Research Article

Adaptive Zoning for Transport Mode Choice Modeling

Alex Hagen-Zanker and Ying Jin

Department of Architecture, University of Cambridge

Abstract

Adaptive zoning is a recently introduced method for improving computer modeling of spatial interactions and movements in the transport network. Unlike traditional zoning, where geographic locations are defined by one single universal plan of discrete land parcels or 'zones' for the study area, adaptive zoning establishes a compendium of different zone plans, each of which is applicable to one journey origin or destination only. These adaptive zone plans are structured to represent strong spatial interactions in proportionately more detail than weaker ones. In recent articles, it has been shown that adaptive zoning improves, by a large margin, the scalability of models of spatial interaction and road traffic assignment. This article confronts the method of adaptive zoning with an application of the scale and complexity for which it was intended, namely an application of mode choice modeling that at the same time requires a large study area and a fine-grained zone system. Our hypothesis is that adaptive zoning can significantly improve the accuracy of mode choice modeling because of its enhanced sensitivity to the geographic patterns and scales of spatial interaction. We test the hypothesis by investigating the performance of three alternative models: (1) a spatially highly detailed model that is permissible to the maximum extent by available data, but requires a high computational load that is generally out of reach for rapid turnaround of policy studies; (2) a mode choice model for the same area, but reducing the computational load by 90% by using a traditional zone system consisting of fewer zones; and (3) a mode choice model that also reduces the computational load by 90%, but based on adaptive zoning instead. The tests are carried out on the basis of a case study that uses the dataset from the London Area Transport Survey. Using the first model as a benchmark, it is found that for a given computational load, the model based on adaptive zoning contains about twice the amount of information of the traditional model, and model parameters on adaptive zoning principles are more accurate by a factor of six to eight. The findings suggest that adaptive zoning has a significant potential in enhancing the accuracy of mode choice modeling at the city or city-region scale.

1 Introduction

Understanding how people choose among different transport modes, and how investment, pricing and regulation can modify this behavior, has been a central concern in sustainable transport policy (Banister 2000). Mode choice modeling, and in particular the discrete choice model (Domencich and McFadden 1975, Ben-Akiva and Lerman 1985, McFadden 2007) is therefore a crucial element in many land-use and transport interaction models. Those

Address for correspondence: Ying Jin, Department of Architecture, University of Cambridge, 1-5 Scroope Terrace, Cambridge CB2 1PX, United Kingdom. E-mail: yj242@cam.ac.uk

Acknowledgements: This research took place as part of the Energy Efficient Cities initiative, which is a project funded by the UK Engineering and Physical Sciences Research Council (EPSRC) at the University of Cambridge (see http://www.eeci.cam.ac.uk). We wish to thank Mr Dale Campbell and his colleagues at Transport for London, who provided the 1991 LATS data, and Mr Gavin Gaunt and his colleagues at the UK Department for Transport who granted the access to Version 3.0 of the LASER model. We also wish to thank Professor Marcial Echenique and Professor Koen Steemers of Cambridge University, Professor Mike Batty of University College London and Professor Michael Wegener of Spiekermann and Wegener Urban and Regional Research for encouragement and comments on this research. We thank the anonymous reviewers for their useful feedback. As usual we are solely responsible for any views, remaining errors and omissions.

integrated models are computationally expensive, however: running one simulation may take days (Wegener 2001, Jin et al. 2002, Arentze and Timmermans 2004). In practice, compromises on geographical coverage and resolution are required to ensure acceptable model run turnaround time, not only for policy simulation and appraisal, but also for calibration, validation and sensitivity analysis. These compromises may be particularly detrimental to mode choice modeling; and strategic zone design is one option to optimize the trade-off between computational cost and precision.

The problem of zone system design has long been associated with the Modifiable Area Unit Problem or MAUP (Openshaw 1978, Putman and Chung 1989, Fotheringham and Wong 1991, Fotheringham et al. 1995). MAUP is not a single problem, but rather describes a range of issues that occur when discrete zones represent continuous geographical space. One issue is that for particular models, the size of zones determines the scale of the analysis and thereby the processes that can be represented or observed. Tobler (1989) argues that the occurrence of this problem indicates nothing but a poorly specified model. In mode choice models, as in this article, this problem does not occur as the scale of the modeled processes is determined by parameters applied on distance terms. A second issue is that the spatial granularity may be insufficient to represent certain processes. This problem is related to the Nyquist rate in information theory (Landau 1967); it means that in order to observe a process at a certain frequency it is necessary to have a sampling system of at least double the frequency. Spatially, this means that the characteristic length of zones must be smaller than the distance over which interactions take place. A third issue is the spatial aggregation error. Aggregation error occurs when the spatial variability of a variable becomes misrepresented; for instance when the variable is considered to be homogeneously distributed within a zone or, alternatively, concentrated in a single point. Spatial aggregation error is a major concern in transport modeling (Miller 1999), where the practice of concentrating all trips to and from a zone in a single node can lead to spurious levels of modeled congestion. In transport modeling, the problem is usually referred to as Transport Analysis Zone design (Ding 1998; Chang et al. 2002; Martinez et al. 2005, 2009; Viegas et al. 2009). Spatial aggregation also affects the precision with which distances between locations are represented, since the measured distance between two zones is only an approximation of the actual distance between locations, objects or agents within the zones (Hillsman and Rhoda 1978, Current and Schilling 1990). A fourth problem is the more general problem of making statistical inferences from aggregated variables (Moulton 1990), which affects much zone based analysis (Williams 1976; Fotheringham and Wong 1991).

Except for the first, all issues can be mitigated by using smaller zones. However, having smaller zones implies having more zones and an increased computational load. When the computational load is a limiting factor, it becomes attractive to employ zone systems that are crafted for efficiency. Common strategies are to control the shape of zones, because compact shapes have smaller aggregation errors, and to concentrate geographic detail where it improves results most strongly; i.e. to optimize the zone system to the model at hand (Openshaw 1977a, 1977b; Ding 1998; Martinez et al. 2009). With improved computing power, MAUP has become a less pertinent issue in the GIS literature. However, transport and land use modeling remain severely constrained by computing capacity and improvements in efficiency are still hard-sought.

Study area boundaries are also known to affect study results (Hartell 2007), although this issue has received less attention in the literature, with the positive exception of the field of Landscape Ecology (Saura and Martinez-Millan 2001, Karau and Keane 2007).



Figure 1 Distribution of commuting trips by distance and modal share in England 2001 (Source: UK Census 2001, excluding trips over 60 km and minor transport modes)

The tension between extent and resolution comes to the fore when modeling mode choice: some modes require representation in fine zones, whereas others require a large extent. Slow modes such as walking and bicycling concentrate in the short distance range (under 10 km) and need a fine-scaled zone system for realistic modeling. Suburban trains are predominantly used for long distance trips (especially beyond 10 km) and therefore require a larger study area. Notably, cars have a significant share across all distances. Other forms of public transport (principally buses) have a fair spread across the 2–20 km range (as an example see England's commuting statistics in Figure 1). Differences in distance that may be trivial for car travel can be decisive for travel on foot or by bicycle. Figure 1b illustrates that understanding both short and long distance transport patterns is important beyond the issue of mode choice: the majority of commuting trips take place over short distances (0–5 km), whereas long distance trips (>30 km) are a small fraction of the total. Yet, the mere 5% of commuting trips further than 30 km do account for 25% of the total passenger km in England.

Current trends are adding to the urgency of the above problem:

- In most city regions, the catchment of day-to-day travel has been growing strongly as a result of the provision of motorways, higher speed rail and regional air services, rising income and transition towards a knowledge economy; the appropriate spatial extent of study areas for sustainable transport policy has been increasing as a result.
- There is an increasing interest in walking and cycling as health-enhancing, green and economical alternatives to motorized transport (Pucher et al. 2010).
- A large pent-up demand exists for expanding the model size, in terms of spatial extent, resolution and thematic detail. For example, microscopic simulation of travel behavior, movements and traffic in complex urban networks can reveal significant additional insights

into the effects of travel demand management policies such as road pricing (Eliasson and Mattsson 2001).

Adaptive zoning is a method for optimized zone design specifically aimed at representing location-to-location relationships. Like traditional methods of zone design, it seeks to optimize computational efficiency by using compact zones and by providing spatial detail where it improves the model most. Adaptive zoning however, creates a distinct zone plan for each interaction origin (or destination, as will be discussed) called its *neighborhood*. Efficiency is gained when the zone composition of the neighborhoods is optimized to the patterns of interaction with the associated origin. Since spatial interactions typically diminish with distance, this optimization generally means that the neighborhood will consist of small zones at short distances and large zones at longer distances. Compared with the traditional approach, adaptive zoning reduces the number of modeled zone pairs from n^2 to $n \times s$, where n is the number of atomic zones and s is the number of zones in each neighborhood. Since s can be independent of other model dimensions, adaptive zoning based models typically scale linearly with study area extent, making them particularly attractive for large-scale applications.

This article follows up on the earlier work that introduced the concepts of adaptive zoning and demonstrated the potential on a highly simplified spatial interaction model. The contribution of the current article is to confront the method of adaptive zoning with an application of the scale and complexity for which it was intended. The hypothesis is that adaptive zoning is more efficient in representing complex spatial interaction patterns than using traditional zoning systems, especially when the model is required to realistically represent interactions over different geographical scales and facilitated by different transport networks. This hypothesis will be tested by a systematic comparison between traditional and adaptive zoning based models. Besides this critical test, this article will evaluate crucial implementation decisions, notably alternative aggregation criteria at the core of the method.

This article tests the hypothesis by comparing three zone systems for the same study area at the metropolitan scale. The first presents the study area in fine spatial detail using a traditional zone system. The second is a traditional zone system but with larger, aggregated zones. The third system is an adaptive zone system that defines its origin zones based on the first zone system and has the same number of zone pairs as the second system. We compare the performance of traditional and adaptive zoning based aggregations and treat the first model as 'ground truth'. The first comparison metric is information theoretic entropy (Shannon 1948), which expresses how much of the information present in the detailed model remains in the aggregations. The second comparison investigates how the associated levels of precision in measuring distance and travel time, affect the calibration of a discrete choice model of mode choice. The case study uses data from the London Area Transport Survey 1991 (London Research Centre 1994) and considers trips within London by car, train, underground, bicycle and on foot.

2 Methods

2.1 Adaptive Zoning

This section describes the rationale and methods for generating an origin-based adaptive zone system; a destination-based system can be created analogously. The concepts of adaptive zoning for generic spatial interaction modeling are introduced by Hagen-Zanker and Jin (2012). The rationale of adaptive zoning is to minimize the aggregation error associated with an origin zone by adopting the size and shape of its destination zones. The shape of the

destination zones is determined to minimize aggregation errors; having more compact zones reduces location uncertainty, which in turn reduces the aggregation error. The size of zones is determined aiming for an even distribution of aggregation error over origin-destination pairs. Here the rationale is to avoid making an effort to reach levels of accuracy in one place that are lost in another, which is an idea common to many compression methods (Kieffer 1971). The amalgamation of zones is organized according to a simple hierarchical structure on top of an atomic zone system. This sharply limits the computational overhead of accounting for the aggregated zones, because the number of aggregated zones in the hierarchy is by definition smaller than the number of atomic zones. In line with this rationale, the algorithm for creating an adaptive zoning system takes two stages: in the first stage a zone hierarchy is created, and the second stage combines zones from different levels of the hierarchy to create a set of destination zones for each origin zone; i.e. its neighborhood.

The first stage of zone hierarchy creation follows a process of incremental clustering from an existing zone plan that forms the base of the zone hierarchy. Besides the consideration of compactness, the joining criterion also aims to maintain at each level of aggregation an even distribution in zone sizes, because this will mean that the zone hierarchy will contain a balanced variety of zone sizes for each location; this is relevant because it provides more flexibility to the next stage of neighborhood creation. The algorithm starts by placing all atomic zones in a list of joining candidates. Then, the algorithm selects two zones from the list, according to a joining criterion (to be defined below) that accounts both for size and compactness. It removes both zones from the list of joining candidates and adds a new zone that amalgamates the two selected zones. In the zone hierarchy the new zone becomes the parent of the two selected zones. The process of selecting and joining zones is repeated until a single zone remains that encompasses the entire study area. Figure 2a illustrates this process for a synthetic example where the study area consists of nine zones.

The second stage of the algorithm is to generate neighborhoods for each origin zone in turn. The algorithm follows a process of incremental refinement. In each step of refinement, one zone in the neighborhood is selected using a splitting criterion (to be defined below) to find the neighborhood zone with the highest associated aggregation error. The selected zone is then removed from the neighborhood and its children are added. Initially each neighborhood consists of only one zone that covers the whole study area. The process of splitting continues once a halting criterion is reached. Here, the halting criterion is a fixed number of zones (*s*) in the neighborhood of each origin zone. By definition, each neighborhood covers the whole study area, and each split replaces one parent zone for two child zones that cover the same area. Figure 2b illustrates the process of neighborhood generation for the same synthetic example as before; due to the small dimension of this example, the possible neighborhoods could be exhaustively listed.

The aforementioned joining and splitting criteria are pivotal to realizing the rationale of the adaptive zoning method. Hagen-Zanker and Jin (2012) detail how the criteria can be derived from estimated aggregation errors in a spatial interaction model. The criteria, then, are functions of the distances between zones and the population distribution. Alternatively, the criteria can be directly derived from trip rates, instead of the idealized spatial interaction model (Hagen-Zanker and Jin 2011a). These two options are the main alternatives available in the context of land use and transport modelling. This article calls these distance- and trip-based criteria for short.

The distance-based criteria are based on the error due to spatial aggregation that occurs in a spatial interaction model. The aggregation error can be seen as the product of two factors: the strength of the interaction, and the relative error due to spatial aggregation. In the model,



Figure 2 Schematic representation of adaptive zoning algorithm. First, a base zone system is stepwise aggregated to form a zone hierarchy. Secondly, one neighborhood is generated for each base zone, by subdividing the study area until the desired number of neighborhood zones is reached

the strength of the interaction increases with the (population) size of zones and diminishes with distance; whereas the relative aggregation error increases with uncertainty of location within a zone, which can be measured by the intrazonal distance. The rationale of the joining criterion is to merge the pair of zones that causes the smallest increment in the estimated error. It takes the following form:

$$c_{a,b}^{join,distance-based} = D_{a\cup b}e^{\beta d_{a\cup b,a\cup b}} - D_a e^{\beta d_{a,a}} - D_b e^{\beta d_{b,b}}$$
(1)

where the algorithm joins the pair of zones *a* and *b* with lowest value, D_a measures the size of destination zone *a* (here the number of trips destined for that zone), and β is the distance sensitivity parameter of a best-fitting model.

The rationale of the splitting criterion is to distribute the spatial aggregation error evenly over the destination zones. For each origin zone i; it is achieved using the following equation:

$$c_{i,j}^{\text{split,distance-based}} = O_i D_j e^{-\beta d_{i,j}} \left(e^{\beta (d_{i,i} + d_{j,j})} - e^{-\beta (d_{i,i} + d_{j,j})} \right)$$
(2)

where the algorithm splits the zone j in the neighborhood of i with the highest value for $c_{ij}^{split,distance-based}$, and O_i measures the size of origin zone i (here the number of trips emanating from that zone).

In the road transport assignment problem, the trip matrix is given and the task is to allocate the given traffic between zone pairs to links in the network. The distance-based criteria which derive from an idealized trip distribution model (the spatial interaction model) seem less appropriate when the trip distribution is given. Therefore Hagen-Zanker and Jin (2011a) introduce alternative criteria based on the specified trip rates only. Here we apply a variation that uses the specified trip rates as the strength of interaction and the intrazonal distance of the aggregated zone as the relative uncertainty. The equations follow the same rationale as the distance-based criteria, and take the following form:

$$c_{a,b}^{join,trip-based} = d_{a \cup b,a \cup b} \sum_{i} T_{i,a \cup b} - d_{a,a} \sum_{i} T_{i,a} - d_{b,b} \sum_{i} T_{i,b}$$
(3)

$$c_{i,j}^{split,trip-based} = T_{i,j}d_{j,j} \tag{4}$$

where $T_{i,j}$ is the number of trips from *i* to *j*.

The joining and splitting criteria make use of distances between zones as well as distances within zones, whereby zones can be highly variable in size. It is therefore important to use a distance metric that is consistent in both cases and robust to variations in zone size. The average distance metric (Okabe and Miller 1996) is most appropriate. However, the average distance has no convenient numerical solution. Therefore, we followed a Monte Carlo integration procedure that estimates the average distance between two zones as the average over a number of random point pairs drawn – one for either zone – from a spatially uniform distribution within the zone boundary (Hagen-Zanker and Jin 2011b). The distance between points is the Euclidean distance. Using average distances allows the following equation for amalgamating destination zones:

$$d_{i,a\cup b} = \frac{A_a}{A_a + A_b} d_{i,a} + \frac{A_b}{A_a + A_b} d_{i,b}$$
(5)

where $d_{i,j}$ is the distance from zone *i* to *j*, A_a is the area of zone *a*, and $a \cup b$ is the zone that amalgamates *a* and *b*.

Adaptive zoning affects how interactions are represented. Traditional OD matrices are square and tabulate interaction between all $n \times n$ zone pairs, where n is the number of zones; whereas adaptive zoning tabulates interactions between atomic origins and aggregate destinations. The total number of destination zones is 2n - 1 (since there are n atomic zones to begin with and n-1 aggregation steps that create a new zone by amalgamating two existing zones), hence the size of the matrix is $n \times (2n - 1)$. The matrix is sparse however, since each atomic zone only interacts with a selected set of aggregated zones. Therefore, the total number of



Figure 3 A comparison of the pattern of aggregation for using traditional aggregation and adaptive zoning, corresponding to the synthetic example of Figure 2. The pattern of aggregation affects the OD matrix as well as the measurements of distances.

non-zero values is $n \times s$, where s is the number of aggregated destination zones in the neighborhood of each atomic zone.

Adaptive zoning also affects the precision of measuring distances between zones. The precision in measuring the distance between zones is negatively dependent on the area of the zones. The area of destination zones, typically, is larger at further distances from the origin zone, and therefore the error in distance measurements will be larger for these zones as well. Figure 3 illustrates the difference in matrix structure for adaptive zoning based and traditional aggregation, as well as the effect on measured distances and aggregated OD matrices.

2.2 Information Theoretic Evaluation

The aggregation of atomic zones implies a loss in information compared to the atomic origindestination matrix; the atomic matrix contains $n \times n$ elements, adaptive zoning reduces this to $n \times s$ non-zero elements, whereas traditional aggregation leaves $m \times m$ elements where m is the number of zones after aggregation. The loss of information is not merely a function of the number of elements in the matrix, because; it also depends on the distribution of values over the amalgamated elements. More information will be retained when the pattern of aggregation correlates with the pattern of the trip distribution; i.e. when higher values in the matrix are subject to a proportionally smaller degree of aggregation. This can be gauged by the information theoretic measure of entropy (Shannon 1948). It quantifies the amount of information in a signal and is proportional to the number of bytes necessary to store the information:

$$H(P) = -\sum_{i} p_i \log p_i \tag{6}$$

where H(P) is the entropy of discrete distribution P ={p₁, p₂, ... p_n}, which is normalized such that $\sum_{i} p_i = 1$. When the signal consists of values in an interaction matrix, the equation for entropy can be rewritten as:

$$p_{ij} = \frac{T_{ij}}{\sum_{i} \sum_{j} T_{ij}}$$
$$H(T) = -\sum_{i} \sum_{j} p_{ij} \log p_{ij}$$
(7)

where T_{ij} is the strength of interaction between zones *i* and *j*. In effect, low entropy values are found for uneven distributions and high entropy values for evenly distributed values. In terms of aggregation this confirms that selectively aggregating cells with lower values of T_{ij} reduces the loss of information.

In the example given in Figure 3, it is apparent that the adaptive zoning based aggregation has a more even distribution of cell values. The measure of entropy confirms this, yielding 1.53 for the 9×9 matrix; a reduction of 25% to 1.14 for the 6×6 matrix; and a reduction of 13% to 1.33 for the 9×4 matrix. Hence, the loss of information by traditional aggregation in this example is double that by adaptive zoning.

2.3 Discrete Choice Model

The use of traditional aggregation or adaptive zoning affects the measurements of distances. The increased zone size implies a loss in precision. For traditional aggregation this loss of precision, in absolute terms, is independent of distance. On the other hand, adaptive zoning has a stronger loss in precision at further distances. Figure 4 illustrates the effect of both aggregation methods on the measurement of distances.

The degree of precision in the measurement of distances will propagate in models making use of those distances. For example, calculating the average trip length using the matrices in Figure 3 yields 13.1 km for the 9×9 case, the 6×6 matrices deviate by 14% yielding 11.3 km, and the 9×4 matrices deviate by 7% yielding 14.0 km. Hence, the spatial aggregation error under traditional aggregation is twice that of the adaptive zoning based method in this example. The case study tests the propagation of aggregation errors in the realistic case of a mode choice model that is dependent on both travel time and road network distance.



Figure 4 Stylized description of the effect of traditional and adaptive zoning based aggregation on the representation of distances

A logit-based discrete choice model (Ben-Akiva and Lerman 1985) is set up that models mode choice as a function of the travel time and mode specific constants for car, bus, underground metro and suburban train. For bicycling and walking the model is based on network distance instead of travel time. Each mode has a mode specific constant, which is normalized through setting the constant for the car mode to zero. The model uses the following equations for trip utility on a given origin-destination zone pair:

$$V_{car,i} = \beta_{time} * t_{car}$$

$$V_{bus,i} = \beta_{time} * t_{bus} + \beta_{bus}$$

$$V_{LU,i} = \beta_{time} * t_{LU} + \beta_{LU}$$

$$V_{train,i} = \beta_{time} * t_{train} + \beta_{train}$$

$$V_{cycle,i} = \beta_{cycledistance} * d_{cycle} + \beta_{cycle}$$

$$V_{walk,i} = \beta_{walkdistance} * d_{walk} + \beta_{walk}$$
(8)

where LU stands for underground metro and all other modal subscripts are self-explanatory. $V_{car,i} V_{bus,i}$, etc. are the observable utilities associated with a mode choice for individual *i*; t_{car} , t_{bus} , etc. are travel times for specific modes; all β values are parameters to be estimated. β_{bus} , β_{LU} , etc. are mode specific constants. β_{time} gives the utility associated with each minute of travel time; and $\beta_{cycledistance}$, $\beta_{walkdistance}$ are distance parameters for cycling and walking.

Under the assumption that the unobservable utilities for the different modes follow the iid or Gumbel distribution, the probability for an individual to use mode m follows from the logit equation:

$$p_i(m) = \frac{e^{V_{m,i}}}{\sum_{m'} e^{V_{m',i}}}$$
(9)

where $p_i(m)$ is the probability of individual *i* to use mode *m* for a particular trip and *m'* iterates over all available modes.

The model is estimated using the Maximum Likelihood Estimation procedure provided by the BIOGEME software (Bierlaire 2003). The software reports parameter estimates as well as (robust) standard errors for each model parameter.



Figure 5 Study area map

Group	LATS Modes
Car	Car driver, Car passenger
Bus	Bus: public, Bus: work/school, Coach
LU	LUL/DLR
Train	BR train
Cycle	Pedal cycle
Walk	Walk
Other	Motorcycle driver, Motorcycle passenger, Plane/boat/other, Medium van driver, Medium van passenger, Other van/lorry driver, Other van/lorry passenger, Small van driver, Small van passenger, Taxi: London, Taxi: Other

 Table 1
 Grouping of LATS 1991 transport modes into six main classes and other

3 Data

The data that is used for the experiment is from the London Area Transport Survey 1991 (London Research Centre 1994). This is an extensive cross-sectional survey detailing trip mode, origin and destination zone, trip purpose, trip time of day for trip, car ownership, household size and income in and around the Greater London area (Figure 5). For the purpose of mode choice modelling we have grouped the modes according to Table 1.

For the purpose of the discrete choice modelling, a subset of the observations is selected that is expected to improve the population homogeneity, yet still has a substantial sample size. The analysis is therefore limited to the morning peak (7–10 a.m.), home-to-work journeys for the higher income segment (> \pm 30,000/year). The subset is covered by 10,716 observations out

		Waightad	Maan distance
	Observations	$[\times 10^3]$	[km]
Car	5,719	347	12.7
Bus	674	42	8.0
LU	1,632	111	15.2
Train	1,525	97	21.1
Cycle	209	13	7.9
Walk	647	40	3.0
Other	310	19	14.1
Overall	10,716	665	13.4

 Table 2
 Summary of the data sample for model calibration

Note: As shown in Table 1, Mode 'Other' includes of a diverse range of trips in distinct modes with very small market shares, and is therefore dropped in the mode choice modelling.

of the initial 354,983. The majority of trips (5,719) is by car and the relationship between trip length and mode choice is consistent with UK Census statistics (Figure 1). Table 2 summarizes the headline statistics.

The trips cover the area in and around London within the M25 ring motorway, which is slightly larger than the geographic coverage of the Greater London Authority. It captures the main London catchment in terms of commuting journeys. The dataset does not include commuting from further afield in South England, for which there is no compatible data. This study area is subdivided into 1,019 zones.

Estimating the model requires the travel time for the chosen mode as well as the other modes for each observation. Since this information is not included in the LATS dataset, we used the LASER land use and transport model for London and surrounding regions as an additional source; LASER version 3.0 has a systematically calibrated and validated multi-modal transport network for 1991 contained within its Base Year model (Jin et al. 2002).

4 Results

4.1 Zone Systems

The fully detailed zone system is aggregated to create both a coarser dimensioned traditional zone system and an adaptive zone system (Figure 6). The aggregation is such that the number of zone pairs reduces by 90% in both cases; in the fully detailed model there are $1,019 \times 1,019$ zone pairs, in the traditional aggregation 309×309 and in the adaptive zoning based system $1,019 \times 98$. When comparing Figures 6b and c it appears that trip- and distance-based aggregation produce very similar results. It seems that the distance-based zones are on average more compactly shaped, but the distinction is minor. The neighborhoods are more substantially different, as can be seen by comparing Figures 6d and e. The distance-based neighborhoods are more small zones, particularly in the centre of London. Figures 6e and f both present trip-based neighborhoods, Figure 6f, however is based on the distance-based zone hierarchy. The patterns of both approaches are very similar, but it does appear that neighborhoods based on the trip-based zone hierarchy contain more irregular shaped zones.



e) Three samples from trip volume based adaptive zone system (build on trip-based hierarchy)



f) Three samples from trip volume based adaptive zone system (build on distance-based hierarchy)

Figure 6 Aggregated zone systems; the dot marks the location of the origin zone, whereas the zone boundaries indicate destination zones

	Entropy					Loss of entropy					
	full	trad -d	trad -t	adap -d	adap -t	adap -m	trad -d	trad -t	adap -d	adap -t	adap -m
train	9.0	3.5	6.2	8.3	8.7	8.7	61%	31%	7%	3%	3%
lu	9.5	2.6	5.3	9.0	9.1	9.1	73%	44%	5%	4%	4%
car	10.1	5.1	4.6	9.9	9.8	9.8	50%	55%	2%	3%	3%
bus	9.5	3.4	4.6	9.3	9.2	9.1	64%	52%	2%	3%	3%
cycle	7.9	3.9	3.8	7.9	7.8	7.8	50%	52%	0%	1%	1%
walk	8.4	3.2	3.7	8.2	8.1	8.1	62%	56%	2%	3%	3%

Table 3 Comparison of information content under traditional and adaptive zone aggregations

full: fully detailed (1,019 × 1,019 zones); trad-d: distance-based traditional (309×309 zones); trad-t: tripbased traditional (309×309 zones); adap-d: distance-based adaptive zoning (1,019 × 98 zones); adap-t: tripbased adaptive zoning (1,019 × 98 zones); adap-m: mixed adaptive zoning (1,019 × 98 zones)

4.2 Information Theoretic Comparison

When evaluating the modes in terms of information content, it is no surprise that the fully detailed model contains the most information for all modes (Table 3); any other result would indicate an erroneous calculation. The same table shows that there is a marked difference in the loss of entropy, calculated as $(H_{agg} - H_{full})/H_{full}$, where *agg* can be any of the five aggregation methods; i.e. distance-based traditional aggregation (*trad-d*), trip-based traditional aggregation (*trad-d*), distance-based adaptive zoning (*adap-d*) or trip-based adaptive zoning (*adap-t*), and mixed adaptive zoning (*adap-m*), which uses the distance-based zone hierarchy to make trip-based neighborhoods.

The matrices based on traditional aggregation lose between 31% and 73% of information per mode, and the variation has no obvious correlation with the modes. For distance-based adaptive zoning the range is 0 to 7%; again no strong association with trip lengths is perceived, although cycling is most efficiently aggregated and train and underground the least. When comparing distance- and trip-based aggregations, it appears that trip-based gives substantially better results for the train mode and similar results for others.

4.3 Discrete Choice Model Estimates

The discrete model for mode choice has been estimated six times, once for each set of matrices. In all cases, all parameters of the discrete choice models proved statistically significant (p < 0.01). The fully detailed model serves as the benchmark for assessing the traditional and the adaptive zoning based aggregations; i.e. we expect the model estimation results will be poorer for all aggregated models, and the extent of deterioration due to the methods of aggregation can be quantified through a comparison with the fully detailed model. Hence, Table 4 gives the values of estimated parameters based on aggregated matrices, relative to the fully detailed estimation. The distance-based traditional aggregation produces better results than the trip-based traditional aggregation, and for the adaptive zoning based methods the mixed method produces the best results. Of these best-in-their-class results, traditional aggregation leads to deviation in parameter estimates of up to 34% and on average 12%; for adaptive zoning these numbers are 4% and 2%, representing an improvement of precision by a factor of 6 to 8. The

	Parameter						
	full	trad-d	trad-t	adap-d	adap-t	adap-m	
β_{time}	-0.0504	-11%	-13%	-3%	-4%	-2%	
β_{bus}	-1.07	9%	11%	2%	3%	2%	
eta_{cycle}	-2.64	-3%	-3%	-1%	-1%	-2%	
β_{LU}	0.752	-21%	-19%	-4%	-6%	-4%	
eta_{train}	0.621	-35%	-38%	-6%	-8%	-2%	
$eta_{\scriptscriptstyle walk}$	-0.849	-6%	-1%	4%	-4%	-4%	
$eta_{ ext{cycledistance}}$	-0.171	-6%	-6%	-1%	-2%	1%	
$eta_{\scriptscriptstyle walkdistance}$	-0.364	-7%	-9%	-3%	-2%	-1%	

 Table 4
 Estimated parameters and standard errors of parameter estimates

β_{time}	Robust standard error of parameter						
	0.00208	-8%	-6%	-1%	-3%	-1%	
eta_{bus}	0.0615	-4%	-3%	0%	-1%	-1%	
β_{cycle}	0.181	3%	4%	-3%	3%	0%	
β_{LU}	0.0642	-1%	3%	1%	1%	0%	
$eta_{ ext{train}}$	0.0726	-4%	-2%	-2%	0%	1%	
$eta_{\scriptscriptstyle walk}$	0.2	-6%	-9%	-2%	1%	-2%	
$eta_{ ext{cycledistance}}$	0.0181	2%	2%	-6%	3%	-2%	
$eta_{\!\scriptscriptstyle walkdistance}$	0.0365	-15%	-18%	-4%	-2%	-5%	

full: fully detailed (1,019 × 1,019 zones); trad-d: distance-based traditional (309×309 zones); trad-t: tripbased traditional (309×309 zones); adap-d: distance-based adaptive zoning (1,019 × 98 zones); adap-t: tripbased adaptive zoning (1,019 × 98 zones); adap-m: mixed adaptive zoning (1,019 × 98 zones)

relative performance regarding the standard errors of the parameter estimates is similar although less pronounced. It must be noted however, that the standard error for most parameter estimates *reduces* due to the aggregation. This may be unexpected considering that the aggregation of zones adds noise to the measurement of distances; however, it also reduces the variability of the distances, which in turn results in lower standard errors. Thus there are two competing effects of aggregation: its arbitrary distortion of the measurement of distances increases standard errors, whereas its structuring quality reduces standard errors. Williams (1976) investigated similar effects for regression models of aggregated variables. We do not attempt to disentangle those two effects, but note that the standard errors do not indicate goodness-of-fit. Table 5 gives the overall fit statistics. As expected, the fully detailed model has the best fit, followed by adaptive zoning based models, of which the mixed variation has the best fit. The traditional aggregations follow only at considerable distance.

5 Discussion

Adaptive zoning follows the rationale of representing strong spatial interactions with proportionally more detail than weak interactions. Considering the typical distance decay in trip intensities it is therefore logically expected that the method performs better on short distances

Model	Likelihood ratio	R ²
Full	11151	0.417
trad-d	10860	0.405
trad-t	10813	0.403
adap-d	11046	0.413
adap-t	11036	0.411
adap-m	11106	0.414

Table 5	Overall fi	t statistics
---------	------------	--------------

full: fully detailed (1019 × 1019 zones); trad-d: distance-based traditional (309×309 zones) ; trad-t: tripbased traditional (309×309 zones); adap-d: distance-based adaptive zoning (1019×98 zones); adap-t: tripbased adaptive zoning (1019×98 zones); adap-m: mixed adaptive zoning (1019×98 zones)

and by extension on modes that are typically used over short distances. The information theoretic comparison confirms this pattern by finding stronger loss of information for the underground and train modes and weaker loss of information in particular for the bicycle modes (walking somehow seems exempt from this logic, possibly because of the very large share of within-zone trips). In theory, such differentiated advantage of adaptive zoning may present a dilemma: should we try to represent better slow modes (i.e. bicycles and walking) at the expense of fast modes (i.e. LU and trains)? Fortunately, this dilemma proved to be a non-issue in the current case. Adaptive zoning is more efficient than the traditional zoning approach, for all modes and by a wide margin.

The better retention of information in the trip matrices appears to benefit the estimation of the mode choice model too. However, it is not a universal relation. Note in particular how the trip-based traditional aggregation contains more information than the distance-based variation, but performs worse on the overall fit statistics of the mode choice model. The probable explanation is that while trip-based aggregation retains more information by aggregating those matrix cells that have low values, it may do this at the expense of the compactness of the aggregated zones. Reduced compactness, in turn, underlies a greater error in the measurement of distances. This realization motivated us to extend the analysis with the mixed method adaptive zoning. This mixed method uses the attractively compact aggregated zones of the distancebased hierarchy in combination with the trip-based neighborhood algorithm that reflects observed trip patterns rather than a model approximation. Figure 6f presents the effects visually: the mixed method approach is equally capable in representing the idiosyncratic trip patterns as the trip-based method, but does so with slightly smoother, rounder zones. The best-ofboth-worlds solution is reflected in superior likelihood ratios and R² compared to either alternative, and there is only a tiny additional loss of information compared to the trip based method.

Of the three adaptive zoning methods, the mixed method is most efficient. If a base year trip matrix is available (as is often the case for multi-modal transport studies), that will be the preferred approach. When that is not available, the distance based method would preferably be applied using distance matrices that accurately reflect transport geography – for instance network structure, service levels, travel speeds, and reliability – instead of the Euclidean distance used here. Nonetheless, even the method of Euclidean distance-based adaptive zoning outperformed the traditional aggregations method by a wide margin and may still represent a superior option in data-poor environments.

Just like the traditional zoning method, the adaptive zoning approach faces the problem of accommodating multiple patterns, e.g. trip patterns by mode, socio-economic group, time of day, etc., within a single zoning system. For example, residents within the pedestrian catchments of rail stations have markedly different modal preferences from the remainder of the population (Cervero and Kockelman 1997). However, just as applying trip-based adaptive zoning accounts for more of the idiosyncrasies in the transport network than the distance-based method, applying subset-specific trip volumes (e.g. by mode, socio-economic group, time-of-day) may help to account for such distributional aspects without having to identify those patterns explicitly. It is not obvious, however, how to combine various subset-specific zone systems, or to use multiple systems in a single model. One option of accounting for subset-specific patterns is to replace the governing criteria – joining zones in the creation of the hierarchy and splitting zones in the creation of the neighborhood – for new criteria that are a weighted combination of the subset criteria.

6 Conclusions

Adaptive zoning offers a new type of interaction geography. Its capability of addressing key problems in spatial interaction modelling, notably the archetypical doubly constrained gravity model and the computational bottleneck of road traffic assignment, has been demonstrated in previous papers (Hagen-Zanker and Jin 2011a, 2012). In this article we have investigated its capability to represent efficiently the multimodal interaction patterns over a wide range of distances up to 60 km, a typical distance range for day-to-day commuting in a large number of metropolitan areas. The tests presented in this article show that for a given computational load, a model that is based on adaptive zoning contains more than twice the information when measured in terms of entropy, and that mode choice models calibrated on adaptive zoning principles are more accurate by a factor of six to eight. The findings suggest that adaptive zoning has a significant potential in enhancing the accuracy of mode choice modelling at the city or city region scale, especially where walking and cycling are considered important components of the transport system.

Further methodological developments may focus on alternative criteria for creating the adaptive zone system. Most notably there is potential to maintain the same algorithmic structure for adaptive zone plan creation while redefining the implementation details using mode-specific or demand segment-specific trip patterns.

References

- Arentze T A and Timmermans H J P 2004 A learning-based transportation oriented simulation system. *Transportation Research Part B* 38: 613–33
- Banister D 2000 European Transport Policy and Sustainable Mobility. London, Spon Press
- Ben-Akiva M E and Lerman S R 1985 Discrete choice analysis: Theory and application to travel demand. Cambridge, MA, MIT Press
- Bierlaire M 2003 BIOGEME: A free package for the estimation of discrete choice models. In *Proceedings of the Third Swiss Transportation Research Conference*, Ascona, Switzerland
- Cervero R and Kockelman K 1997 Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D* 2: 199–219
- Chang K T, Khatib Z, and Ou Y M 2002 Effects of zoning structure and network detail on traffic demand modeling. *Environment and Planning B* 29: 37–52
- Current J R and Schilling D A 1990 Analysis of errors due to demand data aggregation in the set covering and maximal covering location-problems. *Geographical Analysis* 22: 116–26

Ding C 1998 The GIS-based human-interactive TAZ design algorithm: Examining the impacts of data aggregation on transportation-planning analysis. *Environment and Planning B* 25: 601–16

Domencich T A and McFadden D 1975 Urban Travel Demand: A Behavioral Analysis. Amsterdam, Elsevier

- Eliasson J and Mattsson L G 2001 Transport and location effects of road pricing: A simulation approach. Journal of Transport Economics and Policy 35: 417–56
- Fotheringham A S, Densham P J, and Curtis A 1995 The zone definition problem in location-allocation modeling. *Geographical Analysis* 27: 60–77
- Fotheringham A S and Wong D W S 1991 The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning A 23: 1025–44
- Hagen-Zanker A and Jin Y 2011a Improving geographic scalability of traffic assignment through adaptive zoning. In *Proceedings of the 2011 Computers in Urban Planning and Urban Management Conference*, Lake Louise, Alberta
- Hagen-Zanker A and Jin Y 2011b Reducing aggregation error in spatial interaction models by location sampling. In *Proceedings of the Eleventh International Conference on GeoComputation*, London, England
- Hagen-Zanker A and Jin Y 2012 A new method of adaptive zoning for spatial interaction models. *Geographical Analysis* 44: 281–301
- Hartell A 2007 Methodological challenges of environmental justice assessments for transportation projects. Transportation Research Record 2013: 21-9
- Hillsman E L and Rhoda R 1978 Errors in measuring distances from populations to service centers. Annals of Regional Science 12: 74–88
- Jin Y, Williams I, and Shahkarami M 2002 A new land use and transport interaction model for London and its surrounding regions. In *Proceedings of the European Transport Conference*, Cambridge, England
- Karau E and Keane R 2007 Determining landscape extent for succession and disturbance simulation modeling. Landscape Ecology 22: 993–1006
- Kieffer J 1971 Data Compression. New York, Wiley Online Library
- Landau H J 1967 Sampling, data transmission, and the Nyquist rate. Proceedings of the IEEE 55: 1701-06
- London Research Centre GBDoT 1994 Travel in London: London Area Transport Survey 1991. London, HMSO
- Martinez L M, Viegas J M, and Silva E A 2005 Modifiable areal unit problem (MAUP) effects on traffic analysis zones (TAZ) delineation. *Modelling and Simulation* 2005: 313–23
- Martinez L M, Viegas J M, and Silva E A 2009 A traffic analysis zone definition: A new methodology and algorithm. *Transportation* 36: 581–99
- McFadden D 2007 The behavioral science of transportation. Transport Policy 14: 269-74
- Miller H J 1999 Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geographical Analysis* 31: 373–99
- Moulton B R 1990 An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics* 72: 334–38
- Okabe A and Miller H J 1996 Exact computational methods for calculating distances between objects in a cartographic database. *Cartography and Geographic Information Science* 23: 180–95
- Openshaw S 1977a Geographical solution to scale and aggregation problems in regionbuilding, partitioning and spatial modeling. *Transactions of the Institute of British Geographers* 2: 459–72
- Openshaw S 1977b Optimal zoning systems for spatial interaction models. *Environment and Planning A* 9: 169-84
- Openshaw S 1978 An empirical study of some zone-design criteria. Environment and Planning A 10: 781-94
- Pucher J, Dill J, and Handy S 2010 Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine* 50S: S106–S125
- Putman S H and Chung S H 1989 Effects of spatial system design on spatial interaction models: 1, The spatial system definition problem. *Environment and Planning A* 21: 27–46
- Saura S and Martinez-Millan J 2001 Sensitivity of landscape pattern metrics to map spatial extent. *Photogrammetric Engineering and Remote Sensing* 67: 1027–36
- Shannon C E 1948 A mathematical theory of communication. Bell System Technical Journal 27: 623-56
- Tobler W R 1989 Frame independent spatial analysis. In Goodchild M F and Gopal S (eds) Accuracy of Spatial Databases. London, Taylor and Francis: 115–22
- Viegas J M, Martinez L M, and Silva E A 2009 Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. *Environment and Planning B* 36: 625–43
- Wegener M 2001 *The IRPUD Model.* Dortmund, Germany, Institut für Raumplanung, Fakultät Raumplanung, Universität Dortmund
- Williams I 1976 Optimistic' theory validation from spatially grouped regression: Theoretical aspects. *Transactions of the Martin Centre for Architectural and Urban Studies* 1: 113–45