GRADUAL RASTERIZATION: REDEFINING THE SPATIAL RESOLUTION IN TRANSPORT MODELING

Rolf Moeckel (Corresponding Author)
Faculty Research Associate
Center for Smart Growth
University of Maryland
1112 Preinkert Field House
College Park, MD 20742
rolf.moeckel@udo.edu
Phone: (917) 215-4174

Rick Donnelly
Vice President
Parsons Brinckerhoff
6100 Uptown Blvd. NE, Suite 700
Albuquerque, NM 87110, USA
donnellyR@pbworld.com
Phone: (505) 878-6524

Words 5,155
Figures 9 (2,250)
Tables 1 (250)
Total 7,655

Presented at the 93rd Annual Meeting of the Transportation Research Board in Washington, D.C.

ABSTRACT

Finding the appropriate spatial resolution in modeling is a serious challenge at the beginning of every modeling project. The paper presents a methodology to adjust the spatial geography to the resolution of a network. Based on the quadtree algorithm, raster cells are generated that are dynamic in size. Smaller raster cells are used in urban areas and larger raster cells are used in low-density, rural areas. Trip tables of a travel demand model for the State of Georgia, U.S. are disaggregated to this new zone system of raster cells, and assignment results validate significantly better than when using the original zone system.

1 INTRODUCTION

Defining the most suitable level of spatial resolution in modeling often is considered to be more art than science. The challenge to create the ideal zone system exists equally for transportation, land use, environmental impacts or the economic modeling, and each field requires a different level of resolution. Current ad hoc methods of defining zones and networks, and especially the right level of balance between the two, have rarely been studied systematically. In many cases, the zone system is simply based on whatever zonal definition was available, regardless of the purpose for developing the original zone system.

It tends to be labor-intensive to change the zone system, and hence, is done rarely. Defining which neighborhoods belong together, selecting the appropriate zone size in different regions, and deciding which river or street warrants splitting a zone into two zones usually is done manually. After defining the zone system, an almost equal amount of labor is necessary to collect socio-economic data for every zone, commonly including households by income and household size, employment by industry, and other relevant data required for the model to be implemented. As a consequence, zone systems are rarely changed.

Keeping a constant zone system across several applications bears a couple of potential drawbacks. A zone system that is not detailed enough will be unable to capture relevant information. For example, a travel demand model with zones that are too coarse will generate too many intrazonal trips that are neglected in the assignment. This makes the trip length frequency distribution difficult to calibrate, as all intrazonal trips get assigned the same average intrazonal trip length. Matching the observed number of transit trips becomes nearly impossible if several transit stations are within the same zone and transit skims between these stations are set to 0 or a non-zero but arbitrary intrazonal travel time/distance. Land-use modeling becomes inapt if neighborhoods with very different characteristics are lumped together in the same zone. Agglomeration effects for business location choice modeling cannot be represented properly, as the distance between two firms within a zone is assumed to be the static intrazonal distance. If these two firms are in two neighboring zones, they may be located next to each other just across the zone boundary or they may be at the opposite side of each zone.

All above-mentioned problems are reduced with a finer zone system. Moving down to the parcel level, as shown in several research-oriented model setups [such as 1, 2, 3], would eliminate these issues largely. However, there are computational and conceptual limits to reducing the zone size. Logit models for land-use location choice decision or transportation destination choice decisions become instable if the number of alternatives is too large. A work-around could be implementing a two-layer decision making, in which first a neighborhood is selected and secondly a zone is chosen. However, with too many zones, several layers of decision-making would be needed, which imposes challenges due to lack of data, incomplete theory for multi-layer decision-making as well as computational limitations. Computational issues may also make the assignment unfeasible if too many zones are used. As convincingly shown by Wegener [4], more detail is not always better. A balanced level of resolution ought to be chosen in modeling.

2 STATE OF THE ART

Raster cells have been applied in urban simulations since the advent of computers. Commonly, they are easier to handle than amorphous zones. If the dimensions are known, the raster cell's size can be calculated quickly, the centroid can be easily determined, and it is fairly easy to identify neighboring zones. Using amorphous zones, such calculations are significantly more complex. An early example of using raster cells instead of zones is Schelling's Self-Forming Neighborhood Model [5], which shows segregation patterns due to households seeking neighborhoods of their own income level, race or political preferences. Andersson et al. [6] developed an urban growth simulation for a synthetic city, where locations are represented as raster cells. Other examples of using raster cells for a simulation experiment include the business location choice model of Khan et al. [7] or a test bed to analyze differences between simulating firms and simulating employees [8].

Hagen-Zanker & Jin [9] have developed a method they call adaptive zoning. Separately for every origin zone, destination zones are aggregated into larger zones. Zones close to the origin are left as they are, while zones further away from the origin are aggregated into larger zones. Areas with higher population densities tend to be kept more disaggregated than rural areas. The method has been applied to a commute trip model in England. The aggregation of destination zones reduces the number of OD pairs by 96% and the computation time by 70%, without a significant change of model results.

Openshaw [10] describes the Modifiable Areal Unit Problem (MAUP), which shows that spatial analysis depends on the zone sizes chosen. If neighborhoods are aggregated differently, the average income, for example, will show different results. Ophenshaw proposed the Automatic Zoning Procedure (AZP), which – despite its name – is a fairly manual process to aggregate neighboring zones iteratively until statistical measures express the purpose of the analysis most clearly. It is more common, that zone systems are not revised for a specific application. Usually, it is cost-prohibitive to modify the zone system, as methods are missing to systematically adjust the zone topology, and the time required to populate a revised zone system with socio-economic data and other attributes required by the model is far beyond the schedule and budget of most projects. Viegas et al. [11] analyzed MAUP in detail testing various zone systems and their impact on intra-zonal trips and zero-trip zones. They also explain well that a smaller zone sizes may increase the percentage of noise, defined as cells with a small number of trips that fall below the confidence interval.

Spiekermann and Wegener [12] developed a methodology to disaggregate zones to raster cells of 50 by 50 meters. Using a land use layer with a finer geography than the zone system, probabilities for population and employment are calculated for every raster cell, and socio-economic data are allocated to raster cells using Monte-Carlo Sampling. The method is used to calculate emission, which requires a more detailed spatial resolution than other model components. This methodology enables the modeler to quickly move from a fairly coarse zone layer to a detailed grid of raster cells.

Schwarze [13] uses hexagons instead of raster cells to calculate accessibilities. The area covered is likely to be better represented by a hexagon, because the distance from the centroid any location within the shape is smaller than for a square-shaped raster cell. Yet, hexagons can be used to cover an entire area without gaps. Comparing a hexagon and a square of equal size, the diameter of the hexagon is about 14 percent smaller than the diameter of the square. Hence, the area under a hexagon is likely to be more homogenous than the area under a raster cell. While the concept offers theoretical advantages, hexagons are computationally more complex than raster cells. This is probably the reason why hexagon applications are fairly rare in modeling practice.

It is rather uncommon to develop automated algorithms to create zone systems. An exception worth reading is the procedure developed by Martínez et al. [14]. Their approach defines travel patterns based on a household travel survey, and a GIS algorithm creates a zone system that minimizes intrazonal trips and aims at creating zones with a rather homogenous trip generation patterns. Another rare approach of automatically generating zones was developed by Ding [15]. His GIS/Fortran tool aims at creating zones that are homogeneous, exclusive, unique, complete and carry “equity” in trip generation. He evaluated eight different zone systems for South Korea by share of intrazonal trips and change of congestion in a travel demand model.

The literature review revealed a long history of analyzing zonal systems for modeling. While the need for zone systems that provides an appropriate level of detail has been demonstrated repeatedly, methodologies are limited to systematically generate zone systems that offer sufficient detail without leading to excessive model runtimes. Existing approaches are computationally complex and tend to create irregular shapes with some long and narrow zones. It has been shown that varying levels of detail may be appropriate for different modeling tasks, but only few studies dealt with more than one zone system. As the impact of the zone system on model results is substantial, there is a need for new approaches to redefine geography in modeling.

3 GEORGIA STATEWIDE TRAVEL DEMAND MODEL

To analyze various transportation improvements along Interstate Highway 75 from Atlanta GA to Chattanooga TN, the statewide model of the Georgia Department of Transportation (GDOT) was applied. The GDOT model is a state-of-the-art statewide model that covers the state of Georgia with 1715 fairly detailed zones and the rest of North America with 527 zones that become successively larger the further they are way from Georgia. The GDOT model covers both person travel and freight trips, separating short-distance and long-distance trips. Mode choice is calculated for both person and freight travel. The model is documented by Atkins [16].

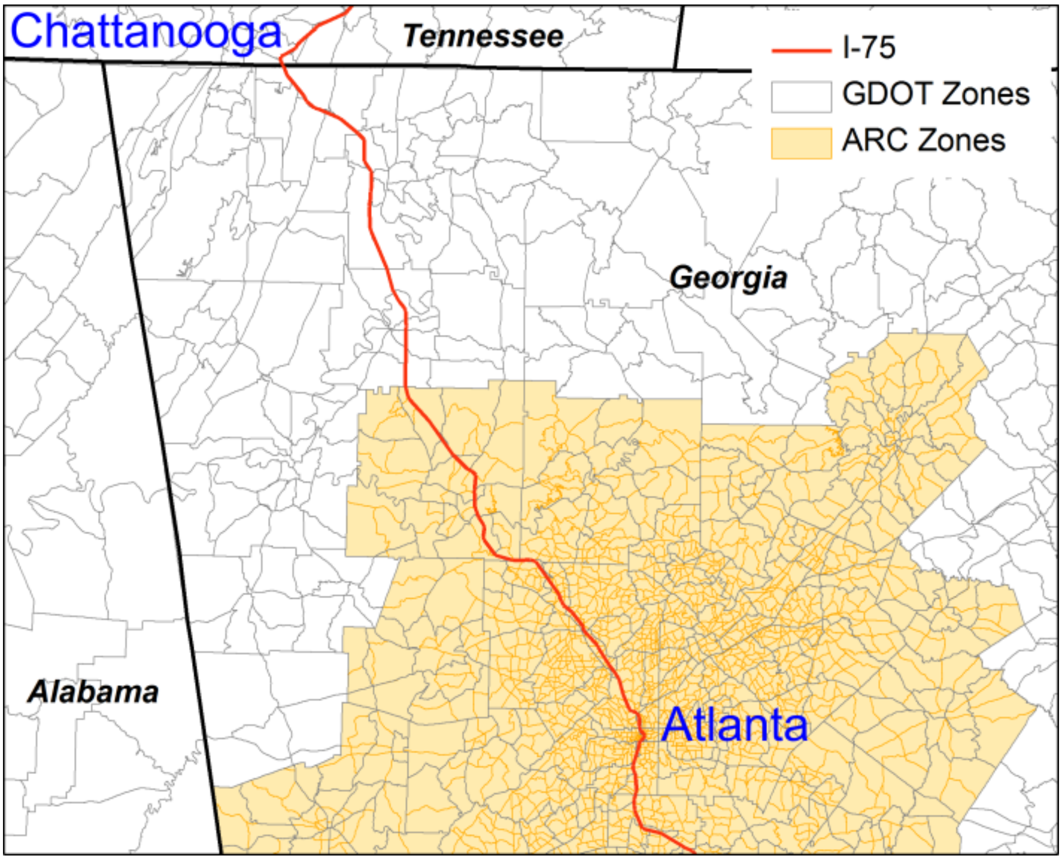

FIGURE 1: GDOT and ARC zones

The corridor is shown in Figure 1, where GDOT zones are shown in grey and zones of the urban model for the Atlanta Metropolitan Council (ARC) are shown in yellow. Figure 2 shows how well the GDOT model validates along the I-75 corridor from Atlanta (left side) to Chattanooga (right side). The blue line shows the model volumes, and the red line shows count data, which are not available for every link. While the model matched observed count data in the regions outside of Atlanta quite well, the model overestimated travel demand within the Atlanta metropolitan area on this I-75 corridor quite a bit. The first link on the left side of this graphic shows an overestimation of 45%. In defense of the statewide model, it has to be noted that this model was built for the non-urban/intra-urban areas of Georgia. Staff members of ARC conduct their own travel demand analysis, and the GDOT model was never intended to be used within the ARC area. The ARC model, however, does not cover the entire corridor (of the corridor shown in Figure 2, only the counties Cobb and Bartow are covered by the ARC model). Thus, the GDOT model was used in this study.

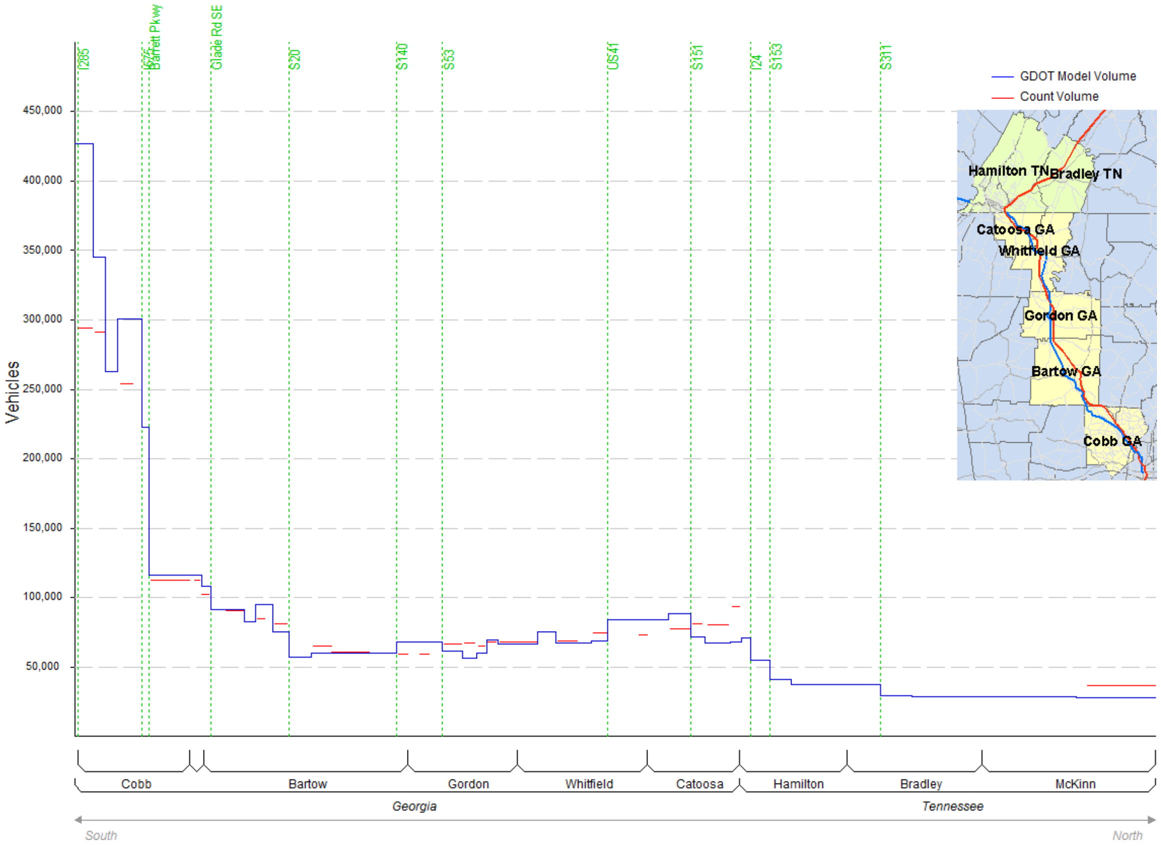

FIGURE 2: Validation of traffic volumes on I-75 corridor

In this particular corridor study, model volumes that match counts reasonably well along the entire corridor were deemed to be necessary to understand long-distance flows. A series of test runs were conducted to better understand why the GDOT model overestimates volumes in the Atlanta region. Tests included:

a) Change of speed on single-lane streets outside MPO areas
b) Scaled population to census data and employment to county business pattern data
c) Scaled GDOT data to ARC data within ARC region
d) Replaced GDOT trip tables with ARC trip tables
e) Added network detail near I-75 and I-285
f) Split selected GDOT zones into smaller zones

None of these test runs resulted in a breakthrough; however, adding more geographic detail in test runs e) and f) showed some improvement. The team decided to systematically add more geographic detail throughout the study area. Instead of manually adding detail to every single zone, an algorithm was developed to automatically scale geographic detail with density. Areas with higher density deserve more detail, as a larger number of trips is generated and attracted by these neighborhoods. Low-density regions, in contrast, can be represented by fairly large zones, as only few trips are affected.

4 RASTERIZATION FOR GEORGIA

Because of their flexibility in size, raster cells were chosen as the spatial unit. As described in the literature review, raster cells offer an enormous flexibility, are easy to use computationally and allow calculating areas and direct distances instantaneously. And in addition, using squared raster cells allows adding four raster cells to one larger raster cell, or vice versa, subdividing one larger raster cell into four equally-sized smaller raster cells.

The methodology developed for this corridor study is shown in a flowchart in Figure 3. Under 'Data preparation', the study area is rasterized into the smallest raster cells under consideration. For computational reasons, the smallest number of raster cells should be a multitude of 2, i.e., the total number of these smallest raster cells in x and in y direction should be either 2, or 4, 8, 16, 32, etc. In this case, the study area was rasterized to 4,096 x 4,096 smallest raster cells. This number was found iteratively to ensure that the desired level of detail in the densest part of the study area could be represented. Using the point-in-polygon algorithm, population is disaggregated from zones to raster cells. To increase the spatial level of detail, the ARC zones were used within the ARC region, while GDOT zones were used outside the ARC region (compare Figure 1). Lacking any further spatial information (such as land use or land cover), population was spread evenly across all raster cells within one zone.

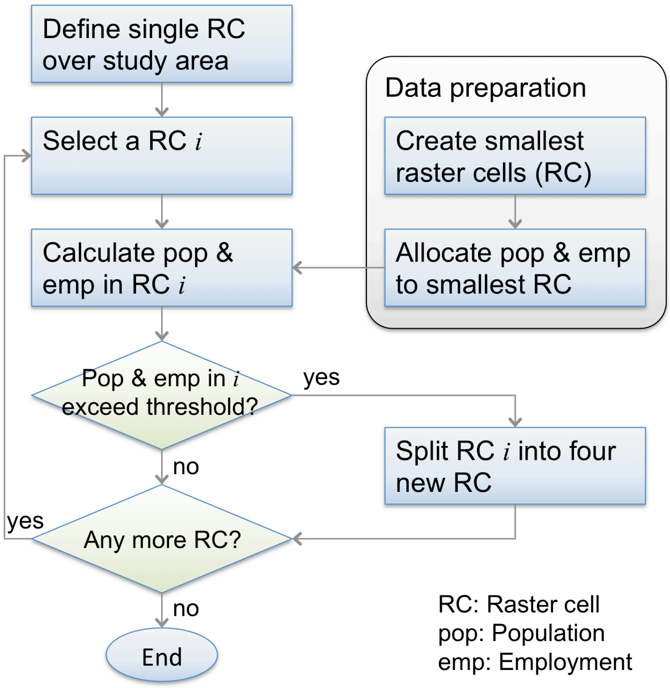

FIGURE 3: Rasterization procedure

The rasterization process starts with one large raster cell covering the entire study area, in this case the state of Georgia. This raster *i* is selected and its total populating and employment is calculated based on the smallest raster cells. Next, the quadtree concept is applied to build raster cells of various sizes. Finkel and Bentley [17] developed the quadtree concept as a computer algorithm to efficiently store data. Samet and Webber [18] first applied this concept to store polygon data. If the population and employment of raster cell *i* is larger than a threshold value (which certainly is the case for this single first raster cell), the raster cell is split into four raster cells of equal size. This step creates the change between the first and the second graphic in Figure 4. Next, the population in each of the four raster cells is compared to the threshold value, and whenever a zone's population exceeds the threshold value, the zone is subdivided into four smaller zones.

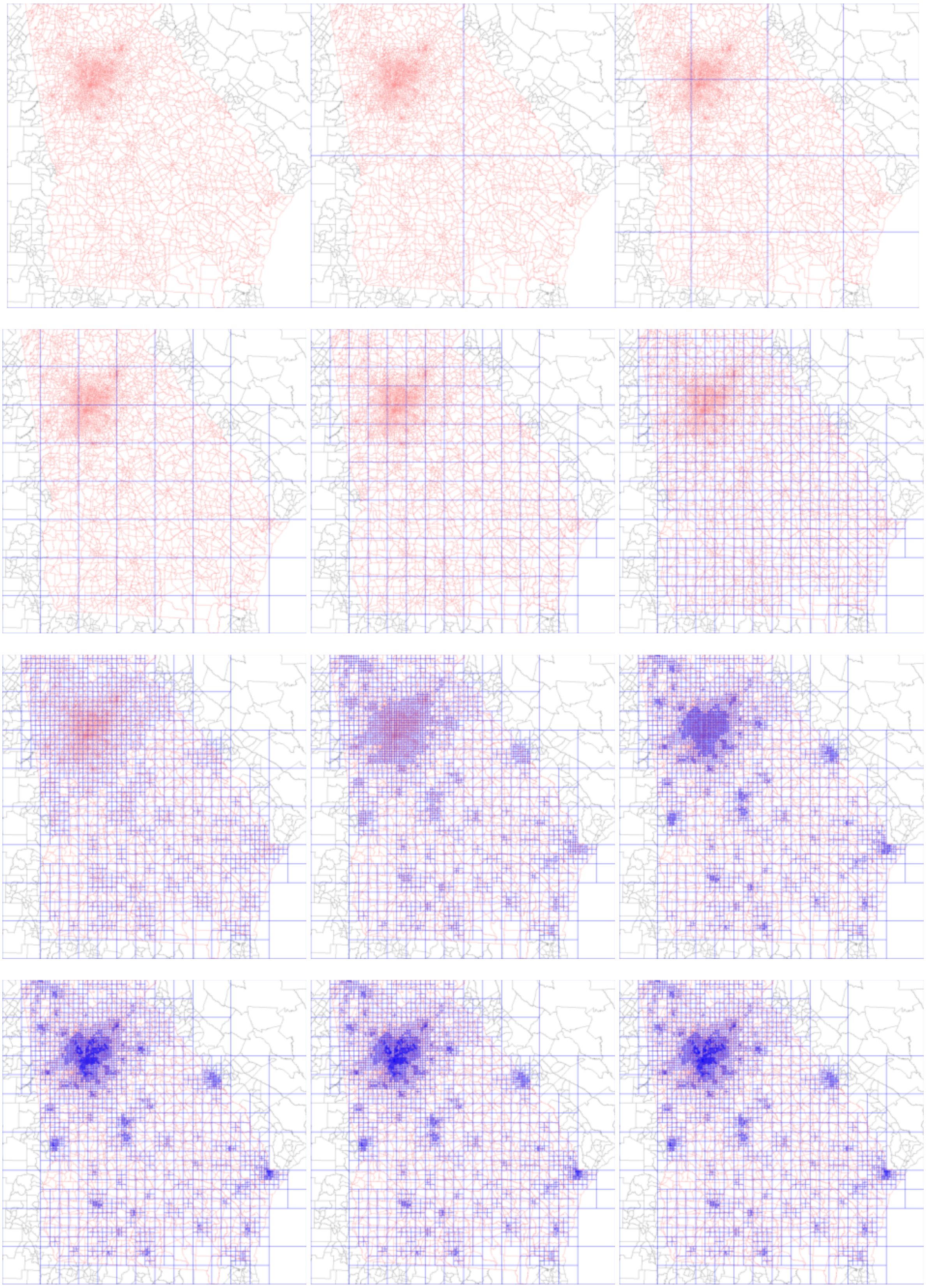

FIGURE 4: Rasterization of zones in the State of Georgia

Figure 4 shows all 12 steps that were necessary to ensure that no single raster cell's population and employment exceeds the threshold value. Areas with greater density, such as Atlanta or Savannah on the Atlantic coast, received a finer grid of raster cells, while rural areas are represented by raster cells that are much larger. The smallest zone size (generated in downtown Atlanta) is 250 x 250 meters. The total number of raster cells is 4,909.

The threshold value used to define if a raster cell shall be split further was defined iteratively. As many trips are generated by households but attracted by employment, using households only would miss subdividing employment centers accordingly. Therefore, population and employment were added, and whenever this sum exceeded the threshold value, raster cells were further subdivided. This helped creating the necessary level of detail at both the origins and the destinations of most trips. The threshold values for [population + employment] that were tested are shown in Figure 5. A negative power curve describes the relationship between the threshold value used and the number of cells generated. For efficiency reasons, it was decided not to exceed 5,000 raster cells. This number was almost reached with a [population + employment] threshold of 5,000. Reducing the threshold further would rapidly increase the number of zones, making the data processing computationally more cumbersome.

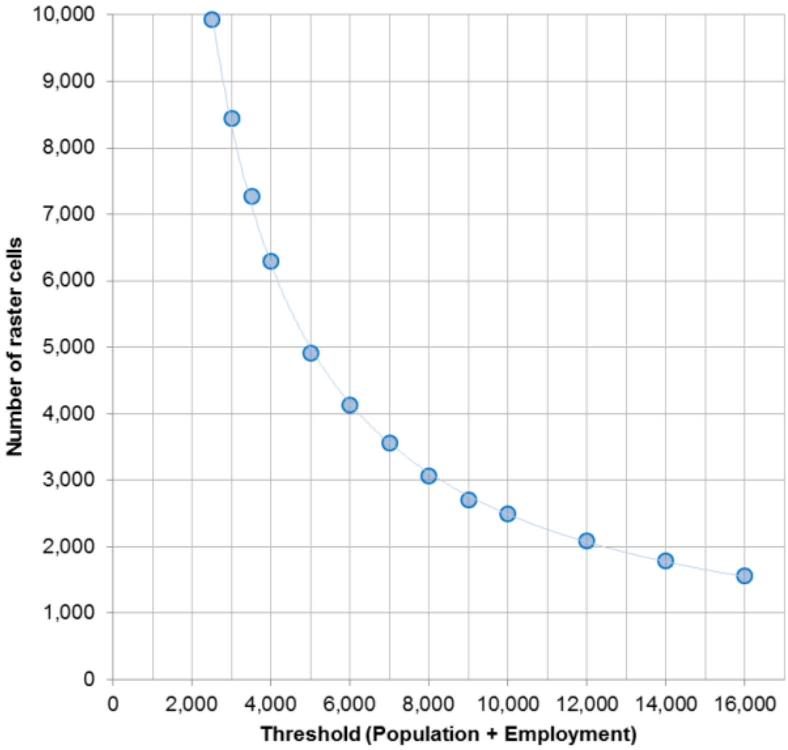

FIGURE 5: Raster cells depending on threshold value

Figure 6 shows a comparison of zones and raster cells for the corridor area. Within the Atlanta region and other urban areas, the zone system became much more detailed. In the rural areas, sizes or raster cells are comparable to zone size. In other words, the rasterization increased the number of zones in urban area (where the GDOT model did not validate as well) and kept about the same number of zones in rural areas (where the GDOT model performed very well).

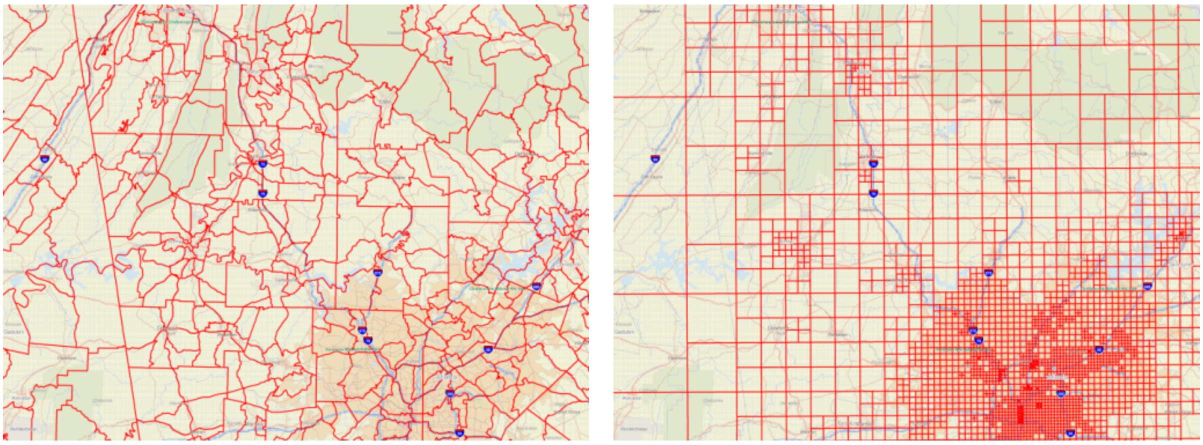

FIGURE 6: Zones and raster cells in Northwest Georgia

5 TRAVEL DEMAND MODEL DISAGGREGATION

After creating this finer zone system of raster cells, the auto trip tables of the GDOT model were disaggregated to this new zone system. It was considered to rebuild the entire model to this finer zone system, though it was concluded that been calibrated to the GDOT zone system, and as such was performing well in most parts of the study area. Instead of building a new model with this revised zone system, the auto trip tables of the GDOT were disaggregated to raster cells of the new zone system. This disaggregation was done for autos and truck separately, and given the focus of this study, only zones within Georgia were disaggregated. Transit trips, though represented in the Georgia statewide model, were disregarded in this study.

Trip ends outside of Georgia are not changed, but trip ends within Georgia are disaggregated from GDOT zones to raster cells. Population and employment by raster cell are used to disaggregate trips. As raster cells do not nest within GDOT zones, the proportional share covered by each zone was used to proportionally allocate population and employment shares to each part of a raster cell covered by different zones. Centroid connectors were built allowing one connector per raster cell and connecting to any link but interstate and freeways. Given the number of raster cells, centroid connectors were allowed to split links. Travel demand was assigned to the network using GDOT's settings for a user equilibrium assignment. The disaggregation of zonal trip tables to raster trip tables runs in approximately 30 minutes (including both autos and trucks). As the GDOT model is a daily model, additional runtime needed for the final assignment is less than ten minutes.

6 INTRAZONAL TRIPS

When changing the zone size, it is important to pay attention to the number of intrazonal trips. When smaller zones are aggregated into one larger zone, trips that used to be interzonal become intrazonal. Vice versa, the disaggregation of zones leads to an increase of interzonal trips. This is particularly relevant if the level of detail of the network is not changed. If trips that used to be intrazonal (and therefore, were not assigned to the network) become interzonal, congestion will increase if the network is not expanded accordingly. In transportation modeling, it is crucial to maintain a balanced level of detail between the zone system and the network.

In this particular application, zones are disaggregated, while the network is not changed. Because of this disaggregation, trips that used to be intrazonal become interzonal, and traffic rises to a level that is not supported by the GDOT network. Adding network detail throughout the state would be a manual process, with a lot of uncertainty how much detail is "enough". Instead, it was decided to keep the network unchanged and adjust the trip table accordingly. Figure 7 shows an example of inconsistent network and zone detail. A trip from raster cell A to raster cell B, that used to be intrazonal, now would have to travel along the dashed green line to follow the available network. In reality, however, there are local streets between A and B that are not represented in the network. Travelers from A to B would use these local roads (shown with the blue arrow) rather than traveling a big detour (shown with the green dashed line).

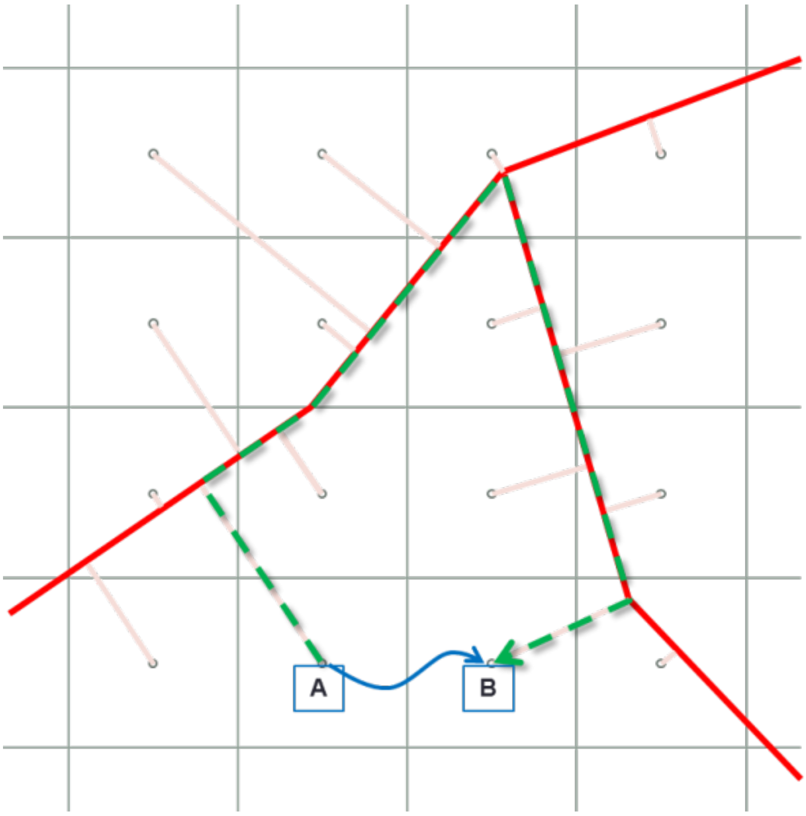

FIGURE 7: Inconsistency between zone size and network detail

To account for this travel behavior on a sparse network, the direct travel time is compared to the travel time on the GDOT network. If the direct path on local roads that are not represented in the GDOT network is shorter than the actual path on the GDOT network, the trip is dropped from the assignment. This way, local trips that used to be intrazonal (and therefore, were not added to the assignment) are dropped and omitted in the assignment step when using raster cells as well. This way, a balance between network density and number of trips in the assignment is retained.

To find out if the trip on the local network not represented by the GDOT network would be shorter than the trip following GDOT network links, a phantom speed on local roads was calibrated. The direct distance from one raster cell to another was divided by the phantom speed to calculate travel time on the local network. The phantom speed was calibrated to ensure that the number of dropped trips was equal to the number of intrazonal trips in the original GDOT model. The phantom speed was calibrated to be 20 mph (or 32 kmh).

The count volume on I-75 just north of I-285 is 294,040 vehicles. The original GDOT model overestimated this value by 31 percent. The ARC model matched this value very well with -4%. The raster cell model achieves and equally good match with +3%.

Figure 8 shows the validation of I-75 along the entire corridor. The grey line shows the original model volume (also shown in *FIGURE 2*), and the blue line shows the volume of the raster cell model. The overestimation of traffic volumes in Cobb County has been resolved, while traffic volumes elsewhere were not affected much.

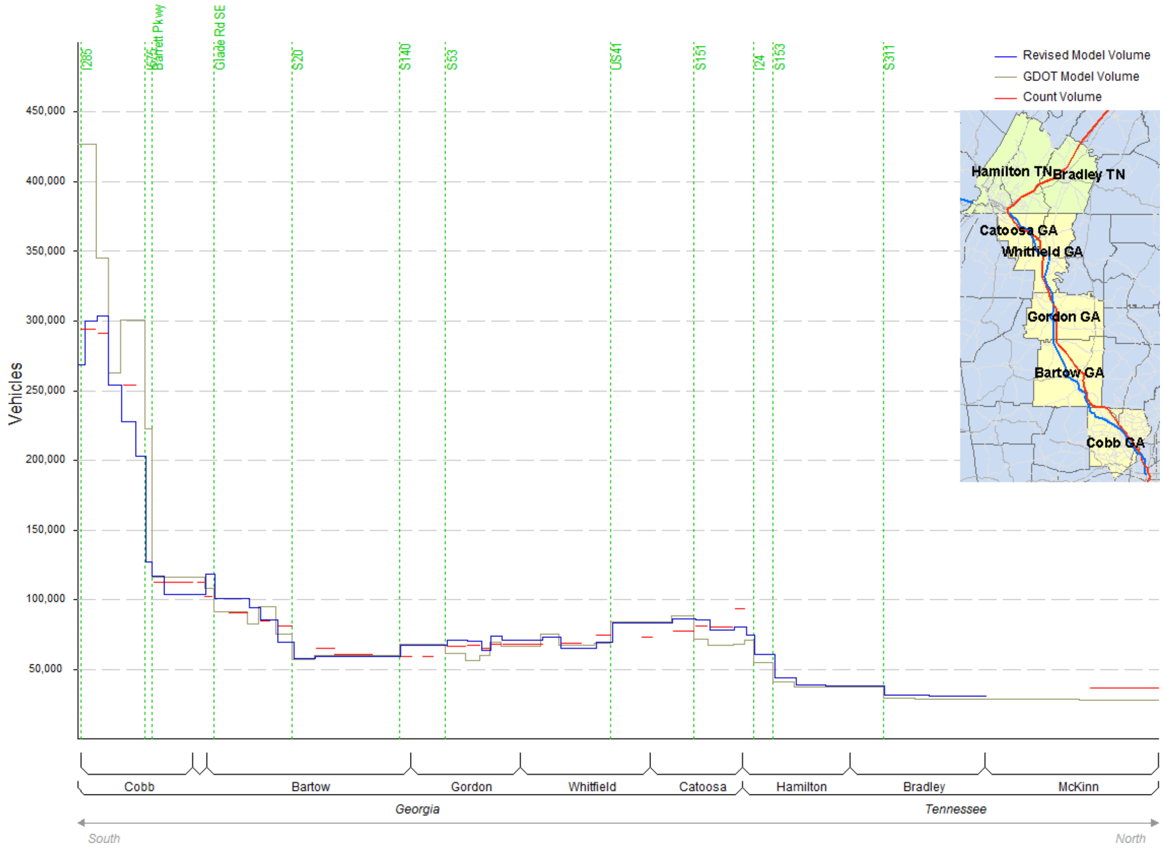

FIGURE 8: Validation of traffic volumes along the I-75 corridor

Table 1 summarizes validation results statistically. The upper section “Northwest” covers count data available within the 15-county area in northeast of Georgia. The 15 counties are shown in yellow in Figure 9. The second part of Table 1 labeled “I-75” shows summaries for all counts on I-75. Using the raster cells improved the percent root mean square error (%RMSE) from 77% to 48% for the northeast of Georgia.

TABLE 1: Model validation in the 15-county area

Region	Model	Sum Counts	Sum Model	Ratio	R^2	RMSE	%RMSE
Northwest	GDOT	10,618,038	11,343,825	1.068	0.897	12,484	77%
	Raster	10,618,038	11,351,367	1.069	0.931	7,810	48%
I-75	GDOT	2,812,130	3,019,897	1.074	0.967	30,743	30%
	Raster	2,812,130	2,822,301	1.004	0.982	9,276	9%

Figure 9 visualizes the percent root mean square error by county, both for the original GDOT model (x axis) and for the raster cell model (y axis). Almost all counties improved substantially. The only county that validated noteworthy worse is Gilmer County, where the

%RMSE increased from 63% to 72%. This is a quite rural county that used to be represented by 6 zones. Even though the raster model assigns approximately 12 raster cells to this county (raster cells do not nest within counties), the zone system with a central zone covering the largest city in this county, Ellijay, gave a better representation for origins and destinations than the coarse raster cells.

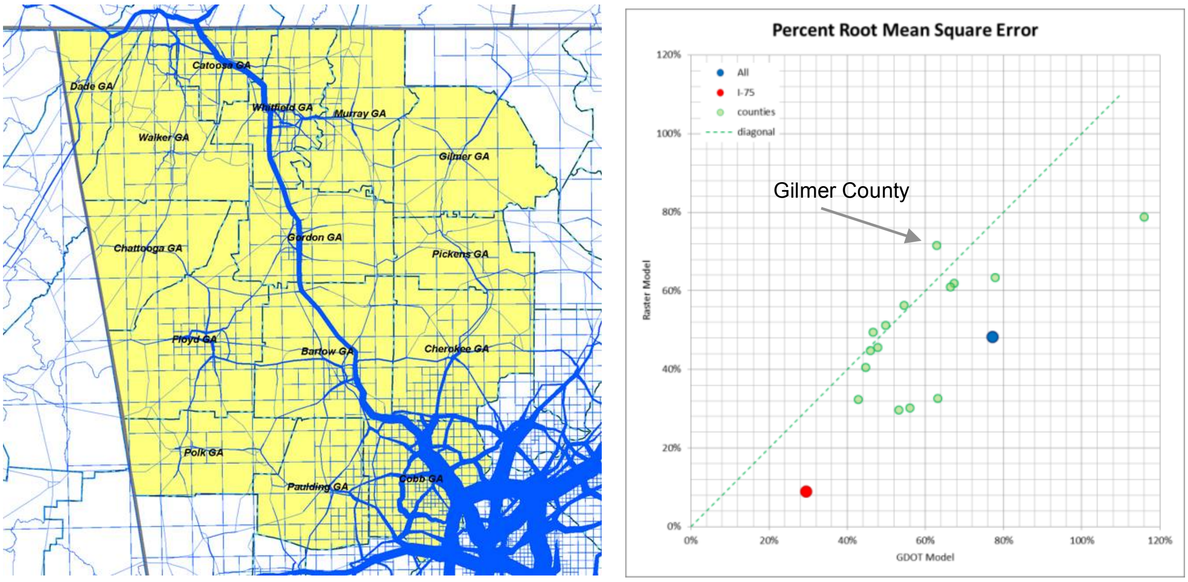

Figure 9: Validation area Northwest Georgia (left) and Root Mean Square Error by county (right)

Most counties, however, improved substantially according to Figure 9. Raster cells have proven to overcome the issue of insufficient resolution for the I-75 corridor in northwest Georgia.

7 CONCLUSIONS

The research presented in this paper shows that model results can be improved substantially by finding an appropriate level of spatial resolution. While this paper does not generalize which level of detail is appropriate, it is remarkable that the model validation improved without changing the model design from trip generation to trip distribution and mode choice. Only the spatial resolution of the assignment step was adjusted, which improved the model performance.

In this particular application, no household travel survey was available. It will be a valuable exercise to apply this methodology for a model for which such survey is available. The micro location provided by the survey may further improve the allocation of trips to raster cells.

The current implementation uses population and employment as a threshold to further subdivide zones. As population and employment are used for trip generation, it might improve the methodology if trip generation of households and employment was used as a threshold value. For example, larger households tend to make fewer trips per person than smaller households. As a consequence, neighborhoods with smaller household sizes may deserve more spatial detail than neighborhoods with larger household sizes, even if the number of residents is identical between these two neighborhoods. Even though the use of trip generation instead of densities is not expected to alter the results substantially (after all the two correlate to a large degree), current research investigates the benefit of using trip generation.

It is conceivable that increasing the spatial unit further might improve the assignment results even more. Current research underway will use the zone systems shown in Figure 5 and assign trips to these different raster cell systems. It is expected that the model results will gradually get better the more raster cells are used. However, there will be a limit after which no further benefits from additional disaggregation are expected. The network does not grow accordingly, and adding more raster cells where there is no network to support feeding these raster cells will not improve results. In addition, run times will increase steeply with additional zonal detail, and at some point the assignment will become unstable. Future research will help defining the most appropriate level of detail for a given resolution of the network.

Vice versa, it would be ideal to develop another methodology that adjust the resolution of the network to fit a given zone system. The authors are thinking about a procedure where the finest network available would be taken as a starting point (such as open street network), and links that would not be necessary for a given zone system would be dropped successively. This procedure certainly would affect the number of intrazonal trips to be removed (as discussed in section 6). Ideally, the network would be just fine-grained enough that no adjustment to intrazonal trips is necessary. Future research will analyze the possibilities to create flexible networks that adjust to the level of resolution of the zone system.

The technique presented in this paper is one possible method of letting the data (settlement patterns) define the appropriate level of spatial detail rather than using arbitrary definitions of zones. Indeed, there is scant evidence that different zonal definitions are compared, for it is laborious to define just one zone system. The presented technique allows varying the size of the cells in order to assess their effect on model performance. Moreover, the method is quick and has parsimonious data requirements, substantially reducing the cost of doing so.

One might argue that the push towards activity-based models may lead to simulations where trips are represented between point locations rather than zones (as demonstrated, for example, in MATSim [19]). A micro-representation of geography might obviate the need for zones in the first place. While modeling without artificial zone systems is an attractive future direction in transportation modeling, conceptional, computational and data-related limitations will continue to call for reasonable zone systems for the time being.

8 REFERENCES

1. Waddell, P., T. Moore, and S. Edwards, *Exploiting Parcel-Level GIS for Land Use Modeling*, in *ACSE Conference. Transportation, Land Use, and Air Quality: Making the Connection*1998: Portland, OR. p. 10.
2. Venter, C.J., T. Lamprecht, and W. Badenhorst. *Demographic and regional economic modeling using stochastic allocation in the City of Johannesburg*. in *Transportation Research Board (TRB)*. 2006. Washington, D.C.
3. Pendyala, R.M., et al., *An Integrated Land Use–Transport Model System with Dynamic Time-Dependent Activity-Travel Microsimulation.* Transportation Research Record: Journal of the Transportation Research Board, 2013. **2303**: p. 19-27.
4. Wegener, M., *From Macro to Micro - How Much Micro is too Much?* Transport Reviews, 2011. **31**(2): p. 161-177.
5. Schelling, T.C., *Micromotives and Macrobehavior*1978, New York, London: W. W. Norton & Company. 252.
6. Andersson, C., et al., *Urban Growth Simulation from 'First Principles'.* Physical Review E, 2002. **66**(2): p. 1-9.

7. Khan, A.S., J.E. Abraham, and J.D. Hunt. *Agent-based Micro-simulation of Business Establishments*. in *Congress of the European Regional Science Association (ERSA)*. 2002. Dortmund.
8. Moeckel, R., *Firm Location Choice Versus Job Location Choice in Microscopic Simulation Models*, in *Employment Location in Cities and Regions. Models and Applications*, F. Pagliara, et al., Editors. 2013, Springer: Heidelberg, New York, Dordrecht, London. p. 223-242.
9. Hagen-Zanker, A. and Y. Jin, *A New Method of Adaptive Zoning for Spatial Interaction Models.* Geographical Analysis, 2012. **44**(4): p. 281–301.
10. Openshaw, S., *The modifiable areal unit problem*, in *Concepts and Techniques in Modern Geography (CATMOG)*, Institute of British Geographers, Editor 1984, Geo Books: Norwich, England. p. 41.
11. Viegas, J.M., L.M. Martinez, and E.A. Silva, *Effects of the modifiable areal unit problem on the delineation of traffic analysis zones.* Environment and Planning B-Planning & Design, 2009. **36**(4): p. 625-643.
12. Spiekermann, K. and M. Wegener, *Freedom from the tyranny of zones: towards new GIS-based models*, in *Spatial Models and GIS. New Potential and New Models*, A.S. Fotheringham and M. Wegener, Editors. 2000, Taylor & Francis: London. p. 45-61.
13. Schwarze, B., *Kleinräumige Erreichbarkeitsanalysen mit GIS*, in *mobil.TUM 2008 – International Conference on Mobility & Transport*2008: Munich, Germany. p. 11.
14. Martinez, L.M., J.M. Viegas, and E.A. Silva, *A traffic analysis zone definition: a new methodology and algorithm.* Transportation, 2009. **36**(5): p. 581-599.
15. Ding, C., *The GIS-based human-interactive TAZ design algorithm: examining the impacts of data aggregation on transportation-planning analysis.* Environment and Planning B-Planning & Design, 1998. **25**(4): p. 601-616.
16. Atkins, *Development of Statewide Model Report. Model v2.0 released 4/15/2011. Draft Report.*, 2011: Atlanta, GA.
17. Finkel, R. and J.L. Bentley, *Quad trees a data structure for retrieval on composite keys.* Acta Informatica, 1974. **4**(1): p. 1-9.
18. Samet, H. and R.E. Webber, *Storing a Collection of Polygons Using Quadtrees.* ACM Transactions on Graphics, 1985. **4**(3): p. 182-222.
19. Balmer, M., et al., *Agend-based simulation of travel demand: Structure and computational performance of MATSim-T*, in *2nd TRB Conference on Innovations in Travel Modelling*2008: Portland, OR.