Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2026) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 17th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 14-16, 2026, Istanbul, Türkiye

Developing an adaptive zoning system for city-scale activity-based travel demand modeling using OpenStreetMap data

Farhan Jamil[a,*], Mohammed Hajouj[a], Luk Knapen[a], Muhammad Adnan[a], Tom Bellemans[a]

[a] *Hasselt University, Transportation Research Institute IMOB, Martelarenlaan 42, 3500 Hasselt, Belgium*

Abstract

City-scale activity-based demand models (ABMs) offer detailed insights into travel behaviors; however, their accuracy is often limited by the coarse spatial zoning used in many zone-based demand models. Conventional Traffic Analysis Zones (TAZs) aggregate data at a level that masks neighborhood-specific variations and misallocates short and non-motorized trips. To enable more realistic urban mobility analysis, this paper introduces a hybrid zoning system that adapts spatial resolution to local activity density. Using Hasselt, Belgium, as a case study, we refine official statistical sectors into high-resolution miniZones. This is achieved by applying constrained k-means clustering to OpenStreetMap building data, followed by Voronoi tessellation. The resulting clusters are transformed into contiguous zones through dissolving Voronoi tessellation. This technique is applied to provide fine-grained detail within the city's scope, reflecting the density of buildings where human activity is high. To keep the refinements of the zoning computationally manageable for a city-scaled regional model, a gradual reduction in detail is incorporated as one moves away from the city. This is achieved by aggregating to coarser official units in the more distant regions. The final result is a hybrid zoning system. This adaptive model approach enhances the representation of trip generation and distribution within cities, providing support for more accurate activity-based travel demand modeling.

Keywords:
Adaptive zoning; Activity-based demand; Traffic analysis zone; OpenStreetMap; Voronoi tessellation; K-means clustering

1. Introduction

Many activity-based models (ABM) specify locations at the traffic analysis zone (TAZ) level. For each TAZ, a number of attributes that describe its attractiveness for particular activities are identified, e.g., the number of people employed, children attending school, employment in the retail or hospitality sector, or total population. However, using coarse spatial units like TAZ's mask important local variations in population, land use, and travel behavior, leading to a loss of intra-urban detail. As a result, many short or non-motorized trips are misallocated, limiting the

* Corresponding author. Tel.: +32-11-26.91.11; Fax: +32-11-26.91.99.
E-mail address: farhan.jamil@uhasselt.be

accuracy of city-scale ABM. Recognizing this, urban planners and modelers push for ever finer spatial resolutions, and there is a clear trend toward smaller zones which subdivide conventional administrative units into much smaller areas. This helps to model traffic behaviors more realistically for regional city-scale ABM applications.

2. Problem Statement

Several obstacles exist to implementing city-scale activity-based demand modeling, particularly in establishing an optimal geographic scale amid the complexity and variability of urban areas. The standard practice of running these models at the Travel Analysis Zone (TAZ) level is limited by traditional zoning systems [8], although providing helpful information, they often aggregate data at a coarser resolution than necessary. A key issue is to avoid trip generation and attraction excessively occurring at a single point (e.g., TAZ centroid), which may lead to unreasonable traffic allocation with some congested links and some empty links. Using smaller TAZ's may mitigate this problem. As a consequence, a zoning system with higher resolution is desired. On the other hand, city-scale activity-based models require covering a large region (due to the extent of trips), but applying high-resolution zoning uniformly is computationally intensive. The complexity of computing LOS matrices is quadratic, as the required computational resources grow with the square of the number of TAZs.

This paper addresses the challenges associated with the spatial resolution by creating a hybrid zoning system that delivers details where necessary, i.e., the city center, and can trade accuracy for efficiency (less detail) in the far off regions.

3. Related Work and Literature Review

Activity-based models (ABMs) require detailed information on individual behaviors and activities, which is often unavailable or incomplete, leading to potential inaccuracies in demand predictions [2, 7, 14]. High spatial and temporal resolution is necessary to assess the impact of policy interventions and design efficient infrastructures. However, achieving this level of detail is challenging due to data limitations and the computational intensity required [7, 17].

Traffic Analysis Zones (TAZs) are a fundamental component in traditional travel demand models, but they pose limitations when applied to city-scale models. These limitations primarily stem from issues related to spatial resolution, data aggregation, and sensitivity to behavioral nuances. In general, a more disaggregated geographical level is supposed to better reflect the traffic behavior and accordingly avoid the trip generation and attraction excessively happening on one point (e.g., TAZ centroid), which may avoid an unreasonable traffic allocation with some congested links and some empty links [1]. TAZs often use coarse spatial units, leading to exaggerated intrazonal trips and biased trip distribution, resulting in high estimation errors [13]. This aggregation can obscure important behavioral distinctions and reduce the model's sensitivity to input changes [9, 10]. The use of TAZs is subject to the Modifiable Areal Unit Problem (MAUP) [15], where the size and shape of zones can significantly affect the outcomes of traffic demand models. Smaller TAZs tend to produce more accurate trip length estimates and lower error rates, but larger TAZs can lead to less detailed and less accurate modeling [3]. Activity-based models require more detailed data and calibration to reflect realistic behaviors and interactions [10, 9]. TAZs are not well-suited for modeling pedestrian and non-motorized travel, requiring finer spatial resolution to accurately capture walking trips and their origins and destinations [4].

These limitations have led to an increasing emphasis on the development of micro-level zoning systems within city-scale activity-based travel demand models. The design of spatial zoning systems for transport modeling faces issues in balancing computational efficiency and spatial resolution. [11] introduced a gradual rasterization method, creating smaller cells in urban areas and larger ones in rural regions, improving model validation. An adaptive zoning system proposed by [5] where destination zone size varies with distance from the origin, reducing computation time by 70% while maintaining accuracy in the study area. [12] further improved the gradual rasterization algorithm by incorporating land use weighting, respecting municipal boundaries, and introducing an iterative process to achieve desired zone numbers. These approaches offer innovative solutions to increase spatial resolution in urban areas without unsustainably increasing the number of zones, enhancing the effectiveness of spatial modeling.

However, these methods present limitations when applied to city-scale activity-based travel demand modeling. In [5] the work focuses on spatial interaction modeling and and employs origin-specific zoning, making it less suited for activity-based models that require a single, consistent geographic framework and finer urban detail. [11]'s rasterization approach does not respect administrative boundaries, complicating integration with official socioeconomic data. While [12] addresses municipal alignment, it remains raster-based, producing zones with geometric uniformity that may not align with actual building clusters.

4. Proposed Methodology

The proposed methodology aims to create a transportation analysis zoning system optimized for city-scale activity-based modeling. The study area considered for detailed zoning is Hasselt, located in the Flanders region of Belgium. The core challenge is to balance spatial detail with computational efficiency in terms of processing the number of zones for activity-based demand generation. To address this, we move away from a uniform raster approach. Instead, we implement an adaptive zoning scheme where the granularity of zones dynamically adjusts based on the density of human activity. The guiding principle is that areas of intense activity require finer zoning to accurately capture trip generation, attraction, and short-distance travel behaviors. Conversely, areas with lower activity can be represented with coarser zones to reduce model complexity. We use building footprint data as a primary proxy for activity density.

4.1. Adaptive Zoning

Our adaptive zoning system is constructed in a hierarchical manner. We use an official Belgian zoning structure known as the statistical sectors (SS) (Figure 1_b). Statistical sectors are the smallest territorial units used by the government statistics office Statbel for the collection, analysis and dissemination of demographic, social and economic data [1].

The core of our method involves disaggregating the base SS into detailed zones, called *miniZones* in this paper. This disaggregation is achieved by applying clustering to building data sourced from OpenStreetMap (OSM). The inherited building contours, coded as closed ways in OSM, are used to calculate the centroids of each building. The centroid represents the building and is used in the clustering process. The underlying assumption is that building density correlates strongly with trip generation rates. Therefore, within the study area, clusters of buildings are used to support the generation of miniZones.

Within the city's perspective of Hasselt, the clustering algorithm considers zones at a distance of 8km around the city center. The distance value is a parameter that can be adjusted based on the city's size and the neighboring cities that influence trips within the city. For the broader influence region, we divide the Flanders region into three distance bands. Until 8km from the city center, we have miniZones; then, from 8 to 20km, we have statistical sectors; and for further regions, we consider the aggregation of statistical sectors to municipalities. This results in a hybrid zoning layer (Figure 1_c) that provides detail where it matters most for urban analysis while maintaining performance over a large region.

4.2. Refinement of Building Data for Activity Relevance

As clustering is based on OSM information, we note that raw OSM data alone struggles to distinguish human activity locations. A naive use of all OSM building tags can introduce noise, as OSM catalogs all physical structures, not exclusively those associated with frequent human activity. This leads to some observed boundary cases of erroneous zone creation in rural or greenfield areas around isolated, low-activity structures (e.g., tool sheds, field shelters, or detached garages), which misrepresent true trip patterns. To create a sensible zoning scheme, we refine the building data by filtering OSM tags to include only points suitable for human activities. This involves removing tags associated with structures that are poor proxies for trip generation. Examples of buildings that are filtered out include those tagged as `shed`, `roof`, `farm_auxiliary`, `garage`, `garages`, `carport`, `shelter`.

[1] https://statbel.fgov.be/en/news/statistical-sectors-2025

4.3. MiniZone Delineation via Constrained Clustering and Tessellation

We implement the delineation of miniZones by clustering building centroids. We apply the commonly used k-means clustering, because

(1) the Euclidean distance between two geographical points is in our case a good candidate for the *clustering distance* function (2) and it offers a scalable and computationally efficient algorithm (since we need to repeat the process to a large number of statistical sectors).

The clustering is constrained by (1) the maximum number of miniZones allowed within a single SS and (2) a maximum surface area `max_cluster_area` setting for each miniZone. The clustering happens inside each unique SS, ensuring the original boundaries are always respected (the SS is covered entirely, and each miniZone is part of exactly one SS). The area constraint is essential to prevent geographically oversized miniZones in cases where buildings are strongly clustered and the algorithm initially identifies too few clusters. The process is adaptive: the allowable range for the number of clusters n_{clust_min} to n_{clust_max}. is determined based on the building density of the input SS. This allows for finer segmentation that already possesses a more detailed administrative structure, which often corresponds to regions closer to the center of our study.

We calculate the Silhouette coefficient, which measures cluster separation [16], to find the number of clusters k within the n_{clust_min} to n_{clust_max} range that provides the best cluster definition. The k with the highest Silhouette score is selected as the candidate optimal solution. This candidate solution is then validated against a geographic constraint: if any resulting cluster exceeds the `max_cluster_area`, the candidate is rejected. To find a valid solution, the minimum number of clusters n_{clust_min} is incremented by one, the clustering process is repeated, and the area criterion is rechecked. This iterative loop ensures that the final result satisfies the geographic sensibility (compact zones), thereby preserving local detail for travel modeling.

The output of this process is a set of building points, each assigned a unique cluster ID. It is necessary to transform these labeled points into a contiguous zoning system that partitions the SS within the city. We define the spatial extent of each cluster through a two-step geometric procedure. For each statistical sector, a Voronoi tesselation [6] is constructed using individual building centroids as seed points. This produces a highly detailed spatial partitioning where each Voronoi cell represents the area closer to its building than to any other within the same sector. Finally, all Voronoi cells whose seed points share the same cluster ID are dissolved into a single, contiguous polygon. These final polygons represent the miniZones Figure 1_e. MiniZones geographically partition statistical sectors (i.e. cover them exactly). This meets the requirement of the next stage in modeling where data about SS need to be disaggregated to miniZones.

5. Results and Discussion

The application of the proposed adaptive zoning methodology to Hasselt, Belgium, successfully generated a hybrid, multi-resolution zoning system. The 117 statistical sectors within the city's 8km core were disaggregated into 943 miniZones using constrained k-means clustering on filtered OSM building points. This resulted in a nearly eight fold increase in zone density within the urban core, providing a significantly finer spatial resolution where human activity is concentrated. The final miniZones are contiguous, nested within official administrative boundaries, and their granularity adapts to local building density (see Figure 1_d and Figure1_e).

The primary achievement of this work is the creation of a high-resolution zoning structure that is directly tied to human activity proxies. By filtering OSM data to exclude non-residential structures (e.g., sheds, garages) and constraining clusters by both statistical coherence (as measured by the Silhouette score) and maximum geographic area, the method ensures that increased spatial resolution is applied where trip generation potential is highest. This addresses the core limitation of traditional TAZs, which misallocate short trips by aggregating origins and destinations to a single centroid. The hybrid framework—transitioning from miniZones (0–8 km) to statistical sectors (8–20 km) to municipalities (20+ km)—provides the necessary detail for modeling intra-urban travel while maintaining regional stability and computational feasibility.

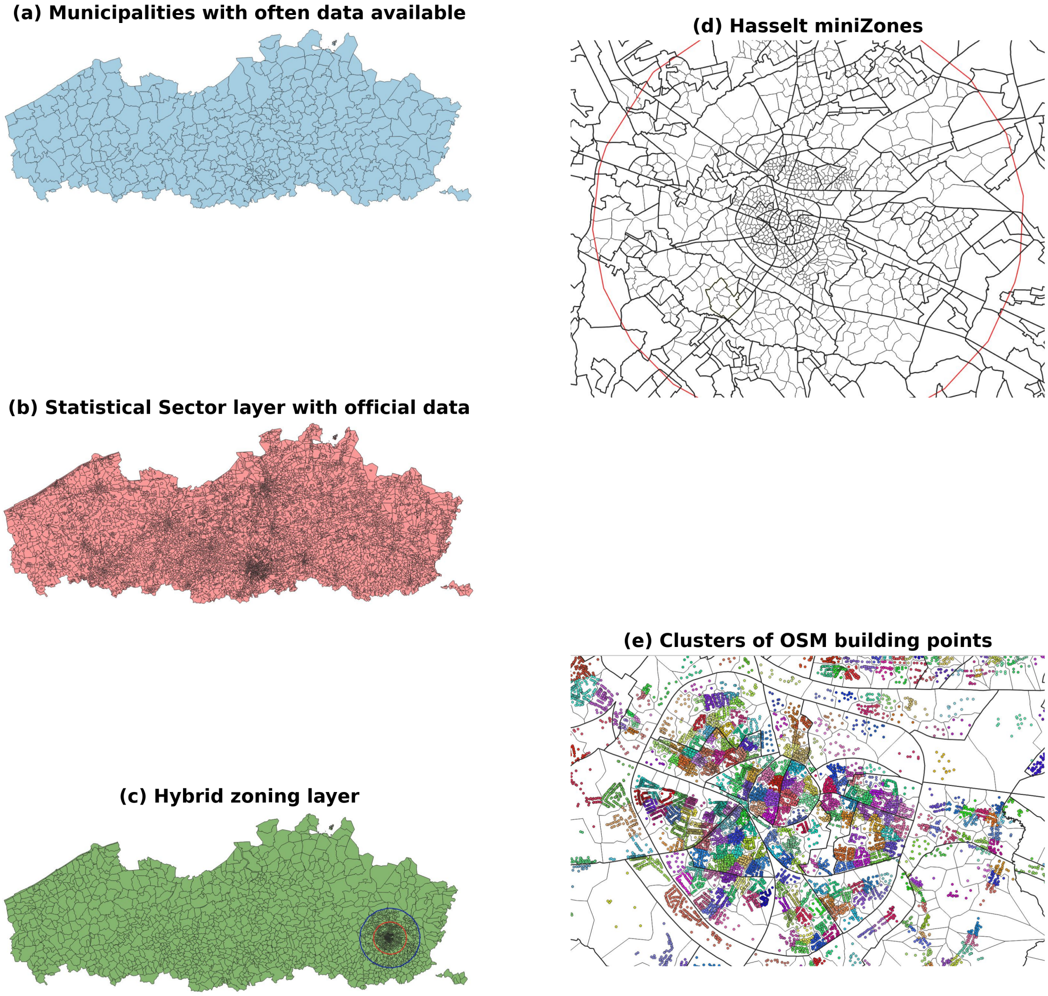

Fig. 1. Comparison of zoning and spatial units used in the study.(a) Municipalities, (b) Statistical Sectors with official data, (c) Hybrid zoning layer combining multiple scales, (d) high-resolution miniZones for Hasselt, and (e) clusters of OSM building points for fine-grained spatial representation.

5.1. Zone Area Distribution

Using miniZones induces an eight-fold increase in spatial resolution while maintaining complete coverage of the study area (157.85 km²). This disaggregation achieves a dramatic reduction in zone dimensions: mean area decreases by 87.6% (from 1.35 km² to 0.17 km²) and median area by 86.7% (from 0.85 km² to 0.11 km²). The minimum and maximum areas also shift toward smaller values under the miniZone system, reflecting the algorithm's emphasis on producing many small areas in dense parts of the city. Although miniZones have a much smaller absolute standard deviation in area (0.1965 km² vs. 1.3597 km² for Statistical Sectors), we complement these absolute measures with the coefficient of variation (CV). Using the statistics in Table 1, the CV for the SS is 100.79% while the CV for the miniZones is 117.38%. This indicates that, although miniZones are much smaller on average (an advantage for

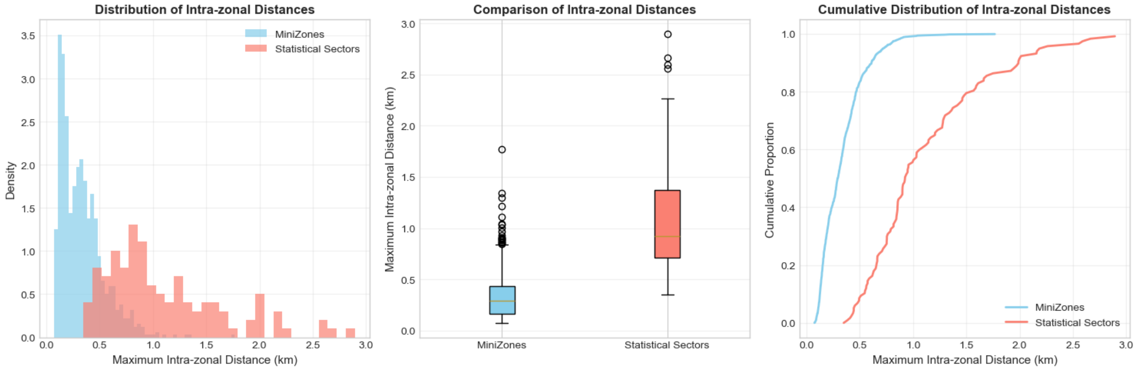

Fig. 2. **Intra-zonal geometric comparison between Statistical Sectors and MiniZones.** (a) Distribution of maximum straight-line distances. (b) Categorical shift showing MiniZones concentrate in short-distance categories. (c) Relationship between zone area and maximum intra-zonal distance, illustrating how reduced zone sizes enable tighter spatial containment.

capturing local heterogeneity), their proportional dispersion around that smaller mean is larger than for Statistical Sectors. The practical interpretation is that adaptive zoning produces, on average, a lower building count in the zones (because we increase the spatial resolution by replacing SS with miniZones) and increases the variation, reflecting that we capture heterogeneity.

Table 1. Zone area statistics for Statistical Sectors and MiniZones. Adaptive zoning reduces the mean zone area by 87.6% and the median by 86.7%, with MiniZones exhibiting a more uniform size distribution (lower standard deviation).

Statistic	**Statistical Sectors (km^2)**	**MiniZones (km^2)**
Mean	1.3491	0.1674
Median	0.8463	0.1125
Std Dev	1.3597	0.1965
Min	0.1198	0.0086
Max	6.7248	1.9480
CV	100.79%	117.38%

5.2. Intra-Zonal Distance Reduction and Spatial Resolution

The adaptive zoning methodology achieves substantial geometric refinement, reducing the median zone area by 86.7% (from 0.846 km² to 0.112 km²) and decreasing the maximum straight-line distance within zones by 68.5% (from 923m to 291m) Figure 2_a. This transformation creates 943 MiniZones where 83.4% have spatial extents under 500m, compared to only 9.4% of the original 117 Statistical Sectors Figure 2_b. Critically, 395 MiniZones (41.9%) now feature extents under 250m—a distance relevant for walking trips while no Statistical Sectors achieved this threshold.

These geometric improvements translate directly to enhanced travel modeling capabilities Figure 2_c. The tighter spatial bounds reduce centroid aggregation error and minimize the misclassification of trips that should cross zone boundaries. By eliminating zones with problematic extents exceeding 1km (reduced from 43.6% to 0.8%), the framework enables more accurate representation of short-distance urban mobility. The eight-fold increase in zone count, combined with the concentration of detail where human activity is highest, provides a spatial foundation that better reflects the fine-grained nature of urban travel patterns, which are essential for activity-based modeling.

6. Data Integration, Limitations, and Future Work

The geometric refinement enabled by the hybrid zoning system also supports practical data integration for modeling. The methodology facilitates the disaggregation of official socio-economic data (e.g., population, employment) from coarser administrative units to the miniZone level using building footprint as a weighting factor or some regression model-based disaggregation. This produces a more realistic distribution of trip generators and attractors across the network, which is critical for the accuracy of subsequent activity-based demand modeling stages, particularly for non-motorized and short-distance trips.

A key limitation of the current approach is its dependency on the quality and taxonomic detail of OSM building tags. While the filtering of non-residential structures improves relevance, the method does not inherently account for variations in trip-generation intensity between different building types (e.g., a large apartment complex versus a single-family home). Another limitation is that while OSM buildings are suitable for delineating zones in person mobility models, they are insufficient for freight transport modeling, which requires specific data on warehouses, logistics centers, and commercial loading areas not consistently available in OSM.

Future work will therefore focus on integrating the new zoning system into a full activity-based model for Hasselt to quantitatively assess its impact on key outputs such as trip length distribution, mode choice, and network assignment accuracy. Further refinement could involve weighting building points by type, volume, or other attributes to better proxy activity potential.

Acknowledgements

This research is supported by the VLAIO cSBO project STRAUSS HBC.2023.0008.

References

[1] Bao, Q., Kochan, B., Shen, Y., Creemers, L., Bellemans, T., Janssens, D., Wets, G., 2018. Applying FEATHERS for Travel Demand Analysis: Model Considerations. Applied Sciences 8, 211. URL: https://www.mdpi.com/2076-3417/8/2/211, doi:10.3390/app8020211.

[2] Carrese, F., Cantelmo, G., Fusco, G., Viti, F., 2019. Leveraging GIS Data and Topological Information to Infer Trip Chaining Behaviour at Macroscopic Level, in: 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), IEEE, Cracow, Poland. pp. 1–10. URL: https://ieeexplore.ieee.org/document/8883329/, doi:10.1109/MTITS.2019.8883329.

[3] Chang, K.T., Khatib, Z., Ou, Y., 2002. Effects of Zoning Structure and Network Detail on Traffic Demand Modeling. Environment and Planning B: Planning and Design 29, 37–52. URL: https://journals.sagepub.com/doi/10.1068/b2742, doi:10.1068/b2742.

[4] Clifton, K.J., Singleton, P.A., Muhs, C.D., Schneider, R.J., 2016. Representing pedestrian activity in travel demand models: Framework and application. Journal of Transport Geography 52, 111–122. URL: https://linkinghub.elsevier.com/retrieve/pii/S0966692316301302, doi:10.1016/j.jtrangeo.2016.03.009.

[5] Hagen-Zanker, A., Jin, Y., 2012. A New Method of Adaptive Zoning for Spatial Interaction Models. Geographical Analysis 44, 281–301. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.2012.00855.x, doi:10.1111/j.1538-4632.2012.00855.x.

[6] Ito, Y., 2015. Voronoi Tessellation, in: Engquist, B. (Ed.), Encyclopedia of Applied and Computational Mathematics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1546–1547. URL: https://link.springer.com/10.1007/978-3-540-70529-1_315, doi:10.1007/978-3-540-70529-1_315.

[7] Keirstead, J., Sivakumar, A., 2012. Using Activity-Based Modeling to Simulate Urban Resource Demands at High Spatial and Temporal Resolutions. Journal of Industrial Ecology 16, 889–900. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1530-9290.2012.00486.x, doi:10.1111/j.1530-9290.2012.00486.x.

[8] Knapen, L., Adnan, M., Bellemans, T., 2023. Location disaggregation for zone based travel plans. Procedia Computer Science 220, 283–290. URL: https://linkinghub.elsevier.com/retrieve/pii/S1877050923005720, doi:10.1016/j.procs.2023.03.037.

[9] Lemp, J.D., McWethy, L.B., Kockelman, K.M., 2007. From Aggregate Methods to Microsimulation: Assessing Benefits of Microscopic Activity-Based Models of Travel Demand. Transportation Research Record: Journal of the Transportation Research Board 1994, 80–88. URL: https://journals.sagepub.com/doi/10.3141/1994-11, doi:10.3141/1994-11.

[10] McWethy, L.B., 2015. Comparing microscopic activity-based and traditional models of travel demand : an Austin area case study. URL: https://repositories.lib.utexas.edu/handle/2152/32263, doi:10.15781/T20K8Z.

[11] Moeckel, R., Donnelly, R., 2015. Gradual rasterization: redefining spatial resolution in transport modelling. Environment and Planning B: Planning and Design 42, 888–903. URL: https://journals.sagepub.com/doi/10.1068/b130199p, doi:10.1068/b130199p.

[12] Molloy, J., Moeckel, R., 2017. Automated design of gradual zone systems: An iterative algorithm to design optimally sized spatial zones suitable for spatial modeling, while respecting municipal boundaries. Open Geospatial Data, Software and Standards 2, 19. URL: http://opengeospatialdata.springeropen.com/articles/10.1186/s40965-017-0032-5, doi:10.1186/s40965-017-0032-5.

[13] Mustafa, R., Christie, J., Hildebrand, E., 2020. Developing an "intelligent" high-fidelity GIS-based travel demand model framework for improved network-wide traffic estimation. Canadian Journal of Civil Engineering 47, 885–897. URL: http://www.nrcresearchpress.com/doi/10.1139/cjce-2019-0085, doi:10.1139/cjce-2019-0085.

[14] Nouvel, R., Zirak, M., Coors, V., Eicker, U., 2017. The influence of data quality on urban heating demand modeling using 3D city models. Computers, Environment and Urban Systems 64, 68–80. URL: https://linkinghub.elsevier.com/retrieve/pii/S0198971516304306, doi:10.1016/j.compenvurbsys.2016.12.005.

[15] Openshaw, S., 1998. Towards a more computationally minded scientific human geography. Environment and Planning A: Economy and Space 30, 317–332. URL: https://doi.org/10.1068/a300317, doi:10.1068/a300317, arXiv:https://doi.org/10.1068/a300317.

[16] Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. URL: https://linkinghub.elsevier.com/retrieve/pii/0377042787901257, doi:10.1016/0377-0427(87)90125-7.

[17] Wong, C.H.H., Cai, M., Ren, C., Huang, Y., Liao, C., Yin, S., 2021. Modelling building energy use at urban scale: A review on their account for the urban environment. Building and Environment 205, 108235. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360132321006363, doi:10.1016/j.buildenv.2021.108235.