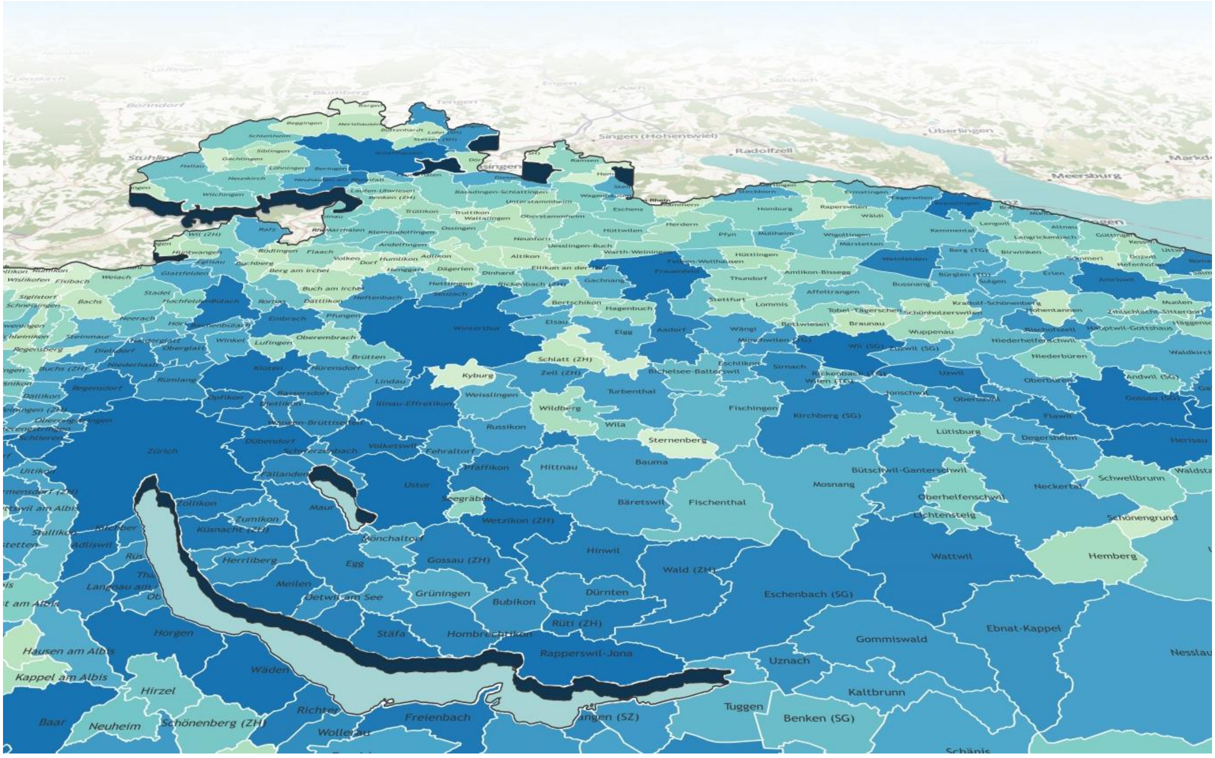


FaLC Transport Simulation Module: How accurate can simplified travel time calculations be?

A Zeiler, regioConcept
G Sarlas, IVT
M Kuliowksy, ESMO
BR Bodenmann, regioConcept
B Sanchez, regioConcept
J Bode, regioConcept
P Furtak, ESMO
KW Axhausen, IVT

Conference paper STRC 2014

STRC

14th Swiss Transport Research Conference
Monte Verità / Ascona, May 14-16,2014

FaLC Transport Simulation Module: How accurate can simplified travel time calculations be?

A Zeiler
BR Bodenmann
B Sanchez
J Bode

G Sarlas
KW Axhausen

M Kuliowsky
P Furtak

regioConcept AG
Schmiedgasse 33
CH-9100 Herisau

Institut für Verkehrsplanung
und Transportsysteme (IVT)
Stefano-Franscini-Platz 5
CH-8093 Zürich

ESMO Žilina a.s.
Rosinská cesta 8
SK-010 08 Žilina

www.regioconcept.ch

www.ivt.ethz.ch

www.esmo.sk

April 2014

Abstract

This paper presents, the assumptions, simplifications and accuracy of different approaches to estimate travel times in transport simulation. Additionally, the paper focuses on the assumptions in the simplified transport simulation module in FaLC as well as its methodology to achieve efficient results. To validate different approaches, the paper compares two different results for travel time estimations (calculation based on speed regression or maximum allowed speed) with the results of the Swiss National Passenger Transport Model (NPVM) calculated in VISEM.

Keywords

Travel time, transport simulation, land-use simulation, speed regression, FaLC, IVT, regioConcept.

1. Introduction

The effect of land-use changes on use of transport infrastructure – and vice versa – is an important task of all integrated land-use and transport simulation tools. Indeed, developers of a land-use simulation tool like FaLC are well advised to focus on certain questions, crucial for the land-use modelling. This certainly involves the calculation of accessibility, a mayor input for spatial planners, cost benefit analyses and indicators regarding environment and climate. In contrast, transport planning issues and simulation are subjects for specialised transport modelling tools like MATSim or VISUM/VISEM.

The FaLC (Facility Location Choice Simulation Tool) is an integrated transport and land use simulation tool (LUTI) developed in a joint project of the Institute for Transport Planning and Systems (IVT) at ETH Zurich, regioConcept and ESMO. The idea of FaLC is to develop a tool to answer different questions for scientific as well as planning purposes (see Bodenmann et al. 2014).

FaLC provides in a first step an integrated Transport Simulation Module. Indeed, this tool is used in FaLC in order to have (very) fast calculations of accessibility variables and (in the future) rough estimations of traffic flows. All this, by implementing (and simplifying) the three transport networks (private, public and soft-mobility) including some important variables like length, potential and maximal velocities or travel time, among others.

This paper presents, the assumptions, simplifications and accuracy of different approaches to estimate travel times in transport simulation. Additionally, the paper focuses on the assumptions in the simplified transport simulation module in FaLC as well as its methodology to achieve efficient results. To validate different approaches, the paper compares two different results for travel time estimations (calculation based on speed regression or maximum allowed speed) with the results of the Swiss National Passenger Transport Model (NPVM) calculated in VISEM.

2. Data

2.1 OpenStreetMap (OSM)

To dispose of a freely available base for the street network, OpenStreetMap (OSM) was chosen. For the area of Switzerland, it offers a good alternative to other solutions with property restrictions and/or costs. For the whole country a Shape-file containing all recorded roads can be downloaded via GEOFABRIK[1]. As OSM data is available all over the world with a consistent data structure, this offers the possibility to easily get and exchange the street network for future study areas. The disadvantage of OSM lies in its nature to be made out of volunteer. The completeness as well as the quality of the data always depends on the person digitalising it (Mondzech and Sester, 2011; Haklay, 2010).

The feature class “roads” contains the attributes:

- osm_id
- name
- ref
- type
- oneway
- bridge
- tunnel
- maxspeed

As the OSM street network data also includes street types like footway, lift, steps and so on, only the relevant subset is extracted. Indeed, only the following street types were considered to be relevant:

- motorway and motorway_link
- trunk and trunk_link
- primary and primary_link
- secondary and secondary_link
- tertiary and tertiary_link

After this selection some streets are not connected to the rest of the network anymore. To avoid connecting the centre of the zones to disconnected part of the network, this streets were

[1] http://download.geofabrik.de/

deleted. The centres of the FaLC zones are than connected to the nearest edge on of the network.

The Shape-file of the base street network (including connectors) consists of 55’696 features, containing 567’769 edges and 557’322 vertices.

2.2 Swiss National Passenger Transport Model (NPVM)

To be able to validate the results of the distances and times calculated based on the OSM network and as the FaLC locations correspond to the NPVM zones, the distance and travel time tables provided by the Swiss National Passenger Transport Model (NPVM 2005) are used to compare the results. The street network used in the NPVM is initially built on the GIS model MicroDrive (status 1999). Till the status of 2005, changes in the network have been added and errors were corrected (ARE, 2006; ARE, 2010; ARE, 2013). In total four different tables are imported by FaLC including:

- travel distances for private motorized transport
- travel times for private motorized transport
- travel distances for public transport
- travel times for public transport

3. Modelling Speed and Travel-times

3.1 OSM with Maximum Speed (OSM_{MS})

Based on the fact, that by default no information about the real travel time on the OSM network is given, in the simplified transport model of FaLC it is supposed that everyone is travelling with the maximum allowed speed. As mentioned above, the OSM network contains an attribute called “maxspeed” but unfortunately the value is not set for all segments and some of the speed values are not reasonable. Thus some completions as well as corrections need to be done. According to the street type a minimum and maximum “maxspeed” value is defined as shown in Table 1. So if a value is below the minimum it is increased to the minimum level defined. In the same way to high levels are decreased. For the cases, where no value is set, the defined maximum “maxspeed” is set.

Table 1 OSM network generalisations

Road type	Min Speed [km/h]	Max Speed [km/h]
Motorway/Trunk	100	120
Motorway/Trunk access	60	80
Primary	60	80
Secondary	50	60
Tertiary	30	50
Link to Location	50	50

Sources: own estimations

3.2 OSM with Speed Regression (OSM_{SR})

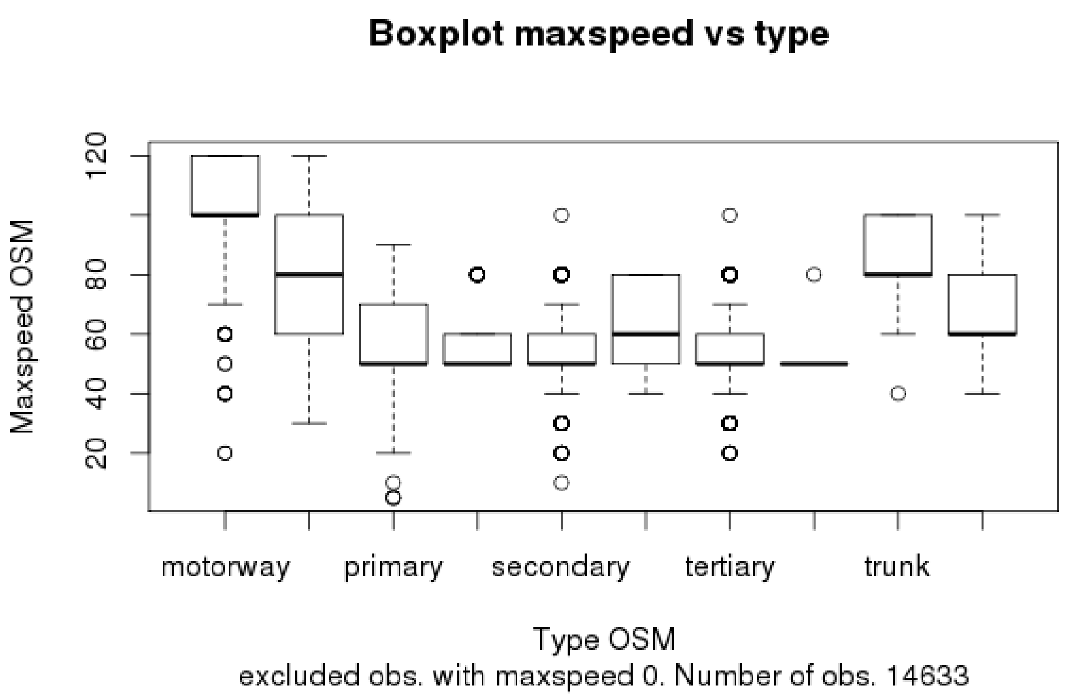
As it is expected that taking only the maximum allowed speeds leads to an underestimation of travel times, a more sophisticated but still simplified way of estimating velocities for each link of the study network is needed.. As such, regression modeling is employed as the mean to accomplish that. The main advantage of the choice of regression modeling, in comparison to more advanced and complex demand models (eg four-step travel demand modelling approach, agent-based approaches), is that it can provide a structural equation of the speeds in association with the characteristics that affect the demand for travel but in a significantly less cumbersome way than the aforementioned approaches.

The purpose of the regression model is to be applied to predict the speeds on the links of the entire network of study. The average daily speed for a typical weekday is the dependent variable of interest. The regression yields two speed components; first, the average road speed per road type, is a non-spatial quantity. Spatial variation is added to the link speed estimates in the second component via the spatially resolved explanatory variables. Spatially resolved road and public transport network densities represent the effect of road supply on speed. Spatial data on population and employment is taken to be indicative of the intensity of local activities, reflecting travel demand locally (Hackney et. al, 2007).

3.2.1 Network characteristics and actual speed values

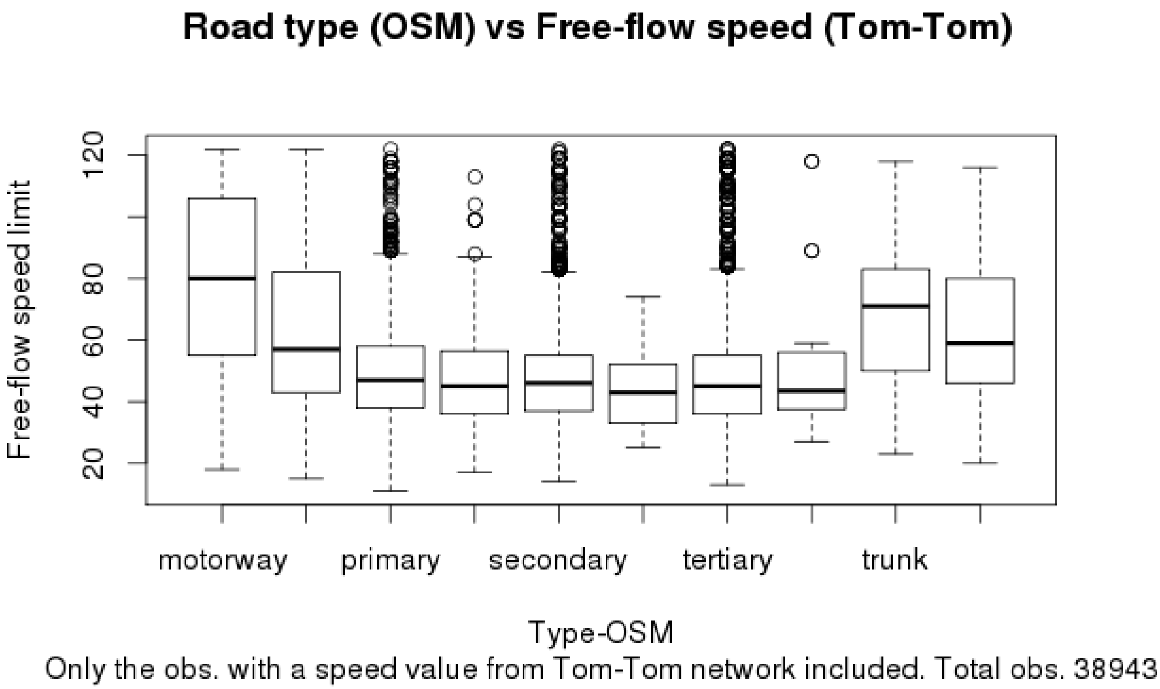
As mentioned before, the network that was employed for the current project is an open-street map (OSM) network of Switzerland. In order to assess the quality of the classification of the links in the OSM network, the maximum speed per link type is plotted in order to obtain an accurate idea of the consistency of the classification of the links based on their reported max-speed. As it can be seen in Fig.1, the deviation of the max-speed and the average value per type can be characterized as reasonable and in line with our expectations. Based on the above, it can be concluded that the reported classification of the links of the network (at least of those with reported maximum speed different than zero) seems to have been accomplished in a consistent way, with the exception of very few links (represented as circles in Fig.1).

Figure 1: Boxplot of maximum speed per reported type of link, excluding links with zero reported maximum speed



In order to proceed to the estimation of the speed regression model, a set of actual speeds is prerequisite to be used as target values for the regression. The target values were available in a different network (commercially provided by Tom-tom), where the average daily speed for weekdays was estimated based on GPS measurements, on a link level. In order to overcome the limitation of different networks' definition, the employed network was spatially joined with the Tom-tom network in order to transfer the actual speed values of each link to the OSM network. An approximate of 22% of the links included in the network of study, are not spatially joined to a link from the Tom-tom network, or are joined to a link that doesn't have an estimated average speed.

Figure 2: Boxplot of free-flow speed per reported type of link after the spatial join of the networks



In Figure 2 the free flow speed of the spatially joined links of the OSM network (transferred from Tom-tom network) is plotted against the type of the links. Attempting a comparison with the plot in Fig.1 and assuming that the free flow speed variable is the closest equivalent to the maximum speed variable of OSM network, it can be seen that the average free flow speed is a bit lower than in Fig.1, and also the deviation from the average values is larger than before. This difference can be explained mainly by the different network definitions and also up to an extent can be attributed to wrongly spatially joined links. The significantly larger number of observations than before contributes to the larger deviation. Additionally, it cannot be overlooked the nature of the OSM network where the variables, and thus the maximum speed, are registered by individual users, who might perceive as higher the actual speed than it really is, and thus report it mistakenly. In summary, the same trend as before can be seen regarding the average speeds, and thus it can be concluded that even though the outcome of the spatial join process is not perfectly accurate since links' definition varies between the two networks, however it can be considered to be sufficiently accurate for the purposes of this project.

3.2.2 Explanatory variables

As mentioned before, the second speed component consists of spatially resolved explanatory variables. Spatially resolved road and public transport network densities represent the effect of road supply on speed. The corresponding road (line) densities were calculated in ArcGIS and the results are stored in cells (pixels) of 100 x 100 meters dimensions to coincide with the cells from the socio-demographic data (hectare data). The results are expressed in meters (length) per square kilometers. In order to calculate the density of the OSM network, the full OSM network of Switzerland was fetched for that reason (date: 30 September 2013). The line density was calculated over different radii R of 100 meters, 500 meters, 1 kilometer, and 5 kilometers. Apart from the total road network density, the density per link type was calculated as well to examine the impact of specific road types' density on the speed, over radii of 500 meters and 1 kilometer. The same density measurements were also calculated for the Tom-tom network and the Navteq network (commercial navigational networks) to compare the results and assure that the density calculation is made in a correct way.

In the case of the public transport network, it is considered more appropriate to calculate the density of the number of stop points as more representative measure. As such, the point density was calculated using ArcGIS software over the same radii as the previous densities. Each point was taken into account multiple times equal to the number of lines that serve each particular stop.

In the same way, the spatial explanatory variables of population and employment densities were calculated. More specifically, the variables of the total population and the full-time equivalent employment positions per hectare were created by aggregating the corresponding values. In this particular case, both variables' densities were also kernel weighted to capture their diminishing impact on the speeds over distance. The same radii as before were used for the calculation and they are expressed in population/ employment positions per square kilometer.

Population and employment data, aggregated per hectar, are available from the Swiss Federal Statistical Office (BFS: Bundesamt für Statistik). The population data correspond in the year 2011, were taken from the "Statistics of Population and Households 2011" ("Statistik der Bevölkerung und der Haushalte 2011", date of version: 30 August 2012), while the employment data were taken from the "Federal Business Census" of 2008 (Eidgenössische Betriebszählung 2008, date of version: 29 March 2008).

The spatial explanatory variables need to be associated somehow to the links of the network. Thereupon, each link of the network was associated with the hectare (cell) values of each spatial variable, closest to the upstream endpoint of the link.

Apart from the above spatially explanatory variables, the average gradient (slope) of each link is an important variable that its inclusion in the model should be tested. A priori, it is expected that uphill results in reductions on the speeds, especially in the case of lower classified roads (e.g. secondary streets). A terrain model with a resolution grid of 25 x 25 meters was used in order to infer the average gradient of each link of the network by intersecting the terrain model with the nodes of the network.

3.2.3 Specification and estimation of the model

The next step after the preparation of the set of variables was to proceed to the specification of the regression model and the estimation of its corresponding parameters. The estimation of the regression model was conducted in R, by means of ordinary least squares estimation (OLS). The estimated values of its regressors are presented in table 2. Based on the results from different tested specifications, and more specifically due to the fact that some of the regressors of the different types of link (average speed) have almost identical values, some link types were merged together (eg motorway and motorway link), resulting to a reduction in the number of variables in a coherent and consistent way (six types of links instead of 10).

Table 2: Estimated parameters of the speed regression model

Coefficients	Estimate	Std. Error	t-value	Pr(>|t|)
Type: motorway	8.50E+01	3.74E-01	227.054	< 2E-16
Type: motorway link	7.33E+01	4.41E-01	166.052	< 2E-16
Type: primary	5.73E+01	2.46E-01	232.771	< 2E-16
Type: secondary	5.57E+01	2.39E-01	233.551	< 2E-16
Type: tertiary	5.39E+01	2.12E-01	253.659	< 2E-16
Type: trunk	7.13E+01	5.52E-01	129.258	< 2E-16
Ramp (dummy)	-9.30E+00	4.03E-01	-23.095	< 2E-16
PuT stops 0.5km	-4.36E-01	2.75E-02	-15.859	< 2E-16
Employm.Normal 0.5km	-2.25E-04	3.13E-05	-7.189	6.64E-13
Population kernel 0.5km	-6.45E-04	4.47E-05	-14.433	< 2E-16
Osm residential density 0.5km	-5.59E-01	1.76E-02	-31.8	< 2E-16
Osm line-density 0.5km	-9.68E-02	9.25E-03	-10.467	< 2E-16
Type secondary: gradient	-6.09E-02	4.89E-02	-1.244	0.21338
Type tertiary: gradient	-9.55E-02	3.58E-02	-2.668	0.00764
Type trunk: gradient	-1.67E-01	1.40E-01	1.40E-01	0.23242

Number of observations: 38943

Mult. Adjusted R-square: 0.9099

AIC: 32434.8

The spatial explanatory variables that were included in the model reflect that the impact of the socio-demographic and network data is rather localized. More specifically, the variables of densities over a radius of 500 meters were found to give the more reasonable and statistically significant results. Population was taken into account as kernel weighted density, exhibiting a diminishing impact on the link speeds over the distance while the full-time equivalent employment positions and public transport stops were taken into account as normal densities (point densities). Regarding the road densities per type of links, it was found that the residential roads density variable should be included in the model. A finding which is in accordance to the fact that residential roads dominate over the rest road types (about 1/5 of roads are classified as residential ones), and thus the use of their density as a stand-alone variable can be considered justifiable. The inclusion of ramps density variable was checked as well but it was found not to have the expected impact on the regression (opposite sign), and thus it was excluded.

A special attention was paid on the inclusion of the gradient of the links in the regression model. Gradient as a stand-alone variable resulted to regressor values with opposite sign, compared to the expectation. As a consequence, gradient of the links was taken into account as an interaction with the link type, though only for the links with lower classification (secondary, tertiary, and trunk links).

It is useful to put into perspective the range of spatially resolved explanatory variables values to comprehend their impact on the speed regression model.

Table 3: Range of values of the non-dummy explanatory variables

Variable	Range of values
Density of PuT stops in 0.5 km radius	0-33.1 stops/ sqr.km
Density of employment positions in 0.5 km radius	0-46564 employm. / sqr.km
Kernel weighted density of population in 0.5 km radius	0-27428 residents /sqr.km
Density of residential links (OSM) in 0.5 km radius	0-33.5 m/ sqr.km
Density of all links (OSM) in 0.5 km radius	0-97.8 m/ sqr.km
average link gradient	[-8,8] %

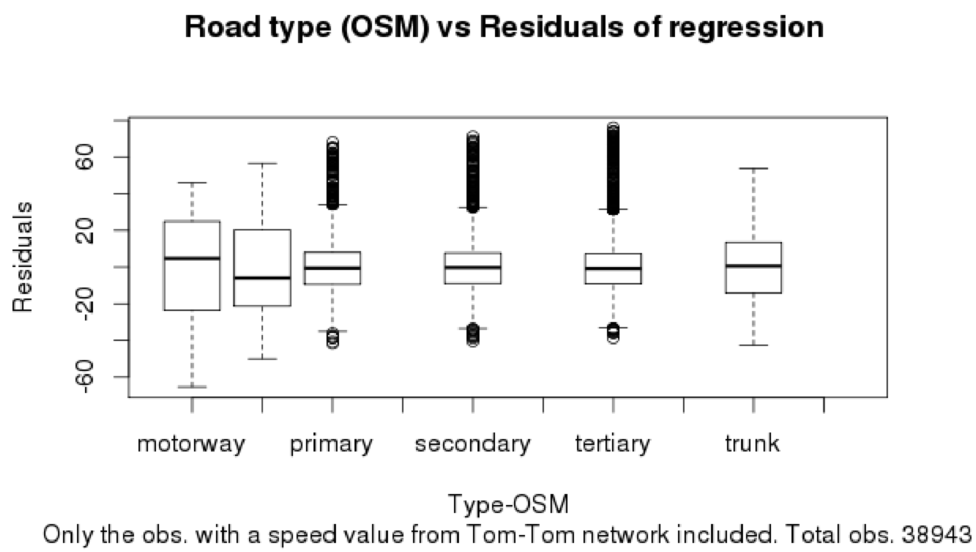
The correlation of the non-dummy included variables was calculated to measure the linear association between them. As it can be seen in table 4, correlation values lie between 0.31 and 0.66, indicating a positive correlation. The magnitude and the sign of the correlation values come as a consequence of the fact that these variables are reflective of the urban density. However, it was considered that each variable has descriptive power that cannot be overlooked and justify their omission.

Table 4: Correlation between the non-dummy explanatory variables

	Employment	PuT stops	Population	Links' density	Residential links density
Employment		0.626	0.481	0.662	0.313
PuT stops	0.626		0.650	0.662	0.500
Population	0.481	0.650		0.683	0.623
Links' density	0.662	0.662	0.683		0.643
Residential links density	0.313	0.500	0.623	0.643	

An analysis of the residuals of the estimated regression model was conducted in order to evaluate the model and comprehend its weaknesses and strengths. As it can be seen in the following plot, the cases of underestimation (positive residuals) were more often in the cases of links classified as primary, secondary, and tertiary. This finding makes apparent that the classification of the links of the OSM network might not have been conducted in a fully consistent way at the first place, and thus the links within the same type might not be homogeneous with respect to their free flow speed, and subsequently with their average daily speed. Consequently, the wrong classification might be the underlying cause of the sign and the magnitude of the residuals. In addition, the magnitude of the residuals raises some concerns that they might be attributed to wrongly matched links during the spatial join procedure. In the cases of the rest types of links, the residuals range lies between -20 and 20 kilometers roughly.

Figure 3: Residuals of regression per road type



3.3 Public Transport

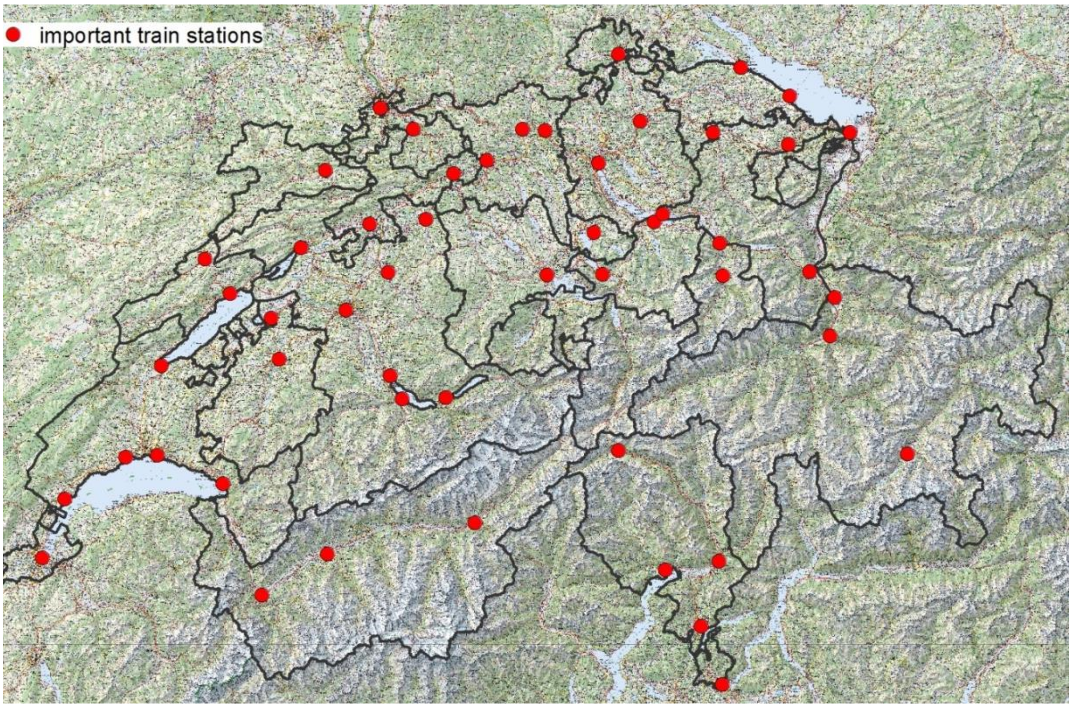
The aim of the set-up of a public transport network in FaLC is to achieve a network that can be treated in the same way as the street network. This implies:

- All relevant information (travel time, frequency, product etc.) is saved in the links of the network
- The network is flexible i.e. the possibility to change the network easily is given
- Shortest path queries are possible
- All FaLC-zones are connected
- Travel times are representative for home-work journeys (and vice versa). For that morning peak hours on a workday are considered (e.g. 07 to 09 on June 12th)
- The travel time between location and railway or bus station depends on the distance between the centre of the municipality and the station.

To get a test-network for the public transport, train and bus connections publicly available in the internet were used. As it would have been very time expensive to search the connections from each zone to each other zone, it was decided to just consider only connections from a zone A to all zones within a radius of 10km and to all important train stations within 50km. Doing so, the amount of data as well as the redundancy in the data can be reduced. The important train stations where chosen considering intercity-connections as well as the importance for transfers (Figure 4).

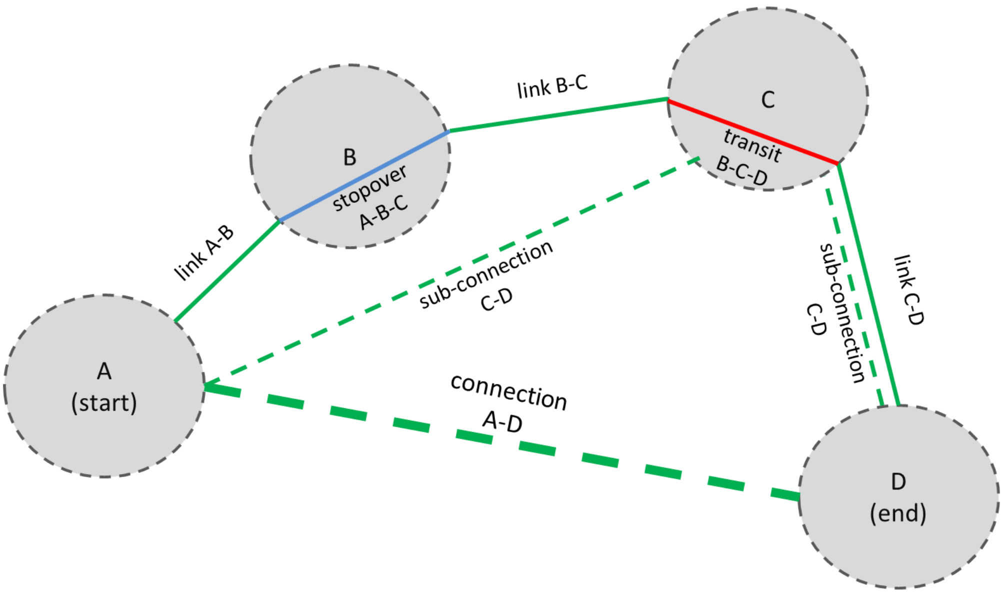
Looking at the structure of a journey by public transport, one will always find the same/similar pattern. There is a start station as well as an end station. The connection departs at a certain time, arrives at a certain time and this consequently results in a total travel time. Furthermore, searching for a connection within a time window, normally a number of connections are available and the best (fastest and fewest transfers) is preferred. Looking a little bit closer, we recognize that there is probably a need to change the train/bus. Doing so, one might have to wait at the transfer station. Looking at what happens between two transfer stations it is seen, that the train stops at various stations to let people get in and out. Having all this information, the structure of a journey by public transport is given as illustrated in Figure 5.

Figure 4 important train stations



Source: www.openstreetmap.org (base-network)

Figure 5 structure of public transport journey



When querying data from the web, only the best connections in terms of time and time-cost within the time window are saved to an xml-file. To detect the best connection considering the time cost, every transit is given a penalty of 21min and the waiting time at a transit is weighted by a factor of 0.5 (Vritc and Axhausen, 2002). Following, the connection with the minimum weighted travel time is considered to be the best. The structure of the xml-file corresponds to the structure defined above (Figure 5). To get the network needed in FaLC, in a next step segments are created

- between two stop stations (no transfer occurring) → called links,
- representing a transits → called transits,
- representing a stopover → called stopovers.

So in total three types of segments can be differentiated: links, transfers and stopovers. Thereby the segments are built out of all best-connections containing the respective connection from a stop A to a stop B. Every segment then contains the following information:

- x- and y-coordinates
- from and to station names (and transit/stopover station for transits and stopovers)
- minimum travel time (considering only the best connections)
- mean travel time (considering only the best connections)
- maximum travel time (considering only the best connections)
- capacity for 1. class (no value at the moment) → only for links
- capacity for 2. class (no value at the moment) → only for links
- type (link/stopover/transfer)
- product (IC/Bus/S/…) → only for links
- maximum frequency → only for links
- interval → only for links

To ensure, that only the possible transfers can be used, every transfer is represented as link within the station. Therefor around the punctual station a circle with a radius of 50 meters is built. The links as well only reach to the bounds of this circle so that two links are always connected through a transfer or stopover (which is represented in the same way as a transit).

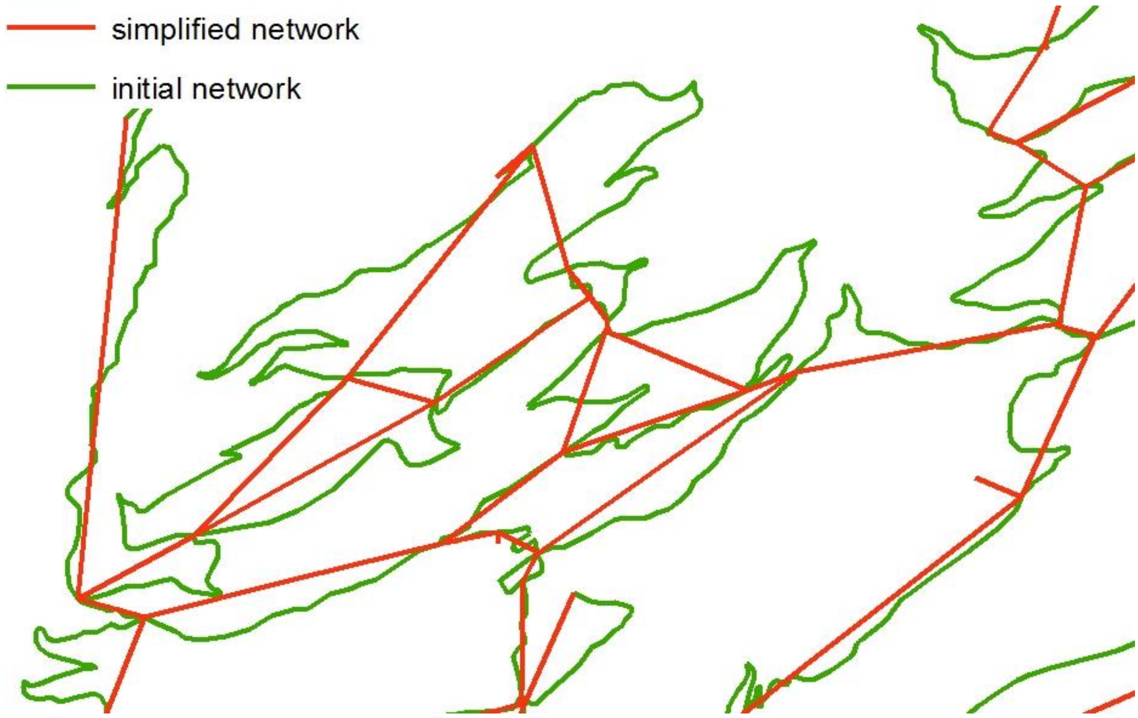
The travel time between location and railway or bus station depends on the distance between the centre of the municipality and the station. Therefor the average pedestrian speed is assumed to be 5km/h. As the considered beeline is not representative for most of the routes, the speed is decreased to 3km/h.

3.4 Integration in FaLC

3.4.1 Simplification of the OSM network

As performance is a very crucial issue in FaLC and the geometry of the OSM-network is far to detailed, the network has to be simplified as much as it is possible without losing important information like distance or street-type while decreasing the number of segments significantly. Doing so, the used OSM network segments including the estimated speed-values are given topological information about their neighbours and the vertices are categorized according to the number of segments they are connecting. After this preparation-step, all consecutive segments between two junctions or endpoints, having the same street-type (motorway, primary, etc.) and, in case of the street network with only maximum speed values, the same speed, are grouped. Thereafter only one segment, reaching from the start point of the first segment to the endpoint of the last segment is created. The segment is given a "length"-attribute, which consists of the sum of all lengths of the segments in the group.

Figure 6 network simplification



Source: www.openstreetmap.org (base-network)

In the case where the speed is not restricting the grouping process (and consequently not all segments have the same speed value), the new segment receives the weighted (by length) mean speed of all segments of the group. In this case, where the speed is not considered as a limiting value during the simplification process, the number of segments can be reduced by a factor of about 15 from 567’784 to 37’275 segments. In the other, more restrictive case, where speed is a determining factor of the simplification, however a reduction of segments by a factor of about 4 can be achieved. Regarding the calculation time of the shortest path algorithm (for the network with estimated maximum speed) for the reading and calculating part (no writing to the database) the simplification results in a time reduction by a factor of ,as well, nearly 15 (from about 15 min to 1min).

3.4.2 Shortest Path Calculations

To calculate the shortest paths form each FaLC location to each other location, a standard Dijkstra algorithm is used. As the number of shortest paths that have to be calculated is quite high, calculation time is the most critical issue to be addressed. Because the performance of the algorithm depends on the min-priority queue performance, several data structures where tested. The Fibonacci heap structure turned out to deliver the best performance results and is so used. Another issue related to the performance concerns the optimal network representation in memory as well as the access of all data. Therefor a forward star data structure is used, which contains:

- Mapping table of source vertices to indexes and reversed one
- Array of pointers to current vertex edges (index to edges array)
- Additional arrays with edge information (distance, time)

The graph representation is static and thus stays the same during the whole calculation with no dynamic changes. This allows a processing of the shortest paths with multiple threads – every thread is calculating shortest paths from different start vertex. Doing so, the total performance of the task can be improved.

The shortest path implementation in FaLC calculates the shortest paths and shortest times for three networks:

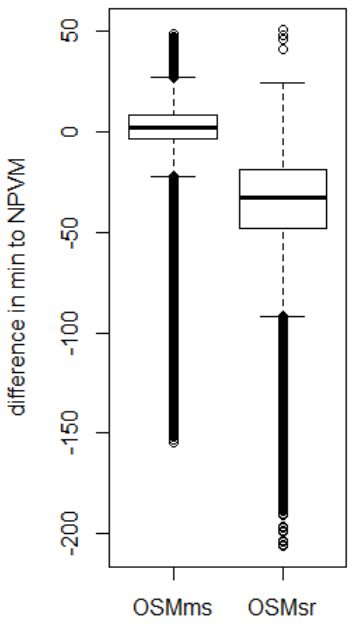
- general car network,
- public transport network and
- bicycle paths network (created as car network subset with limited max. speed).

4. Results

In FaLC environment, three different networks have been integrated as a distance database (from location “A” to location “B”). The user/modeller is therefore able to switch between one network to the other.

The following Figure 7 shows the distribution of the differences between both OSM based travel times and the NPVM in form of boxplots as well as descriptive statistics. Results based on maximum speed assumptions fit quite well with NPVM results: the median of the time differences shows that OSM_{ms} is only 2.5 minutes faster than NPVM. In contrast, estimated times in OSM_{sr} are considerably slower (32.5 minutes in median).

Figure 7 Differences between the travel times based on OSM and NPVM



	Difference NPVM - OSM_{ms}	Difference NPVM - OSM_{sr}
average	1.9	-35.1
median	2.5	-32.5
min.	-154.8	-206.0
max.	48.9	51.0
stdv.	10.9	21.4

4.1 $NPVM_{MIV}$ – OSM_{MS}

The implemented Swiss Case Study currently uses the OpenStreetMap Maximum Speed network for 2013. The following Table 2 describes the differences between the National Transport Model (NPVM) and OSM (maximum speed) in minutes and kilometres for seven selected exemplary trips. As we saw in Figure 7 most travel times are very close to the NPVM; but, there are also some extreme outliers with very big differences up to 155 min.

Table 2 Distance and time comparison between NPVM and OSMms

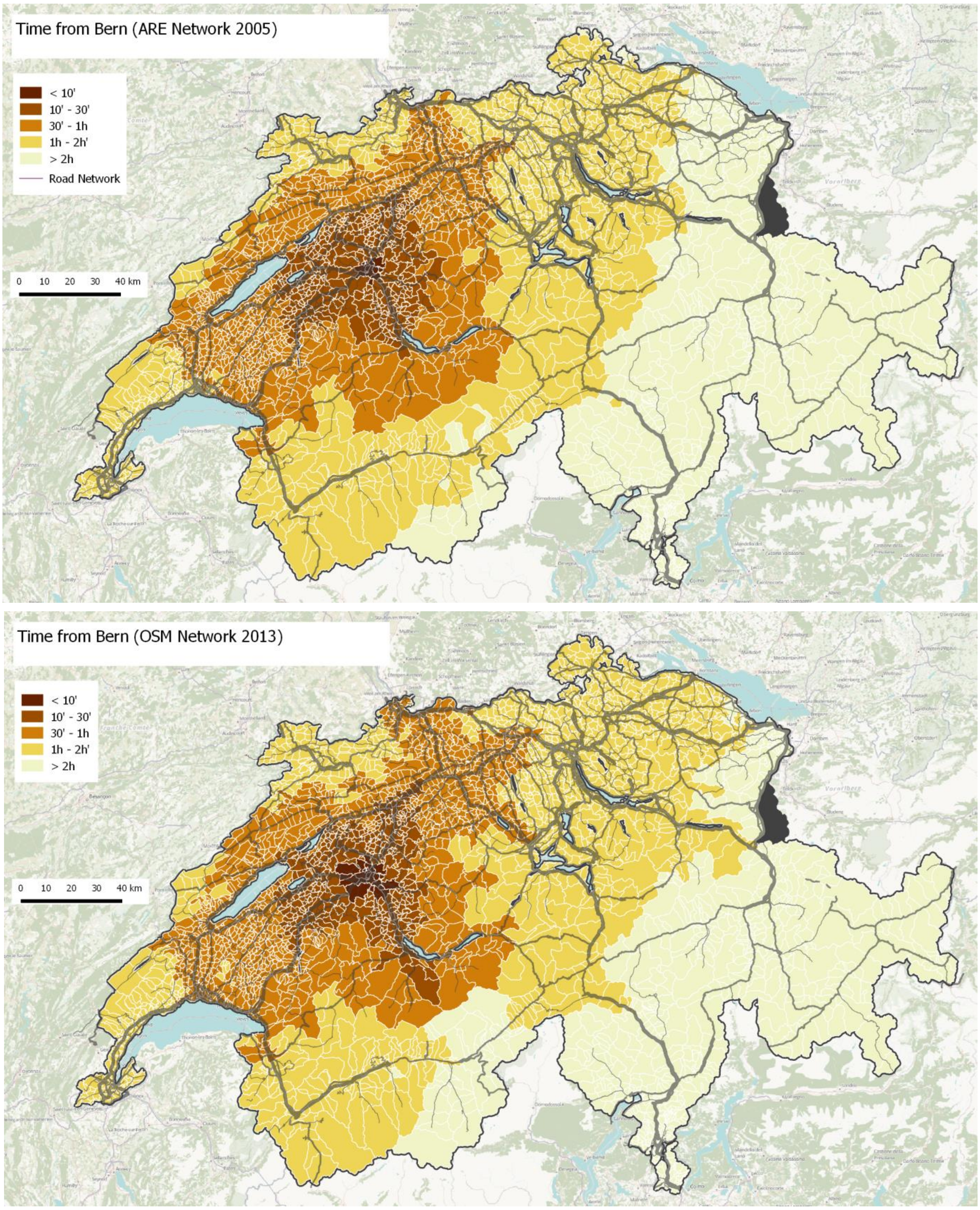
	NPVM - 2005		OSM Max Speed - 2013		NPVM - 2005		OSM Max Speed - 2013	
Trips with motorways	km	%	**km**	**%**	min	%	**min**	**%**
Zurich - Chur	117	100%	**116**	**99%**	70	100%	**66**	**94%**
Lausanne - Zurich	222	100%	**205**	**92%**	138	100%	**124**	**90%**
Bern - Brig	107	100%	**164**	**153%**	115	100%	**124**	**108%**
Trips without motorways	km	%	**km**	**%**	min	%	**min**	**%**
Zurich - Meilen	15	100%	16	**107%**	25	100%	17	**68%**
Lausanne - Lucens	29	100%	29	**100%**	29	100%	30	**103%**
Bern - Vechigen	11	100%	12	**109%**	15	100%	14	**93%**
St. Gallen - Appenzell	15	100%	16	**107%**	17	100%	20	**118%**

The reason for these big differences is assumed to lie in an incompleteness of the OSM network. Nevertheless, in total the travel times calculated with the OSM_{ms} network are on average (median) only about 2 min faster than the NPVM travel times. As the OSM network contains no information about traffic loads and other resistances it was expected that the OSM network will be faster than the NPVM network.

As the differences were expected to be much bigger, this result shows that the very simple assumptions already lead to quite good and usable travel times. The differences concerning the travel distances mainly show the differences between the geometry/completeness of the two networks. It can be observed that in most cases the completeness of the OSM network is higher than the one of the NPVM network and so shorter paths can be found. Indeed the overestimation of the maximum speed on smaller streets can also be a reason to choose a shorter path. However the case of "Bern – Brig" shows a missing part in the OSM network. Comparing the two networks, it is seen that in the NPVM the Lötschberg tunnel (car shuttle train linking a part of Bern – Brig trip) makes part of the street network. However, as in the OSM network only "real" streets were considered the Lötschberg tunnel is missing. Looking at the short trips, where no motorways need to be used (Table 2), for "Zürich – Meilen" it is seen that the travel time is more underestimated than in other cases. As this route is typically very likely to be affected by traffic jams, this shows the need to incorporate loads.

The following set of maps shows graphically the travel time differences in minutes from Bern to the rest of locations. In this case we can also observe that OSM network generally (except the part reachable through the Lötschberg tunnel) covers a larger area in less travelled time.

Figure 8 Travel time differences from Bern using NPVM network (top) and OSM maximum speed network (bottom)



4.2 NPVM$_{MIV}$ – OSM$_{SR}$

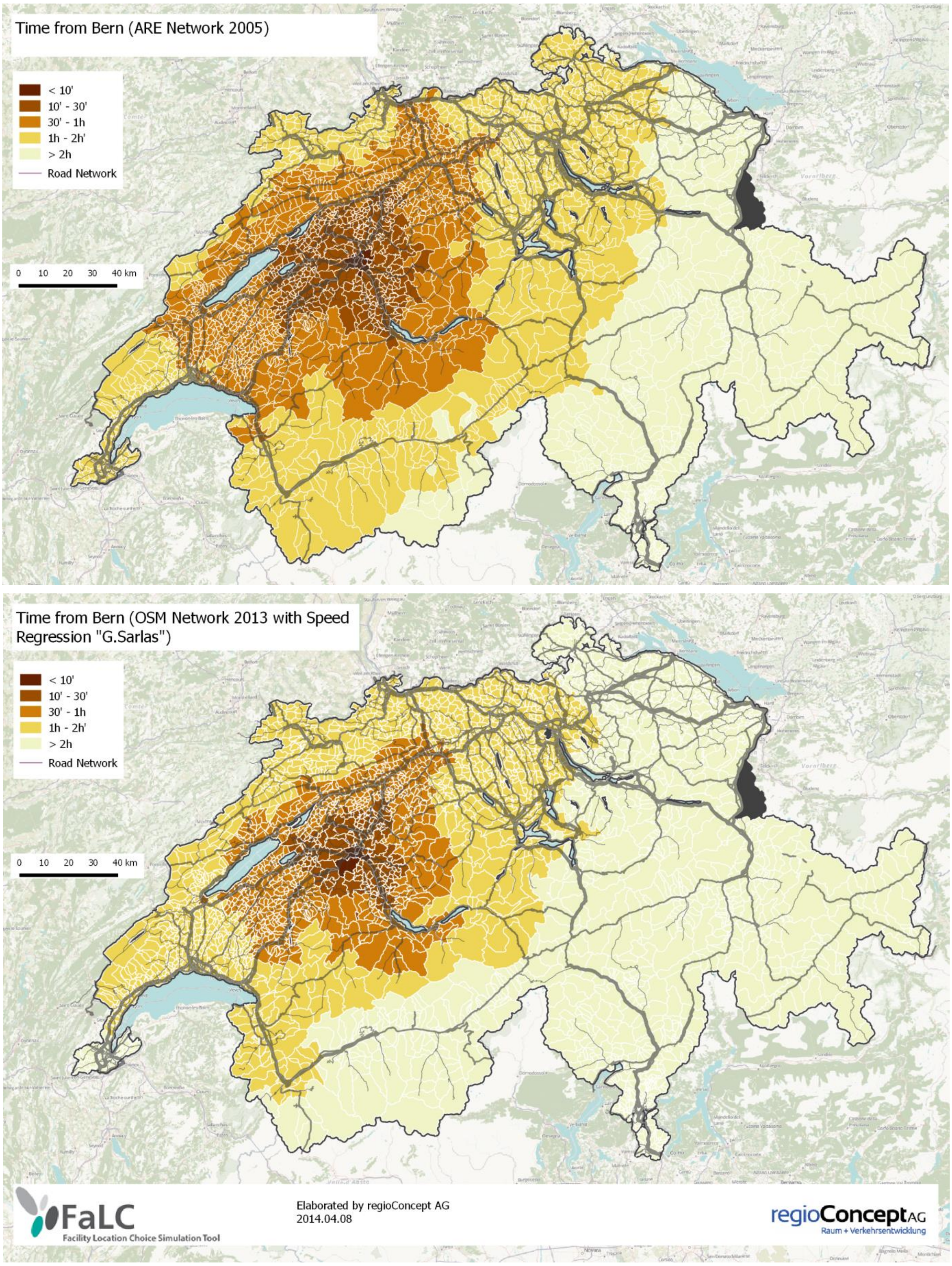
When NPVM and OSM$_{SR}$ (with speed regression) are compared, the result show very different values. With the implementation of the speed regression, we can state that even if the distances are the same as in the previous comparison (same street network) the travelled time is highly increased resulting in large differences to the NPVM results. Like this, on average the travelled time is 32 minutes bigger than in the NPVM. This might be caused by a yet missing calibration of the speed values and thus an underestimation of the latter.

Table 3 Distance and time comparison between NPVM and OSMsr

	NPVM - 2005		**OSM Speed Reg - 2013**		**NPVM - 2005**		**OSM Speed Reg - 2013**	
Trip	km	%	**km**	**%**	min	%	**min**	**%**
Zurich - Chur	117	100%	**116**	**99%**	70	100%	**99**	**141%**
Lausanne - Zurich	222	100%	**205**	**92%**	138	100%	**183**	**133%**
Bern - Brig	107	100%	**164**	**153%**	115	100%	**174**	**151%**
Trip (avoiding motorways)	km	%	**km**	**%**	min	%	**min**	**%**
Zurich - Meilen	15	100%	16	**107%**	25	100%	23	**92%**
Lausanne - Lucens	29	100%	29	**100%**	29	100%	36	**124%**
Bern - Vechigen	11	100%	12	**109%**	15	100%	18	**120%**
St.Gallen - Appenzell	15	100%	16	**107%**	17	100%	21	**124%**

The following set of maps confirms this difference in minutes from Bern to the rest of locations. In this case we can observe that OSM speed regression network covers a shorter area in more travelled time.

Figure 9 Travel time differences from Bern using NPVM network (top) and OSM maximum speed network (bottom)



4.3 Validation with Google and ViaMichelin

Google Maps and ViaMichelin are the two well used routing machines in the internet (this holds at least for regioConcept). Thus the aim is to compare and validate the distances and time obtained in the two previous networks (OSM without and with speed regression). Table 4 and Table 5 show the results of this comparison, including a) the travel distance and time of the routing machine, b) the percentage of this results compared to OSM with maximum speed, and c) the percentage with speed regression. Percentage values of less the 100% can be interpreted as the OSM networks gives faster results than the routing machine – and vice versa.

E.g. the comparison of Zurich – Chur in Table 4 shows that estimations in OSM_{ms} result in a slightly smaller distance (98% of Google Maps Distance) and 11% less travel time (89% of Google Maps travel time).

Table 4 Distance and time comparison between Google Maps and OSM Networks

	a) Google Maps		b) OSM Max Speed - 2013		c) OSM Speed Reg - 2013	
Trip	**km**	**min**	**% km**	**%min**	**% km**	**%min**
Zurich - Chur	118	74	**98%**	**89%**	**98%**	**134%**
Lausanne - Zurich	224	141	**92%**	**88%**	**92%**	**130%**
St.Gallen - Luzern	143	87	**80%**	**89%**	**80%**	**132%**

Table 5 Distance and time comparison between ViaMichelin and OSM Networks

	a) ViaMichelin		b) OSM Max Speed - 2013		c) OSM Speed Reg - 2013	
Trip	**km**	**min**	**% km**	**%min**	**% km**	**%min**
Zurich - Chur	119	82	**97%**	**80%**	**97%**	**121%**
Lausanne - Zurich	225	164	**91%**	**76%**	**91%**	**112%**
St.Gallen - Luzern	144	97	**79%**	**79%**	**79%**	**119%**

Comparing the two online routing services, we observe that estimated times obtained by ViaMichelin generally are considerably slower than the times delivered by Google Maps. This result is unexpected, as searching day times were the same in both routing services. Indeed, this shows, that estimation of travel time is difficult and uncertain (also with according real data of navigation systems). One reason for this are the countless factors influencing travel times in reality, and another reason certainly are model assumptions and business policies of the according services.

As we can observe, the **distances** tracked in the two OSM networks are shorter than the ones provided by the two online-routing calculators. Some of these differences may are caused by different centre definitions.

Regarding **travel times**, there are considerable differences in the results between the OSM networks. The travel time in OSM_{ms} is around 75% and 90% lower than the online services. However, in the case of OSM with Speed Regression, the travelled time is 110% and 135% higher than the online-routing calculators.

5. Conclusion

The comparison of the estimated travel times show that OSM street network offers a good, freely available alternative to licenced data sets. Additionally, the consistent data structure of OSM offers the possibility to easily switch the network for future case study areas.

Against any expectations, the very simplified assumption of taking just maximum allowed speed as a velocity in the shortest path calculation, lead to quite good results, similar to the distances and time tables provided by the Swiss National Passenger Transport Model NPVM. Of course, using only maximum speed results generally in too fast travel times, but, compared to NPVM only in a range of some percentages.

In contrast, the estimated speeds using the speed regression model yet yield to an underestimation of the speed compared to the other results (NPVM, OSM_{MS}, Google Maps and ViaMichelin). But, after an according calibration process, also very good results are expected – presumably better results than just using maximum speed.

The comparison of the distance and time tables achieved by using the OSM street network with Google Maps and ViaMichelin shows that the results of the two latter applications lie between the OSM_{MS} and the OSM_{SR} values. As experiences of the authors show, in Switzerland Google Maps offers very good travel time estimations. As the OSM_{ms} travel times are not too far away of this values they are assumed to be reasonable. In total, the internal transport model of FaLC already provides reasonable and usable distances and times using the OSM_{MS} implementation. However the calibration of speed values gained by the regression model in a further step will certainly lead to even better results and is therefore highly recommended by the authors.

Calculation speed is a very important issue for FaLC as multiple runs are needed to stabilize the results and reducing white noise due to Monte Carlo Simulations in FaLC. Bodenmann et al. (2014) used the mean values of the results of 50 to 75 simulation runs to get fairly reliable results and they suggest doing even more (100 runs). So, for a 10 year simulation, only one additional minute for each year will cause 17 hours of additional calculation time. This is the reason why each minute for the calculation affects the work of the modellers significantly.

The simplification of the initial OSM network reduces the calculation time by a factor of about 15. This is about 14 min less calculation time per run. Therefore, when repeating the calculation often, simplifying the network is important.

The efficient implementation in FaLC, using optimal data structures and network representations as well as a multithreading approach results in a very fast calculation of the distances and times from every of the 2949 locations in FaLC to every other location. Finally, the whole process of reading, calculating and writing distances and travel times to the database only takes about 6 min: 1 minute for the calculation and 5 minutes for writing the risults in the database. This is the reason, why writing results in the database has to be reduced to the minimum.

The comparison of the distances and travel times received by using the OSM and NPVM network showed also some weaknesses of using the OSM-netwurk as such: e.g. car shuttle trains, like through the Lötschberg or the Vereina tunnel, need to be added to the OSM network in a further step.

As for the public transport no distances and times could be calculated yet, one of the next steps will include the calculation of the distance and time tables using the public transport network.

6. References

Bodenmann, BR, B. Sanchez, J. Bode, A. Zeiler, M. Kuliowsky, P. Furtak, KW Axhausen (2014) FaLC Land-Use and Transport Interaction Model for Switzerland: first results, paper presented at *14th Swiss Transport Research Conference*, Ascona, May 2014.

Bundesamt für Raumentwicklung ARE (2006) *Erstellung des nationalen Personenverkehrsmodell für den öffentlichen und privaten Verkehr – Modellbeschreibung*, UVEK, Bern.

Bundesamt für Raumentwicklung ARE (2010) *Nationales Personenverkehrsmodell des UVEK, Basismodell 2005*, UVEK, Bern.

Bundesamt für Raumentwicklung ARE (2013) *Nationales Personen- und Güterverkehrsmodell des UVEK, Aktualisierung auf den Basiszustand 2010*, UVEK, Bern.

Vrtic, M. und K.W. Axhausen (2002) Modelle der Verkehrsmittelwahl aus RP- und SP Datengrundlage, in FGSV (Hrsg.), Heureka ´02, 293-310, FGSV, Köln.

Haklay, M. (2010) How good is volunteered geographical information? A comparative study on OpenStretMap and Ordnance Survey datasets. *Environment an Planning B: Planning and Design*, **37**, 682-703.

Mondzech, J. und Sester, M. (2011) Quality Analysis of OSM Data Based on Application Needs. *Cartographica*, **46** (2), 115-125.