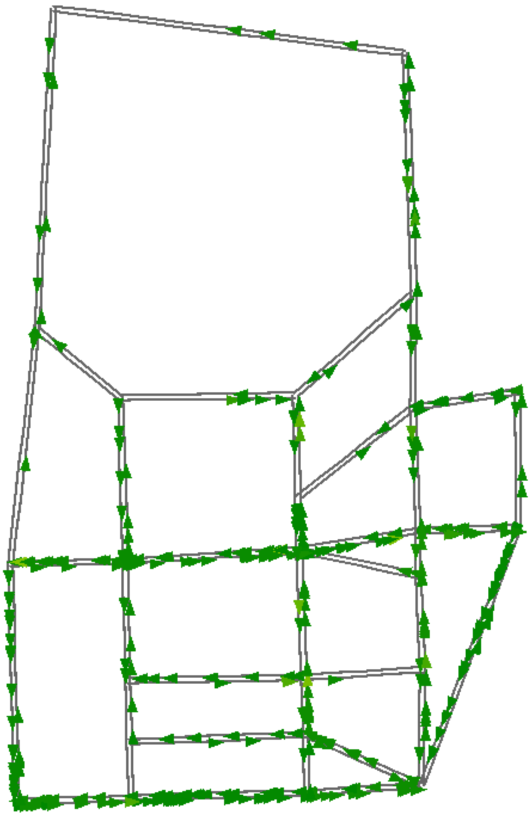


Enriched Sioux Falls Scenario with Dynamic And Disaggregate Demand

Artem Chakirov
Pieter J. Fourie

March 2014

(FCL) FUTURE CITIES LABORATORY 未来城市实验室

(SEC) SINGAPORE-ETH CENTRE 新加坡－ETH研究中心

IVT Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Enriched Sioux Falls Scenario with Dynamic And Disaggregate Demand

Artem Chakirov
Singapore – ETH Centre
Future Cities Laboratory
1 CREATE Way
#06-01 CREATE Tower
Singapore 138602

chakirov@ivt.baug.ethz.ch

Pieter J. Fourie
Singapore – ETH Centre
Future Cities Laboratory
1 CREATE Way
#06-01 CREATE Tower
Singapore 138602

fourie@ivt.baug.ethz.ch

March 2014

Abstract

This paper presents an enriched, agent-based small scale scenario with dynamic demand and an integrated public transport system based on the commonly used Sioux Falls road network. The scenario aims to provide a realistic, fully dynamic demand with heterogeneous socio-demographic users and a high degree of spatial resolution. Real world survey and land-use data is used to generate a diverse synthetic population and accurate activity locations. The socio-demographic characteristics include age and sex on individual and income on household levels. The assignment of home and work locations employs land-use and building information, census data from the City of Sioux Falls, South Dakota as well as commonly used static OD-matrices from LeBlanc *et al.*, 1975.

This enriched Sioux Falls scenario can serve as a convenient test-case for the study of different transportation policies as well as a test bed for the extension and development of agent-based simulation frameworks. It is important to note that the scenario does not aim to replicate the real City of Sioux Falls, SD, and remains a fictitious test-case scenario. In this work, we use Multi-agent Transport Simulation (MATSim) in order to evaluate the stochastic user-equilibrium of the compiled supply and demand.

The scenario data, as well as simulations input and output, can be obtained by following the instructions at *www.matsim.org/scenario/sioux-falls*.

Keywords

Sioux Falls, scenario, dynamic demand, MATSim, agent-based modelling, disaggregate demand

Preferred citation style

Chakirov, A. and P. Fourie (2014) Enriched Sioux Falls Scenario with Dynamic and Disaggregate Demand, Working paper, Future Cities Laboratory, Singapore - ETH Centre (SEC), Singapore.

1 Introduction

Since the early days of transport planning, simple and robust scenarios were required in order to test, demonstrate and compare methods and algorithms. A simplified road network based on the city of Sioux Falls, South Dakota, became very popular within the transport research community, as it was readily available. The scenario was first introduced by Morlok *et al.* (1973) as a traffic equilibrium network and consequently used and later adapted as a benchmark and test scenario in many publications, i.e. LeBlanc *et al.* (1975), Abdulaal and LeBlanc (1979a), Suwansirikul *et al.* (1987), Friesz *et al.* (1992), Meng *et al.* (2001), Bar-Gera *et al.* (2013) to name but a few. A more exhaustive list of Sioux Falls variants used as a benchmark continuous network design problem can be found on the webpage related to transportation network test problems, managed by Prof. Hillel Bar-Gera from Ben-Gurion University of the Negev (Bar-Gera, 2014).

With the emergence of dynamic activity-based modelling approaches and the development of agent-based simulation tools such as MATSim and TRANSims, new small-scale scenarios with dynamic demand and a high level of disaggregated information became necessary to test and demonstrate the new approaches. The challenges are that such scenarios have to stay computationally manageable and be compiled out of freely available data sources in order to ensure comparability and free public access. Furthermore, the socio-demographic characteristics and spatial distributions should resemble real world scenarios, but without necessarily exactly replicating a particular place.

In this paper, we extend the classic static Sioux Falls scenario and enrich it with dynamic demand and land-use information. The goal of this work is the development of a scenario featuring various layers of disaggregated information on the demand and supply side, which can serve users and developers of agent-based simulation tools as a convenient test-case to study different transportation policies as well as a test bed for new software extensions. Multi-Agent Transportation Simulation framework (MATSim) is used to run the developed scenario, but as all necessary information is stored in database tables, the scenario is easily adaptable to other activity- and agent based simulation tools. As a starting point, we use the classic Sioux Falls network, supplement it with an extensive bus network and make use of real micro-census and land-use information from the real city of Sioux Falls, to create diverse demand with varying socio-demographic characteristics.

On the demand side of the activity-based model only two simple activity chains are initially included, in order to keep the scenario accessible and to facilitate interpretation and understanding of the possible effects of policies under study. Modelling only *home – work – home* and *home – other – home* activity chains, destinations are assigned using a parameter-free radiation model (Simini *et al.* 2012).

In the following, we present the generation of supply-side, road and public transport networks as well as the demand side, synthetic population generation and assignment of home and work locations, in greater detail. The developed scenario is tested using the agent-based simulation framework MATSim (www.matsim.org). The scenario data, as well as simulations input and output, can be obtained by following instructions at *www.matsim.org/scenario/sioux-falls*.

In the remainder of the paper, we distinguish between *Sioux Falls* (the scenario) and the *City of Sioux Falls* (the actual city in South Dakota, United States).

2 Supply generation

A transportation test network should ensure sufficient complexity of travellers' choice dimensions while limiting computational effort. To this end, the Sioux Falls test network was introduced by Morlok *et al.* (1973) as a starting point. The structure of this network captures the major arterial roads of the City of Sioux Falls, but was never intended to replicate the real city or all characteristics of its transportation system, such as travel times and mode share. The network, comprised of 76 arcs, 24 nodes and 552 origin-destination pairs, was extended with a bus network consisting out of 5 bus lines by Abdulaal and LeBlanc (1979b), who used it to evaluate methods for combining modal split and equilibrium assignment models. Adjusting the capacities of the roads and placing the bus stops for the 5 bus lines in regular intervals, we substantially build upon the aforementioned authors' network design.

As we are interested in using the network within a dynamic traffic assignment and queue-based traffic models, flow capacities of single links are a major parameter of interest. In their study, LeBlanc et. al. (1975) used parameters for the travel time function dependent on flow for each of the 76 links, an approach fundamentally different to a queue-based model, which is used in MATSim. Furthermore, in order to apply real land-use information, a resemblance of the network shape and node locations to the existing road network is crucial. Therefore, using a map of the City of Sioux Falls, nodes of the toy Sioux Falls network are matched to the major intersections of the city, so that the 76 links roughly overlap with the major arterial roads of the city (Figure 1). For setting physical road parameters such as road length, number of lanes and associated maximum car flow rates, 2 types of road links are defined: highways and urban roads. The following characteristics have been chosen:

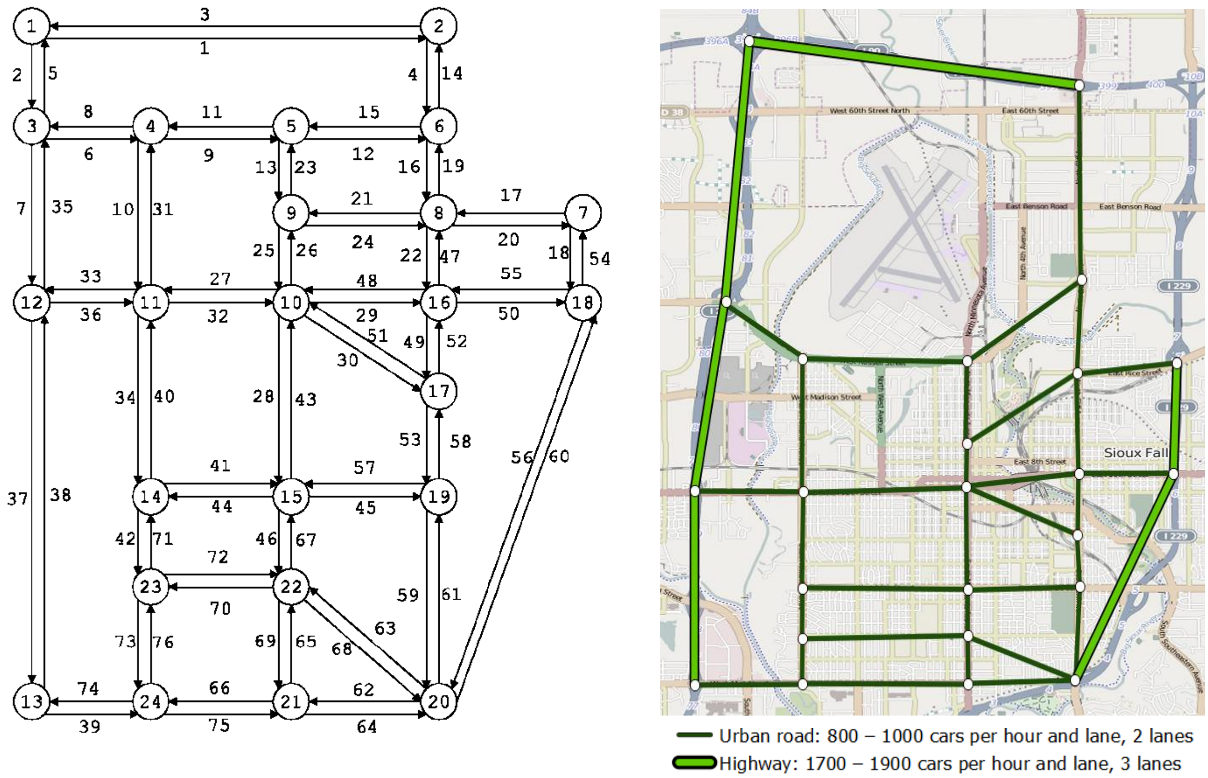
Urban road: 2 lanes with flow capacity 800 – 1000 cars per lane per hour;

Highway: 3 lanes with flow capacity 1700 – 1900 cars per lane per hour.

The length of all links is set equal to the Euclidian distance between intersections. As the urban links represent only major roads of the city, 2 lanes per direction is a good approximation of real conditions. For a few sections of highway, 3 lanes per direction might be a slightly exaggerated value. The corresponding flow capacities were chosen according to values indicated in literature such as the Highway Capacity Manual (TRB, 2000) or other related research publications (e.g. Ng and Small, 2011).

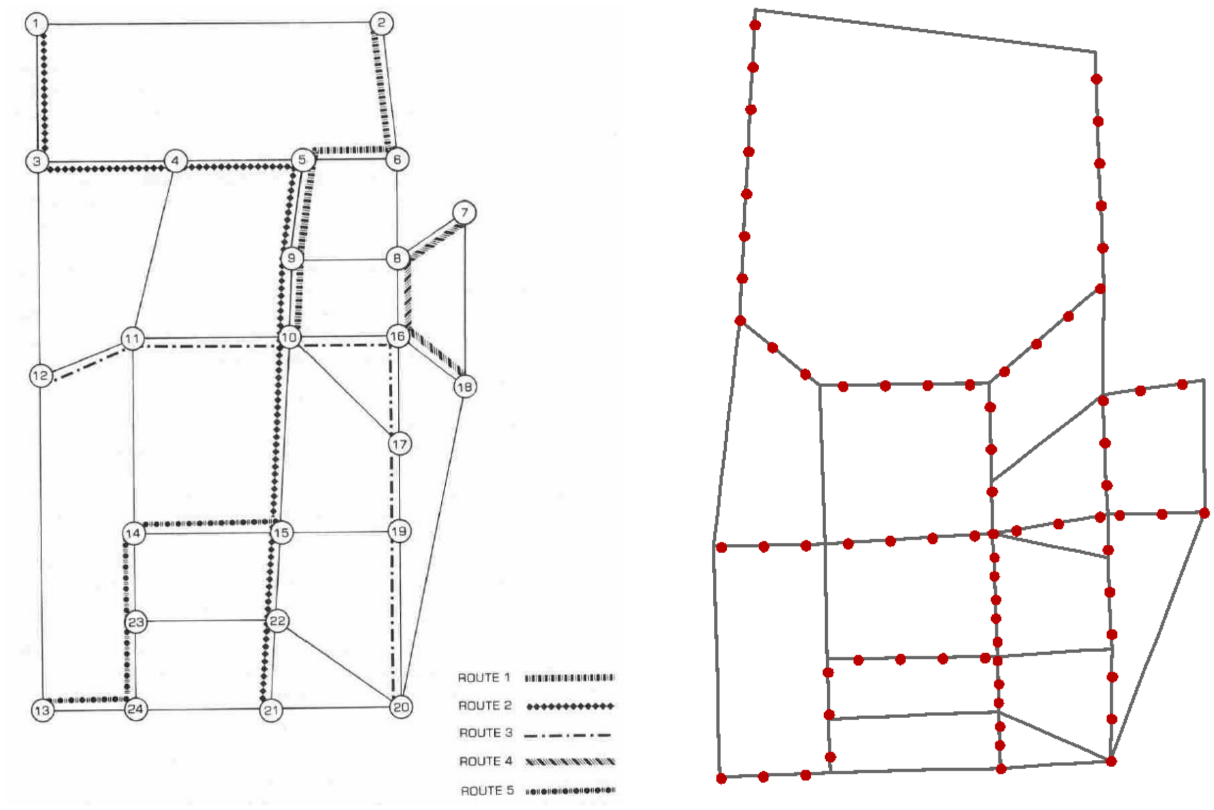
Setting the number of lanes to minimum 2 lanes per direction will allow more flexibility in using the scenario for different policy studies, e.g. investigation of effects from conversion of one car lane into a bus-only lane.

Figure 1 Comparison of the Sioux Falls Network by LeBlanc *et al.* (1975) with the adjusted geometry network (graph on the left from Meng and Yang, 2002; background on the right from www.openstreetmap.org).



The public transportation network, shown in Figure 2, was implemented as a bus network consisting out of 5 bus lanes, with the routes proposed by Abdullal and LeBlanc (1979). Bus stops were located 600m apart, and offset 5m from the road link and are displayed in Figure 2 on the right (in two cases on the Route 2 the distance between the consecutive stops was reduced in order to use the same stop facilities as by Route 1, where the two lines run parallel on same links). Furthermore, drawing from the feature set of MATSim, each bus stop was assigned to have a bus bay (basically a separate bus-only link, which can hold infinite number of buses), so that no road lanes are blocked during the boarding and alighting process at the bus stops. In order to make the bus service a competitive mode of transport with a significant number of users, a relatively short headway of 5 minutes was chosen for our tests.

Figure 2 Comparison of the Sioux Falls Network with public transport lines from Abdullal and LeBlanc (1979) with the new Sioux Falls Network with bus stops located 600m apart.



Due to the design of MATSim's queue simulation, agents are only handled at the beginning and end of each network link, and cannot enter or leave a link along its length. They enter and leave the network at links ending at the points (nodes) closest to their origin and destination coordinates. Therefore, origins and destinations located along very long links will lead to a loss of spatial detail, as all origins and destinations along the length of the link are effectively assigned the same coordinate. Consequently, to improve the level of spatial detail, all links were evenly split into smaller links with maximal length of 500 meters each. Following this operation, the network detail increased from 27 to 282 nodes and the number of links increased from 76 to 334.

The link:node ratio decreases from 76/27 = 2.81 to 334/282 = 1.18, because newly created links do not share nodes. So, for two links going in opposite directions between the same two nodes, each will be split into n newly created links with n-1 newly created nodes. Consequently, the effective network topology remains unchanged, since the number of intersections are the same as for the original network (27), as well as the number of paths between intersections (76).

3 Demand generation

The generation of a disaggregated, agent-based demand description, which closely resembles reality, not only in terms of trip origins and destination, but also with respect to associating travel patterns with socio-demographic characteristics is difficult.

LeBlanc *et al.* (1975), as well as some of the later papers using the Sioux Falls test network, employed a pre-defined OD-Matrix without further specification of its origin. Though a rather simple disaggregation and transformation of this demand to an agent-based model is possible (Balmer *et al.* , 2006), additional data is still required to assign socio-demographic characteristics to each person in a meaningful way.

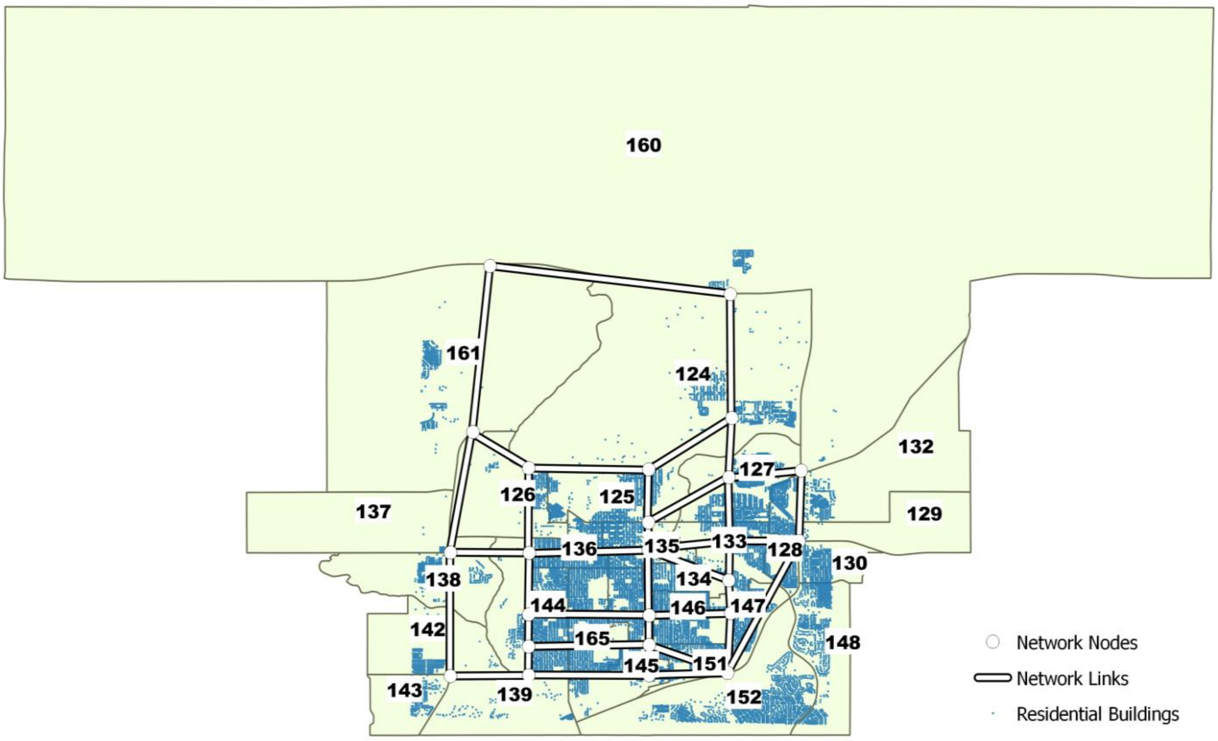
A realistic socio-demographically heterogeneous demand population is crucial for unlocking the potential of agent-based simulation; allowing the use of the scenario for further software development, as well as detailed policy studies with a focus on user heterogeneity. Therefore, in order to represent a household structure, demographic profile and income distribution as realistically as possible, a synthetic population using census data from census districts in and around City of Sioux Falls was generated. In the following sections, we address the process of synthetic population generations, home and work location assignment as well as choice of trip chains and temporal activity restrictions in detail.

3.1 Synthetic population generation with Entropy Maximization

A realistic population of individuals grouped into households is essential to an agent-based test scenario. Demographic attributes and household roles and responsibilities affect various aspects of the daily activity/travel schedule, and are strong predictors of choices such as mobility tool ownership, home and activity location and transportation mode, to name a few.

As a full census of the study area is not publicly available, we produced a synthetic population of households that matches the aggregate distribution of demographic attributes (age, sex and household income) recorded during the 2010 US Census for the 27 census tracts inside and adjoining the city centre of Sioux Falls (Figure 3). The population is composed of household and person records taken from the (anonymous) 5-year sample (2007-2011) of the American Community Survey, covering 5.0% of all households in Sioux Falls.

Figure 3 2010 US Census Tracts with geometry adjusted Sioux Falls road network and residential buildings located within the rectangular study area around it.



Based on a survey of the state of the art of population synthesis in (Müller and Axhausen 2011), we followed the *synthetic reconstruction* approach proposed by Beckman et al. (1996), which is composed of two steps: fitting and generation. In the fitting step, a weight is assigned to each household record in the sample, such that the sum of the weights for each demographic category will match the (control) total for that category from the census. In the generation step, the weights are used to determine the probability of a household being sampled in a process of sampling with replacement that continues until the total number of households for a census tract is reached. The process is repeated for each of the 27 census tracts, to produce a total of 107,486 persons in 43,936 households.

While the process of sampling with replacement is simple and well understood, various approaches for the fitting step have been proposed (Müller and Axhausen 2011). We are interested in finding a set of household weights that will retain the joint distribution of demographic attributes in the sample as closely as possible, while matching all control totals from the census. The joint structure would be fully retained in the case where all household weights were exactly the same (uniform weights), therefore the fitting procedure should aim to minimize the difference between assigned weights and uniform weights.

The entropy optimization approach proposed by Bar-Gera *et al.* (2009) minimizes the entropy of assigned weights relative to uniform weights, subject to the constraint that the sum of weights in each category should match the household and person control totals. We applied this method to the Sioux Falls scenario, using a Python implementation by Kirill Müller (Institute for Transport Planning and Systems (IVT), ETH Zurich). Weighted sampling with replacement was performed using the sample() method, in the base package of the R system of statistical computing (R Core Team, 2013).

The resulting synthetic population is composed of repetitions of 1,671 out of the 1,991 households in the sample; households with an undeclared/unknown household income were excluded from the process. Figure 4 shows the deviation from the ideal case of uniform weights for the synthetic population, i.e. where each household in the sample occurs 43,936/1,671 = 26.3 times in the synthetic population. The maximum cumulative deviation from uniform weights is 6061 households, occurring at the 651st observation in the ordered list of sample households.

The accuracy of the synthesis is shown in Figure 5, where the number of persons or households per control category is plotted for each census tract, against the corresponding control total from the census. The (normalized) root mean square deviation (RMSD / NRMSD) of the synthetic population for each control category and census tract combination appears in Table 1.

Figure 4 Cumulative number of households in the synthetic population, plotted against observation number sorted by decreasing number of occurrences of each sample household in the synthetic population.

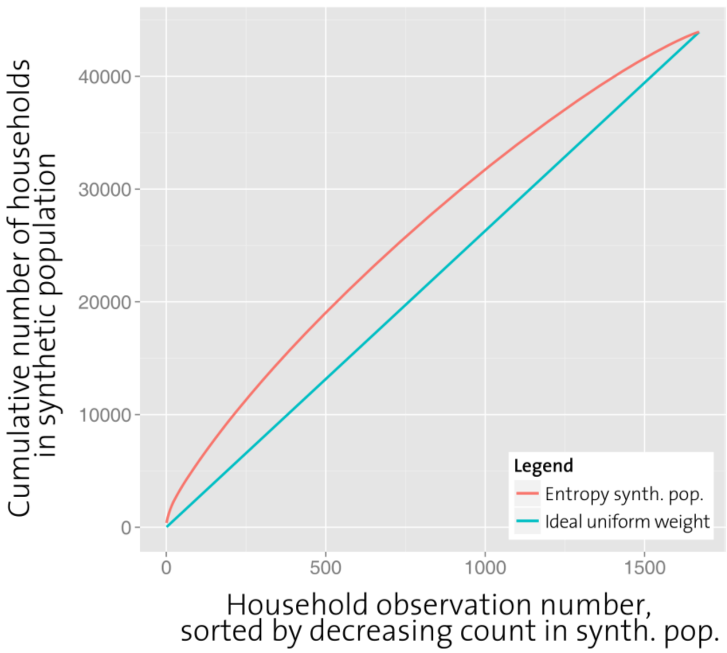


Figure 5 Number of persons/households per control category in synthetic population for each census tract versus the corresponding control total from US Census 2010.

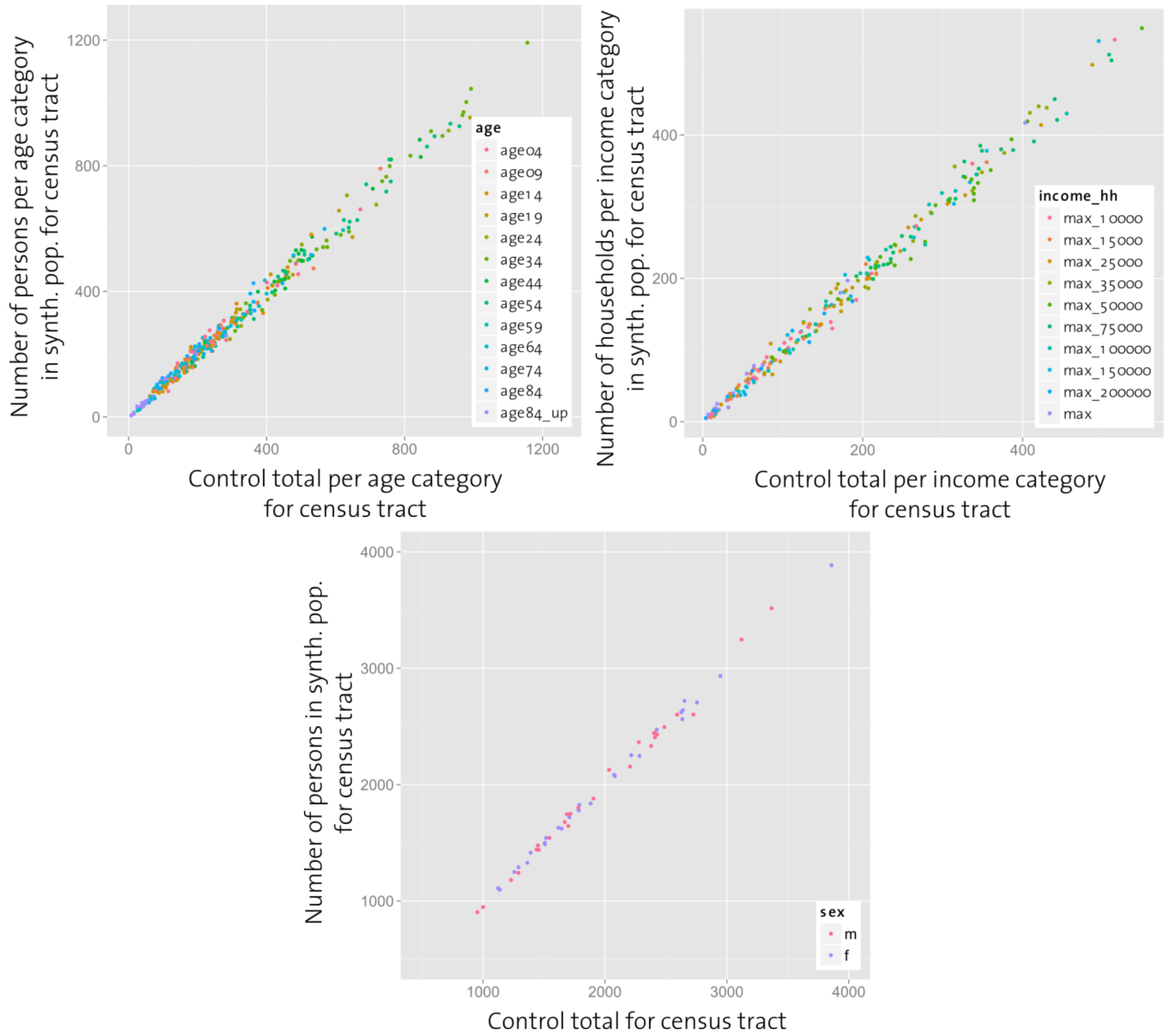


Table 1 (N)RMSD of synthetic population for each control category / census tract combination.

	Age	Sex	Household Income
RMSD	21.56	47.94	12.14
NRMSD	0.019	0.017	0.022

Table 2 shows the overview of number of households, number of persons as well as averages of socio demographic characteristics for each tract. As can be seen, the population shows noticeable spatial diversity, especially related to household income.

Table 2 Synthetic population characteristics per census tract.

Census tract	Number of households	Number of persons	Avg. household size	Avg. age	Avg. household income (USD)
124	1162	5137	4.42	35.12	61056
125	1866	4527	2.43	37.28	54576
126	1058	2509	2.37	31.39	40656
127	1552	3505	2.26	36.85	48794
128	1185	2995	2.53	37.71	58274
129	1749	5267	3.01	30.62	60273
130	1306	3043	2.33	37.77	59542
132	1283	3616	2.82	35.15	98774
133	1540	3215	2.09	38.22	46654
134	1263	2533	2.01	36.32	51239
135	2125	4570	2.15	34.20	38785
136	1673	3471	2.07	36.12	44133
137	1278	3217	2.52	36.90	54528
138	1916	4660	2.43	30.70	57329
139	2586	4521	1.75	38.78	38479
142	1444	3483	2.41	38.45	72536
143	2111	4718	2.23	41.05	57607
144	2190	4918	2.25	43.18	100613
145	2073	5309	2.56	34.48	49609
146	1305	2984	2.29	38.03	94239
147	992	2002	2.02	41.98	58211
148	2674	7133	2.67	36.72	99832
151	794	2058	2.59	38.73	93973
152	2123	5322	2.51	40.51	135434
160	1655	4680	2.83	35.65	83964
161	1184	3106	2.62	29.41	65089
165	1849	4987	2.70	37.00	71031

The distribution of income for all household of the synthetic population is shown in Figure 6. It exhibits an approximately log-normal shape as is to be expected for income. The distinctive peaks in income groups of 30k-40k USD and 60k-70k USD can be attributed to the categories used in the reporting procedure. Figures 7 and 8 visualize the population density and spatial income distribution using a 100x100m grid. As the population density stays fairly constant with only a few cells with densities significantly above average, the income distribution shows a clear spatial pattern, with high income households located in the south-eastern part of the city.

Figure 6 Household income distribution (in USD).

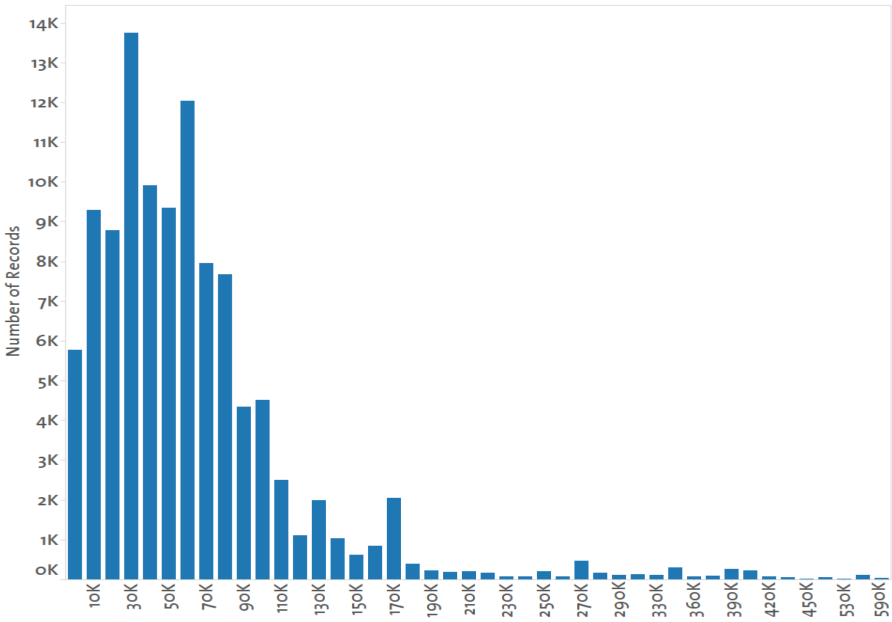


Figure 7 Home locations of synthetic population (total persons per 100m x 100m cell).

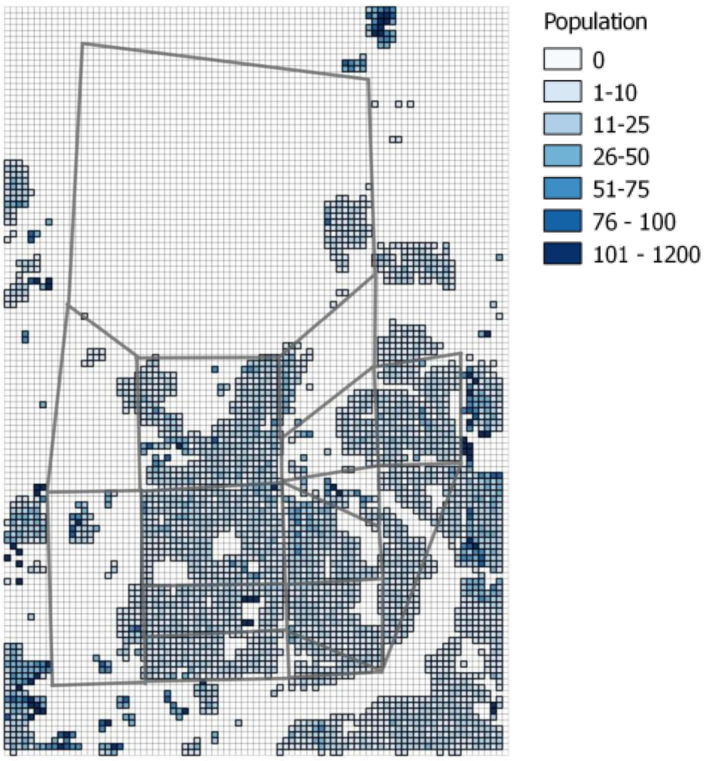
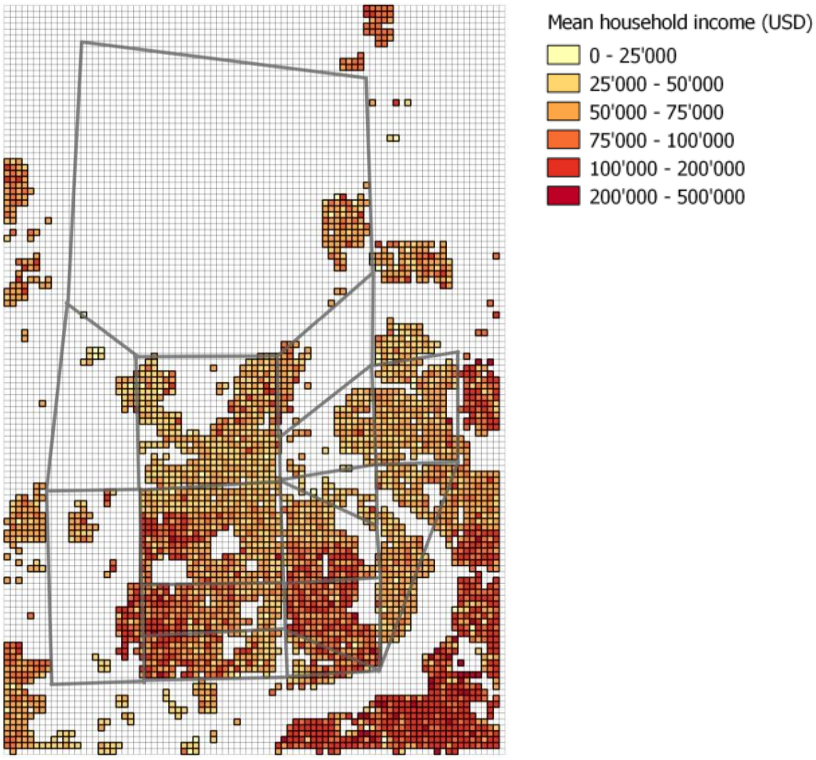


Figure 8 Mean household income (per 100m x 100m cell).



3.2 Home location assignment

The population synthesis, as presented in Section 3.1, provides the total number of households and persons at the census tract level. As the agent-based model allows for spatial resolution at the building level, we are required to assign a specific home location to each household.

For this assignment, we used the dataset of buildings, located in close proximity to the Sioux Falls road network. This dataset from May, 2013, contains information on all residential buildings, including number of units in each building. It was kindly provided by the City of Sioux Falls GIS division.

In the process, each household is randomly assigned to a residential unit within the tract it belongs to. In case the number of households exceeds the number of units in a specific tract, we allow for multiple occupants per unit after all units are full. This is particularly relevant for the peripheral tracts (e.g. 129, 132, 160), as these tracts extend beyond the rectangular study area around the road network and therefore contain buildings and units which are located outside this area. This results in a high density of home locations on peripheral parts of the study area, right outside the road network. However, as the majority of the road network in these areas represents highways, the increased demand from these regions can be considered as representative of traffic influx into the city centre from the outskirts and is in line with the goal of generating a high demand on the network during the peak hours.

3.3 Trip chains and work and secondary location assignment

In order to keep the scenario clear, only two trip chain types are selected for the initial population: *home – work – home* and *home – other – home*. The survey data used to generate the synthetic population contains the mode of transport to work for persons not working at home. Assuming that the employment status of a person correlates with their income, and given that we have controlled for income during population synthesis, we assign for all repetitions of a person occurring in the synthetic population the same employment status as was recorded for that person in the survey.

We thus identified a total of 56904 commuting workers. For secondary activity trip chains, the population without a work trip and older than 21 years was chosen, resulting in a total of 27206 persons performing secondary activities.

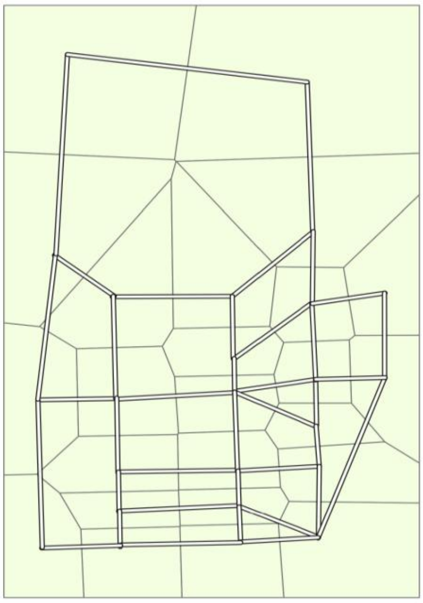
The OD-Matrix presented by LeBlanc *et al.* (1975) contains total of 360600 trips (the trip table printed in LeBlanc *et al.* (1975) actually contains 3.6 million trips, but was later revised by Abdullal and LeBlanc (1979b) as erroneously labelled with thousands of vehicles per day; the correct units are hundreds vehicles per day; see Appendix 1). This is approximately twice as many trips as the number of trips resulting from the two preceding steps in our demand generation, which sum up to 168220 trips in total. The significant difference in total numbers has several roots. Firstly, our model does not take into account any other trip types such as educational, commercial, freight and through traffic. Secondly, the assumption about the number secondary trips are rather arbitrary and e.g. do not take into account secondary trips after work activities. And thirdly the origin and assumptions behind the OD-Matrix from LeBlanc *et al.* (1975) are not stated in any publications reviewed.

As no information on the real number and distribution of work places within the relevant area was available to us, the OD-Matrix from LeBlanc *et al.* (1975) was taken as an indicator of the number of work locations in each zone. Subsequently, the assignment of work places to individual workers was performed using a parameter free radiation model presented by Simini *et al.* (2012).

Although the total number of trips differs, we can use the OD-Matrix of LeBlanc as an indicator of workplace attraction for each zone. Given the number of home-based trips originating in each zone, we try to resemble the total share of trips originating in each zone as indicated in LeBlanc’s OD-Matrix. Therefore the number of work places in each zone is assigned with a goal to match the share from total trips.

In LeBlanc's network, each zone is represented by a node in the network. In the case of our scenario, where the origins and destinations are represented by single buildings, we apply Voronoi decomposition to define zones around each node and accordingly assign a zone number to each building. The resulting zoning from Voronoi decomposition within the study area is shown in Figure 9.

Figure 9 24 zones created by Voronoi decomposition method.



As described above, in order to match the proportions of trips from each zone in the Sioux Falls OD-Matrix, we initially neglect the secondary activity trips and assume that the OD-matrix only contains *home* – *work* – *home* activity chains. As home-based trips represent exactly 50% of all trips, we apply the following simple calculation for share of workplaces in a zone z:

$$weight_z = \left(\frac{2 * \ trips_z^{out}}{totalTrips} - \frac{workers_z}{totalWorkers}\right) \tag{1}$$

$$if\ weight_z < 0\ then\ weight_z = 0 \tag{2}$$

$$workplaces_z = \frac{weight_z}{\sum_z weight_z} * totalWorkers \tag{3}$$

$trips_z^{out}$	Number of trips originating in zone z (OD-Matrix from LeBlanc *et al.* (1975), see Appendix 1)
$totalTrips$	Total number of trips (OD-Matrix from LeBlanc *et al.* (1975))
$workers_z$	Workers in zone z, as identified generated in the synthetic population

$totalWorkers$	Total number of workers within the synthetic population
$workplaces_z$	Number of workplaces in zone z

First, a weight for each zone is defined based on the difference between the proportion of all trips starting in the zone and number of workers, which inevitably corresponds to the number of home-work trips originating from this zone (home and work locations within the same zone are not permitted). As the total number of workplaces corresponds to the number of workers in the synthetic population, the number of workplaces in each zone is defined using the calculated weights.

For the assignment of secondary activity locations for the 27206 persons, no capacity restrictions are enforced. All buildings that have commercial use or are marked as community facilities, are considered as potential activity locations. The location assignment is again performed with a parameter free radiation model (Simini *et al.*, 2012).

Figure 10 shows the Euclidian distance distribution between home - work and home – secondary activity locations, as a result of assignment by the radiation model. Both distributions show a very similar pattern, with the home – secondary activity distribution having a slightly larger number of longer trips. The y-axes, which indicates number of trips, have different scales, as the number of work trips is approximately twice the number of trips to secondary activity locations.

Figure 10 Distribution of Euclidian distances between *home – work*) and *home – secondary* locations.

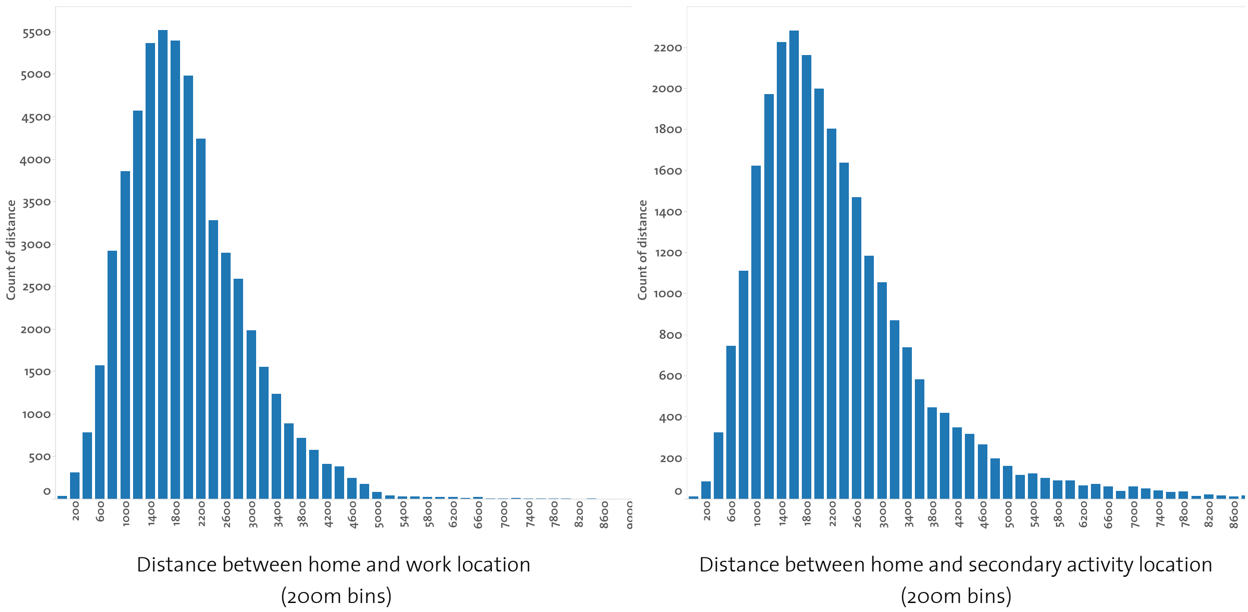


Figure 11 Number of work locations displayed per Voronoi node cell / zone and using a heatmap visualization.

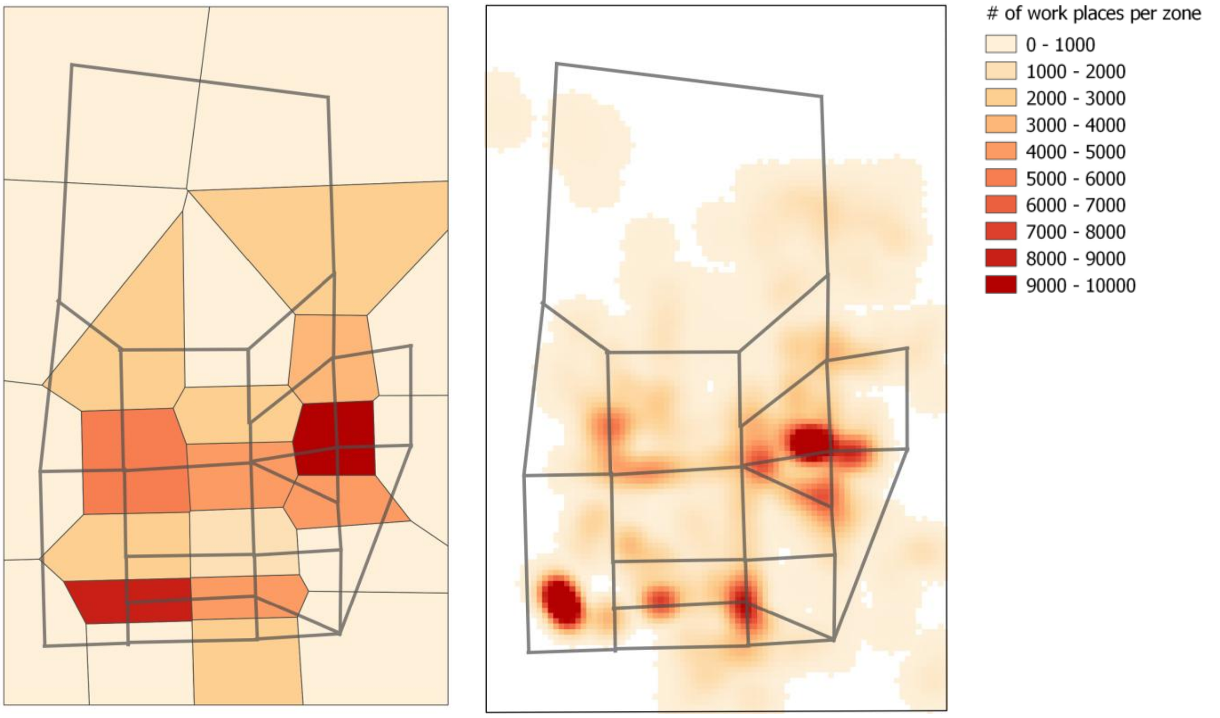
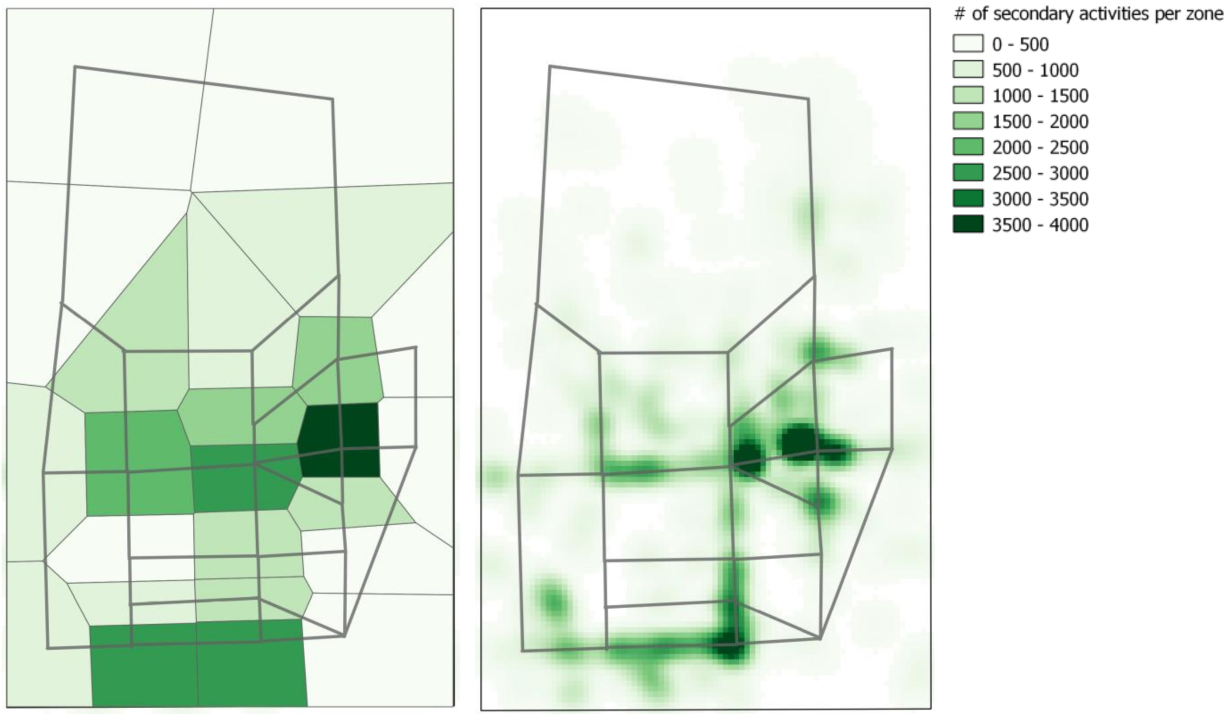


Figure 12 Number of secondary activity displayed per Voronoi node cell / zone and using a heatmap visualization.



The spatial distribution and resulting numbers of work places and secondary activity locations in each zone are displayed in Figures 11 and 12. Both work and secondary activity locations show a significant correlation in their spatial distribution, with high densities in the east-central and southern parts of the city.

3.4 Activity timings and trip start distribution

The evolutionary nature of the MATSim framework allows agents to adapt their departure times from iteration to iteration. The performance of modified plans is evaluated considering the constraints for each activity type. These constraints can include earliest and latest activity start times; typical and minimal activity duration; and opening times of facilities where the corresponding activities are performed. Starting with rather simple assumptions of normally distributed departure times for home – work - home trips during morning and evening peak-hours and uniform distribution of secondary activity trips through the day, we use MATSim's evolutionary approach to create a relaxed, stable state demand for these constraints. The initial departure time distribution, including its parameters, is shown in Figure 13 and Tables 3 and 4. The constraint parameters used in the simulation and the resulting departure time distribution after reaching an equilibrium will be discussed in detail in the following section.

Figure 13 Initial activity end (departure time) distribution (5 min bins).

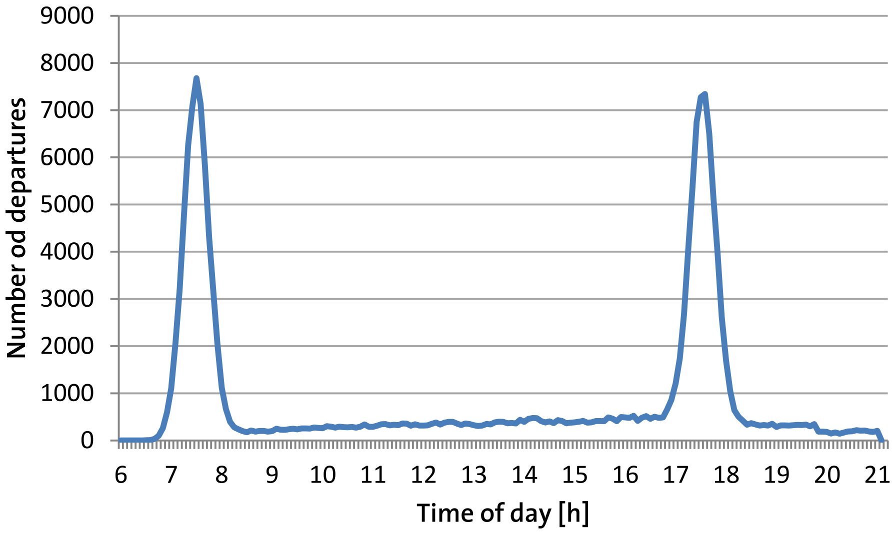


Table 3 Initial distribution of work activities.

Departure from activity	Distribution	Mean	Std. Deviation
Home – Work – Home			
From Home	Normal	7:30am	15min
From Work	Normal	17:30pm	15min

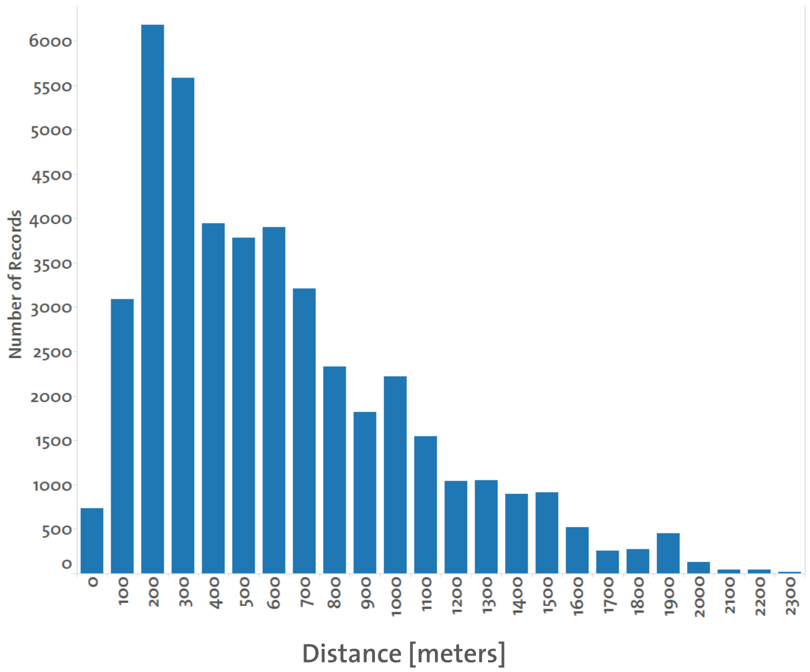
Table 4 Initial distribution of secondary activities.

Departure from activity	Distribution	From	Till
Home – Secondary – Home			
From Home	Uniform	7:45am	7:45pm
From Secondary	Uniform	1h 15min after departure from home	

3.5 Car ownership

In order to exploit the full potential of a disaggregated demand and to add another degree of realism to the scenario, we apply a car ownership model on the household level to the synthetic population. From the large number of available car ownership models (see Jong *et al.* (2004) for comprehensive overview), we have chosen an ordered probit model estimated by Giuliano and Dargay (2006) based on the US Nationwide Personal Transportation Survey (NPTS) 1995. All parameters of this model are available within the synthetic population and land-use information of the Sioux Falls scenario. Next to socio-demographic characteristics of a household (number of adults, children, pensioners, household income), the model uses attributes of residential location (population density, public transport access and dwelling type). Including public transport accessibility as a variable allows to account for characteristics of the scenario with its added area-wide bus network (Figure 14).

Figure 14 Public transport accessibility: distribution of distances between households residential location and closest bus stop (Euclidian distance).



Giuliano and Dargay (2006) define a latent ordinal preference index C^* which is used to determine the number of cars owned based on estimated constant threshold. It is defined as

$$C^* = \beta X + \varepsilon \tag{4}$$

with X being a vector of explanatory variables and β the vector of parameters related to these variables. The number of cars owned by a household follows from the rule:

$$\begin{aligned} C &= 0 \quad && \text{if} \quad && C^* \leq 0 \\ &= 1 && \text{if} && 0 < C^* \leq \mu \\ &= 2 && \text{if} && \mu < C^* \end{aligned} \tag{5}$$

with zero, one or two vehicles per household. The estimated parameters for Unites States by Giuliano and Dargay are used to model car ownership in Sioux Falls scenario are shown in Table 5.

Table 5 Estimated parameters by Giuliano and Dargay (2006).

Variable	Parameter value
Intercept	2.69
1 adult in household	− 1.21
3 or more adults in household	0.39
Children in household	0.11
Pensioner in household	− 0.37
Income < $20K	− 0.51
Income $40K - $55K	0.44
Income > $55K	0.76
Density < 1K pers/sqm*	0.17
Density 1K – 4K pers/sqm*	− 0.09
Density > 10K pers/sqm*	− 1.18
Transit within 0.5 miles*	− 0.24
Town/row house	− 0.27
Apartment	− 0.57
Threshold parameter	1.79

*1 mile = 1.61 km; 1sqm = 2.59 sqkm;

The reference category is a household with 2 adults, no children, income of $20K - $40K, living in a single family detached house and area density of 4000 – 10,000 persons per square mile. The model is applied using following definitions: adults age $\geq$ 18, children's age $\leq$ 18, pensioners age $\geq$ 65; population density in the whole study areas < 1K pers/sqm; residential buildings with 1 unit are single family detached houses, buildings with 2-6 units are equivalent to town/row houses and buildings with more than 6 units are apartment complexes.

Applying this model to the synthetic population of Sioux Falls scenario leads to total car population of 69517 cars, with 6.5% of households owning no car, 28.8% owning one car and 64.7% owning two cars. These ownership rates are also in line with overall US household car ownership rates as presented by Giuliano and Dargay (no car: 8%; 1 car: 30.2%; 2 cars: 61.9%).

The intra-household car assignment prioritizes employed members of the household over the one with secondary activities. First, the available cars are assigned to the agents having *home – work – home* trip chains. In case of a household consisting out of more employed members as there are cars available, the selection members using the cars is performed randomly. Afterwards, the car availability for agents performing secondary activities depends on number of cars still left unused within the household. The resulting number and share of agents having a car available for their trips is shown in Table 6.

Table 6 Car availability on person level.

Persons trip chain	Car available	No car available
home – work – home	48116 (84.6%)	8788 (15.4%)
home – secondary – home	17642 (64.8%)	9564 (35.2%)

This translates to a maximum of 65758 cars on the road as 3759 cars were assigned to households, which does not have members performing work or secondary activity, as identified in Section 3.3. This leads to the maximal possible car mode share of 78.2%, in case every available car would be chosen as a mode of transportation.

4 Simulation results based on dynamic traffic assignment

The newly augmented Sioux Falls scenario described above, is implemented within the Multi-Agent Transport Simulation (MATSim) framework, which integrates travel demand based on activity schedules with simulation-based dynamic traffic assignment. In the following sections we give a brief overview of the MATSim framework and provide references to in-depth descriptions of MATSim features. We describe the behavioural models and predefined constraints that were chosen for the simulation. The behavioural parameters are crucial for the outcome of the simulation and the final equilibrium state. Finally the outcomes of the simulation with respect to mode shares, travel times and travel distances are presented and discussed.

4.1 MATSim framework

Multi-Agent Transport Simulation (MATSim) framework uses an iterative approach for agent-based dynamic traffic assignments. Integrated simulation of private and public transport based on the queuing model allows time-dependent calculation of travel times accounting for spill-over effects and direct interaction of private and public transport. Based on a co-evolutionary algorithm, agents alter their behaviour from iteration to iteration, trying to find optimal routes, modes and departure times and therefore maximize the total utility of their daily activity schedule. Following each iteration of the queue-based network assignment, the choices of each agent are evaluated and scored, allowing agents to select more successful options for execution of their schedule in the next iteration. The selection of travel alternatives from the choice set of each agent is performed based on a random utility model which, after a number of iterations, leads to a convergence of individual and total utilities and therefore to an agent-based Stochastic User Equilibrium (SUE) (Nagel and Flötteröd, 2009).

As MATSim features a modular architecture, different functions of the simulation package as well as various behavioural features can be enabled or disabled. For the simulation presented, we use the integrated public transport simulation, where buses share the road space with cars and are affected by congestion. The following individual choice dimensions were open for agents to explore:

1. Route choice: In case an agent is selected for rerouting, new routes for all trips in its daily schedule are determined based on time dependent link travel times from previous iterations and without changing modes. In the case of private transport, the classical Dijsktra Algorithm is used. Details on the multi-node Dijkstra routing algorithm of public transport can be found in (Rieser, 2010).

2. Mode choice: In case an agent is selected for mode choice, it can change the mode of its journeys. As the mode choice has to be consistent (taking bus in the morning to work and car back home is improbable, due to non-availability of the vehicle at the work place), mode choice is altered at a sub-tour level (round-trip). For our scenario this means that each agent changes the mode of all his trips, as all agents undertake only one tour per day (home – work – home or home – secondary – home). Available modes are car, public transport (bus) and walk.

3. Departure time choice: This choice dimension enables agents to alter their departure times . Selected agents modify their departure times and activity durations of a daily plan randomly within a pre-defined time window. For the simulation results presented in the following chapter, we used a time window of +- 60min.

More details on the architecture and functionality of modules handling these choice dimensions within the MATSim framework can be found in Balmer *et al.* (2006), Balmer *et al.* (2009) and "MATSim-T Multi Agent Transport Simulation" (2013).

4.2 From database entries to agents with daily activity plans

The result of the synthetic population generation process, described in section 3.1, is a population table with a unique entry for each person, including information on the household it belongs to and individual socio-demographic characteristics. In sections 3.2 – 3.4 we presented methods that were used to assign home, work and secondary activity locations for each agent, according to their employment status and age.

In order to translate these tables into data suitable for simulation input, a daily activity plan for each agent has to be created and written out in XML-format (for detailed specifications of MATSim plan files see www.matsim.org). The plan file contains the activity locations, planned departure and activity start times for each agent and modes and routes for each trip. In case a trip has no routes assigned, as in the case of the initial demand produced for the simulation, MATSim routes each trip using one of the common routing algorithms (in our case we have chosen Dijkstra’s algorithm (Dijkstra, 1959). Once all trips have been routed, the dynamic demand is loaded on the transportation network and a time frame of 24-30h is simulated using MATSim’s queue-based traffic simulation and time step of 1 second.

4.3 Activity constraints and behavioural parameters

One of the core characteristics of agent-based models is the decision-making heuristics of individual agents. Even simple rules on individual level lead to emergence of complex collective behaviour and convoluted system dynamics (see Helbing (2011) for a general overview).

As mentioned above, the MATSim framework uses an agent-based SUE formulation, which relies on co-evolutionary iterative learning and a utility-based approach to model behaviour on individual level. Following a *trial and error* approach, every agent optimizes its daily activity schedule from iteration to iteration and accumulates a certain number of possible daily schedules, also called plans, in its memory. This is analogous to choice set generation and is followed by a probabilistic approach for selection of a plan from the choice set of plans using a multinomial logit model. During the choice set generation phase, agents change routes, modes or departure times from iteration to iteration, as specified in the configuration, and evaluate the outcome of the executed daily activity chains after the network loading. The ultimate goal of each individual agent is the maximization of its total utility gained through the day, which eventually leads to the convergence of the system to a state that can be considered as the stochastic user equilibrium. Detailed discussion of this approach is presented by (Nagel and Flötteröd, 2009).

Choice dimension and choice set generation

The process of choice set generation based on utility maximization for each agent is highly dependent on the chosen utility functions. In MATSim the total utility of each agent is a sum of 2 major utility functions: the utility gained by performing activities and (dis-)utility of traveling:

$$U_{total} = \sum_{i=1}^{m} U_{act,i} + \sum_{i=1}^{m} U_{trip,i}$$

where U_{total} is the utility of a given, daily plan; m is the total number activities and trips. The number of activities and plans is assumed to be the same; for the sake of consistency, the first- and last activity should be of the same type and are counted as one. $U_{act,i}$ is the (positive) utility from performing activity i; $U_{trip,i}$ is the (negative) utility from undertaking a trip i.

Utility from performing activities

The utility earned from activity performance is calculated using a logarithmic approach, presented by (Charypar and Nagel, 2005):

$$U_{act,i}\left(t_{act,i}\right) = \beta_{act} \cdot t_{*,i} \cdot \ln\left(\frac{t_{act,i}}{t_{0,i}}\right)$$

where $U_{act,i}$ is the total utility gained from performing activity i for the time $t_{act,i}$; β_{act} is the marginal utility of an activity at its typical duration; $t_{*,i}$ is the typical duration of the activity i; $t_{0,i}$ is a scaling parameter, which becomes relevant in case the activities can be dropped from the daily schedule.

Disutility of traveling

The disutility of traveling is calculated for each trip and based on the mode of transportation:

Car	$U_{car,i}(t_{tr,i}, c_i) = \beta_{0,car} + \beta_{tr,car} \cdot t_{tr,i} + \beta_{c,car} \cdot c_{i,car}$
Public transport	$U_{pt,i}(t_{tr,i}, c_i) = \beta_{0,pt} + \beta_{tr,pt} \cdot t_{tr,i} + \beta_{c,pt} \cdot c_{i,pt}$
Walk	$U_{walk,i}(t_{tr,i}) = \beta_{tr,walk} \cdot t_{tr,i}$

where U is the total disutility from trip i, which depends on the trip duration $t_{tr,i}$ and the monetary c_i cost caused by the trip. The monetary cost can contain fixed cost per trip, as in the case of a flat fare for public transport usage; or distance dependent cost, such as petrol and maintenance cost in case of car. β_0 is an alternative specific constant, which allows to capture effects neglected by the model, e.g. additional cost resulting from searching and paying for parking.

Activity constraints

The activity constraints define the time intervals and typical utility duration parameter within the activity utility function, changing the response of the logarithmic function along the utility axis. Consequently, the relation between the time window for an activity and its typical duration affect how peaked or spread out the demand in the peak-hours will be and therefore the degree of congestion of the network.

For the evaluation of the Sioux Falls scenario, we use two different constraint sets to demonstrate the consequences of these constrains on demand peak spreading and the stability of the network under the resulting distinct peak-loads. The defined activity constrains are listed in Table 7.

Table 7 Activity constraints.

Activity	Typical duration	Opening Time	Closing Time
Case 1: Rigid work constraints			
Home	13 h	always open	always open
Work	9 h	8 a.m.	**18 p.m.**
Secondary	1 h	8 a.m.	8 p.m.
Case 2: Soft work constrains			
Home	13 h	always open	always open
Work	9 h	8 a.m.	**20 p.m.**
Secondary	1 h	8 a.m.	8 p.m.

In case an agent arrives at the activity location before the opening time or departs after closing time, the "idle" time is indirectly penalized by opportunity cost of time – utility which could be gained by performing an activity longer. MATSim also offers an option of adding additional disutility for early and late arrivals, which was not used in this case.

Behavioural parameters

The behavioural parameters used in utility functions of activities and trips are based on the estimated demand model for Sydney by Tirachini, Hensher, and Rose (2012). The time-related parameters derived in their model have to be adjusted for application in the activity-based context. In order to provide a value for marginal utility of performing an activity, the travel mode with smallest disutility is set as a baseline, under the assumption that travelling with this mode is as good or as bad as idling and doing nothing. Therefore, the corresponding parameters are split into opportunity costs of time and a mode-specific disutility of travelling, as was done previously by Kickhöfer *et al.* (2011), Kickhöfer *et al.* (2013), Kaddoura *et al.* (2013), Kickhöfer and Nagel (2013), who employed the same parameters within their MATSim based studies.

Previous studies (Tirachini *et al.*, 2012, Kickhöfer and Nagel, 2013, Kaddoura *et al*. 2013) also used a pre-defined modal split for the calibration of the mode specific constants. In our case such a target does not exist. Matching the modal split of the real City of Sioux Falls would lead to a public transport mode share below 1%, which is not the objective of this scenario. Therefore, in case of car, we define the mode specific constant based on assumptions for access/egress times of parking locations of total 10 min and associated parking costs of $6[1] per trip. The mode-specific constants of public transport and walk are assumed to be zero. Furthermore, MATSim provides option for additionally penalizing the waiting time at transit

[1] As initially provided by Tirachini *et al.* (2012), all monetary values ($) in this paper are stated in Australian Dollar (AUS). 1 AUD = 0.88 USD (21.01.2014).

stops. Waiting time thereby is defined as the time between agents arrival at the transit stop and boarding a public transport vehicle. As Tirachini *et al.* (2012) do not explicitly provide a parameter for waiting time at the bus stop, we assume that it is perceived in the same way as the travel time on the bus, which is a rather optimistic assumption. An overview of the parameters employed and the assumed monetary costs is given in Tables 8 and 9 respectively.

Table 8 Behavioural parameters.

Estimated parameter by (Tirachini, Hensher, and Rose 2012)		Adjustment for activity-based framework	**Sioux Falls Scenario**
$\hat{\beta}_{act}$	n.a.	$\beta_{act} = -\hat{\beta}_{tr,car}$	+0.96 [utils/h]
$\hat{\beta}_{tr,car}$	-0.96 [utils/h]	$\beta_{tr,car} = \hat{\beta}_{tr,car} - \beta_{act}$	0 [utils/h]
$\hat{\beta}_{tr,pt}$	-1.14 [utils/h]	$\beta_{tr,pt} = \hat{\beta}_{tr,pt} - \beta_{act}$	-0.18 [utils/h]
$\hat{\beta}_{tr,walk}$	-2.1 [utils/h]	$\beta_{tr,walk} = \hat{\beta}_{tr,walk} - \beta_{act}$	-1.14 [utils/h]
$\hat{\beta}_{wait,pt}$	-	$\beta_{wait,pt}$	-0.18 [utils/h]
$\hat{\beta}_{c}$	-0.062 [utils/$]	$\beta_{c} = \hat{\beta}_{c}$	-0.062 [utils/$]
$\hat{\beta}_{0,car}$	0 [utils]	$\beta_{0,car} = \beta_{tr,walk} \cdot 10\,\text{min} + \beta_{c} \cdot 6\$$	-0.562 [utils]
$\hat{\beta}_{0,pt}$	-2.080 [utils]	$\beta_{0,pt}$	0 [utils]
$\hat{\beta}_{0,walk}$	-0.092 [utils]	$\beta_{0,walk}$	0 [utils]

Table 9 Monetary cost overview (EUR = Euro; $ = Australian Dollar (AUD)).

	Tirachini *et al.* (2012)	Kadoura *et al.* (2013)	Kickhöfer and Nagel (2013)	**Sioux Falls Scenario**
PT Fare	2.1 $	0 - 8.9 $	not specified	2 $
Car cost per km	0.14 $/km	0.4 $/km	0.30 EUR/km	0.4 $/km
Parking cost	not specified	not specified	not specified	6 $/trip

4.4 Simulation results

Simulation procedure and simulation parameters

Starting with the initial demand, 2900 iterations are performed, where agents modify their routes, modes and departure times. The re-planning rate, which determines the number of agents mutating their daily plan after each iteration were set to 1% for each choice dimension. After a relaxed system state with (almost) constant average travel times and constant average realized utility from iteration to iteration is reached, all re-planning modules are switched off and an additional 100 iterations without re-planning are performed, where agents select one of the plans saved in the memory using the multinomial logit approach. The number of plans which each agent can keep in its memory is limited to 6. After a total of 3000 iterations, the outcome of the last iteration is considered to be in SUE and is used for the analysis.

As a consequence of the stochastic nature of the simulation resulting from application of the random utility model, the outcome of the simulation is dependent on the initial random seed. This issue has been addressed in detail by Horni, Charypar, and Axhausen (2011), where the variability in link volumes showed little variability on a daily basis but substantial variability when the link volumes were compared hourly. Though a detailed evaluation of variability and distribution of outcome lies outside the scope here, we conduct 5 runs with different random seeds and address differences in the outcomes in the evaluation below. Analysis of travel time and travel distance distribution is performed only for the simulation with the default MATSim random seed.

MATSim specific parameters used for the simulation can be found in Appendix 2.

Stochastic User Equilibrium

It can be observed from Tables 10, 11 and 12 that the investigated indicators such as mode share, travel times and travel distances show only minimal variability depending on the random seed. This holds for both cases with different activity constraints as described in Section 4.3, indicating robust convergence of the system towards the Stochastic User Equilibrium. Comparing the both cases, the less pronounced demand peaks of case 2 lead to less variability throughout all indicators than the more peaked demand in case 1. Though no explicit investigation of the variability of results from iteration to iteration in the neighbourhood of the relaxed system state is performed in this work, the variability for different random seeds appears to lie within the same magnitude as the inter-iteration variability and is almost indistinguishable from it.

For both sets of activity constraints, car is the dominant mode of transportation with a mode share of ca. 64% (Table 10). According to the car availability rates set in Section 3.5, this implies that ca. 14% of agents, who have a car available, prefer not to use it. From Tables 10 and 11 as well as Figures 15 and 16 can be observed that the car mode produces the shortest travel times and serves as the dominant means of transportation, especially for longer trips. This relates to the average car speed of 42 km/h and average bus speed of 35 km/h for case 1 (car: 53 km/h, bus 37 km/h for case 2), which indicated a relatively uncongested network conditions. The walking speed is defined as a constant simulation parameter and is set to 3km/h. The length of the walking trip is defined as Euclidian distance times a factor of 1.3.

Mode share

Table 10 Mode shares in the relaxed system state (3000 iterations) for 5 different random seeds and 2 sets of activity time constrains (see Section 4.3).

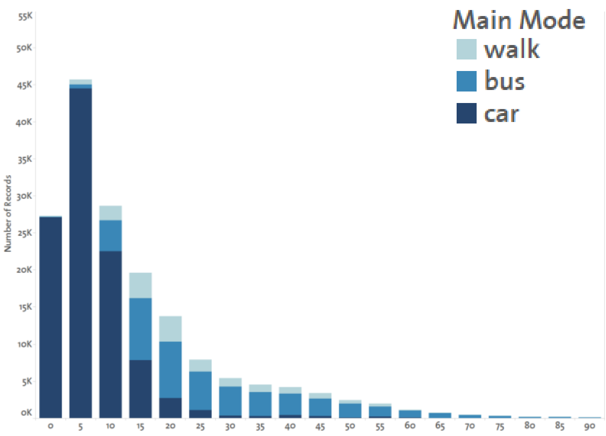
Mode	Seed 1 (default: 4711)	Seed 2 (956450029)	Seed 3 (737696784)	Seed 4 (10327469)	Seed 5 (483123812)	Mean	Std. deviation
Case 1							
Car	63.86%	63.95%	63.89 %	63.88 %	63.94 %	**63.90%**	**0.035 %**
Bus	26.45%	26.41 %	26.44%	26.47%	26.31%	**26.42%**	**0.056 %**
Walk	9.69%	9.64 %	9.68%	9.65%	9.75%	**9.68%**	**0.039 %**
Case 2							
Car	64.85%	64.84%	64.83%	64.87%	64.88%	**64.85%**	**0.019 %**
Bus	27.23%	27.24%	27.26%	27.22%	27.21%	**27.23%**	**0.017 %**
Walk	7.92%	7.92%	7.91%	7.92%	7.91%	**7.92%**	**0.005 %**

Travel times

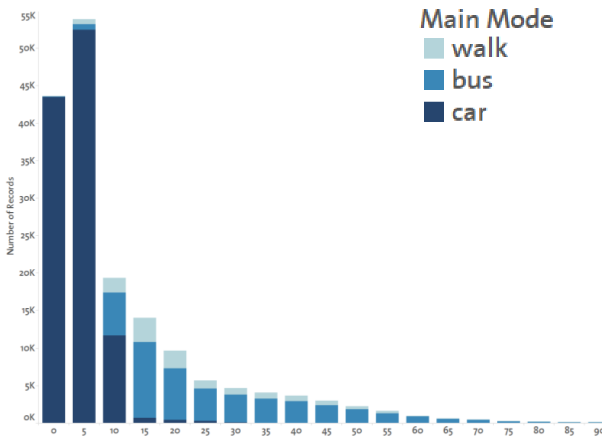
Table 11 Average travel times in the relaxed system state (3000 iterations) for 5 different random seeds and 2 sets of activity time constrains (see Section 4.3). 9′06″ = 9 minutes, 06 seconds.

Mode	Seed 1 (default: 4711)	Seed 2 (956450029)	Seed 3 (737696784)	Seed 4 (10327469)	Seed 5 (483123812)	Mean	Std. deviation
Case 1							
Car	9′06″	9′03″	8′58″	9′00″	9′02″	**9′02″**	**3.0″**
Bus	31′21″	31′39″	31′33″	31′30″	31′34″	**31′31″**	**6.7″**
Walk	26′34″	26′33″	26′31″	26′33″	26′41″	**26′34″**	**3.8″**
Total	16′41″	16′42″	16′38″	16′39″	16′41″	**16′40″**	**1.6″**
Case 2							
Car	6′11″	6′11″	6′12″	6′10″	6′14″	**6′11″**	**1.4″**
Bus	29′59″	29′56″	29′57″	29′57″	29′57″	**29′57″**	**1.1″**
Walk	25′39″	25′40″	25′38″	25′42″	25′44″	**25′41″**	**2.4″**
Total	14′12″	14′12″	14′13″	14′11″	14′14″	**11′13″**	**1.1″**

Figure 15 Histograms of travel times.



Case 1, Seed 1, 5min bins



Case 2, Seed 1, 5min bins

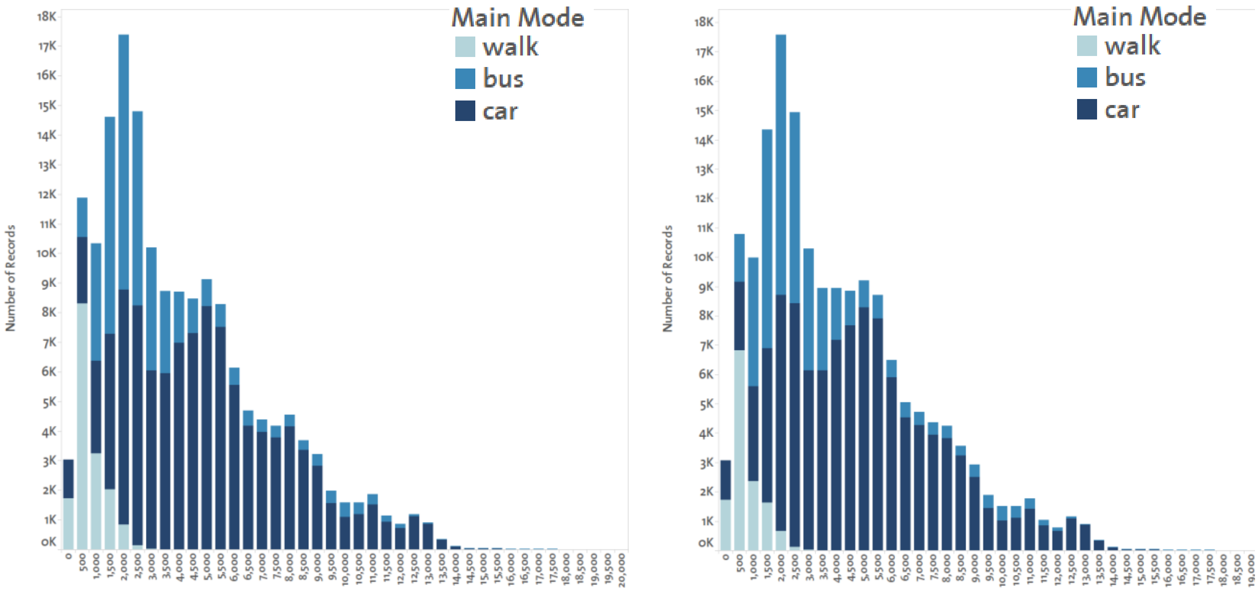
Travel Distances

Table 12 Average travel distances in the relaxed system state (3000 iterations) for 5 different random seeds and 2 sets of activity time constrains (see Section 4.3).

Mode	Seed 1 (default: 4711)	Seed 2 (956450029)	Seed 3 (737696784)	Seed 4 (10327469)	Seed 5 (483123812)	Mean	Std. deviation
Case 1							
Car	5309m	5310m	5311m	5309m	5311m	**5310m**	**1.0m**
Bus	3352m (2422m)	3355m (2422m)	3353m (2421m)	3351m (2420m)	3358m (2427m)	**3354m** **(2422m)**	**2.8m** **(2.7m)**
Walk	1021m	1020m	1019m	1021m	1026m	**1021m**	**2.7m**
Total	4376m	4380m	4378m	4377m	4379m	**4378m**	**1.6m**
Case 2							
Car	5245m	5245m	5245m	5245m	5245m	**5245m**	**0.0m**
Bus	3326m (2396m)	3326m (2396m)	3327m (2397m)	3327m (2397m)	3325m (2396m)	**3326m** **(2396m)**	**1.0m** **(0.6m)**
Walk	986m	987m	986m	988m	989m	**987m**	**1.3m**
Total	4386m	4385m	4385m	4386m	4386m	**4386m**	**0.6m**

*The values in brackets for the bus mode show the in-vehicle distance

Figure 16 Histograms of travel distances.



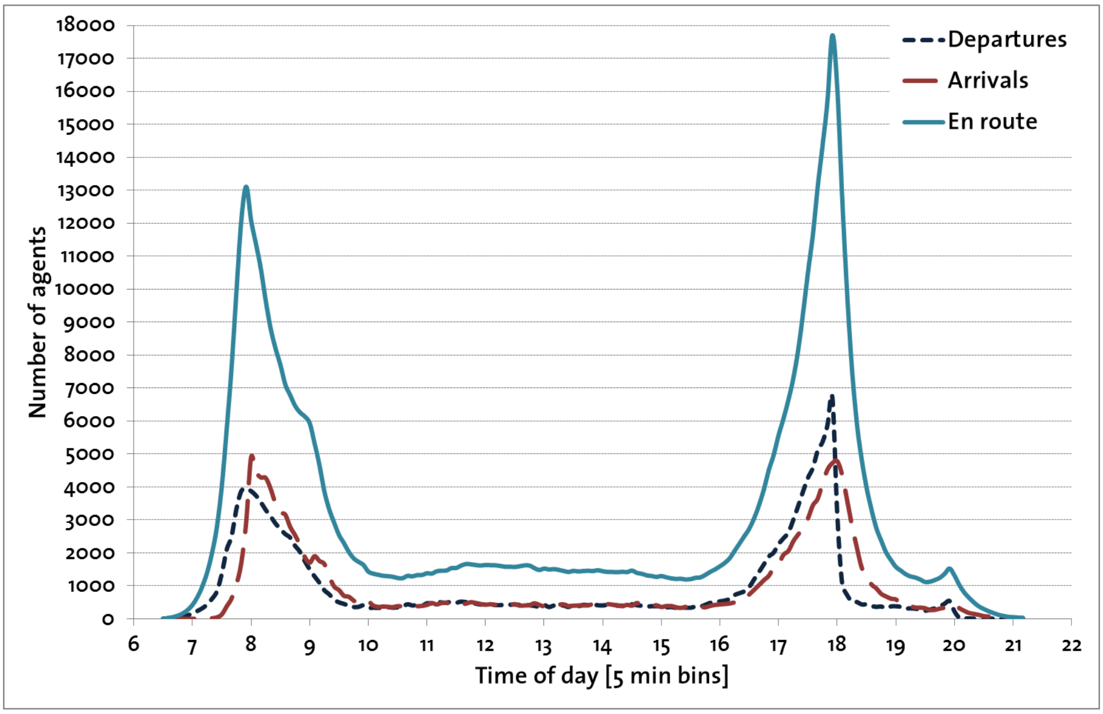
Case 1, Seed 1, 500m bins

Case 2, Seed 1, 500m bins

Trip time distribution

Figure 17 shows the distribution of departure and arrival times as well as number of agents en route in the relaxed system state, after 3000 iterations. The rather rigid activity constraints, as set for case 1, lead to distinct rush hour peaks. Especially in the evening, the number of departures and with it also agents on the road peaks at 6pm, the end time of the work activity. Such distinct peaks might appear uncommon in the American or European context, but can be observe in cities with a rigid working hours culture, e.g. Singapore (Chakirov and Erath, 2011).

Figure 17 Departure Time, Arrival Times and Number of Agents En Route distributions (Case 1, Seed 1, 5 min bins)



Macroscopic Fundamental Diagram

Geroliminis and Daganzo (2007, 2008) have demonstrated that some urban areas exhibit a "Macroscopic Fundamental Diagram", where the number of vehicles in the network is related to the space-mean flow of the network. Importantly, they showed that this MFD is a property of the network itself, and not very sensitive to changes in demand.

In order to be able to reproduce a MFD, the area has to be of a certain minimal size and mainly consist of similar road types. From the perspective of this study, the interesting question is if the Sioux Falls scenario provides a sufficient size in order to demonstrate MFD properties, as stated by Geroliminis and Daganzo (2007, 2008).

Figure 18 shows the density/flow MFD of the relaxed state of case 1 with the default random seed (Seed 1). Here only urban road links (see Figure 1) have been taken into account. As it can be seen, the diagram shows only a part of a typical MFD, as the oversaturated part, where the critical capacity of the network is reached and the flow decreases, does not appear. But when we look at the MFD of the system after first iteration, with average travel times of about 52 minutes and highly congested network in the rush hour, a full MFD can be observed (Figure 19). The large number of vehicles in the network produces a drop in space-mean flow that persists from 7:30 until noon, with significant hysteresis in the time following maximum vehicle density. The effect is less pronounced in the evening, when vehicles face increasing capacity as they move away from the city centre.

The disappearance of the full MFD with increasing iterations is interesting; it appears that through the individual time, mode and route adjustments of the individual agents, the system finds a close to maximum space mean flow that can be accommodated (approx. 1300 veh/lane/hr).

Figure 18 MFD of relaxed system (Case 1, Seed 1, 3000 it, 10 min bins).

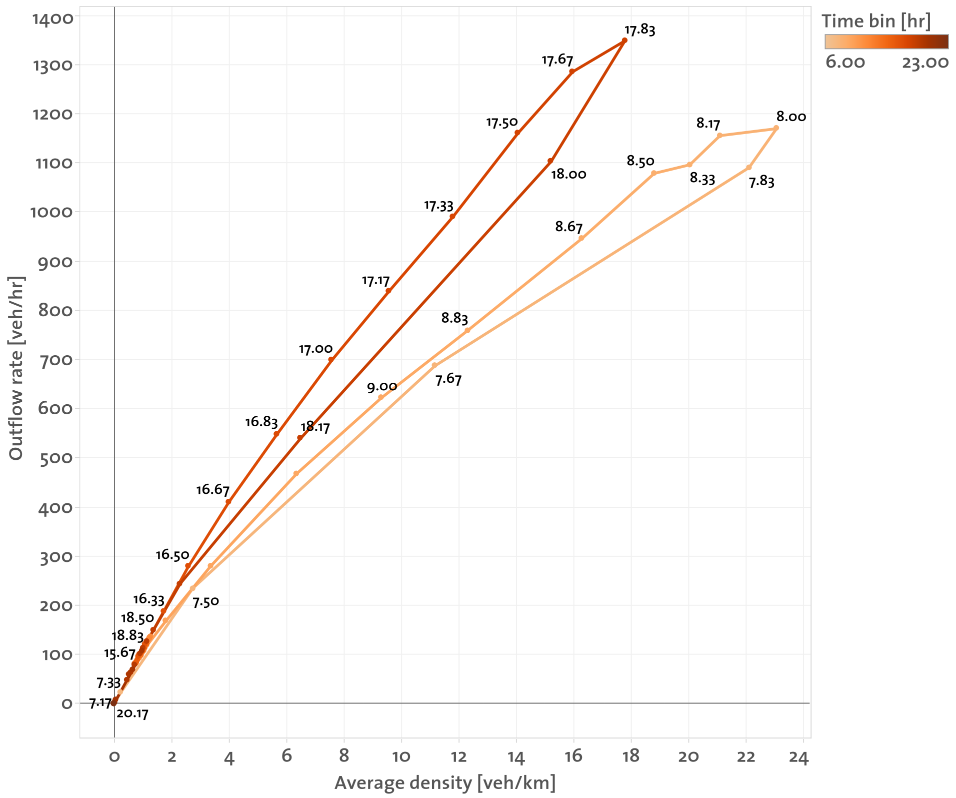
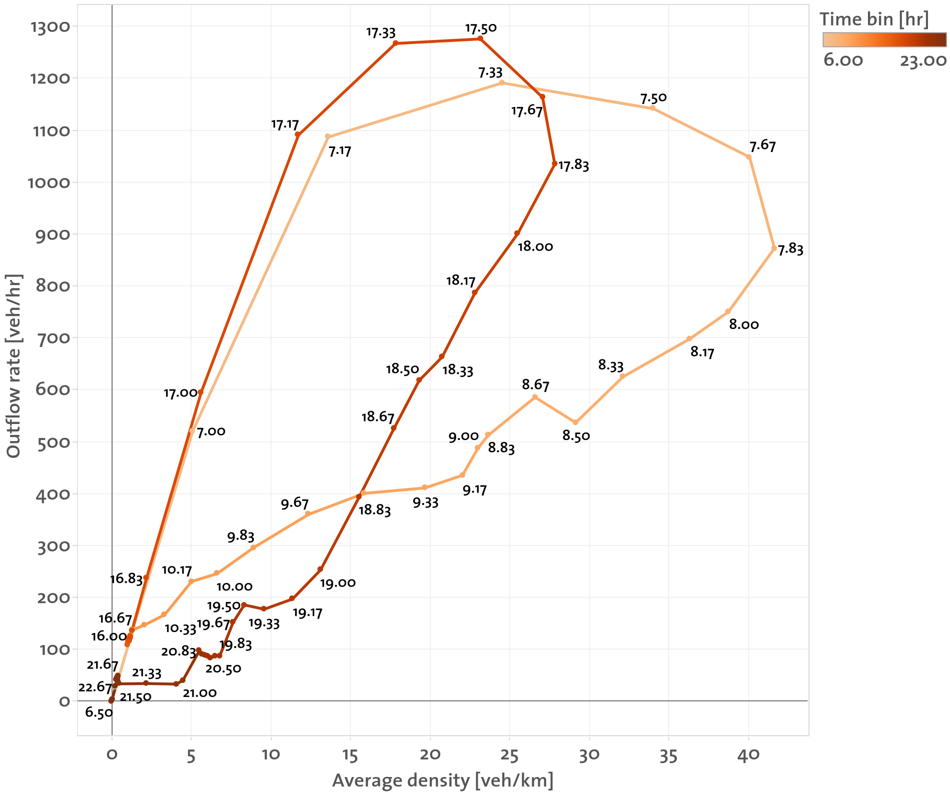


Figure 19 MFD of initial (un-optimized) system (Case 1, Seed 1, 1 it, 10 min bins).



5 Conclusion

In this work, we present an extended Sioux Falls test scenario with an integrated private and public transportation network and an agent-based, dynamic demand. To our knowledge, this is the first attempt to create a publicly accessible test scenario, with realistic, socio-demographically diverse and spatially distributed demand. This scenario can be used for development, testing and evaluation of models and tools within an agent-based simulation framework. Especially for concepts dependent on households' and individuals' socio-demographics, such as value of time, or for initial implementation of computationally intensive methods, such as certain optimization techniques, the scale and detail of this scenario prove highly advantageous. The future extension of this scenario might include more detailed modelling of secondary activities, including more complex activity chains and incorporation of educational activities.

We hope that this work will contribute to the further development of detailed test and benchmark scenarios, in order to further facilitate comparability and tool development in the fields of transport planning and transportation economics.

Acknowledgments

We would like to acknowledge the City of Sioux Falls GIS division for kindly proving data on the current land-use and Ihab Kaddoura and Benjamin Kickhoefer from TU Berlin, who provided feedback on potential bugs during the scenario generation and testing process. We also thank Prof. Kay Axhausen and Dr. Alexander Erath for valuable input to the generation of this scenario.

References

Abdulaal, M. and L.J. LeBlanc (1979a) Continuous Equilibrium Network Design Models, *Transportation Research Part B: Methodological*, **13** (1) 19–32.

Abdulaal, M. and L.J. LeBlanc (1979b) Methods for Combining Modal Split and Equilibrium Assignment Models, *Transportation Science,* **13** (4) 292–314.

Balmer, M., K. Axhausen and K. Nagel (2006) Agent-Based Demand-Modeling Framework for Large-Scale Microsimulations, *Transportation Research Record: Journal of the Transportation Research Board*, **1985**, 125–34.

Balmer, M., M. Rieser, K. Meister, D. Charypar, N. Lefebvre, K. Nagel and K. Axhausen (2009) MATSim-T: Architecture and simulation times, in A. L. C. Bazzan and F. Klügl (eds.) *Multi-Agent Systems for Traffic and Transportation Engineering*, 57–78, Information Science Reference, Hershey.

Bar-Gera, H. (2014) Transportation Network Test Problems, accessed January 27. http://www.bgu.ac.il/~bargera/tntp/.

Bar-Gera, H., F. Hellman and M. Patriksson (2013) Computational Precision of Traffic Equilibria Sensitivities in Automatic Network Design and Road Pricing, *Procedia - Social and Behavioral Sciences*, **80** 41–60.

Chakirov, A. and A. Erath (2011) Use of Public Transport Smart Card Fare Payment Data for Travel Behaviour Analysis in Singapore, paper presented at *16th International Conference of Hong Kong Society for Transportation Studies,* Hong Kong.

Charypar, D. and K. Nagel (2005) Generating Complete All-Day Activity Plans with Genetic Algorithms, *Transportation*, **32** (4) 369–97.

Dijkstra, E. W (1959) A Note on Two Problems in Connexion with Graphs, *Numerische Mathematik,* **1** (1) 269–71.

Friesz, T.L., H.-J. Cho, N.J. Mehta, R.L. Tobin and G. Anandalingam (1992) A Simulated Annealing Approach to the Network Design Problem with Variational Inequality Constraints, *Transportation Science*, **26** (1) 18–26.

Geroliminis, N. and C.F. Daganzo (2007) Macroscopic Modeling of Traffic in Cities, paper presented at *86th Annual Meeting of the Transportation Research Board,* Washington, D.C..

Geroliminis, N. and C.F. Daganzo (2008) Existence of Urban-Scale Macroscopic Fundamental Diagrams: Some Experimental Findings, *Transportation Research Part B: Methodological*, **42** (9) 759–70.

Giuliano, G. and J. Dargay (2006) Car Ownership, Travel and Land Use: A Comparison of the US and Great Britain, *Transportation Research Part A: Policy and Practice* **40** (2) 106–24.

Helbing, D. (2011) *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior*. Springer Verlag.

Horni, A., D. Charypar and K.W. Axhausen (2011) Variability in Transport Microsimulations Investigated with the Multi-Agent Transport Simulation Matsim, *Arbeitsberichte Verkehrs- und Raumplanung*, **692**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Jong, G. de, J. Fox, A. Daly, M. Pieters and R. Smit (2004) Comparison of Car Ownership Models, *Transport Reviews*, **24** (4) 379–408.

Kaddoura, I., B. Kickhöfer, A. Neumann and A. Tirachini (2013) Optimal Public Transport Pricing: Towards an Agent-Based Marginal Social Cost Approach, *hEART* 2013, 13–09, Stockholm.

Kickhöfer, B., D. Grether and K. Nagel (2011) Income-Contingent User Preferences in Policy Evaluation: Application and Discussion Based on Multi-Agent Transport Simulations, *Transportation*, **38** (6) 849–70.

Kickhöfer, B. and K. Nagel (2013) Towards High-Resolution First-Best Air Pollution Tolls, *Networks and Spatial Economics*, 1–24.

LeBlanc, L.J. (1975) An Algorithm for the Discrete Network Design Problem, *Transportation Science*, **9** (3) 183–99.

LeBlanc, L.J., E.K. Morlok, and W.P. Pierskalla (1975) An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem, *Transportation Research*, **9** (5) 309–18.

MATSim-T Multi Agent Transport Simulation (2013), accessed April 30. http://matsim.org/.

Meng, Q., H. Yang and M.G.H. Bell (2001) An Equivalent Continuously Differentiable Model and a Locally Convergent Algorithm for the Continuous Network Design Problem. *Transportation Research Part B: Methodological*, **35** (1) 83–105.

Morlok, E.K., J.L. Schofer, W.P. Pierskalla, R.E. Marsten, S.K. Agarwal, J.W. Stoner, J.L. Edwards, L.J. LeBlanc and D.T. Spacek (1973) Development and Application of a Highway Network Design Model, Volumes 1 and 2, Final Report: FHWA Contract Number DOT-PH-11, Northwestern University, Evanston.

Müller, K. and K.W. Axhausen (2011) Population Synthesis for Microsimulation: State of the Art, paper presented at *90th Annual Meeting of the Transportation Research Board*, Washington, D.C..

Nagel, K. and G. Flötteröd (2009) Agent-Based Traffic Assignment: Going from Trips to Behavioral Travelers, paper presented at 12th International Conference on Travel Behaviour Research (IATBR), Jaipur.

R Core Team, R (2005) A language and environment for statistical computing, R foundation for Statistical Computing, Vienna, Austria. http//www.R-project.org/.

Rieser, M. (2010) Adding transit to an agent-based transportation simulation, Ph.D., TU Berlin, Berlin.

Simini, F., M.C. González, A. Maritan and A.-L. Barabási (2012) A Universal Model for Mobility and Migration Patterns, *Nature* **484** (7392) 96–100.

Suwansirikul, C., T.L. Friesz and R.L. Tobin (1987) Equilibrium Decomposed Optimization: A Heuristic for the Continuous Equilibrium Network Design Problem, *Transportation Science*, **41** (4) 254–63.

Tirachini, A. D.A. Hensher and J.M. Rose (2012) Multimodal Pricing and Optimal Design of Public Transport Services: The Interplay between Traffic Congestion and Bus Crowding, in Proceedings of the Kuhmo Nectar Conference on Transportation Economics, Berlin.

Appendix 1

Figure 20 Matrix of trips between each node pair from LeBlanc *et al.* (1975). The units are hundreds of vehicles per day (in LeBlanc *et al.* (1975) the trip table was initially erroneously labelled as thousands of vehicles per day and later corrected in Abdulaal and LeBlanc (1979b) to hundreds of vehicles per day).

	Destination Nodes																				
Origin Nodes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0	1	1	5	2	3	5	8	5	13	5	2	5	3	5	5	4	1	3	3	1
2	1	0	1	2	1	4	2	4	2	6	2	1	3	1	1	4	2	0	1	1	0
3	1	1	0	2	1	3	1	2	1	3	3	2	1	1	1	2	1	0	0	0	0
4	5	2	2	0	5	4	4	7	7	12	14	6	6	5	5	8	5	1	2	3	2
5	2	1	1	5	0	2	2	5	8	10	5	2	2	1	2	5	2	0	1	1	1
6	3	4	3	4	2	0	4	8	4	8	4	2	2	1	2	9	5	1	2	3	1
7	5	2	1	4	2	4	0	10	6	19	5	7	4	2	5	14	10	2	4	5	2
8	8	4	2	7	5	8	10	0	8	16	8	6	6	4	6	22	14	3	7	9	4
9	5	2	1	7	8	4	6	8	0	28	14	6	6	6	9	14	9	2	4	6	3
10	13	6	3	12	10	8	19	16	28	0	40	20	19	21	40	44	39	7	18	25	12
11	5	2	3	15	5	4	5	8	14	39	0	14	10	16	14	14	10	1	4	6	4
12	2	1	2	6	2	2	7	6	6	20	14	0	13	7	7	7	6	2	3	4	3
13	5	3	1	6	2	2	4	6	6	19	10	13	0	6	7	6	5	1	3	6	6
14	3	1	1	5	1	1	2	4	6	21	16	7	6	0	13	7	7	1	3	5	4
15	5	1	1	5	2	2	5	6	10	40	14	7	7	13	0	12	15	2	8	11	8
16	5	4	2	8	5	9	14	22	14	44	14	7	6	7	12	0	28	5	13	16	6
17	4	2	1	5	2	5	10	14	9	39	10	6	5	7	15	28	0	6	17	17	6
18	1	0	0	1	0	1	2	3	2	7	2	2	1	1	2	5	6	0	3	4	1
19	3	1	0	2	1	2	4	7	4	18	4	3	3	3	8	13	17	3	0	12	4
20	3	1	0	3	1	3	5	9	6	25	6	5	6	5	11	16	17	4	12	0	12
21	1	0	0	2	1	1	2	4	3	12	4	3	6	4	8	6	6	1	4	12	0
22	4	1	1	4	2	2	5	5	7	26	11	7	13	12	26	12	17	3	12	24	18
23	3	0	1	5	1	1	2	3	5	18	13	7	8	11	10	5	6	1	3	7	7
24	1	0	0	2	0	1	1	2	2	8	6	5	7	4	4	3	3	0	1	4	5

Appendix 2

beelineDistanceFactor	1.3
walkSpeed	3 km/h
stuckTime	3600 sec
timeStepSize	1 sec
maxAgentPlanMemorySize	6
ChangeExpBeta	0.7
TimeAllocationMutator – MutationRange	3600 sec

MATSim specific parameters used for the simulation