



The 12th International Conference on Ambient Systems, Networks and Technologies (ANT)
March 23 - 26, 2021, Warsaw, Poland

Address-based computation of intra-cell distances for travel demand models

Matthias Heinrichs^{a,*}, Rita Cyganski^a, Daniel Krajzewicz^a

^aGerman Aerospace Center - Institute of Transport Research, Rudower Chaussee 7, 12489 Berlin, Germany

Abstract

Intra-cell travel is a difficult topic in zone-based travel demand models. In the past, different strategies were applied to approximate the average intra-cell distances by using zone size, diameter, distances to neighbors and other techniques. In this work we examine a new approach which uses address coordinates and a high performance routing program to achieve more accurate results. We apply our method to three different zoning schemes for the city of Berlin, Germany. The result consists of average travel distances for each zone based on all potential destinations. Surprisingly, the ratio of network distance and beeline distance is independent from the three different zoning schemes tested. This detour-factors depend only on the street network. For beeline distances larger than 400 m, we identify a simple transfer function between beeline and travel distance, which extends the classic transfer functions between different zones to intra-zone distances. We show that many distances below 100 m belong to a second class of intra-cell routes, which corresponds to the routing in a Manhattan-block grid. These two classes describe the intra-cell distances in our example quite well and we expect this approach to be helpful to examine other areas.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the Conference Program Chairs.

Keywords: demand modelling ; intra-zone travel ; travel times ; zoning ; accessibility

1. Introduction

The use of zone-based accessibility and impedance indicators is common practice in the analysis of spatial relationships, accessibility and demand modelling. Despite ongoing efforts to increase the spatial resolution in demand modelling (see e.g. [8], [11]), the usage of zone-based indicators widely applies not only to aggregated demand models – often referred to as four-step-models [17]. Albeit disaggregate, microscopic demand models often provide coordinate-based input data on the population or destinations, the usage of a zoning system and corresponding travel time matrices remains to be the standard procedure due to the high computational burdens associated with address-specific routing. This comes, however, at a price as well. Determining travel times or distances between two zones

* Corresponding author. Tel.: +49-30-67055-174

E-mail address: matthias.heinrichs@dlr.de

on basis of a single routing result – usually between the geographical centers of the two zones – is blind to specific local situations within those zones. Irregular grids, natural barriers such as rivers or peninsulas and their man-made counterparts such as train tracks or one-way streets are examples of situations that might result in substantial deviations of the averaged routing results from the “real” travel time. This is especially relevant for shorter distances and thus adjacent zones. Additionally, travel times and distances for trips starting and ending within the same zone have to be provided. Especially when modelling daily trips conducted by foot or bike, a substantial share of them have a length that makes them prone to be intra-cell trips. For instance, more than 60% of the trips reported in the German national household survey on mobility called “*Mobilität in Deutschland 2017*” have a trip length of 1000 m or below [18]. Consequently, the quality of modelling results is heavily depending on the provision of the adequate distance and travel-time information for intra-cell trips. In this paper we present a method to compute all possible intra-cell routes for a large-scale area and analyze the effects of different geographic aggregations. We show that in our example the intra-cell traffic between two coordinates can be approximated by a simple transfer function, which is independent of the cell-size and shape. We show that if explicit routing takes too long or is unavailable the trip length of short trips can be estimated well even when starting with the beeline distance only.

2. State of the art

The generation of travel times matrices is a standard task when it comes to providing the input data for travel analysis and demand modelling. Accordingly, several procedures for the calculation of the impedance measures do exist (see e.g. [16], [12],[4]), and are sometimes even integrated in matrix manipulation tools being part of demand modelling software suites, e.g. PTV Visum. [12] provides a recent overview of established as well as seemingly more ad hoc approaches for the calculation of intra-cell impedance values. Depending on the main idea, they might roughly be split into three groups:

1. calculations on the basis of the area or perimeter of the zone
2. calculations based on the distance to a defined number of neighboring zones
3. calculations of the average crow-fly or network-based distance retrieved for a small random sample of coordinate-pairs within the zone

The area-based approaches can further be distinguished according to the geometrical figure used for the calculations. Common practice is the transformation of the zone area into a circle, where the distance is calculated as a function of the circle radius (see e.g. [9]). Alternatively, minimal bounding boxes or squares can be used, and the distance is then calculated as a function of the side length. Generally, as [10] emphasize, the derivation of intra-cell travel times based on Euclidean distances and circle deduction may be appropriate for very disaggregate zoning schemes but widely ignores intra-cell variations.

Variation can be taken into account to some degree when considering the travel times to the n -nearest neighbors instead – an approach rather common in transport modelling. As [16] note, it is often assumed in such cases that the average travel time within a zone ought be roughly half of the average time needed to visit the centroids of the n neighboring zones. The deviation between this approximation and the actual travel times within a zone using this calculation method is prone to be dependent on the number of considered neighbor cells and the topological situation of the centroids.

Sampling size and specifics are also highly relevant for the quality of results obtained with the third type of calculation approaches. Indeed, when comparing the travel times calculated for specific examples using a variety of specific implementations of the above outlined three general approaches, [12] found strong dependencies between size, layout and network topology on one side and the quality of calculation results on the other – with the special challenges varying with the approach. Interestingly and contrary to [10], the author concludes based on his case studies that the use of an equal-area-circle approach is preferable to the calculation of impedance measures using the neighboring cells, which is well established in transport modelling.

Overall, it remains to be concluded that the predominant approaches for the generation of intra-cell travel information appear to be rather coarse, variable in the quality of results and dependent on the size of the individual zones in



Fig. 1: Three different spatial zoning schemes of Berlin, Germany

the zoning scheme. They therefore might generally be considered insufficient especially for the high spatial resolution that is becoming more and more standard especially in microscopic demand modelling.

At the same time, computational power has been increasing steadily, making a computation of such matrices at the level of single origins and destinations possible in reasonable time. Also, fine-grained, even mode-specific road networks as well as GPS-based positions of origins and destinations that can be used within a high-level resolution routing became available. For instance, many of the required information for an address-based calculation of trip-distance matrices can be obtained from the OpenStreetMap (OSM) project [19]. OSM also includes information about buildings. While OSM data is easily obtainable, resolution, reliability and time reference of the data can vary between areas of interest (see e.g. [1]). Alternatively, information about buildings is often available from federal agencies, such as the Federal Agency for Cartography and Geodesy of Germany (Bundesamt für Kartographie und Geodesie – BKG). Increasingly, information about public transport lines and schedules is available in the GTFS format [7].

Given these prerequisites as well as tools supporting such a fine-grained computation, for instance the tool UrMoAC [14] described in the next section, it seems meaningful to consider the computation of all distances at address rather than at aggregate level or via sampling.

3. Proposed method

The proposed method is quite simple: First, we collect all possible destination coordinates within one zone. Then, we compute routes and measure their length for every pair in each zone. Finally, we take the median of all travel times and distances to represent the intra-cell values. In the following we explain each step in detail.

In the following, we use the city of Berlin as demonstration area. We use three systems for spatial zoning: a) the official 1223 traffic analysis zones (TAZ) of Berlin [20], which use roads as cell-borders, b) the official 195 statistical areas, which are an aggregation of the 1223 TAZ to meet a minimal population and maximal area, and c) the 1 km INSPIRE grid of Europe [6], which is a pure geometrical subdivision (See Fig. 1).

To locate all possible destinations, we use the coordinates of all postal addresses in Berlin from the Federal Agency for Cartography and Geodesy of Germany [3]. Furthermore, we added coordinates of parks, forests and other destinations without a postal address, which were calculated mainly using the German Digital Landscape Model (ATKIS Base-DLM) available from the BKG [2]. Next, we assigned all destination coordinates to the zone that surrounds them. Usually, there is more than one destination per zone.

The subsequent computations used the open source “Urban Mobility Accessibility Computer” (UrMoAC) tool. The tool is a command line application designed to compute a large variety of accessibility measures, including travel time and distance matrices, on a most fine-grained level. As such, UrMoAC reads a road network, e.g. imported from OSM, as well as a list of origins and a list of destinations, both given on the level of single buildings or addresses. For each of those combinations of an origin and a destination, the direct access to the next road, usable by the regarded mode of transport, is computed and included later in the generated measures.

Table 1: Size comparison of the different zoning in km²

	TAZ	Stat. Area	Grid
min	0.01	0.394	1
max	14.535	25,81	1
average	0.728	4.5674	1
median	0.452	3.056	1
number	1223	195	1016

Using this tool, we computed the distance and duration for walking between every pair of coordinates and additionally the beeline distance for each pair. As an average walking speed, we propose 1 m/s (3.6 km/h), which represents a normal walking speed reduced by 10% for crossing streets.

The measures are computed using a plain Dijkstra routing algorithm [5], yet with some extensions that increase the resolution of the outputs. Besides considering the position of the sources/destinations at the road they were assigned to, UrMoAC e.g. also considers finding a destination on the “opposite side of the road” – instead of routing over the next intersection. The tool distinguishes between walking, bicycling, using a car, as well as combinations of these modes with public transport. It supports different limits, such as a maximum travel time, a maximum distance, the number of potential destinations or routing to the nearest destination as well as a large variety of output options. UrMoAC is available as open source and can be easily adopted for specific areas and research questions [13]. For routing, we use the OSM road network dated to the 18th of October 2019.

This procedure can be extended to other modes. Frankly, this causes a lot of additional considerations, such as searching for a parking lot for cars, locations to lock your bike and assignment of public transport stops for each TAZ. In this paper, we focus on walking distances, because walking dominates the intra-cell traffic in our example of Berlin, as the TAZ we use are rather small (see Table 1).

After computing the routes between each pair of destinations, there are several possibilities to use the so obtained data. First, we computed the mean values of the travel times and the distances of all connections within one zone. If there is only one or there is no destination in a zone, we prohibit intra-cell traffic. The intra-cell travel information for all connections can be averaged and directly used as intra-cell information in travel simulations. Second, the median of the detour factors and its standard deviation can be computed for each zone. Using this detour factor, we can estimate the walking distance from the beeline distance between two destinations without the need for computing the routes again. This speeds up the travel demand simulation as the time consuming route computation between thousands of places can be avoided.

4. Results

The computed intra-cell travel distances for the TAZ zoning are shown in Figure 2a. One sees that the intra-cell distances rise with the size of the area of the zone, which was expected and corresponds to the findings of previous work [12]. The distribution of the computed intra-cell distances are shown in figure 3. This plot shows that most of the intra-cell distances are below 1000 m and some very large zones have a median up to 3500 m. These zones are the huge zones in the peripheral areas of Berlin dominated by forests and lakes.

The computed intra-cell travel distances and times are difficult to evaluate. If the collected destinations are complete, we have all possible routes. But only a few connections are used in reality and there is no ground truth available. Household travel-surveys as [18] do not provide valid information for short trips at this geographic resolution. Therefore, we decided to examine the stability of this approach over different spatial zoning systems, to evaluate whether the zoning system affects the length distribution of routed trips.

To test the proposed method, we applied it to the three different spatial zoning schemes for the city of Berlin named before. First, we computed the lengths and durations by routing and additionally the beeline distance for every pair of addresses within one zone. Next, we computed the detour factors for every connection:

$$detour = \frac{distance_{routing}}{distance_{beeline}} \quad (1)$$

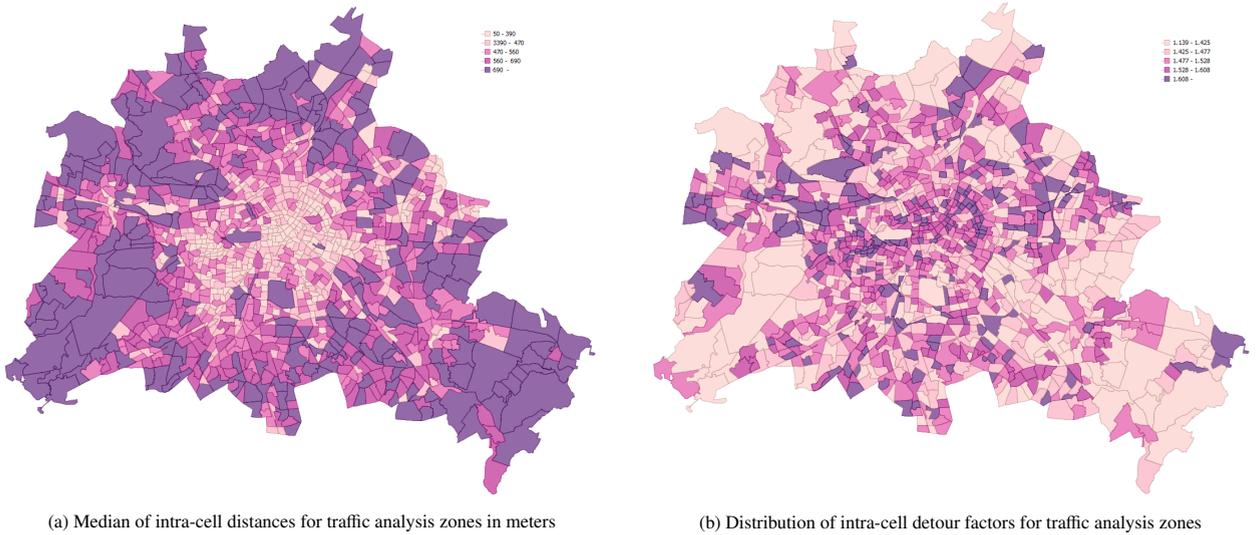


Fig. 2: Results for the city of Berlin

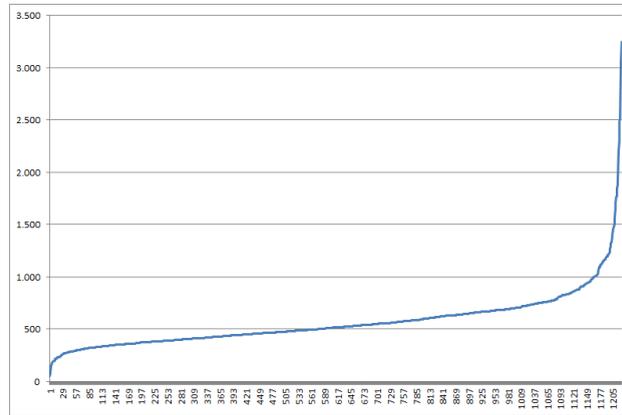


Fig. 3: Distribution of intra cell distances for traffic analysis zones in meters

The results of the TAZ-based detour factors are shown in fig. 2b. The map shows a scattered plot of different detour factors. However, the lowest detour factors are computed for the peripheral zones with a high share of forests and water areas in the south west and south east as well as in the northern parts of Berlin.

To analyze the results in more detail, we computed a histogram of detour factors for every beeline distance. The bin-size was 0.01 for the detour factors and 1 m for the beeline distances. The histogram was plotted as a heat map. In addition, the red lines show the median of detour factors and for every 100 m, a box-plot is integrated. The boxes show the range of one quartile of the data above and below the median. The whiskers represent the data which are considered as inliers. Additionally, the green line shows a synthetic function based on the recommendations from [15] (see equ. 2). The parameters for this function are the same for all different zoning schemes. Yet, being derived empirically, it should be assumed that they will differ between different regions.

$$detour(beeline) = \frac{75m^{0.65}}{beeline^{0.65}} + 1.12 \tag{2}$$

The maximum beeline distance for the 1000 m grid is naturally limited to $\sqrt{2km^2} \approx 1414m$. The other grids have intra-cell distances up to 5000m. However, distances over 1500m are very seldom for the TAZ-grid (see fig. 3).

To compare all three histograms obtained using the named zoning systems, we limit the maximum beeline of the histograms to 1500 m. The results can be seen in Fig. 4a, Fig. 4b and Fig. 4c.

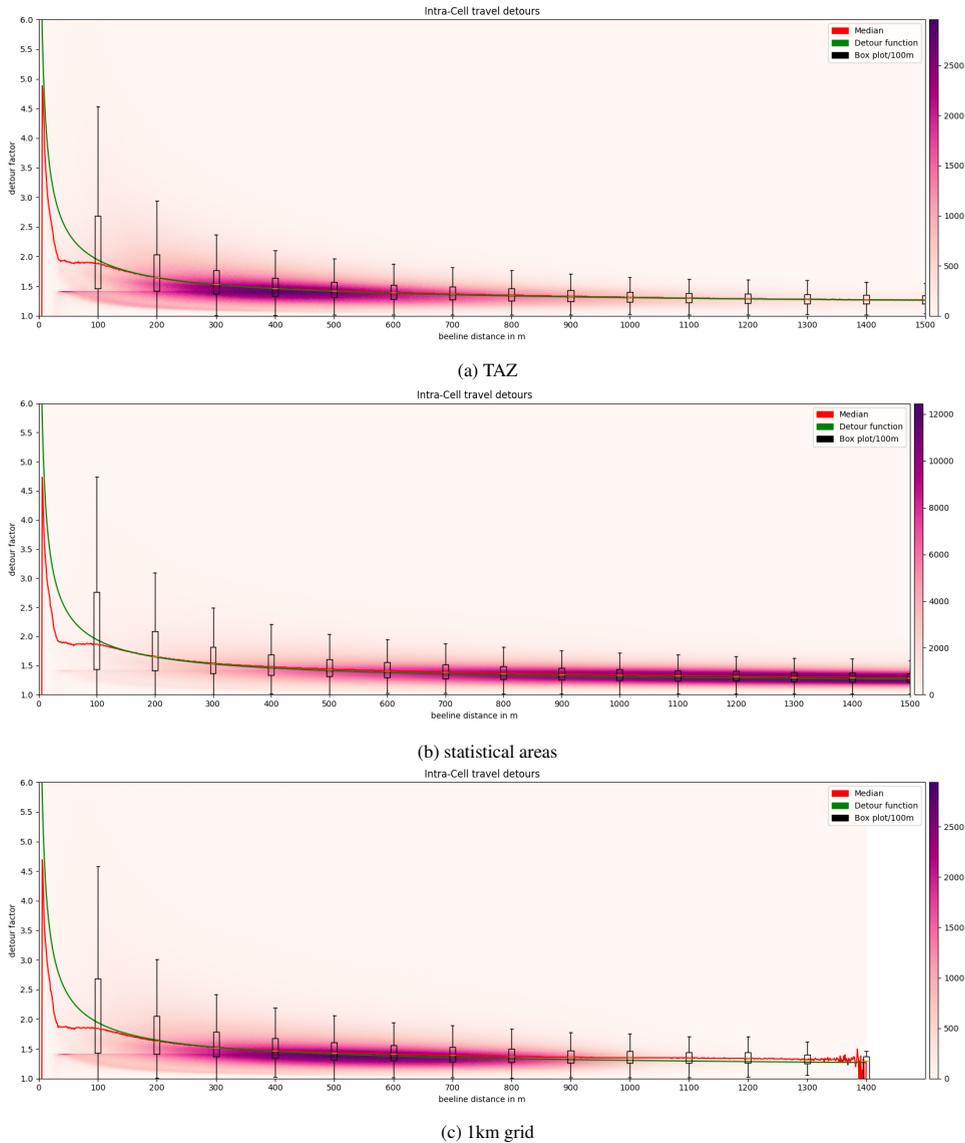


Fig. 4: Detour-factors for different zoning schemes

Comparing the results, one can see that the median line has nearly the same behavior for all three different zoning schemes: For distances between 1 m and 25 m, the median falls very steep from 4.7 to 1.9 and remains constant until a beeline distance of 100 m. From there, the median runs asymptotic to a value of 1.12. The heat-map basically shows where the majority of connections is located. As expected, the heat-map areas with the highest counts move to longer beeline distances, when the size of the zoning increases. However, there is a prominent line at $\sqrt{2}km^2 \approx 1414m$ in all three figures. This artefact derives from city parts, which fulfill the classic street layout of regular Manhattan blocks. This explains why the median is more or less constant until the dominant block size of 100 m. For beeline distances longer than 100 m, the detour factors are very close to the synthetic function given by equ. 2. Below this distance the detour factors are highly variable.

One may wonder about detour factors above 2, a value to be found when going "around the block". The reason herefor is the usage of the origin/destinations' centroids as starting points and including the access/egress to/from the respective building to the next road in the results produced by UrMoAC, visualised in Fig. 5. This behaviour can be disabled, yet it is to be discussed whether including these access/egress times yields in more exact results or not.

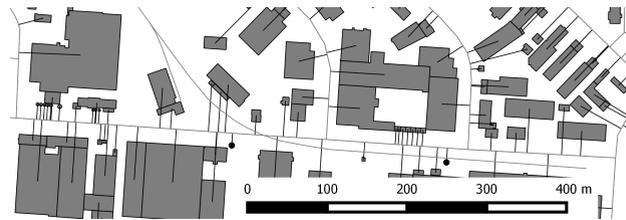


Fig. 5: Visualisation of the access/egress paths (black) between buildings (grey areas) and bus stops (black circles) and the respectively next roads (grey).

5. Conclusions

In this work, we show that it is possible to directly calculate the intra-cell travel times for all possible connections. Furthermore, we find that a simple transfer function can estimate the relation of beeline distances to network distances, even on an intra-cell level.

As illustrated, the presented approach allows for the calculation of travel impedance information for usage in accessibility analysis and travel demand modelling on basis of data at a high spatial resolution. It is therefore capable of accounting for local specifics in the street network topology such as natural or man-made barriers.

The results show stable behavior of the relation between the detour factor and the beeline distance across very different spatial zoning schemes for our example region. This shows that the intra-cell traffic distances can be computed independently from the selected zoning just by using a transfer function for a specific research area.

However, the results indicate that there are two prominent routing schemes in a city: One corresponds to the classic Manhattan-block distance, the second to a more complex routing in areas with curved roads, due to natural borders or man-made, often historic, barriers. To distinguish between these two schemes, we could try to classify a zone, based on the knowledge whether its streets are forming a grid or not.

There are, of course, various extensions and improvements that we are planning to address in future work. First of all, the work presented here focuses on the calculation of travel times for walking trips, because this is the main mode in urban intra cell traffic. An extension to other modes, especially public transport, is planned.

Second, the computation is based on unweighted links. Since locations have different importance, e.g. levels above ground, households per building or cinema seats, we plan to extend our approach on weighted links.

Furthermore, it is planned to further verify the spatial transferability of the detour factors derived with the approach. The results presented here aimed at testing the robustness of the approach and independence of the results from the zoning sizes used. Extensive testing of the influence of topology differences on the quality of the results — particularly the derived detour factors — is yet to come.

References

- [1] Ali, A.L., Schmid, F., 2014. Data quality assurance for volunteered geographic information, in: *International Conference on Geographic Information Science*, Springer. pp. 126–141.
- [2] Bundesamt für Kartographie und Geodäsie, 2020a. Digitale landschaftsmodelle und landbedeckungsmodelle von deutschland. URL: <https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/digitale-landschaftsmodelle.html>.
- [3] Bundesamt für Kartographie und Geodäsie, 2020b. Georeferenzierte adressdaten. URL: <https://gdz.bkg.bund.de/index.php/default/georeferenzierte-adressdaten-ga.html>.
- [4] BURGSTALLER, B., PILLICHSHAMMER, F., 2009. The average distance between two points. *Bulletin of the Australian Mathematical Society* 80, 353–359. doi:10.1017/S0004972709000707.
- [5] Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numerische Mathematik* 1, 269–271. doi:10.1007/BF01386390.

- [6] for spatial Information in Europe, E.C.I., 2020. Inspire geographic grid of europe. <https://inspire.ec.europa.eu/theme/gg>.
- [7] Google, 2020. Gtfs format specification. <https://developers.google.com/transit/gtfs>.
- [8] Griesenbeck, B., Garry, G., Hossack, S., Porter, R., Penmetsa, R., 2009. Small is beautiful: Why you should get rid of zones and start using parcels in your travel demand model, in: Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington, D.C., USA. p. 20.
- [9] Guth, D., Holz-Rau, C., Maciolek, M., Scheiner, J., 2010. Beschäftigungssuburbanisierung, siedlungsstruktur und berufspendelverkehr: Ergebnisse für deutsche agglomerationsräume 1999-2007. *Raumforschung und Raumordnung* 68, 283–295.
- [10] Horner, M.W., Murray, A.T., 2002. Excess commuting and the modifiable areal unit problem. *Urban Studies* 39, 131–139. doi:10.1080/00420980220099113.
- [11] Horni, A., Nagel, K., Axhausen, K.W., 2011. High-resolution destination choice in agent-based demand models. *Arbeitsberichte Verkehrs- und Raumplanung* 682. Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau, Eidgenössische Technische Hochschule Zürich.
- [12] Klein, T., 2020. Mittlere binnendistanz räumlicher aggregate: Gegenüberstellung unterschiedlicher ansätze zur schätzung der mittleren binnendistanz von räumlichen aggregaten am beispiel deutscher verwaltungsgebiete. *AGIT – Journal für Angewandte Geoinformatik* , 330–346URL: <http://doi.acm.org/10.1145/1219092.1219093>, doi:10.14627/537698032.
- [13] Krajzewicz, D., 2020. Urmoac github page. <https://github.com/DLR-VF/UrMoAC>.
- [14] Krajzewicz, D., Heinrichs, D., Cyganski, R., 2017. Intermodal contour accessibility measures computation using the 'urmo accessibility computer'. *International Journal On Advances in Systems and Measurements* 10, 111–123. URL: <http://www.thinkmind.org/index.php?view=instance&instance=SysMea+v10+n34+2017>.
- [15] Lohse, D., Schnabel, W., V, D., 2011. Grundlagen der Straßenverkehrstechnik und der Verkehrsplanung: Band 2 - Verkehrsplanung. Beuth Studium, Beuth Verlag GmbH. URL: <https://books.google.de/books?id=I2f-bSsW04MC>.
- [16] Martin, W.A., McGuckin, N.A., 1998. Travel Estimation Techniques for Urban Planning. NCHRP Report 365. National Research Council, Washington, DC.
- [17] McNally, M.G., Rindt, C., 2000. The activity-based approach, in: Hensher, D.A., Button, K. (Eds.), *Handbook of Transport modelling*. Pergamon Press, pp. 53–69.
- [18] Nobis, C., Kuhnimhof, T., 2018. Mobilität in Deutschland – MiD: Ergebnisbericht. Technical Report. Institute of Transport Research. URL: <https://elib.dlr.de/125879/>.
- [19] project, O., 2020. Openstreetmap project. <https://www.openstreetmap.org/>.
- [20] Senatsverwaltung für Stadtentwicklung Berlin, o.J. Teilverkehrszellen und verkehrszellen: Informationen. URL: <https://www.berlin.de/sen/uvk/verkehr/verkehrsdaten/teilverkehrszellen-und-verkehrszellen/>.