THE END OF TRAVEL TIME MATRICES? OR: WHY WE SHOULD USE INDIVIDUAL TRAVEL TIMES

Nico Kuehnel (corresponding author)
Department of Civil, Geo and Environmental Engineering
Technical University of Munich
Arcisstr. 21, 80333 Munich
Germany
ORCID: 0000-0003-0527-8653
nico.kuehnel@tum.de

Dominik Ziemke
Transport Systems Planning and Transport Telematics
Technische Universität Berlin
Sekr. SG12, Salzufer 17-19, 10587 Berlin
Germany
ziemke@vsp.tu-berlin.de

Rolf Moeckel
Department of Civil, Geo and Environmental Engineering
Technical University of Munich
Arcisstr. 21, 80333 Munich
Germany
ORCID: 0000-0002-6874-0393
rolf.moeckel@tum.de

Kai Nagel
Transport Systems Planning and Transport Telematics
Technische Universität Berlin
Sekr. SG12, Salzufer 17-19, 10587 Berlin
Germany
nagel@vsp.tu-berlin.de

Submission date: August 1, 2019
6,917 words (text) + 2 tables (500 words) = 7417 words

Abstract

To reduce inaccuracies due to insufficient spatial resolution of models, it has been suggested to use smaller raster cells instead of larger zones. Increasing the number of zones, however, increases the matrix size of skim tables. Those become difficult to create, to store and to read, while most of the origin-destination pairs are calculated and stored but never used. At the same time, such approaches do not solve inaccuracies due to lack of temporal resolution. This paper proposes to store and process travel times at the finest spatial resolution possible (at x/y coordinates) and a highly detailed temporal resolution. The approach is tested in the context of an integrated land use/transport model (ILUT) where travel times affect, among others, household relocation decisions. In this paper, person-level individual travel times are compared against traditional skim-based travel times. It was shown that skim-based travel times fail to capture the spatial and temporal variations of travel times on a microscopic scale. While skims provide acceptable averages in the case of car travel times if a dense network and small zones are used, transit travel times are heavily affected by temporal and spatial aggregation. When looking at travel-time-dependent relocation decisions in the land use model, transit captive households tend to react more sensitively to the level of service in transit when individual travel times are used. The results suggest that individual travel times can improve the spatial and temporal accuracy of models.

Keywords: integrated land use/transport models, microsimulation, agent-based models, travel time matrices

INTRODUCTION

In 2000, Spiekermann & Wegener (*1*) published an article with the title "Freedom from the tyranny of zones." The idea was to use small raster cells instead of zones to reduce the error caused by the lack of spatial resolution. For matrix-based travel time skims, however, raster cells proved to be impractical. The matrix grows by a factor of n^2, where n is the number of zones. In large zone systems, the matrix becomes difficult to create, to store and to read, while most of the origin-destination pairs are calculated and stored but never used. In addition, every travel time matrix is created for one point of time during the day, which may not represent well travel times for another time of the day.

This paper proposes a new method to store and process travel times that allows the finest spatial resolution possible (at x/y coordinates) and a highly detailed temporal resolution (here applied in 15-minute time bins, but it could be used in smaller time steps likewise). As an example to present the relevance of this high-fidelity representation of travel time, an integrated land use/transport model is used. Results suggest that travel demand models would equally benefit from this microscopic representation of individual travel times.

In traditional integrated land use/transport models, the transport model provides zone-to-zone travel times in form of skim matrices. Those affect accessibilities, and thereby, household relocation decisions. Usually, there is one travel time for each zone-to-zone-relation, which is aggregated in time (e.g. one travel time value for the peak hour) and space (e.g. one centroid per zone). In reality, however, a worker who commutes at 5:00 am experiences a different level of congestion and different mode options than a commuter traveling at 8:00 am. The impact of travel options on accessibility becomes even more complex for households with multiple workers. The temporal aggregation ignores that travel times may vary substantially during the day, which is of interest when trips are not made during peak hours. Especially for transit travel times, the time of day plays an important role due to service hours and frequencies.

The spatial aggregation to zones may also affect results, particularly in larger zones. This is also known as the modifiable areal unit problem (MAUP) first described by Openshaw (*2*). While the problem of spatial biases is well known, “the effects of spatial biases on LUTI models remain largely unexplored and underestimated” (*3*). For transit travel times the distance to the next stop is important when accounting for access and egress times. In Addition, the next stop might be close, but the lines serving this stop will not connect every other zone equally well. A transit stop further away may serve better for a given trip.

Existing aggregate approaches are unable to account for individual travel experiences. This project will overcome this gap by linking land use and transport time microscopically.

In this paper, a new microscopic integrated land use/transport model is used to compare person-individual travel times against traditional skim-based travel times for the feedback from a transport to a land use model. The individual travel times use x/y coordinates and represent traffic conditions at specific times of day. The skim-based approach uses peak-hour travel times and centroid connectors. The goal of this paper is to identify the benefits of individual travel time queries in comparisons to more traditional skim-based travel times.

LITERATURE REVIEW

Disaggregated microscopic models help capturing heterogeneities in travel behavior and household relocation (*4, 5*). High spatial resolutions support the representation of environmental issues(*6*). On the other hand, Wegener (*4*) pointed out that many disaggregate transport models were too slow to be executed multiple times in integrated land-use transport models. Another issue of microsimulations are stochastic variations between model runs that prevent reproducing results. Wegener (*4*) concludes that " 'the more micro the better' may be misleading." The computing time of full-scale microsimulation models can easily exceed days or weeks. Adding too much complexity to simulation model is one of the sins Lee also describes in his "Requiem for Large-Scale Models" (*7*). One should pay attention to not increase complexity of models too much and keep models - as Einstein is said to have said - as simple as possible but no simpler.

Nevertheless, there is a continued interest in increasing the spatial and temporal resolution in models (*8*). Policies that test local impacts (such as transit-oriented development) or time-specific impacts (such as dynamic tolling) require more detailed representations of space and time. To strive for the right level of detail remains a challenge for many transport and land use modelers (*9*).

In a previous study, the land use model SILO was coupled with the transport simulation MATSim (*10*). Here, an agent-based transport model successfully replaced an aggregated transport model for the Maryland region. MATSim proved to reproduce zone-to-zone skim matrices sufficiently well by averaging travel times from sampled coordinates in each zone. This first coupling, however, did not yet include the feedback from the transport model to the land use model. However, it was proposed to implement a query architecture that allows agents in the land use model to query individual travel times from the transport model, such that agents who look for a new job or dwelling can query for travel times between micro-locations at specific times of day. The proposed query architecture has become operational in the meantime and provides the foundation for the research of this paper.

In a review of existing integrated land use/transport studies, Badoe and Miller (*11*) identified several studies that "have worked with zonal-aggregate variables for gross spatial units [...] thus clouding the effects [...]".

The level of zonal aggregation largely affects simulation outcomes. The modifiable areal unit problem (MAUP) which was described by Openshaw (*2*) states that results of spatial analyses are influenced by the chosen zone size. Typically, there is a tradeoff to make between few large zones with coarse resolution and many intrazonal trips and many small zones with a finer granularity but much higher computing times. The MAUP affects the true representation of travel times (*12*). For transport models, zones should be "larger where there is less activity and smaller where there is more activity" (*13*). Previous studies confirmed that smaller zone sizes improve the fit of the model to observed data (*14*). Another study identified that the level of detail should be high for travel time queries to nearby zones and can be lower for more distant zones (*15*). Special attention should be paid to intrazonal travel times. Some of the microscopic simulation frameworks like MATSim do not use zones at all but only work with network graphs and x/y coordinates. The implementation of individual travel times in an integrated land use/transport model is expected to reduce the impact of the chosen zone system while increasing computation times.

THE FABILUT MODELING SUITE

The FABILUT (flexible, agent-based integrated land use/transport) modeling suite consists of the land use model SILO (Simple, Integrated Land-use Orchestrator) (*16*) and the transport simulation model MATSim (Multi-Agent Transport Simulation) (*17*). For travel demand generation, MITO (Microscopic Transportation Orchestrator) (*18*) is used in this study. All three models are open source and written in Java, which allows for a tight integration. For studies with no travel demand model available, the FABILUT modeling suite can also be run with SILO and MATSim only, which e.g. allows to simulate the commute segment of traffic (*10*).

On a year-by-year basis, SILO models demographic events (e.g. birth, marriage, death, etc.), household relocation and real-estate updates, such as construction of new dwellings, renovation, price updates, etc. SILO belongs to the class of land use models that incrementally update an existing synthetic population. For any year selected, MITO and MATSim are run to create travel demand and to simulate traffic on the network. By this process, simulation-based, link- and time-specific travel times are created.

Currently, travel times are used for four reasons in SILO:

- Accessibility calculation: peak hour skims are used to calculate potential accessibilities by car and transit.
- Current **housing satisfaction** of residents: commuting times for all workers in a household are used to assess how satisfied the household is with its current dwelling location.
- Household relocation: a household will **evaluate** multiple **vacant dwellings** by taking into account commuting times of all workers of this household.
- Job search: a person will evaluate vacant jobs based on the expected commute time to the respective region.

Traditionally, a transport model would provide skim matrices with zone-to-zone travel times for a given time of day (sometimes distinguishing peak and off-peak travel times). Such skim matrices aggregate spatially by providing travel times from zone centroid to zone centroid and temporally by providing a limited number of times per day (commonly only one time, such as morning peak). In this research, we explore for the first time the use of individual travel times. We call these travel times individual because

1. they reflect travel times from a micro location to a micro location in x/y coordinates. The size of zones becomes irrelevant, as all locations are stored in x/y coordinates in SILO
2. they reflect travel times for a specific time of day. Someone traveling to work at 5:00 AM in the morning will see different travel times than someone traveling to work at 9:00 AM. Also, the availability of travel modes will differ by time of day.

In the FABILUT modeling suite, MATSim is used to simulate traffic. In MATSim, each person is resolved individually as an agent and has one or more plans. A plan is a chain of activities (e.g. home–work–shop–home), including locations and activity end times. Activities at different locations are connected by trips. MATSim is based on a co-evolutionary algorithm which iterates over the three steps *traffic simulation*, *scoring*, and *replanning* (*17*). In *traffic simulation* (also mobility simulation or mobsim), travel demand is simulated on the physical network. The selected plans of all agents are executed simultaneously in second-by-second steps. The default physical simulation is a queue model, in which every link is modeled as a first-in-first-out (FIFO) queue. This computationally efficient design makes MATSim suitable to simulate large metropolitan regions. A common approach to reduce computing times, is using sampled scenarios where only a sample of the full population of agents is simulated in the transport supply system whose properties are scaled-down correspondingly (*19*).

After traffic simulation, agents score (*scoring*) their plan, partially in reaction to their individual simulated travel experience based on the notion of utility maximization. In final step of every iteration, agents have the chance modify their daily plan (*replanning*) with regard to different choice dimensions (route choice, mode choice, departure time choice etc.). If a new plan is created, it is simulated in the traffic simulation of the next iteration. If not, agents select a plan from their existing plan choice set according to a probability distribution that converges to a multinomial logit model.

In the current setup, MITO is used to model travel demand. MITO is a microscopic transport demand model that creates home-based tours and non-home-based trips of the population residing in the study area including mode and departure time choice. MATSim is used to simulate the tours/trips created by MITO, i.e. sub-segments of full day plans. As replanning strategy, only route choice is enabled, such that the application of MATSim in this study resembles that of a pure dynamic traffic assignment tool. Based on the MATSim transport simulation, travel times that are stored per link in 15-minute-specific time bins are created. This allows for spatially and temporally highly resolved travel time queries by the SILO land-use model.

Under the same technical interface, SILO also queries travel times by public transport from MATSim, which are provided based on the region's public transport schedule. While possible in the MATSim transport simulation, public transport does not have to be explicitly simulated in the context of this study as transit travel times can be requested between micro locations in x/y coordinates for any given point of time during the day based on the transit schedule. The recent implementation of the raptor transit router (*20*), which has significantly reduced the computation times for transit routing, facilitates this task.

In this research, we implemented both skim-based travel times and individual travel times. This allows us to test both approaches and explore the differences between querying skim-based versus individual travel times.

Household Relocation

The representation of travel times is particularly relevant for the household relocation module of SILO. Household relocation can be triggered when households are unsatisfied with their current dwelling, when couples marry/divorce or for children who leave their parental household. Inmigrating households use the same relocation decision rules as well. Relocation is modeled as a two step discrete choice. First, a household evaluates all regions of the study area. Regions are sets of zones, usually grouping them to a higher administrative level (e.g., county). The evaluation of regions takes into account the number of vacant dwellings, racial or nationality shares, region-wide average rent prices and the commute time between the region and employment zones of working household members. For the region evaluation, an approximate travel time is sufficient given the large areal size of regions. The selection of a region uses a multinomial logit choice model in which the probability of choosing a region depends on the utility of a region in comparison of the utilities of all other region alternatives:

$$p(r) = \frac{e^{\beta \times u_r}}{\sum_i e^{\beta \times u_i}} \tag{1}$$

where u_r is the utility of option r and u_i are utilities of all choice alternatives. Once a region has been chosen, a sample of 20 randomly drawn vacant dwellings inside this region is chosen. The household evaluates all choice alternatives and selects a dwelling using equation 1, where r stands for a dwelling. The utility of a dwelling accounts for the size, quality and price of the dwelling and accessibility of the zone where the dwelling is located. For households with workers, the expected commute times from this new dwelling for each worker are included in the evaluation to ensure that a household attempts to find a location within an acceptable commute time for all workers in this household.

The locations in x/y coordinates of the vacant dwellings and job locations are known. Also, the model dataset provides a job start time for every worker in the household. It is hypothesized that individual travel times could improve evaluation of dwellings over the use of skim-based travel times. Both travel time to work by auto and by transit are considered in the evaluation of the travel time to work. The utility component for commuting times for dwelling i is defined as

$$u_{commute,i} = \prod_j e^{-\lambda * tt_{i,j}} \tag{2}$$

where $tt_{i,j}$ is the commute time to work place j from dwelling i. An exponentially decreasing function represents the probability of commuting for the given amount of time. $tt_{i,j}$ is defined as a composite travel time consisting of car and transit travel times, depending on the ratio of cars and workers in the household:

$$tt_{i,j} = \tau \times tt_{i,j,car} + (1 - \tau) \times tt_{i,j,transit} \tag{3}$$

where $\tau = \frac{cars}{workers}$ is the ratio of cars to workers (capped at 0 and 1) and $tt_{i,j,car}$ and $tt_{i,j,transit}$ are car and transit travel times to workplace j from dwelling i. This definition will make households with cars less sensitive to transit travel time while households without cars are considered to be transit captives that rely on transit travel times.

Query Architecture for Individual Travel Times

The implemented query architecture allows agents to query for *expected* individual travel times from and to micro-locations in the form of x/y coordinates at a specific time of day. Agents do not draw on upon *experienced* travel times during the transport simulation as this would require to simulate every individual agent during the transport model simulation and ignore the advantages of MATSim's scaling capabilities, which would unnecessarily increase runtime.

Whenever SILO requires travel times, MATSim's trip router is queried. The router accounts for

actual traffic conditions for auto travel times and for the actual schedule of transit. The router also includes access and egress times as well as transfer times for public transport queries. For car travel time queries, it is assumed that the car is parked very close to origin and destination, resulting into access and egress times that can be neglected.
The query architecture does not compute travel times preemptively as it is done for skim matrices. Rather, it returns individual travel times as they are needed.

STUDY AREA

The zone system for the Munich study area, for which a skim matrix was generated, was developed with an automated zone system generator that creates smaller raster cells in densely populated areas and larger raster cells in rural areas, while respecting administrative boundaries (7). The synthetic population for this study area (*21*) includes household and job locations and was created using iterative proportional updating (*22*).

SKIM MATRICES FOR COMPARISON

The skims are calculated for auto and transit travel times by routing between weighted zone centroids of each zone at a defined and fixed peak hour (once for the morning and once for the afternoon peak). For $4,924$ zones, each skim matrix has $24,245,776$ travel time values, of which many entries are never used. Zone centroids are obtained by geographically averaging the micro-coordinates of dwellings, weighted by their resident's household size. For intrazonal travel times, we consider Z as the set of zones that include the n closest neighbors in terms of travel times. The intrazonal travel time $tt_{i,i}$ of zone i is defined as a given share λ of the average travel time to these closest neighbors:

$$tt_{i,i} = \lambda * \frac{\sum_{j \in Z} t_{i,j}}{n} \tag{4}$$

where $t_{i,j}$ is the travel time from zone i to j and λ is a configurable parameter. By trial-and-error, reasonable estimates are obtained by setting n to 5 and λ to 0.66. In other words, the intrazonal travel time is set to two thirds of the average distance to the next five zones. For individual travel times, all queries ask for explicit origin and destination x/y coordinates, no intrazonal travel times need to be calculated.

For transit travel times skims, all stops in a 1,000 meter radius around the weighted centroid of the origin are routed to all stops in the same radius around the centroid of the destination zone. In cases where no stops are found within the 1,000 meter radius, the (single) closest stop to the centroid at any distance is selected. The most optimistic route is then selected and access/egress times by walk are added between the stops of the selected route and the centroids of zones. In a last step, the resulting zone-to-zone travel time by transit are compared to the direct walk travel time. The shorter option is saved in the skim matrix.

RESULTS

For this analysis, the travel times queried by agents throughout the first simulation year for both the skim and the individual case are compared. A sample of 200,000 queries during housing search of the first simulation year was recorded. The individually routed travel time is compared against the skim travel time. To allow for a fair comparison, both the skim and the individual travel time queries were obtained from the same relaxed MATSim simulation of each scenario. Figure 1 shows a visualization of dwelling evaluations in the study area. It can be seen that the density of queries correlates with the population and employment density which is highest in the five larger cities of the study area.

Comparison of Travel Time Provision Methods

First, four different setups are compared to determine the influence of the car network density and the skim peak hour on travel times. Two networks with different network densities were analyzed with two peak-hour alternatives. The dense network consists of *504,109* links, while the coarse network has *142,703* links. Based on traffic count data, the morning peak hour is set to be 8 AM and the afternoon peak hour to 5 PM. For the Munich use case, the afternoon peak hour is more congested than the morning peak hour. Table 1 shows the root mean squared errors (RMSE) and correlation coefficient (r) between the individually queried travel times and the respective skim query for the four setups. Both setups of the afternoon peak hour show higher RMSE values than their morning peak counterparts. This is expected as the queries from SILO use job start times as their query time, and the majority of workers starts their job in the morning hours. When comparing network density, the dense network setups exhibit more congruent results for both peak-hour alternatives. This can be understood as another variant of the MAUP problem. The accuracy of routing decreases with less realistic networks. At the same time, there will be fewer route alternatives for congested route segments. This increases the impact of congestion and leads to higher fluctuations. Additionally, the coarser network is less connected, which leads to high under- and overestimation of travel times depending on the actual queried coordinate or centroid.

Figure 2 shows scatter plots for the four scenarios. The setup that uses the morning peak and the dense network shows the best match between skim and individual travel times. For both plots of the morning peak, there are point clouds to right of the diagonal that represent queries that were underestimated by the skim. Those queries are mostly from households in which the workers start work at untypical times (e.g. afternoon or evening). As congestion typically occurs inbound to the cities in the morning and outbound in the afternoon, the skim does not predict high congestion for people that commute into the city in the afternoon, thus underestimating their travel times.

Figure 3 depicts the comparison of the transit case. Results are neither affected by car network density (because transit is routed on a separate, congestion-free network based on a planned schedule) nor the peak hour used for the skim because service times are almost identical in the afternoon and morning peaks. It can be seen that the spread between individual and skim travel times is much larger than for auto travel times. The RMSE for the transit comparison is 66.45 minutes, the correlation coefficient is 0.84. The RMSE is rather high, because skim-based and individual travel times tend to be more different, especially in the range of longer travel times. This is plausible

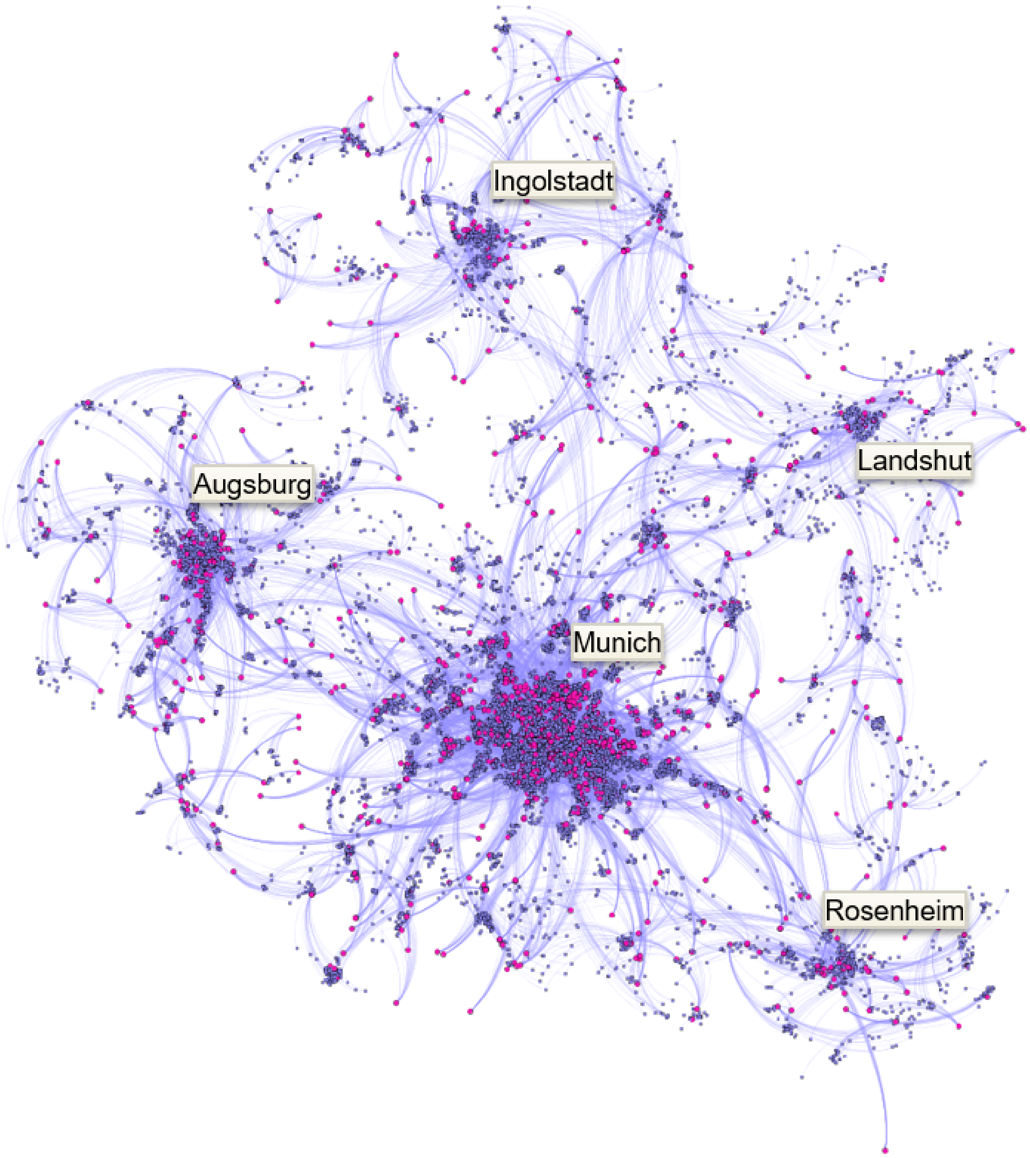

FIGURE 1 : Visualization of dwelling searches simulated by SILO (sample of 25,000 searches shown). Red dots indicate job locations of workers of the household looking for a new dwelling. Purple dots represent vacant dwellings that were evaluated by these households. The lines show which dwellings were assessed in terms of commuting times.

as those queries are usually between more rural zones, which tend to be also larger zones. There, transit accessibility is low and the correct actual distance to the next stop is more decisive. In the skim case, the transit travel times are the same for the whole zone, which can be very inaccurate for large, rural zones. The correlation coefficient is relatively high as most of the queries are from households which live in one of the major cities in the study area, where zones are small. Overall, there seems to be no systematic bias to under- or overestimate transit travel times. Compared to car travel times, transit travel times are generally higher as expected. The error for transit travel times is higher than for auto travel times as the car network is much more connected than the transit network, which makes it less crucial to query from/to specific points (i.e. stops) in the network.

In the following sections, only the morning peak hour and the dense network will be considered to analyze auto travel times. This is the setting where skim-based and individual travel times are most similar. By choosing this setup, we give the skim-based approach the best possible performance in comparison to individual travel times. One should keep in mind, however, that skims will perform worse in many applications than presented below.

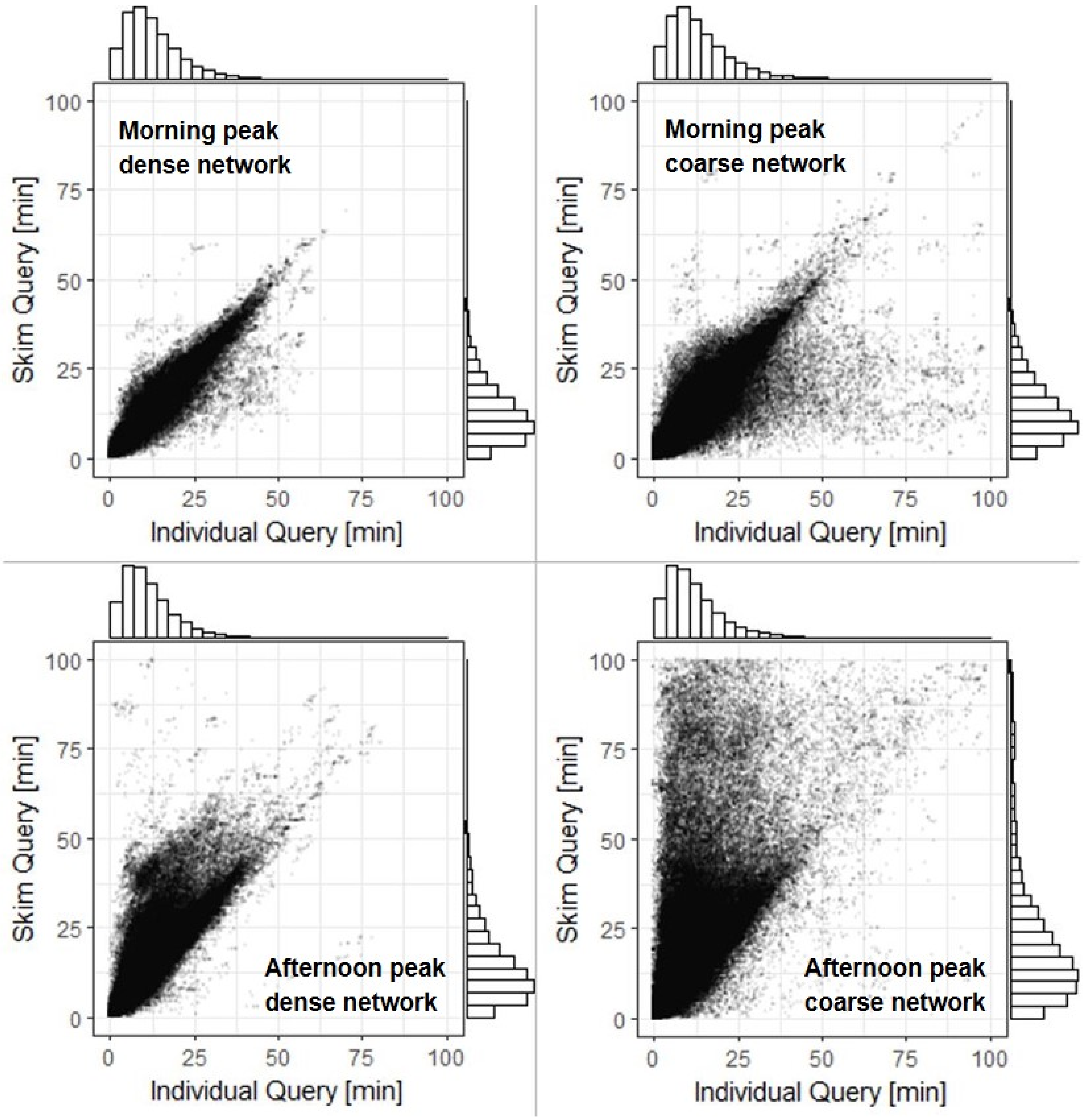

FIGURE 2 : Comparison between individual and skim-based travel times for four different setups: morning peak - dense network (top left), morning peak - coarse network (top right), afternoon peak - dense network (bottem left) and afternoon peak - coarse network (bottom right).

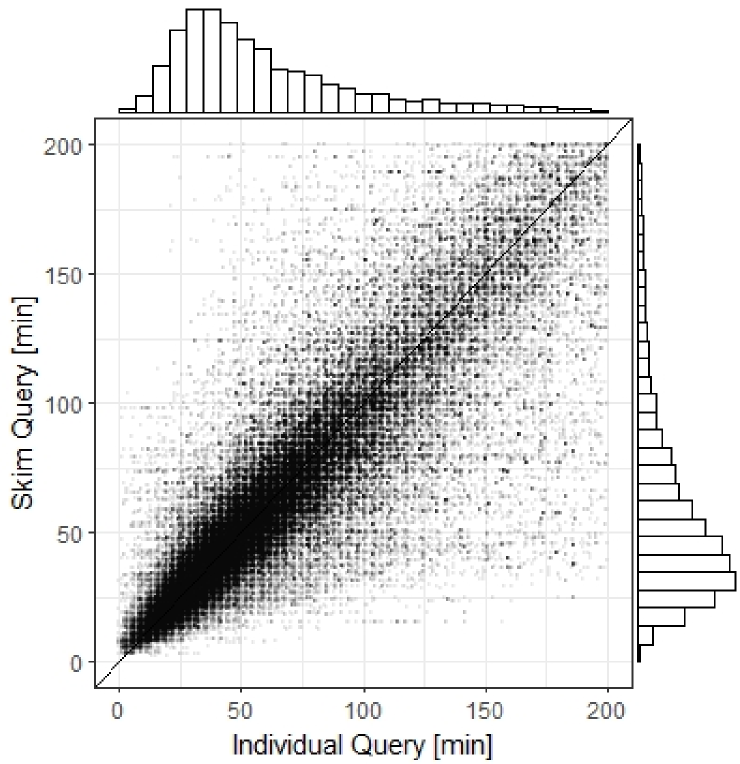

FIGURE 3 : Query result comparison between individual and skim based travel times for transit modes.

TABLE 1 : Root mean square errors (in minutes) and correlation coefficients between individual and skim-based travel times for different setups.

	Dense network	Coarse network
Morning Peak	RMSE = 3.139 r = 0.929	RMSE = 8.339 r = 0.709
Afternoon Peak	RMSE = 7.094 r = 0.817	RMSE = 26.731 r = 0.487

Spatial Influence

To analyze the effect of the spatial aggregation for the skim, the comparison is repeated with the time of day of the query fixed in the individual case as well, i.e. the same peak hour time as in the skim case is queried. For car travel times, the RMSE reduces to 1.89 minutes (compared to 3.14 minutes without isolation of the spatial influence), which suggests that for the dense network the spatial aggregation is not too inaccurate when the routing is always done at the peak hour. Additionally, the outliers in which the skim travel times underestimated travel times are largely reduced, which supports the hypothesis that those are emerging from queries at untypical times. In the transit case, however, the RMSE hardly drops to 59.26 minutes. This suggests that the spatial aggregation is impacting the difference between individual and skim travel times much stronger than in the car case. Again, this can be explained by the importance of the actual microlocation in relation to stop locations.

The MAUP problem can be seen when comparing the travel time differences against the zone

TABLE 2 : Root mean square errors (in minutes) for car and transit comparisons stratified by zone size *a* of the origin zone (km²).

	a <= 1km²	1km² < a <= 2km²	2km² < a <= 3km²	3km² < a <= 4km²	4km² < a <= 5km²	a > 5km²
Car	3.184	2.370	2.560	2.470	3.278	3.371
Transit	50.522	61.485	108.818	77.436	76.247	92.172

sizes of origin or destination. Table 2 shows the RMSE for the car and the transit comparison for different sizes of the origin zone (i.e. the dwelling zone) of the query. The RMSE stays rather constant around 3 minutes in the car travel time case. For transit, however, the initial RMSE of 50.522 minutes of small origin zones increases with zone size. This is because larger zones are more inaccurate per se and additionally have a lower population density which should also correlate with the transit network density. Lower network densities increase the variation of travel times for exact coordinates inside the zone. Figure 4 shows the RMSE for all origin zones of the queries. One can see that the error is low for the larger cities like Munich. However, the error increases fast when outside city boundaries.

Temporal Influence

The impact of temporal aggregation of skim travel times is analyzed by comparing skim and individual travel time by fixing the zone connectors in the case of individual travel times, but still using individual job starting times. The RMSE in the comparison of car travel times is 2.88 minutes, which suggests that the impact of temporal aggregation is higher than the impact of spatial aggregation. In contrast to the comparison with fixed query times, fixing the spatial component (i.e. zone connectors) reveals the outliers in which the skim underestimates travel times. This confirms that these outliers are an artifact of temporal aggregation which is inaccurate for untypical job start times. This is confirmed by looking at the RMSE throughout the day (see figure 5). The RMSE for queries from 6 am to 10 am is 2.44 minutes. In the afternoon from 3 pm to 7 pm the RMSE increases to 6.55 minutes, with most of the queries underestimated by the skim. In Munich, the congestion in the afternoon peak hour is typically higher than in the morning peak hour. This is not captured when the skim is computed for the morning peak hour. Additionally, there are people with anticyclical behavior that start their job in the afternoon and who have to travel in the opposite direction of the main congestion, e.g. people who live inside the city and go to work outside. In this case the skim does not predict high traffic for going out of the city since it was created for the morning peak. This leads to underestimated travel times.

Contrary to car travel times, the transit travel times seem to be less distorted by temporal aggregation than by spatial aggregation. The RMSE drops to 25.01 minutes when querying from fixed zone connectors which is less than half of the error of the spatial impacts comparison. This can be explained by the fact that transit travel times are routed based on scheduled times on a separate network without congestion. As service is similar in terms of frequencies for most of the day, the time of day does not have a strong effect on transit travel time queries. The RMSE is 59.52 minutes from 6 am to 10 am and hardly changes to 57.40 minutes in the afternoon from 3 pm to 7 pm. The seemingly larger variations in the morning hours in figure 5 can be explained by the fact that the amount of queries is much higher than in the afternoon hours, leading to a higher spread.

However, the RMSE is still high and large differences can be seen during night and off-service

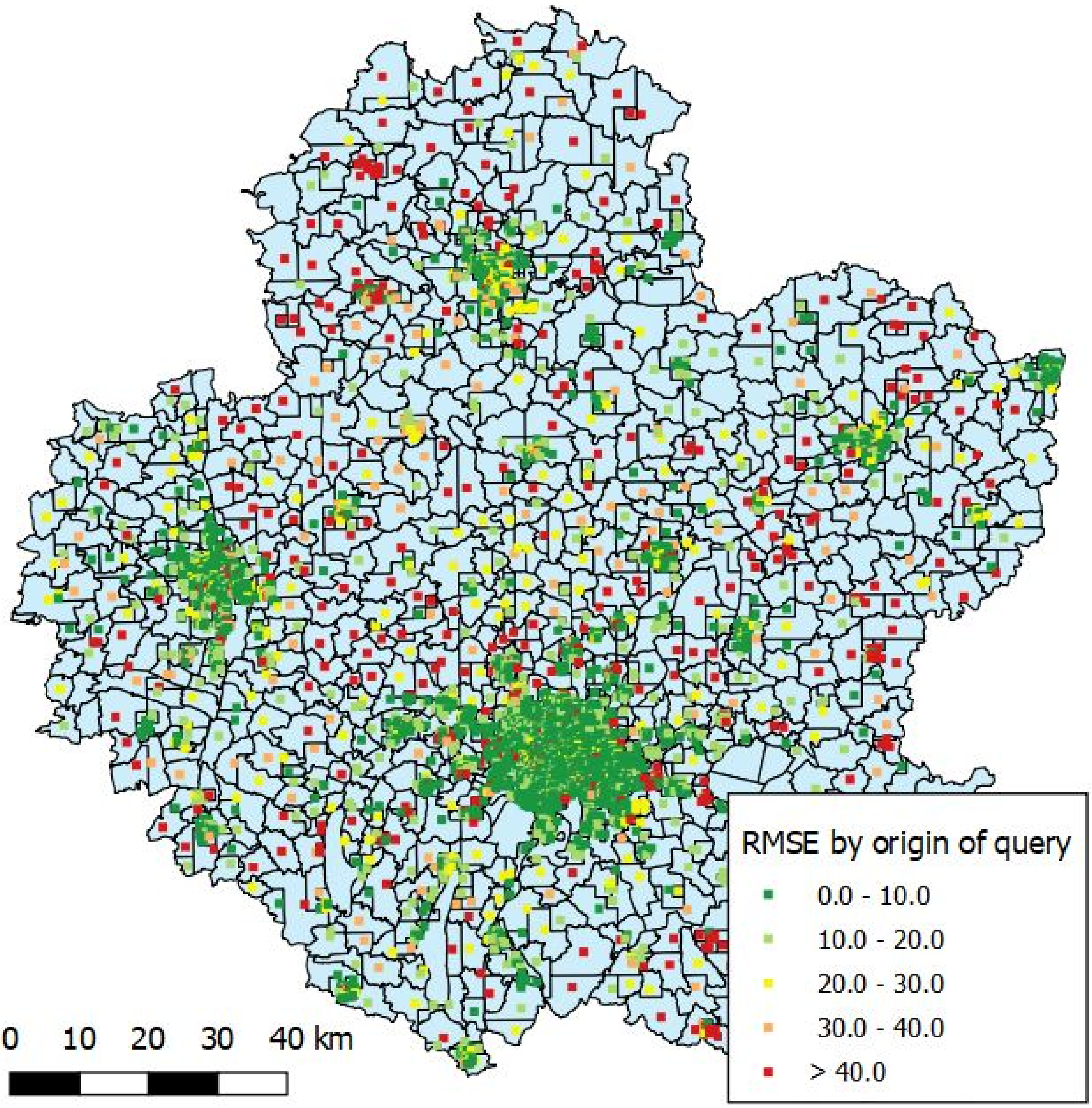

FIGURE 4 : Root mean square error values for the transit comparison by zone centroids of the origin zones of queries.

hours. Figure 5 shows an almost diagonal line in the early night hours until 4 am during which the skim underestimates travel times. As most of the transit services do not operate in those hours, the transit router will return a large direct walking trip, making transit very unattractive. The underestimation of travel times reduces towards the start of the transit operation around 4am as people might as well wait for the start of the service. It can be expected that the error increases when the schedule varies more throughout the day.

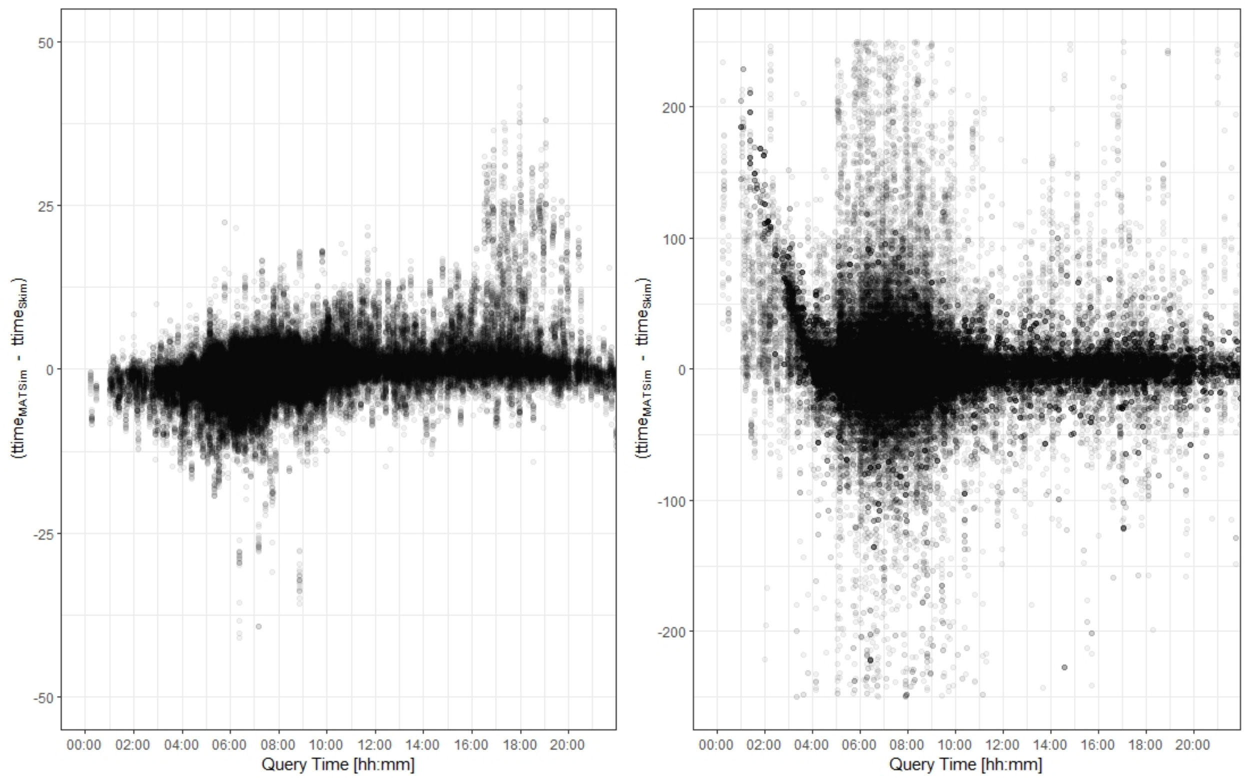

FIGURE 5 : Difference between individual and skim travel times by time of day for car (left) and transit (right). Note different scales.

Impacts on Household Relocation

It is expected that the aggregations in skims lead to inconsistent behavior in microscopic relocation decisions. Transit captive households without cars that only evaluate transit travel times will select more randomly when choosing from dwellings within the same zone in the skim case as the zone-to-zone transit travel times will be the same. In the query analysis it was shown that the spatial impact is high for transit skims. Compared to individual travel times, households should on average move closer to transit stops because the exact microlocation in relation to stop positions is important. To test this, all household moves of the first simulation year of the Munich scenario were recorded for the skim and the individual travel time representation. In both scenarios the households that decided to look for a new dwelling were randomly chosen with a probability of 1%. This is to prevent that the current housing satisfaction, which is also based on current travel times, would lead to completely different households that decide to move. After the simulation, the distances of the new dwelling locations to the nearest stop according to the transit schedule were calculated. When looking at all relocations (28,430 cases), the average distance to the closest transit stop after moving is 1085,45 meters in the skim scenario and 1085,19 meters in the individual travel time scenario. The average distances are the same when looking at all relocations which includes the majority of households which are not "‘transit captives"’. However, when looking at the relocations of households that have no cars and at least one worker who has to commute (2,886 cases), the average distance to the closest stop drops to 605,57 meters for the skim scenario and to 569,51 meters in the individual travel times scenario. One can see that those households correctly show a higher sensitivity to transit accessibility in both scenarios. In the individual travel time scenario households seem to be slightly more sensitive (about 6%) to transit stop distance than in the skim scenario, which is a small effect but confirms the hypothesis of a more random selection of dwellings in the skim case. It is important to note that the nearest stop distance is not necessarily the stop which is served by the actually taken transit line for getting to work.

DISCUSSION

The presented results suggest that the temporal and spatial aggregation of travel times can have a large impact on their accuracy. While this is less true for car travel times if a dense network is used, it becomes even more important in the case of transit travel times which proved to be very unreliable. A disadvantage of the individual query is the extended computation time. Still, a model run of the FABILUT modeling suite with individual travel times run multiple decades into the future can be finished in less than two days for the Munich use case. On the other hand, skim-based approaches that aim to improve accuracies (e.g. using multiple distinct time-of-day-specific matrices to reduce effects of temporal inaccurracies or a very high spatial resolution to reduce effects of spatial inaccuracies) also increase computing times and beyond that lead to memory requirements that can become unwieldy. The relocation impacts of the use case were small but nevertheless clearly confirm the increased accuracy of individual travel times. A reason for the relatively small impact could be that households do not react very sensitive to changes in transit travel times yet as they draw upon the same commuting probability as used for the evaluation of car travel times which tend to be shorter. Another limitation of the current approach is that workers only evaluate their trip going to work and do not care about the return trip.

OUTLOOK

This research showed the relevance for individual travel times for the integration of land use and transport models. With this example, we were able to show that it makes a substantial difference when we introduce the spatial and temporal detail of individual travel times compared to aggregate skim matrices. This finding suggests that it will also make a substantial difference when we replace skim matrices with individual travel times in travel demand modeling. Most destination choice and mode choice models in operation use skim matrices to calculate the utilities of different destinations and various modes. Sometimes, peak hour skims and off-peak hour skims are distinguished. One could imagine, however, that replacing skims with individual travel times may have an equally substantial impact as shown in this paper for land use/transport model integration. For example, the transit schedule differs by time of day, which is likely to affect the tripmaker's mode choice depending on their departure time. Similarly, congestion changes over the course of a day, which may entice them to choose different destinations for trips in the morning than in the afternoon. Last but not least, many travel demand models suffer from a coarse zone system. In a skim-based world, every trip to a larger zone takes the same travel time, no matter whether the final destination is close to the zone centroid or at the outskirts of the zone. Individual travel times allow to overcome this spatial and temporal aggregation.

The implemented query approach can be extended to additionally include person attributes (e.g. age, gender, disability status, value of travel time) or vehicle attributes (e.g. fuel type, noise and pollutant emission rates). Furthermore, policy-relevant network attributes can be implemented (e.g. environmental zones, time-dependent tolls etc).

Another important benefit is that intrazonal travel times and the problems of defining their calculation (*23, 24*) are not needed anymore.
The presented integrated land use transport model approach and the implemented query will also allow to query the transport model directly for additional data like noise or air pollutant emissions. Even in situations where the individual query is not desired because of higher computing times, the presented coupling can also be applied for MATSim to create the skims after each transport model run. This reduces the amount of input data as the skim does not have to be provided at the start of the program, whereas a MATSim simulation scenario of a study region can be comparatively easily be set-up in case a synthetic population of the SILO model already exists for that region (*10*).

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgesellschaft (DFG) under project 5051013. It was completed with the support of the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme [grant number 291763].

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: N. Kuehnel, D. Ziemke, R. Moeckel, K. Nagel; data collection: N. Kuehnel, D. Ziemke; analysis and interpretation of results: N. Kuehnel, D. Ziemke, R. Moeckel, K. Nagel; draft manuscript preparation: N. Kuehnel, D. Ziemke. All authors reviewed the results and approved the final version of the manuscript.

References

[1] Spiekermann, K. and M. Wegener, *Freedom from the tyranny of zones: towards new GIS-based models*, Taylor & Francis Group, London, pp. 45–61, 2000.

[2] Openshaw, S., A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning and Spatial Modelling. *Transactions of the Institute of British Geographers*, Vol. 2, No. 4, 2006, p. 459.

[3] Thomas, I., C. Cotteels, J. Jones, A. P. Bala, and D. Peeters, Spatial challenges in the estimations of LUTI models: some lessons from the SustinCity project. In *Integrated Transport & Land Use Modeling for Sustainable Cities* (M. Bierlaire, A. de Palma, R. Hurtubia, and P. Waddell, eds.), EPFL Press, Lausanne, 2015, chap. 4, pp. 55–74.

[4] Wegener, M. and S. &. Wegener, From Macro to Micro – How Much Micro is too Much? *Published in Transport Reviews*, Vol. 31, No. 2, 2009, pp. 14–16.

[5] Davidson, W., R. Donnelly, P. Vovsha, J. Freedman, S. Ruegg, J. Hicks, J. Castiglione, and R. Picado, Synthesis of first practices and operational research approaches in activity-based

travel demand modeling. *Transportation Research Part A: Policy and Practice*, Vol. 41, No. 5, 2007, pp. 464–488.

[6] Spiekermann, K. and M. Wegener, Environmental feedback in Urban models. *International Journal of Sustainable Transportation*, Vol. 2, No. 1, 2008, pp. 41–57.

[7] Lee, D. B., Requiem for Large-Scale Models. *Journal of the American Planning Association*, Vol. 39, No. 3, 1973, pp. 163–178.

[8] Miller, E. J., D. S. Kriger, and J. D. Hunt, *Integrated Urban Models for Simulation of Transit and Land Use Policies Guidelines for Implementation and Use*, 1999.

[9] Donnelly, R., G. Erhardt, R. Moeckel, and W. A. Davidson, *Advanced Practices in Travel Forecasting. A synthesis of Highway Practice. NCHRP Report 406*, 2010.

[10] Ziemke, D., K. Nagel, and R. Moeckel, Towards an Agent-based, Integrated Land-use Transport Modeling System. In *Procedia Computer Science*, 2016, Vol. 83, pp. 958–963.

[11] Badoe, D. A. and E. J. Miller, Transportation-land-use interaction: Empirical findings in North America, and their implications for modeling. *Transportation Research Part D: Transport and Environment*, Vol. 5, No. 4, 2000, pp. 235–263.

[12] Homer, M. W. and A. T. Murray, Excess Commuting and the Modifiable Areal Unit Problem. *Urban Studies*, Vol. 39, 2002, pp. 131–139.

[13] Molloy, J. and R. Moeckel, Automated design of gradual zone systems. *Open Geospatial Data, Software and Standards*, Vol. 2, No. 1, 2017, p. 19.

[14] Lovelace, R., D. Ballas, and M. Watson, A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. *Journal of Transport Geography*, Vol. 34, 2014, pp. 282–296.

[15] Hagen-Zanker, A. and Y. Jin, A New Method of Adaptive Zoning for Spatial Interaction Models. *Geographical Analysis*, Vol. 44, No. 4, 2012, pp. 281–301.

[16] Moeckel, R., Constraints in household relocation: Modeling land-use/transport interactions that respect time and monetary budgets. *Journal of Transport and Land Use*, Vol. 10, No. 2, 2016, pp. 1–18.

[17] Horni, A., K. Nagel, and K. W. Axhausen (eds.) *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, 2016.

[18] Moeckel, R., N. Kuehnel, C. Llorca, A. T. Moreno, and H. Rayaprolu, Microscopic Travel Demand Modeling: Using the Agility of Agent-Based Modeling Without the Complexity of Activity-Based Models. In *Annual Meeting of the Transportation Research Board*, Washington, DC, 2019.

[19] Ziemke, D., I. Kaddoura, and K. Nagel, The MATSim Open Berlin Scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. *Procedia Computer Science*, Vol. 151, 2019, pp. 870–877.

[20] Rieser, M., D. Métrailler, and J. Lieberherr, Adding Realism and Efficiency to Public Transportation in MATSim. In *18th Swiss Transport Research Conference*, 2018, pp. 1—-21.

[21] Moreno, A. and R. Moeckel, Population Synthesis Handling Three Geographical Resolutions. *ISPRS International Journal of Geo-Information*, Vol. 7, No. 5, 2018, p. 174.

[22] Konduri, K. C., D. You, V. M. Garikapati, and R. M. Pendyala, Enhanced Synthetic Population Generator That Accommodates Control Variables at Multiple Geographic Resolutions. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2563, No. 1, 2016, pp. 40–50.

[23] Okrah, M. B., R. Moeckel, and G. Wulfhorst, Finding the optimal level of spatial resolution for handling non-motorized travel in macroscopic travel demand models, 2017.

[24] Moeckel, R. and R. Donnelly, Simulation of intrazonal traffic flows: The end of lost trips. In *Proceedings of the 11th Conference on Computers in Urban Planning and Urban Management (CUPUM)*, 2009, pp. 16–18.