge.[9]

The paper is structured as follows: In section 2, we describe the data sets used for creating the supply (network) and demand (population) data. Section 3 covers the construction of the simulation model. In section 4, we give some results of parametric simulation runs. In section 5, we discuss some of the issues we faced while constructing the model and directions for future research.

2 Data description

2.1 Road network

The road network data for this scenario is based on OpenStreetMap. The OpenStreetMap data was converted to a simulation network by assigning attributes related to traffic flow to road segments. This is done based on the value of the `highway` tag, a road classification scheme particular to OpenStreetMap.[4] A graph representation of the road network is constructed from the OpenStreetMap data, where each intersection becomes a vertex and each road segment becomes a link. The precise geographic embedding of the road segments is discarded, and only its length is stored as a link attribute, along with its capacity, maximum speed in uncongested state, and number of lanes. The values we used are given in figure 1. The resulting network for the entire country has approximately 22,000 nodes and 63,000 links.

According to the CIA world factbook web site, the length of the road network of Côte d'Ivoire is 80,000 km, of which 6,500 km are paved. The OpenStreetMap documentation features an overview of the international equivalence of highway tags, but no country in Africa is currently included in this overview. The combined length of all edges labelled with any value of the highway tag is 29,300 km. The combined length of all edges labelled with `highway=primary`, `highway=motorway` or `highway=trunk` is 9,000 km, suggesting that these categories together already contain some roads classified as unpaved by the other source, and that the rest of the network should definitely be considered unpaved. Still, the data only accounts for less than half of the reported network length, so it is to be expected that siginificant parts of the network, most of it probably in rural areas, is not available in the model. This contrasts with coverage in the more economically developed countries, where the part of the road network not covered by OpenStreetMap can be considered negligible for traffic modelling purposes.

Highway tag	Lanes	Free speed (km/h)	Capacity (veh/(l*h))
`motorway`	2	120	2000
`motorway_link`	1	80	1500
`trunk`	1	80	1500
`trunk_link`	1	50	1500
`primary`	1	80	1500
`secondary`	1	60	1000
`tertiary`	1	45	600
`minor`	1	45	600
`unclassified`	1	45	600
`residential`	1	30	600
`living_street`	1	15	300

Table 1: Link attributes used for the values of the OpenStreetMap highway tag.

Upon inspection, certain areas of the country seem to have a good coverage, notably the area of the economic capital, Abidjan. For this reason, and because our simulation approach has so far been applied mostly for urban areas, we decided at this point to focus on the city.

2.2 Mobile phone sightings

The mobile phone data under consideration here consist of two sets of individual trajectories collected over a study period of 150 days: One set of high spatial resolution (HSR) data, and one set of long term (LT) data. The HSR set consists of trajectories generated from billing data, and tracks 50,000 individuals. The individuals are drawn from the customer base of one mobile phone operator, which claims five million customers, at an estimated total population of twenty million.

Locations are given by the number of the cell phone tower with which the mobile phone was communicating at the time of the record. Locations are only recorded while the user is in a call. After every two-week period, the population sample is redrawn, so it is not possible to track a single individual over more than two weeks. The LT set tracks another sample of 50,000 individuals, but over the whole study period, at the price of providing much lower spatial resolution, namely on the level of sub-prefectures. In this experiment, the idea was to directly generate a pattern of daily traffic from trajectories, so we use the HSR data set.

3 Traffic simulation model

The simulation is a loop which consists of:

- traffic flow simulation
- scoring

- replanning

This loop operates on an initial population of individiual agents which is generated from the mobile phone data. The following sections describe the phases in detail.

3.1 Initial Demand

Côte d'Ivoire has an estimated population of 20 million, according to the World Bank population data set on `google.com/publicdata`. We generate a synthetic one percent sample population by creating 200,000 synthetic agents. We simulate no particular day, but a typical work day, so we overlay four arbitrary days of mobility traces, each taken from a different sample (size 50,000) of mobile phone customers from the HSR data set. This means that we are combining data from multiple days to build a population for a supposedly average working day. Clearly, this is not the same as having 200,000 samples from one day. It is, however, still better than the alternative, which is expanding the first 50,000 samples to 200,000.

Each agent is equipped with a mobility plan which consists of an alternating list of

- geo-located and timed activities
- leg descriptions, including mode of transport and route

Agents essentially divide their time between conducting an activity, and travelling.

For each agent, an initial mobility plan is devised which is consistent with the data: The agent has to be in cell C_i at time t_i for every reading i. This leaves many degrees of freedom. In particular, it is not known whether the agent is travelling when a reading is taken. Any partitioning of the time into activities and trips which is consistent with the readings is in principle admissible. We start by defining every reading i to be an activity which ends at the time t_i the reading is taken. If several consecutive readings happen at the same cell, only the latest of those is considered. At this time, the agent will start travelling towards the location of the next reading $i + 1$ and, upon arrival, will stay there until the time t_{i+1}. During this time, the agent is considered to be conducting an activity. Activity locations are fixed to a geographical point which is randomly drawn from the Voronoi cell C_i of the tower where the reading was taken. It is assumed that activity locations have direct access to the road network, so they can be positionally identified with links. Each randomly drawn point is therefore snapped to the end of the nearest link, which is considered to be the activity location.

The plan is then checked for initial feasibility. Each leg is routed through the road network on the fastest route at maximum uncongested vehicle speed, as per the link attributes stored in the road network model. The travel time is summed up, and it is checked if the agent would under these assumptions be able to reach the next activity location in time. If this is not the case, the plan is redrawn. This procedure is iterated 20 times, and if no feasible plan has been found by then, the case is considered pathological and discarded. This does not necessarily mean that the input data does not resemble a real trajectory. Incomplete road network data is the most likely reason for these cases.

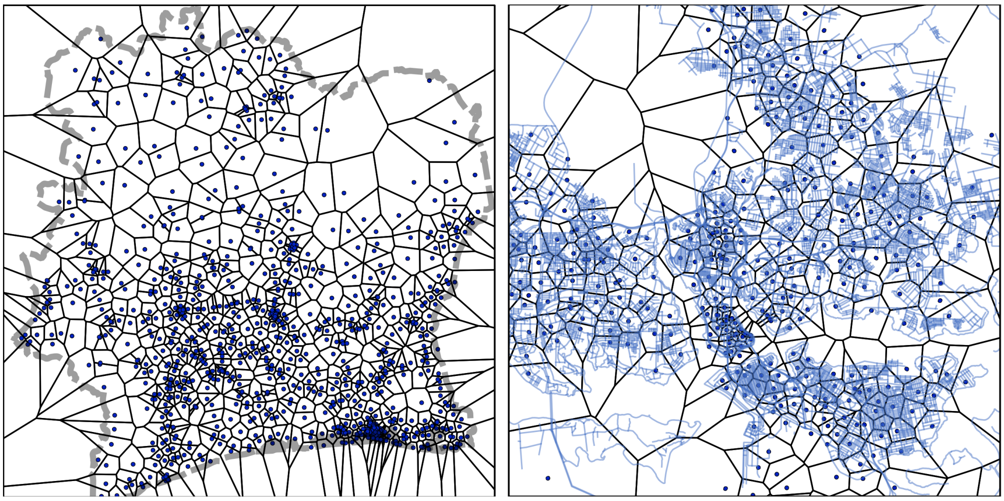

Figure 1: Cell tower locations with their Voronoi cells. Comparing Voronoi cell sizes on a country-wide scale (left, overlayed with national border) and on the scale of the Abidjan urban area (right, overlayed with road network) show the large variation in cell size.

Finally, we filter and keep only those plans with at least one sighting in the Abidjan urban area, which is our study area.

The result of this initial demand generation process is a 1% synthetic population sample where every agent uses a car for every trip and tries to take the freespeed-fastest route through the road network. This initial population is then fed into the simulation loop for relaxation, the steps of which are described in the next paragraphs.

3.2 Traffic flow

The mobility simulation concurrently executes the mobility plans of the agents. Agents leave their activity location at the scheduled activity end time and head for their next destination. The road traffic is simulated using the queuing model of traffic flow[10]. In this model, the limited flow capacity of links is honored, so that in every time step, only as many vehicles can exit a link as specified. If more vehicles enter a link than its flow capacity admits, vehicles will accumulate on the link until its storage capacity is hit, which is determined by the length and width of the link. When a link is full, no more vehicles can enter, causing the congestion to propagate back upstream. The simulation records time use for each agent: The points in time where agents depart from and arrive at their activity locations are fed back to the agents as experience to evaluate the utility of the executed plan. Note that simulated vehicles are considered uniform. In this experiment, in contrast to explicitly modelling different vehicle types and their interaction [11], we model mixed traffic as a combination of uniform car traffic and the possibility to switch to an uncongested mode, which is described below.

```
<person id="10008_1">
 <plan>
  <act type="sighting"
   link="91273255_1059606493_1059606739_R"
   x="-455429.63088982203" y="593487.4664630938"
   end_time="07:42:00" />
  <leg mode="car"
   dep_time="07:42:00"
   trav_time="00:04:15"
   arr_time="07:46:15">
    <route type="links">
     91273255_1059606493_1059606739_R
     91273255_1059606493_1059606739
     91273253_1219665629_1059606739_R
     [...]
     125239948_338881494_338881537</route>
   </leg>
   <act type="sighting"
    link="125239948_338881494_338881537"
    x="-454496.5669239445" y="593286.9239709259"
    end_time="17:36:00" />
   <leg mode="car"
    dep_time="17:36:00"
    trav_time="00:00:31"
    arr_time="17:36:31">
     <route type="links">
      [...]
     </route>
    </leg>
    <act type="sighting"
     link="30630786_338406146_338406157"
     x="-454806.9268885712" y="593254.6327337974"
     end_time="18:35:00" />
 </plan>
</person>
```

Figure 2: Example for an agent plan. This person had 3 call records on the examined day. From the location of the first call record (("sighting", location randomized within the cell) to the second, the free speed travel time is only about 4 minutes, from which the actually experienced, congested travel time may differ enormously. Upon arrival at the second location, the agent will wait until 17:36:00 and depart for its final destination.

3.3 Plan scoring

The agents evaluate the outcome of their plan with a simple utility-based approach. Time spent travelling is considered to contribute negative utility. Since, in this study, we do not have an activity model which would allow comparing the relative utility of time spent at one location with another, we consider time spent in activities to have no contribution to the utility function. Since we have no prior knowledge about the trip strucure, and our modelling decision to split trips at the locations of the GSM readings is somewhat arbitrary in this respect, we consider the disutility of travelling to be linear in total travel time.

$$U = \beta_{trav} \cdot t_{trav} \tag{1}$$

Since in this experiment the utility function does not have any other terms, the value for β_{trav} is arbitrary, as long as it is negative.

3.4 Replanning

After each iteration, the agent population has the opportunity to change their mobility plans in reaction to the outcome of the mobility simulation. In this study, agents have three replanning options. Two of them are creative. Agents chosing these options produce a new plan and execute it in the next iteration of the mobility simulation. These options are route choice and mode choice. The third is switching plans, in which agents retry a previously executed plan from their plan memory based on its previously experienced utility. In each iteration, 10% of the agent population consider their route choice and mode choice, respectively. The remaining 80% can switch plans.

Route choice Agents reconsider their route through the road network. Instead of taking the least-cost path based on free speed travel times, the link travel times computed as an outcome of the last iteration of the traffic flow simulation are used. In the first iteration, the traffic will concentrate on main roads, leading to high traffic volumes and high traffic times. Agent reconsidering their route will divert to smaller roads in the next iteration.

Mode choice In the initial population, all trips are done by car. Our model summarizes all alternatives to driving a car in a second mode. Agents which choose this mode are not routed through the network at all. They experience a travel time calculated from the free-speed car travel time between the origin and destination locations, times a travel time factor which characterizes the mode. These agents do not interact with other agents while travelling. They are not impeded by other travellers and do not contribute to congestion themselves.[12] Depending on the travel time factor, a certain share of the population remove themselves from the road network. Note that the modal split is not part of the input data, but an output of the simulation, dependant in particular on the travel time factor.

Switching plans Every agent has a fixed-size plan memory, set to size 5 in this experiment. Agents which are assigned the option to switch plans pick one of their previously tried plans uniformly at random, and switch to that plan with a probability depending on the difference between the most recently experienced scores of both plans:

$$p_{ij} = 0.01e^{\frac{s_j - s_i}{2}} \tag{2}$$

In this equation, p_{ij} is the probability of switching frm plan i to j, s_i is the current score of plan i, and 0.01 is the probability of switching between equally scored plans. The simulation is iterated until the system reaches a relaxed state. We consider this to be the case as soon as the average agent score (i.e. travel times) and the mode share have stabilized.

4 Implementation and Results

The scenario was implemented using the MATSim agent-oriented transport simulation software (`www.matsim.org`). In order to be able to run experiments on a desktop computer, we decided to simulate a 1% sample of the synthetic population described in the previous section, scaling the network capacity accordingly. In our experience with the process, this is sufficient to pick up large-scale characteristics of the system. One simulation run takes about an hour on a 2.2 GHz Intel Core i7 MacBook. A run consists of 180 iterations of the simulation loop, which we found to be enough for the quantities presented in this paper to stop drifting. For the last 30 iterations, the creative replanning options, namely route choice and mode choice, are disabled, and agents only switch between existing plans. This is done to eliminate the bias introduced by having a large fraction of the agent population take new routes or the alternative mode without regard for their possibly low utility.

We produced several parametric simulation runs, varying the travel time factor for the non-car alternative. As can be seen in figure 3, a travel time factor of 4 already leaves some agents unable to reach the point of their last sighting before midnight, which means that this state of affairs would be clearly inconsistent with the data. Factor 2 still seems admissible. The travel time factor can be interpreted as how long car travel times along a path in the congested network need to become compared to its free-flow state so that agents travelling along that path will be moved towards using the alternative mode.

In figure 4, we plot the resulting share of car drivers over the travel time factor. Since a factor of 4 is already considered inadmissible, the prediction would be a share of not much more than 0.2.

Even with a factor of 4, i.e. a situation which is quite congested according to Fig. 3, no systematic or directed traffic jam patterns emerge. The network just seems too full overall. This is quite different from other similar studies (e.g.[13, 14]), where we always found quite well-structured congestion patterns, in particular into the city during the morning peak. Further inspection of the results leads to the observation that most of the congestion in our model seems to be away from the freeways, on the secondary road network. That is, under congested conditions traffic is unable to get out of the

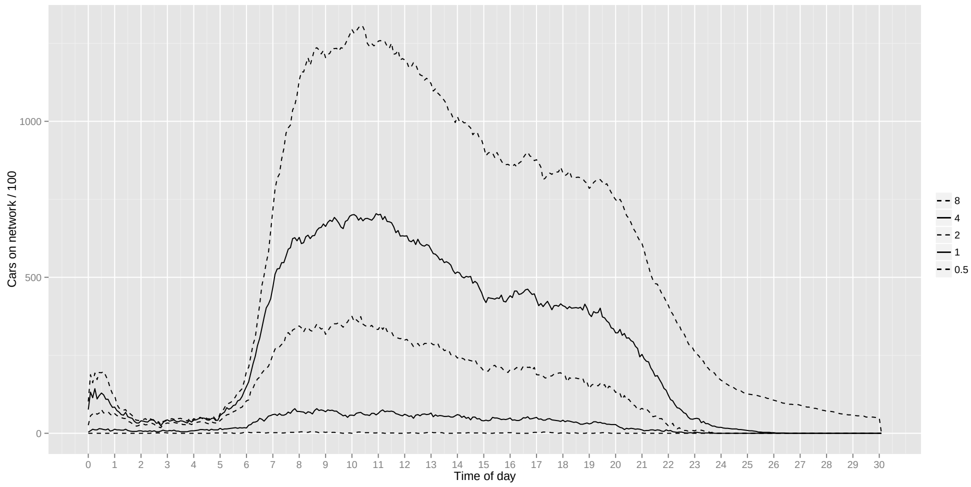

Figure 3: Number of cars en-route over time of day, plotted for different values of the alternative mode travel time factor. One agent represents 100 travellers. Note that as only sightings from a single day are used to construct the artificial population, agents are expected to be at their final location by midnight at the latest. For factor 4, the network is already too full to permit this. If the alternative mode is faster than driving (factor 0.5), the network is, expectedly, cleared.

secondary street network. Once the traffic makes it onto the primary network, the model displays few if any restrictions. Clearly, this statement would need to be verified on the ground before being a possible basis for planning decisions. It could, for example, also be a consequence of the demand generation, which, in particular because of spurious cell handovers, may generate a lot more local traffic than there is in reality. If such verification on the ground would corrobate that the local congestion effects are over-estimated, then methods to remove those spurious cell handovers from the demand generation would need to be inserted into the model.

Figure. 5 displays the probability density of the total travel time per person per day. One notices a peak near 0.2 hours for the car mode, and near 0.6 hours for the non-car mode. While the 0.6 hour value seems plausible for a reasonable walk length of 30 minutes and an average number of daily trips of less than 2 (given in [15]), the 0.2 value seems way too low, suggesting again that we overestimate the number of local (very short) car trips.

5 Discussion and outlook

Modelling road network access In the present scenario, as well as in the MATSim software package, the assumption is that every activity location has direct access to the modelled road network. For rural areas in developing countries, this is clearly not met, if

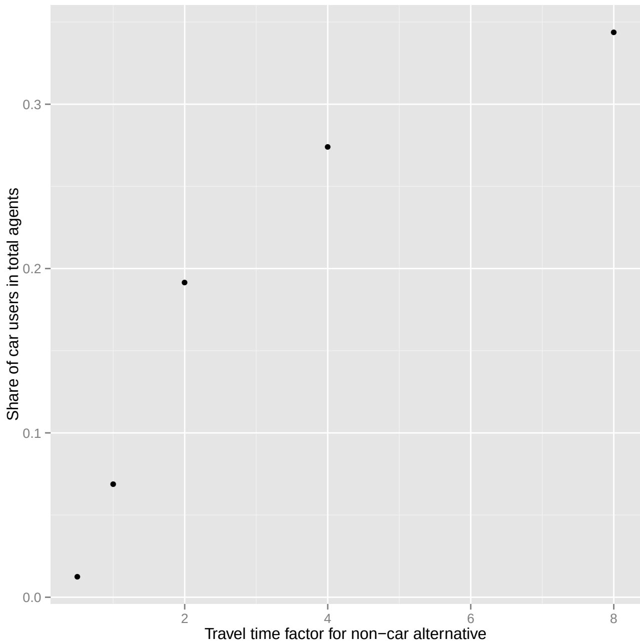

Figure 4: Number of agents using a car, plotted over the alternative mode travel time factor. Note that an agent either does or does not use a car for all trips of the simulated day. We do not consider mode choice for individual trips.

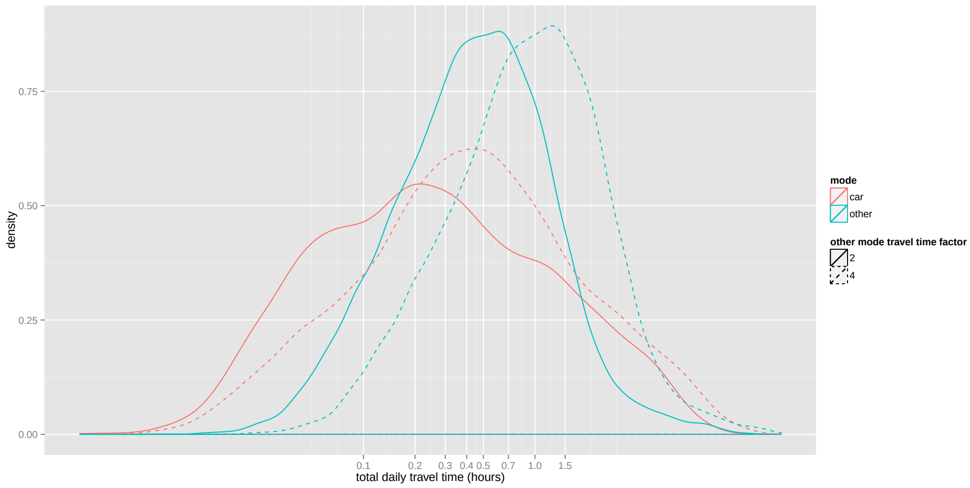

Figure 5: The daily time spent traveling, for the driving and non-driving sub-populations.

only because the OSM-based network model accounts for less than half of the presumed length of the actual road network. In this paper, we focus on the Abidjan urban area, but for a country-wide study, it may be worthwhile to improve on this aspect.

We therefore intend, in future version of MATSim, to explicitly account for the distance between the random location and the network by modelling trips in several stages: network access, network travel and network egress. The travel time for the network stage is determined by the traffic flow simulation, as in the present paper. Travel times for network access and network egress are determined by multiplying Euclidic distance by some factor which represents an unknown mode of travel through unknown terrain with an unknown detour. Conceptually, this would be done by extending the road network with virtual access links which orthogonally connect activity locations to the nearest road network link. This would model a mode of travel composed of several stages, like walking to the next road, being picked up by a motorist, and continuing by car.

Imputing behavioral meaning Most if not all similar studies go the path that they first attempt to create plausible daily activity plans from the mobile phone records and only then move on to an assignment of the traffic onto the traffic infrastructure. For the present investigation, we have deliberately chosen to immediately assign the mobile phone data to the road network, without an intermediate interpretational layer. There were two reasons to do so:

- We believe that plausible traffic patterns can already be obtained without that intermediate step, and that it saves a lot of time in order to get such simulations up and running. This could, for example, be important for situations with limited budget, or with situations with time pressure such as, say, disaster relief. Clearly, the claim that the results are realistic would need to be checked, for example by traffic counts data on the ground. Such data is, however, fairly straightforward to obtain, for example by, on a particular day, employing someone to stand next to the roadside and count vehicles.
- We believe that it is possible to impute the activity chains also *after* the traffic assignment. In fact, we believe that it may be better to do so, since the interpretational layer always means a loss of information that may still have been in the raw data, such as the deletion of seemingly implausible sightings, or certain variations in the temporal structure from one day to the next. With the approach discussed in this paper, one could always keep the original plan based on the mobile phone data, but generate multiple alternative plans for every synthetic traveler that would be consistent with the mobile phone data. For example, it could be assumed that some phone calls would actually be done en-route, or that some activities would carry on after the last phone call at a certain location. Out of these multiple interpretations of the mobile data, the system could converge to a set of interpretations that is most consistent with other data, such as, for example, time-dependent traffic flow data. This will be the subject of future work.
- An alternative approach might be to use a time use survey, which is available in

many countries, as additional data input. Time use surveys are similar to trip diaries in that they follow persons over days, but in contrast to trip diaries they typically do not register locations. Advantages of time use surveys over trip diaries include that they are considerably cheaper to obtain since the geocoding of the locations is expensive, at least with traditional approaches, and they have fewer privacy issues. In consequence, time use surveys are available in many places where trip diaries are not available. In addition, it may even be possible to use a time use survey from a neighboring city or country if the cultures are sufficiently similar. The imputation of activity chains from those time use surveys could be done in ways similar to those pointed out above: For each given sequence of cell phone sightings, one would select all possibly matching activity chains, or a randomly drawn subset. A data assimilation algorithm would then pick those activity chains most consistent with directly and anonymously measured data, such as traffic counts.

Statistical bias Trajectories sampled from mobile phone users alone are most probably biased. For example, not all members of the population have a phone, persons with a phone have vastly different calling patterns, and trips of persons who make fewer calls will be underreported. There is some indication that, for the purpose of mobility studies, such bias may not be as dramatic as it seems [16]. We believe that the approach discussed above, which is to do the data assimilation with the traffic model already up and running, might also help here. More specifically, one could imagine to give different statistical weights to every synthetic person. Based on other data, like for example time-dependent traffic flow data, one could re-weight the synthetic persons in order to bring the simulation closer to the data. This would presumably increase the weights of those types of persons that were under-weighted in the data, and decrease the weights of those types of persons that were over-weighted. Clearly, the approach will not work if certain types of persons are not included at all. We will investigate these issues in future work.

6 Conclusion

- The investigation demonstrates once more that it is possible to use publicly available OpenStreetMap data as the basis for traffic simulations. In the present situation, the coverage of the rural areas was still insufficient. However, one can either assume that this will improve over the years, or one could dispatch special investigations to insert the missing information if that turns out to be necessary for a specific study.

- The investigation also demonstrates that it is possible to obtain traffic patterns from mobile phone sightings without any layer of interpretation whatsoever. The traffic patterns look plausible; however, some verification would be necessary to decide if they are close enough to reality in order to use the model for policy analysis.

- A parametric study demonstrates that the model is sensitive to the performance of non-car modes.
- Sec. 5 discusses a method how the model could be systematically improved further if additional data is available. As explained, such possible data coould consist of, e.g., time use surveys or traffic counts.

References

[1] B. Kickhöfer, F. Hülsmann, R. Gerike, K. Nagel, Rising car user costs: comparing aggregated and geo-spatial impacts on travel demand and air pollutant emissions, in: T. Vanoutrive, A. Verhetsel (Eds.), Smart Transport Networks: Decision Making, Sustainability and Market structure, NECTAR Series, Edward Elgar, 2012, in press, also VSP WP 11-16, see www.vsp.tu-berlin.de/publications.

[2] T. Nicolai, K. Nagel, High resolution accessibility computations, VSP Working Paper 13-02, TU Berlin, Transport Systems Planning and Transport Telematics, see www.vsp.tu-berlin.de/publications (2013).

[3] D. Pearce, C. Nash, Social appraisal of projects: A text in cost-benefit analysis, Wiley & Sons, London, 1981.

[4] M. Zilske, A. Neumann, K. Nagel, OpenStreetMap for traffic simulation, in: M. Schmidt, G. Gartner (Eds.), Proceedings of the 1st European State of the Map – OpenStreetMap conference –, no. 11-10, Vienna, 2011, pp. 126–134, see sotm-eu.org/userfiles/proceedings_sotmEU2011.pdf.
URL `sotm-eu.org/userfiles/proceedings\_sotmEU2011.pdf`

[5] D. Zielstra, A. Zipf, Quantitative studies on the data quality of OpenStreetMap in Germany.
URL `http://www.giscience2010.org/pdfs/paper_187.pdf`

[6] N. Rieser-Schüssler, M. Balmer, K. Axhausen, Route choice sets for very high-resolution data, Transportmetrica iFirst (2012) 10021. doi:10.1080/18128602.2012.671383.

[7] J. Ma, H. Li, F. Yuan, T. Bauer, Deriving Operational Origin-Destination Matrices from Large Scale Mobile Phone Data.
URL `http://www.mygistics.net/docs/MobileOD.pdf`

[8] F. Calabrese, G. Di Lorenzo, L. Liu, C. Ratti. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area [online].

[9] V. D. Blondel, M. Esch, C. Chan, F. Clerot, P. Deville, Data for Development: the D4D Challenge on Mobile Phone Data, arXiv.org.
URL `http://arxiv.org/abs/1210.0137`

[10] C. Gawron, An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model, International Journal of Modern Physics C 9 (3) (1998) 393–407.

[11] A. Agarwal, M. Zilske, K. Rao, K. Nagel, Person-based dynamic traffic assignment for mixed traffic conditions, VSP working paper 12-11, TU Berlin, Transport Systems Planning and Transport Telematics, see www.vsp.tu-berlin.de/publications (2012).
URL https://svn.vsp.tu-berlin.de/repos/public-svn/publications/vspwp/2012/12-11/

[12] M. Rieser, D. Grether, K. Nagel, Adding mode choice to a multi-agent transport simulation, Transportation Research Record: Travel Demand Forecasting 2009 2132 (2009) 50–58. doi:10.3141/2132-06.

[13] K. Nagel, Towards simulation-based sketch planning: Some results concerning the Alaskan Way viaduct in Seattle WA, Tech. Rep. 08-22 (2008).
URL https://svn.vsp.tu-berlin.de/repos/public-svn/publications/vspwp/2008/08-22

[14] K. Nagel, Towards simulation-based sketch planning, part II: Some results concerning a freeway extension in Berlin, Tech. Rep. 11-18 (2011).

[15] World Bank. Non-Motorized Transport in African Cities [online].

[16] A. Wesolowski, N. Eagle, A. M. Noor, R. W. Snow, C. O. Buckee, The impact of biases in mobile phone ownership on estimates of human mobility, Journal of The Royal Society Interface 10 (81). arXiv:http://rsif.royalsocietypublishing.org/content/10/81/20120986.full.pdf+html, doi:10.1098/rsif.2012.0986.
URL http://rsif.royalsocietypublishing.org/content/10/81/20120986.abstract