geographical analysis

Geographical Analysis (2012) ••, ••–••

A New Method of Adaptive Zoning for Spatial Interaction Models

Alex Hagen-Zanker, Ying Jin

Department of Architecture, University of Cambridge, Cambridge, U.K.

Spatial interaction models commonly use discrete zones to represent locations. The computational requirements of the models normally arise with the square of the number of zones or worse. For computationally intensive models, such as land use–transport interaction models and activity-based models for city regions, this dependency of zone size is a long-standing problem that has not disappeared even with increasing computation speed in PCs—it still forces modelers to compromise on the spatial resolution and extent of model coverage as well as on the rigor and depth of model-based analysis. This article introduces a new type of discrete zone system, with the objective of reducing the time for estimating and applying spatial interaction models while maintaining their accuracy. The premise of the new system is that the appropriate size of destination zones depends on the distance to their origin zone: at short distances, spatial accuracy is important and destination zones must be small; at longer distances, knowing the precise location becomes less important and zones can be larger. The new method defines a specific zone map for every origin zone; each origin zone becomes the focus of its own map, surrounded by small zones nearby and large zones farther away. We present the theoretical formulation of the new method and test it with a model of commuting in England. The results of the new method are equivalent to those of the conventional model, despite reducing the number of zone pairs by 96% and the computation time by 70%.

Introduction

Models of spatial interaction, such as transport, migration, commuting, and trade, typically use discrete zone systems to represent the dispatching and receiving ends of interactions. When zones encompass multiple locations, the partitioning causes an aggregation error (Hillsman and Rhoda 1978), which increases with the size of zones. The use of smaller zones, however, comes at the cost of increased computational requirements in terms of processing time, storage, and working memory. In practice, concessions to computational requirements negatively affect the spatial extent, spatial resolution, thematic detail, and precision of models. The depth and rigor of model applications also may suffer under such computational burden. Aggregation errors can cause bias

Correspondence: Alex Hagen-Zanker, Department of Architecture, University of Cambridge, 1-5 Scroope Terrace, Cambridge CB2 1PX, U.K.
e-mail: ahh34@cam.ac.uk

Submitted: December 17, 2010. Revised version accepted: October 31, 2011.

doi: 10.1111/j.1538-4632.2012.00855.x
© 2012 The Ohio State University

(Goodchild 1979; Openshaw 1984), and when zones are larger than a (generally unknown) threshold, models become invalid (Tobler 1989).

During the last three decades substantial increases have occurred in computing capacity, making computational burden a less pressing issue for many applications, whereas other factors such as data availability, concordance with existing zone systems, and the principle of parsimony have become more prominent considerations in model zone design. Over the same period, however, new and often highly detailed data sources have proliferated, along with geographic information system tools that aid the interrogation and presentation of fine-resolution geographic information. These improvements have greatly contributed to socioeconomically and geographically detailed representations in, for instance, land use and transport interaction modeling (Batty 2009) and activity-based modeling (Axhausen and Gärling 1992; Schönfelder and Axhausen 2010). While little ground exists to doubt the modelers' commitment to principles of parsimony, state-of-the-art land use–transport interaction models (Batty 2009), such as UrbanSim (Waddell 2000), MEPLAN (Echenique 2004), TRANUS (de la Barra, Pérez, and Vera 1984), and the IRPUD (Institut für Raumplanung, Universität Dortmund) (Wegener 2011), have amplified socioeconomic and geographic details in their recent model applications (Jin, Williams, and Shahkarami 2002; Arentze and Timmermans 2004; Waddell et al. 2007; Wegener 2009; Sancar 2010). For these models, computation time continues to be a decisive factor in the design of zones. Their long run times, measured in hours to days, are a barrier to the type of extensive model analysis that is considered good modeling practice, including calibration, validation, sensitivity, and uncertainty analysis (van Waveren et al. 1999; Crosetto, Tarantola, and Saltelli 2000; Kocabas and Dragicevic 2006; Hagen-Zanker and Martens 2008). This context explains our motivation to address the computational requirements of spatial interaction models.

Here we propose a new method of spatial partitioning, which we name *adaptive zoning*. Its premise is that the appropriate size of destination zones depends on the distance to their origin zones: at short distances, spatial accuracy is important, and destination zones must be small; at longer distances, knowing the precise location becomes less important and zones can be larger.

We start by using a conventional zoning system that provides the origins for a spatial interaction model and the building blocks for its destination zones. We call these *atomic zones*. We then devise a specific spatial partitioning for each origin zone in turn that is in line with the adaptive zoning principle. This partitioning is achieved by amalgamating atomic zones into successively larger destination zones with increasing distance from a given origin zone. Thus, a specific pattern of spatial partitioning is formed for each origin zone across an entire study area.

We call this specific pattern of spatial partitioning the *neighborhood* of the origin zone, indicating that as many different neighborhoods as origin zones exist in the model. Each origin zone becomes the focus of its own bespoke neighborhood, which is a set of aggregate zones that, when taken together, covers the entire model area exactly once.

The procedure for creating neighborhoods involves two stages. In the first stage, aggregate zones are created using a hierarchical clustering algorithm. In the second stage, a selection from the aggregate zones is made to form each neighborhood. The procedure aims to minimize the spatial aggregation error associated with the adaptive zoning system.

One long-standing effort has been to optimize the use of computing resources in spatial interaction modeling. Our new method echoes such efforts reported in the literature. For example, researchers commonly optimize a zone system for their model and modeled processes (Openshaw 1978; Goodchild 1979; Fotheringham, Densham, and Curtis 1995; Ding 1998). In this

tradition, areas of little interest are represented in less detail; that is, by larger zones. For example, the appropriate zone system for England differs for research focused on commuting to London in the south versus that focused on journeys to school in the northern city of Manchester. Adaptive zoning extends this principle by recognizing that the most appropriate zoning for spatial interaction varies not only from model to model but also within the same model.

Another related methodology is destination sampling as applied in transport modeling (Williams and Lindsay 2002; Miller et al. 2007), which shares the same motivation as adaptive zoning. Destination sampling attempts to reduce the number of spatial interactions as a function of the required spatial detail through a heuristically defined sampling frame. By comparison, the adaptive zoning method has a stronger theoretical base through defining the spatial aggregation error as its operational criterion and is far easier to implement.

In physics and engineering, an analogy can be found in finite element analysis, namely the moving mesh methods of computational fluid dynamics (Berger and Colella 1989). A key difference is that in these fluid dynamic models, interactions take place only over the face of directly adjacent mesh cells, whereas spatial interaction models consider relationships between nonbordering zones.

Another similarity may be found in cellular automata (CA) (Clarke, Hoppen, and Gaydos 1997; White, Engelen, and Uljee 1997; Webster and Wu 1999; Wickramasuriya et al. 2009) and agent-based models (Parker et al. 2003) that are applied to land use modeling. These are bottom-up models that capture land use change processes in terms of local interactions. In CA, cells interact with their neighbors, whereas in agent-based models, agents interact with their environment. The commonality is that local perspectives drive decisions. Recent developments of variable grids in CA land use modeling that apply reduced precision for large distance interactions (Andersson, Rasmussen, and White 2002; Andersson et al. 2002; van Vliet, White, and Dragicevic 2009) inspired the development of the methodology presented in this article. A disadvantage of the CA approach is its implied dependence on Euclidean distances, whereas current combined land use and transport models tend to express distances in terms of minimal path road distance, travel time, or generalized costs. The zone-based approach of this article is consistent with land use and transport interaction models and is open to any distance measure.

We address here a fundamental aspect of spatial interaction models by proposing an alternative for their representation of space. Therefore, this article intends to demonstrate its merits with one of the cornerstones of spatial interaction modeling—the doubly constrained gravity model—with an illustrative test employing commuting data for England. (See Gitlesen et al. 2010 and Gordon 2010 for recent applications and discussions of the doubly constrained gravity model.) We note that the application of adaptive zoning necessitates an extension of the doubly constrained gravity model to accommodate its more complex representation of space. Similarly, the adaptive zoning algorithms also are not independent of the gravity model; they rely on the specification of the model's aggregation error. Accordingly, the next section on method presents the basic structure of the adaptive zone system before discussing the gravity model, followed by a description of the details of the adaptive zoning method.

Method

Principles of adaptive zoning

Traditionally, spatial interaction models make use of discrete zones to represent locations. For a given system, there are n^2 interaction pairs, where n is the number of zones. Adaptive zoning

also is based on a discrete zone system, although each zone interacts only with the reduced number of aggregate zones that form its neighborhood. Hence, it involves only $n \times s$ interaction pairs, where s is the number of aggregated zones in each neighborhood. The zones that form each neighborhood are selected from one single pool of aggregate zones. These aggregate zones are created through a process of hierarchical clustering of atomic zones, which precedes the creation of the neighborhoods. Therefore, hierarchical clustering is the first step of adaptive zoning.

Hierarchical clustering uses an algorithm that starts with a working list of all atomic zones. From that list, the two zones that are "nearest" to each other are selected to create a new aggregate zone. This aggregate zone is the *parent* of the two atomic zones, its *children*. The child zones are removed from the working list and the parent is added. This process is repeated until only one parent zone remains, which logically covers the whole model area. Two crucial operations detailed in the subsection "Adaptive zoning details" are the *comparison operation*, which determines the "nearest pair," and the *merge operation*, which determines the properties of a parent zone as a function of those of its children. This process generates a hierarchy of successive parent–children zones that, when taken together, form the pool of zones for building the neighborhood of each origin zone.

For each origin zone, this neighborhood generation algorithm creates a set of aggregate zones that covers the zone's entire neighborhood. The algorithm is based on incremental refinement, which means that for each origin zone it begins by considering the top zone in the hierarchy (i.e., the whole study area) as the neighborhood; it then iteratively selects the "most offending" aggregate zone and splits it into its children until a termination criterion is met. Here the termination criterion is a fixed number of zones in each neighborhood. The most offending neighborhood zone is the one with the highest associated aggregation error, details of which follow in the subsection "Adaptive zoning details."

The aggregation error may be viewed as the product of the relative error and the intensity of spatial interaction. The algorithm intends to spread the aggregation error evenly, which means that areas with strong spatial interaction have a small relative error, and areas with weak spatial interaction have a large relative error. Where the first law of geography holds—that is, where interactions at short distances are stronger than at long distances (Tobler 1970)—the size of neighborhood zones increases with distance from the origin zone.

Conventionally, interactions from origin to destination zones are stored in n-by-n origin–destination matrices. These matrices get extended under adaptive zoning. For ease of understanding, we arrange the matrices in the conventional way, with rows representing origins and columns representing destinations. For an adaptive zoning model, the number of matrix cells is greater, since the matrix contains both atomic zones and aggregate zones as destinations. The matrix is also sparse, since the origin zones interact only with a reduced number of aggregate zones. Overall, the adaptive zoning-based origin–destination matrix has fewer nonzero elements.

Note that the adaptive zoning system is asymmetric. The preceding description consistently assumes that atomic origin zones interact with aggregate destination zones. The reverse, where atomic destination zones interact with aggregate origin zones, is just as viable. The doubly constrained gravity model that we subsequently extend makes use of both the origin-based and destination-based adaptive zoning systems. Fig. 1 illustrates the key concepts discussed in this section.

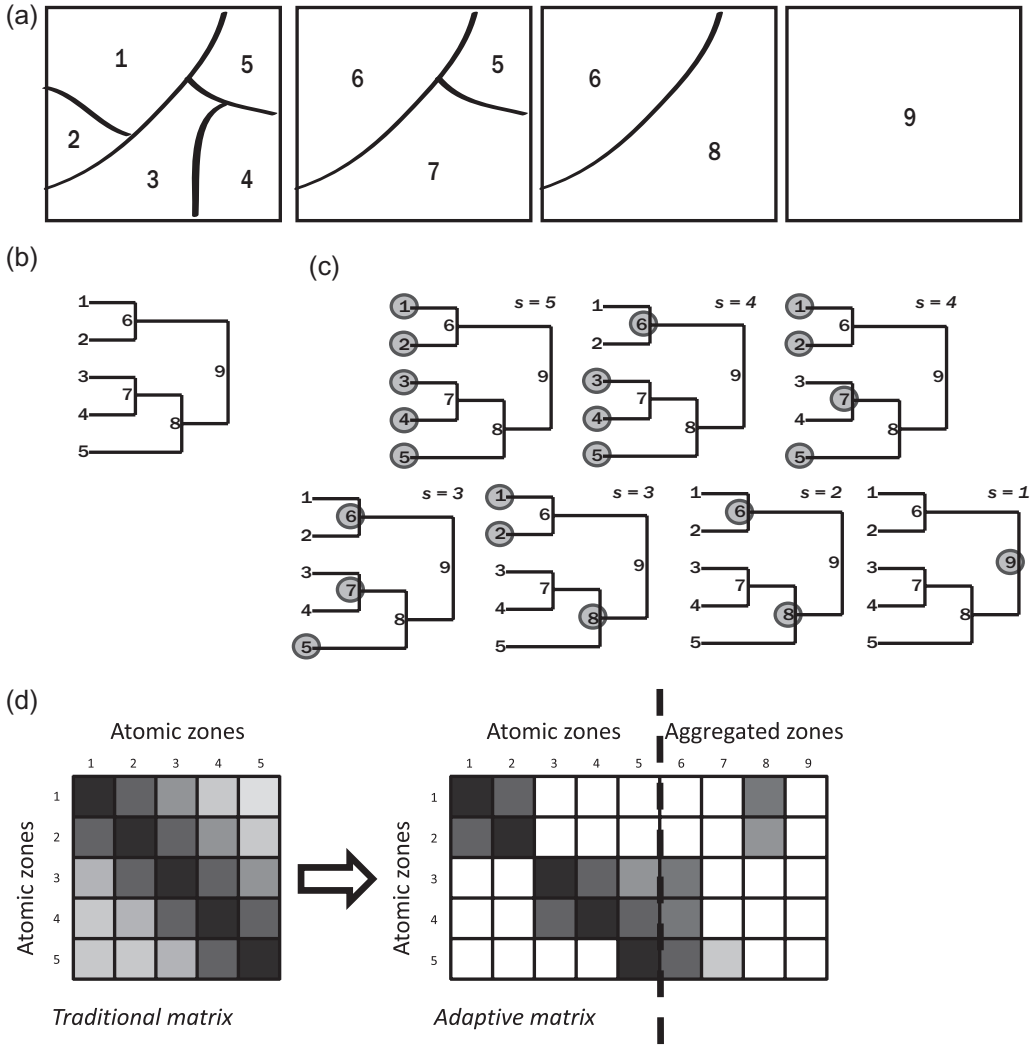

Figure 1. (a) Stepwise hierarchical aggregation of a discrete zone system; (b) Tree representation of zone hierarchy; (c) All possible combinations of zones that cover the whole model area once (i.e., neighborhoods); (d) Transforming a full traditional interaction matrix to a sparse adaptive zoning-based interaction matrix through neighborhood-based aggregation. Gray level indicates cell value, with white cells indicating zero-value cells.

The doubly constrained gravity model

The basic model

The doubly constrained gravity model has the following form:

$$T_{ij} = a_i b_j P_{ij}, \tag{1}$$

where T_{ij} is the interaction between zones i and j, P_{ij} is the prior distribution of interaction between zones i and j, and a_i and b_j are balancing factors, whose value is determined by the constraints on interaction origins and destinations per zone, to satisfy the following equations:

$$R_i = \sum_j T_{ij} \quad and \quad C_j = \sum_i T_{ij}, \tag{2}$$

where R_i is the constraint for the i-th row and C_j is the constraint for the j-th column. Thus, $\sum_i R_i$ must be equal to $\sum_j C_j$. Balancing factors can be found by iteratively applying the following equations (Fratar 1954):

$$a_i = \frac{R_i}{\sum_j b_j P_{ij}}, \quad b_j = \frac{C_j}{\sum_i a_i P_{ij}}, \tag{3}$$

although today they typically are found by including origin- and destination-specific indicator variables in a Poisson regression model specification.

The prior distribution expresses the "gravity" nature of the model and is defined as

$$P_{ij} = O_i D_j e^{-\beta d_{ij}}, \tag{4}$$

where O_i is the size of origin zone i and D_j is the size of destination zone j, d_{ij} is the distance separating zones i and j, and parameter β governs the sensitivity to distance. Origin and destination size may be different from the constraints, but in this article $O_i = R_i \forall i$ and $D_j = C_j \forall j$.

The distance between two zones may be defined as the average distance between any pair of points where one is in each of those zones, that is,

$$d_{ij} = \frac{\int_{p \in i} \int_{q \in j} d_{pq}}{A_i A_j}, \tag{5}$$

where p integrates over all points in zone i and q over all points in zone j. A_i and A_i respectively are the total areas of zones i and j. This definition of distance is consistent for the case of self-distance and implies that the distance of a zone to itself is equal to the average distance between any pair of points within that zone. This article uses Euclidean distances; hence, $d_{pq} = \|p - q\|$.

Sampling-based aggregation

The doubly constrained gravity model makes use of a function of distance between zones to predict the interaction between points. However, because zones cover an area, the precise distance between individual interacting points is not known. Furthermore, because the interaction is a nonlinear function of distance, assuming that the expected interaction between these points corresponds to their expected location may not be correct, even though for small enough zones this may be an acceptable approximation. In the current case, however, aggregate zones can be rather large accounting for the size and shape of zones becomes important. A well-known problem in the closely related field of discrete choice modeling concerns the necessity of accounting for the effect of the size and variability of aggregates of alternatives. The problem is that the size and variability are often imperfectly understood, and the analysis has to depend on judgment, experience, and proxy variables (Ben-Akiva and Lerman 1985, pp. 252–75).

In the context of discrete choice modeling, Train (2009) discusses an alternative approach to dealing with aggregation issues and recommends the use of simulation when formal solutions are

not available. For the doubly constrained gravity model, the simulation approach can be based on randomly sampled locations within zones. In a separate article (Hagen-Zanker and Jin 2011), we present a procedure that finds the random points within each zone boundary polygon. Following that approach, the prior for two aggregate zones can be estimated as the mean prior of sampled priors:

$$\widehat{P}_{ij} = \sum_n w_n P_{nij} = \sum_n w_n O_i D_j e^{-\beta d_{nij}}, \tag{6}$$

where $\widehat{P}_{ij}$ is the estimate of the aggregated prior used by the model. P_{nij} is the prior of the n-th sample, which is a function of the n-th sample distance d_{nij}, and w_n is the corresponding sample weight. In the current case, all samples have the same weight; hence $w_n = 1/N$, where N is the number of samples taken. Because no specific assumptions are made about the location of subjects within a zone, this approach remains consistent with entropy maximization.

In practical applications, such as transport modeling, obtaining distances for a large sample of points often is not feasible. Therefore, an approximation is applied on the basis of Euclidean distances. This approximation assumes that the variability in distances correlates perfectly with the variability in Euclidean distances:

$$d_{nij} = d_{ij} \frac{\left\| r_{in} - r_{jn} \right\|}{\sum_{n'} w_{n'} \left\| r_{in'} - r_{jn'} \right\|}, \tag{7}$$

where r_{in} is the n-th randomly sampled location in zone i, and likewise for j. The index n' iterates over all samples.

Using $\widehat{P}_{ij}$ instead of P_{ij} mitigates the effect of the modifiable area unit problem, as demonstrated by Hagen-Zanker and Jin (2011).

Aggregation errors

The error due to spatial aggregation is estimated as the error in the calculation of the prior based on spatial imprecision. The values of balancing factors a_i and b_j are left out of the estimation because these follow from a calibration and are not known beforehand. Following Guare (1991), the error in the prior can be approximated as follows:

$$\Delta P_{ij} = \left| \frac{O_i D_j e^{-\beta(d_{ij} - \Delta d_{ij})} - O_i D_j e^{-\beta(d_{ij} + \Delta d_{ij})}}{2} \right| = \frac{1}{2} O_i D_j e^{-\beta d_{ij}} \left| e^{\beta \Delta d_{ij}} - e^{-\beta \Delta d_{ij}} \right|, \tag{8}$$

where Δd_{ij} is the spatial imprecision that depends on the size, shape, and relative position of zones i and j. Fig. 2 gives an indication of the geometrical meaning of the aggregation error. An approximation is used based on a perfect circle analogy; if zones were circles, then $\Delta d_{ij} = r_i + r_j$ and $r_i = d_{ii}$. Substituting this result into equation (8) yields

$$\Delta P_{ij} = \frac{1}{2} O_i D_j e^{-\beta d_{ij}} \left| e^{\beta(d_{ii} + d_{jj})} - e^{-\beta(d_{ii} + d_{jj})} \right|. \tag{9}$$

Considering that, in general, $e^{\beta(d_{ii} + d_{jj})} \gg e^{-\beta(d_{ii} + d_{jj})}$, equation (8) can be approximated by a product of three factors associated with i, j, or the ij pair:

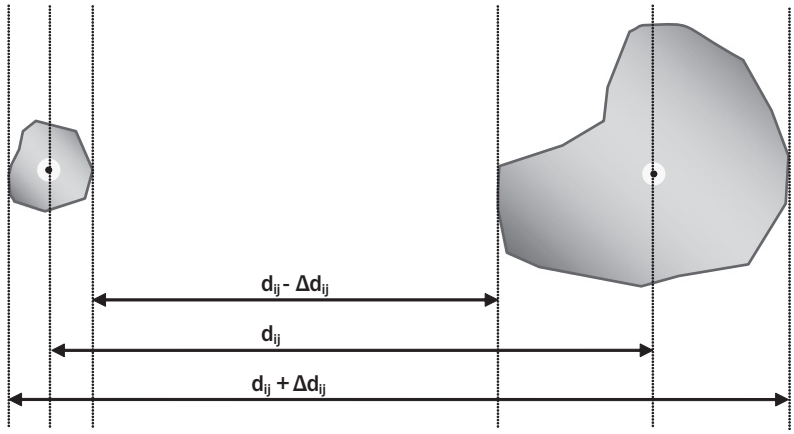

Figure 2. A graphical illustration of the spatial aggregation error of discrete zone distances.

$$\begin{aligned} \Delta P_{ij} &\approx \frac{1}{2} F_i G_j H_{ij}, \\ F_i &= O_i e^{\beta d_{ii}}, \\ G_j &= D_j e^{\beta d_{ij}}, \\ H_{ij} &= e^{-\beta d_{ij}}. \end{aligned} \tag{10}$$

where F_i, G_j, and H_{ij} are the three components of the approximate error.

Accommodating the adaptive zoning system

The doubly constrained spatial interaction model calculates balancing factors for each zone. The adaptive zoning system introduces the complexity of using aggregate zones: balancing factors of parent zones are not independent of those of child zones. The relationship between parent and child zones can be found on the basis of a mean field analysis, which requires that the balancing factor of a parent zones is equivalent to the aggregation of its children; hence,

$$T_{iJ} \approx \sum_{j \in J} T_{ij} \quad \forall i, \tag{11}$$

where J is the parent of zones with index j. This equality can be solved when assuming that $d_{ij} \approx d_{iJ} \forall j \in J$. Thus,

$$a_i b_J O_i D_J e^{-\beta d_{iJ}} = \sum_{j \in J} a_i b_j O_i D_j e^{-\beta d_{iJ}} \Rightarrow b_J = \sum_{j \in J} b_j \frac{D_j}{D_J}. \tag{12}$$

Likewise, when the aggregation takes place for the origin side of the interaction,

$$a_I = \sum_{i \in I} a_i \frac{O_i}{O_I}, \tag{13}$$

where I is the parent of zones with index i.

The constraints can be straightforwardly adjusted to the adaptive zoning system:

$$R_i = \sum_{J \in N_i^C} T_{iJ}^C, \quad and \quad C_j = \sum_{I \in N_j^R} T_{Ij}^R, \tag{14}$$

where T_{iJ}^C is the element in row i and column J of the *column-aggregated* interaction matrix, and N_i^C is the set of zones in the column-aggregated neighborhood of i. Likewise, T_{Ij}^R is the element in row I and column j of the *row-aggregated* interaction matrix, and N_j^R is the set of zones in the row-aggregated neighborhood of j.

One might estimate a_i (and b_j likewise) by

$$a_i = \frac{R_i}{\sum_{J \in N_i^R} b_J P_{iJ}^C}, \tag{15}$$

where P_{iJ}^C is the column-aggregated prior matrix. But, due to the earlier assumption that $d_{ij} \approx d_{Ij} \forall i \in I$ the two aggregated prior matrices are not perfectly consistent, and

$$\sum_j \sum_{I \in N_j^R} a_I b_j P_{Ij}^R \neq \sum_i \sum_{J \in N_i^C} a_i b_J P_{iJ}^C, \tag{16}$$

which implies that all constraints cannot be met simultaneously! Therefore, two additional balancing factors, c and d, are introduced to absorb the discrepancy between the two matrices:

$$\begin{aligned} T_{Ij} &= a_I b_j c P_{Ij}, \\ T_{iJ} &= a_i b_J d P_{iJ}. \end{aligned} \tag{17}$$

Note that the a and b factors are only meaningful in terms of their relative differences; both c and d are singular constants, and hence do not affect the interpretation of the a and b values.

The full iterative sequence for finding the balancing factors becomes

$$\begin{aligned} a_i &= \frac{R_i}{\sum_{J \in nbh_i} b_J P_{iJ}}, \\ a_I &= \sum_{i \in I} a_i \frac{O_i}{O_I}, \\ c &= \frac{T}{\sum_j \sum_{I \in nbh_j} a_I b_j P_{Ij}}, \\ b_j &= \frac{C_j}{\sum_{I \in nbh_j} a_I P_{Ij}}, \\ b_J &= \sum_{j \in J} b_j \frac{D_j}{D_J}, \\ d &= \frac{T}{\sum_i \sum_{J \in nbh_i} a_i b_J P_{iJ}}, \end{aligned} \tag{18}$$

where $T = \sum_i R_i = \sum_j C_j$.

Adaptive zoning details

Having introduced the basic principles of adaptive zoning in the subsection "Principles of adaptive zoning" and extending the spatial interaction model in "The doubly constrained gravity model," we are now in a position to discuss details of the adaptive zoning method, particularly the comparison and merge operations. Consider an extended synthetic example that consists of 400 zones in a regular grid of 20-by-20 cells. Two variants are considered: one with uniform population density, and one where the density varies such that there are two urban centers separated by a "green belt" (Fig. 3a).

The hierarchy of zones is formed by iteratively comparing the distance between zone pairs and merging the "nearest" zone pair. For the purpose of minimizing the associated aggregation error, "nearness" is formally defined as

$$c_{ab} = G_{\{a,b\}} - G_a - G_b, \tag{19}$$

where c_{ab} is the merging criterion for zones a and b, $\{a,b\}$ indicates the merged zone of a and b, and G_j is the aggregation error associated with zone j and refers back to equation (10), which states that $G_j = D_j e^{\beta d_{jj}}$. This criterion drives the algorithm toward compact zones as well as toward a uniform distribution of aggregation errors at each level in the zone hierarchy. This procedure has two related effects: first, having a uniform distribution of aggregation error implies that the aggregation error is distributed evenly across the study area at each hierarchical level; second, for a given location, a variety of aggregation errors is present across the hierarchy. Therefore, the neighborhood algorithm has a full range of error levels from which to choose. The size of the merged zone follows by simple summation:

$$D_{\{a,b\}} = D_a + D_b. \tag{20}$$

Distances of the merged zone are found by area-weighted averaging, which is consistent with the definition of the distance between zones in equation (3):

$$\begin{aligned} d_{i\{a,b\}} &= \frac{A_a d_{ia} + A_b d_{ib}}{A_a + A_b}, \\ d_{\{a,b\}j} &= \frac{A_a d_{aj} + A_b d_{bj}}{A_a + A_b}, \\ d_{\{a,b\}\{a,b\}} &= \frac{A_a^2 d_{aa} + A_a A_b (d_{ab} + d_{ba}) + A_b^2 d_{bb}}{(A_a + A_b)^2}. \end{aligned} \tag{21}$$

Fig. 3b portrays a sample of results of the aggregation algorithm for the two variants in the illustration case.

The neighborhood is formed by incrementally splitting the aggregated neighborhood zone with the highest associated aggregation error. Following equation (9) this quantity is proportional to $O_i D_j e^{-\beta d_{ij}} \left| e^{\beta(d_{ii}+d_{jj})} - e^{-\beta(d_{ii}+d_{jj})} \right|$.

Fig. 3c portrays a sample of results of the neighborhood algorithm for the two variants. This graphic demonstrates that the aggregate zone sizes in the neighborhoods reflect population density as well as distance to the origin zone.

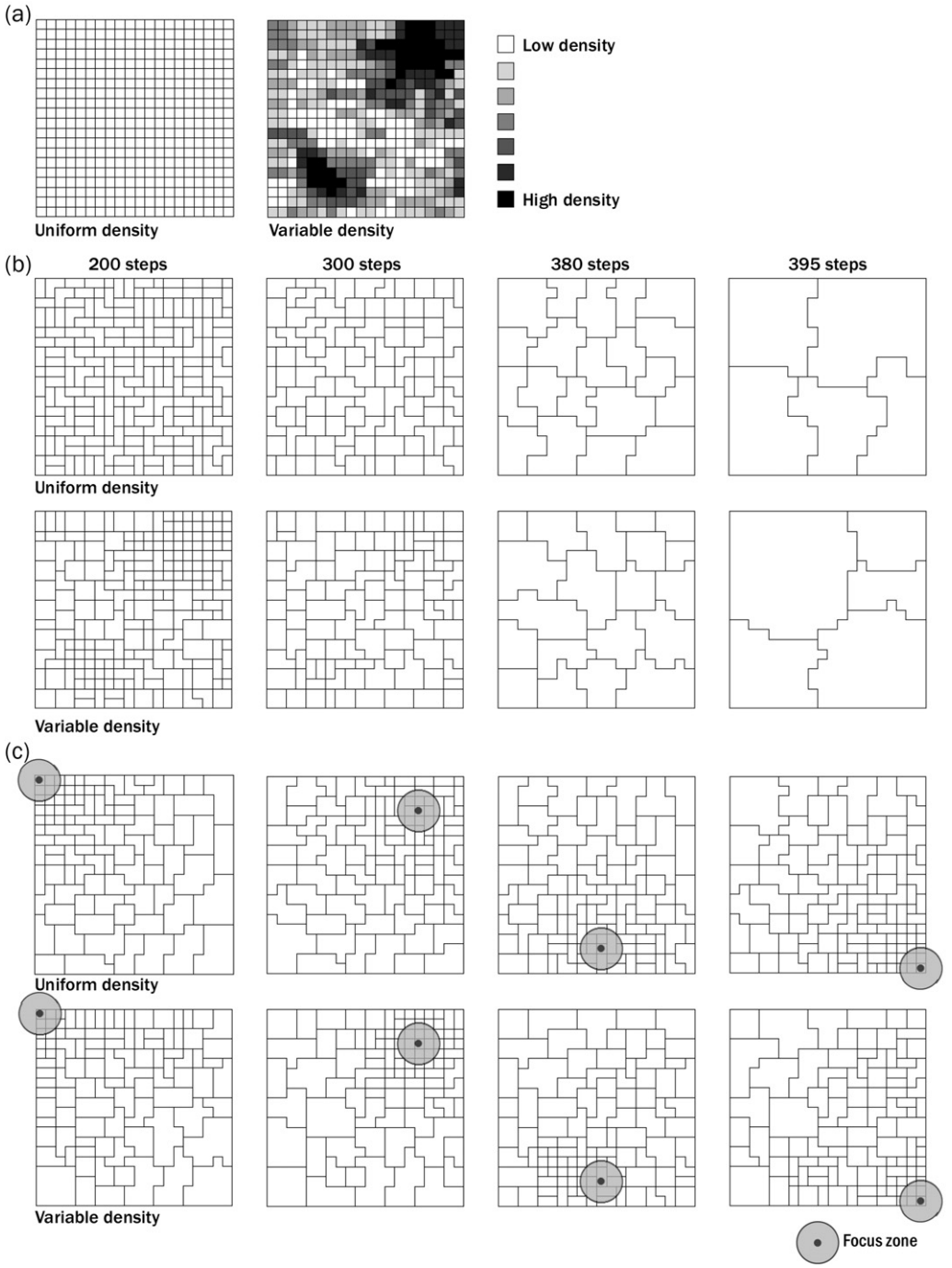

Figure 3. (a) Two example cases: a system of 400 zones with either uniform or variable density; (b) Four levels of spatial aggregation for both cases (arbitrarily selected out of 400 levels); (c) Four neighborhoods for both cases (arbitrarily selected out of 400 neighborhoods).

An information theoretic evaluation

How do we assess the performance of a model using adaptive zoning—for example, when compared with a conventional model with the most detailed zoning (i.e., atomic zoning), which implies best performance but a high computation time? The adaptive zoning model is expected to suffer from aggregation error. Therefore, we rely on information theoretic measures (Shannon 1948) to quantify the information loss owing to aggregation. To bring that information loss into perspective, we also assess divergence between the models and the observed data using the same theoretic framework.

Loss of information is based on the measure of entropy that is central to information theory. It is used to quantify information, is proportional to the number of bytes necessary to store the information contained in a signal, and may be written as follows:

$$H(P) = -\sum_i p_i \log p_i, \tag{22}$$

where *H(P)* is the entropy of a discrete distribution $P = \{p_1, p_2, \ldots p_n\}$, which is normalized such that $\sum_i p_i = 1$. When a signal consists of values in an interaction matrix,

$$\begin{aligned} p_{ij} &= \frac{T_{ij}}{\sum_i \sum_j T_{ij}}, \\ H(T) &= -\sum_i \sum_j p_{ij} \log p_{ij}. \end{aligned} \tag{23}$$

Thus, the loss of information due to aggregation is

$$\Delta H = H\left(T^{full}\right) - H\left(T^{aggregated}\right). \tag{24}$$

The following Kullback–Leibler distance (Kullback and Leibler 1951) expresses how much additional information is required to transform one signal to another:

$$D_{KL}(P \| Q) = \sum_i p_i \log \frac{p_i}{q_i}, \tag{25}$$

where D_{KL} is the Kullback–Leibler distance between the two distributions *P* and *Q*. The adaptation to interaction matrices is analogous to that of entropy, yielding

$$D_{KL}\left(T^{model} \| T^{reality}\right) = \sum_i \sum_j p_{ij}^{model} \log \frac{p_{ij}^{model}}{p_{ij}^{reality}}. \tag{26}$$

A test case

To assess the proposed adaptive zoning method, we test it on a simple yet realistically dimensioned case study of commuting to work in England. The spatial patterns of commuting and their future evolution are of direct policy interest, as is demonstrated by the modeling of commuting to work in the U.K. National Transport Model (Department for Transport 2012) and by a London-focused land use and transport model covering Britain (Jin, Williams, and Shahkarami 2002).

The commuting data are derived from the U.K. 2001 Census at the level of standard table wards ("wards" hereafter; they are broadly based on electoral wards in England but not identical to them). The data used are from Table W201 available from the Centre for Interaction Data Estimation and Research (http://cider.census.ac.uk). Wards are the most detailed geography at which 2001 Census commuting data are made available. England is partitioned into 7,932 wards (commuting into and out of England represents a very small number of trips and therefore is ignored here). The constraints for total commutes into and out of each ward are derived directly from the source table. One complication of the data is the small cell adjustment method (SCAM) that was applied to the matrix before release in order to protect confidentiality. Commuting flows of exactly one or two persons have been reassigned to either zero or three persons using an unspecified procedure by the United Kingdom's Office of National Statistics (Stillwell and Duke-Williams 2007). The data are used here in their published form, and no further attention is paid to the effect of SCAM. A second data source used is U.K. Borders (http://edina.ac.uk/ukborders/), which disseminates digital U.K. boundary data sets; it was utilized to sample Euclidean distances as described in the subsection "Sampling-based aggregation."

Two doubly constrained spatial interaction models have been built: one with conventional zoning at the level of 7,932 wards, and the other with adaptive zoning comprising 300 zones in the neighborhood of each origin or destination zone. The case study is grossly simplified in many ways; for example, by lumping all jobs and workers together in a single category, by depending on Euclidean distances, and by ignoring the effects of spatial autocorrelation and the numbers commuting into and out of England. One needs to keep in mind that the purpose of the case is to test and demonstrate the adaptive zoning system on a realistically dimensioned and representative model without delving into further complexities.

Results

The first step of the analysis was to estimate the β parameter by means of calibration. In a simple bracketing optimization called golden section search (Press 1992), the value of β is found that minimizes the difference between modeled and actual trip distributions measured by $\sum_i \sum_j \left| T_{ij}^{model} - T_{ij}^{reality} \right|$ as proposed by Smith and Hutchinson (1981). This is a relatively naive approach to model estimation, because it does not employ recent techniques to account for spatial autocorrelation (Fischer and Griffith 2008; LeSage and Pace 2008). The resulting value for β is 0.27 km^{-1}, which is used to create the adaptive zoning system and to estimate the model with adaptive zones.

A selection of results of the adaptive zoning algorithm is given in Fig. 4. Each neighborhood is set to 300 zones. This number, though arbitrary, is of the same magnitude as the number of zones in a London-focused model of Britain (Jin, Williams, and Shahkarami 2002). Because we begin with 7,932 wards, the number of zone pairs is reduced by 96%. The zone boundary maps illustrate how the size of neighborhood zones is a function of both distance and population density.

The detected discrepancy between the adaptive zoning-based model and the traditional model has two dimensions. First is the loss of information due to aggregation of the interaction matrix, and second is the propagation of effects when the model is estimated under adaptive zoning.[1] The loss of information due to aggregation is quantified as

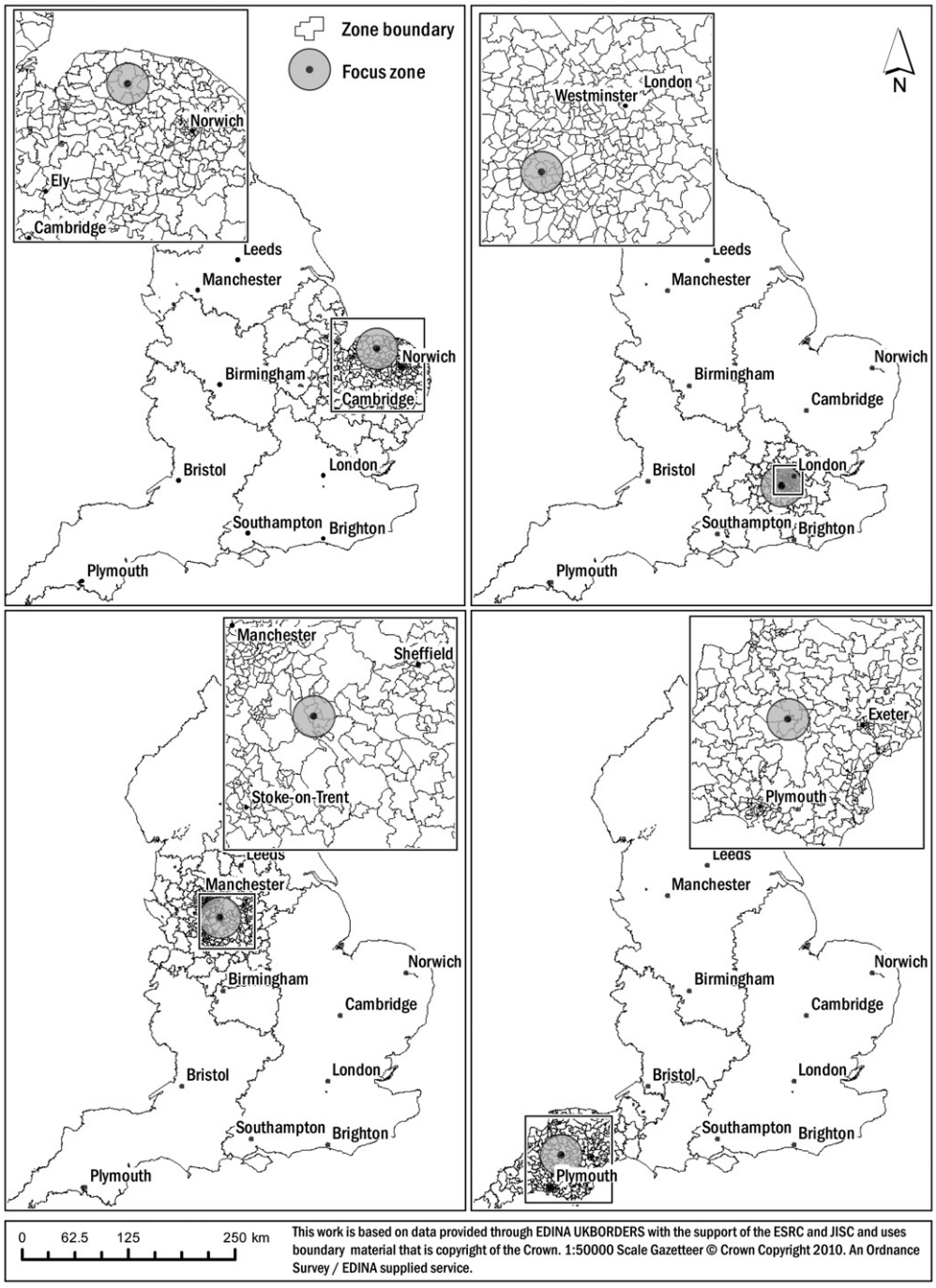

Figure 4. Four neighborhoods for modeling commuting patterns in England (arbitrarily selected out of the full set of 7,932 neighborhoods).

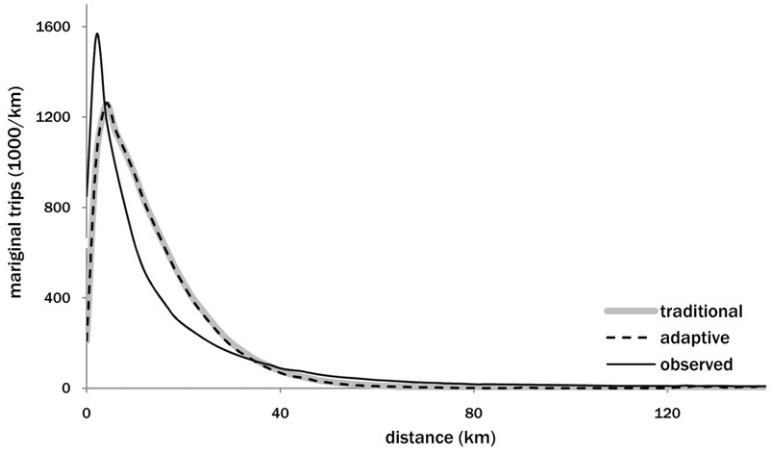

Figure 5. Trip-distance distribution for the traditional model, the adaptive zoning-based model, and the England census data.

$$\Delta H = \frac{H\left(T^{traditional,full}\right) - H\left(T^{traditional,aggregated}\right)}{H\left(T^{traditional,full}\right)} = \frac{12.38 - 12.05}{12.38} = 2.6\%, \tag{27}$$

where $T^{traditional,full}$ is the full trip matrix that results from the traditional model. $T^{traditional,aggregated}$ is the same trip matrix but aggregated using the adaptive zoning system. In this case, the full matrix consists of $7{,}932^2$ cells, whereas the aggregated sparse matrix has $7{,}932 \times 300$ nonzero cells. A reduction in number of cells by 96% has led to a loss of information of 2.6%, indicating a large degree of redundancy in the full matrix.

The second model discrepancy dimension is measured as the Kullback–Leibler distance relative to the total information content, which is

$$\frac{D_{KL}\left(T^{traditional,aggregated}, T^{adaptive}\right)}{H\left(T^{traditional,full}\right)} = \frac{0.023}{12.38} = 0.26\%, \tag{28}$$

where $T^{adaptive}$ is the trip matrix that follows from the adaptive zoning-based model. The divergence between the models is small (0.26%) compared with the direct difference due to aggregation (2.6%). This finding suggests that the accumulation of aggregation errors may be negligible, and modelers can make informed decisions about the degree of aggregation when setting the total zone number of each neighborhood.

Fig. 5 brings the correspondence between the models in context of their fit with the observations. The figure presents the trip-distance distribution for the traditional model, the adaptive zoning-based model, and the recorded data from the U.K. 2001 Census. These results show no substantial difference between the two models. The discrepancy between the models and the recorded commuting pattern is reasonable given the simplistic nature of the model and its known limitations (De Vries, Nijkamp, and Rietveld 2009).

The close relation between the model and the adaptive zoning system presents a chicken-or-egg problem when calibrating the model. To calibrate a model, a zone system is required; yet, to create a zone system, ideally the parameters of the model should be known. To overcome this

problem, either a discrepancy must be accepted between the parameters underlying the zone system and the parameters used by the model, or an iterative procedure must be applied. This iterative procedure begins with an initial best-guess value for β, which then is used to create the adaptive zone system under which the model is calibrated. Next, the resulting value is used to create a new adaptive zone system, and the model is calibrated again. These steps may be repeated until the parameter β converges. For the case of commuting in England, the initial guess of the value for β was 0.1 km^{-1}, which was used to create the adaptive zone system; next, the gravity model was calibrated, yielding a new value of 0.27 km^{-1}. A new zone system was created based on this value, and again the gravity model was calibrated, yielding a value of 0.27 km^{-1} and confirming that the model has converged. Note that the second calibration was much faster than the first because the parameter domain could be more narrowly bracketed. Thus, the iterative procedure displayed very good convergence, needing only a single iteration to find the optimal value of β when given a reasonable initial estimate.

When comparing the computation time for the traditional and adaptive zoning-based model, a number of factors are at play: the direct effect of having fewer zone pairs to process; the additional computation required for the newly introduced balancing factors c and d; a possible difference in the number of iterations necessary to converge on the balancing factors; and the one-shot cost of creating the zone hierarchy and the zone neighborhoods. As such, there is a combination of variable and constant costs, and the comparison should account for both in realistic proportions. Therefore, we compared the computation time necessary for a full calibration, including all the overheads associated with adaptive zoning. The results show a marked reduction in computation time by 70%, from 4,202 to 1,265 s.

Underlying the issue of computation time is the issue of scalability. The cost of the traditional doubly constrained gravity model has a complexity of $O(n^2)$. Under adaptive zoning, this improves to $O(n \times s)$. When s is taken as a constant, this simplifies to $O(n)$, which highlights the theoretical advantage of adaptive zoning. The creation of the zone hierarchy requires a sorted list of all zone pairs, which has an associated complexity of $O(n^2 \log n)$. The creation of neighborhoods requires one sorted list of aggregate zones in each neighborhood, with an associated complexity of $O(n \log s)$; again, considering s constant gives $O(n)$. In other words, adaptive zoning improves the complexity of the doubly constrained gravity model from $O(n^2)$ to $O(n)$, if not for the hierarchy generation, which is $O(n^2 \log n)$. These are theoretical properties, valid only at the limit where $n \to \infty$; nevertheless, the cost of the hierarchy generation remained very reasonable. For even larger cases where the cost of the hierarchy creation may become substantial, different options are available: use an existing hierarchy, such as nested administrative regions and zones; create a hierarchy that nests within an existing hierarchy; or, apply less exhaustive but better scalable clustering techniques (Kaufman and Rousseeuw 2005).

The analysis is similar in terms of storage and dynamic memory requirements: the number of computed elements in the interaction matrices changes from n^2 in the conventional case to $n \times s$ for adaptive zoning. Sparse matrix structures require some additional storage overhead for row-and-column indices, because the matrix values are not stored in plain arrays. Furthermore, an additional matrix is required as both aggregations by origin and by destination are concurrently kept in memory. The memory requirements for the zone hierarchy ($2n$ elements) and the neighborhoods ($n \times s$ elements) also are small compared with the full conventional matrices ($n \times n$ elements). Most storage is required for the distance matrix (n^2 elements), and most working memory for the sorted list of all zone pairs (n^2 elements). These dimensions were not

prohibitive in the applied case study; if they become problematic in other analyses, the solution may be to use existing hierarchies or alternative clustering methods.

Conclusion

This article has demonstrated how an adaptive zoning system can drastically reduce the number of interacting zone pairs in spatial interaction models while maintaining the granularity of the original zone system. In our case study about commuting in England, the results of the adapted zoning-based model are equivalent to the traditional model, despite reducing the number of zone pairs by 96% and the computation time by 70%.

The adaptive zoning system has potential for reducing the computational time of existing spatial interaction models and for bringing much larger spatial scales and finer resolutions into the scope of realistic applications, notably in land use and transport interaction modeling and activity-based modeling, where computational demands still constrain model applications. Improved efficiency can open avenues for systematic analysis, including calibration, validation, sensitivity analysis, and scenario analysis.

However, any advance is subject to the adjustability of the model at hand. Models cannot automatically be executed on the basis of an adaptive zoning system. They require adjustment for the specific structure of the system. In many models, the focus intermittently is on the dispatching and receiving end of interactions (as in the estimation of a and b factors in the worked case study); therefore, double bookkeeping may often be necessary, tabulating interactions according to alternate aggregations of both the origin zones and the destination zones. Reconciling both aggregations in a single model will likely be a recurring theme in adaptive zoning-based modeling. In our case study, this problem was solved by the introduction of the c and d factors that absorb the discrepancy between the two aggregations (equation [17]).

The processes of zone clustering and neighborhood generation are driven by a rationale toward minimizing aggregation error. They do not find the exact optimal solution, and research about a more formal system of error minimization may be warranted. The rationale of error minimization, however, appears quite powerful as the automatically generated neighborhood maps seem rather close to what one can expect from "expert judgment." For example, the London-centered neighborhood in Fig. 4 is reassuringly close to the zoning scheme used for some London-centered transport models (Jin, Williams, and Shahkarami 2002). Based on these results, one might envision a framework of error minimization that extends to, among other things, a broader class of decisions about model dimensioning, including the size of time steps and the number and nature of socioeconomic segments.

The model in this article is static in nature and used under well-known conditions. In practice, conditions are likely to be unknown a priori. For example, here we first estimated the β parameter on the basis of the traditional fully detailed model. In a more realistic case, the fully detailed model may not be available and β may have to be estimated by means of the adaptive zoning system. The very good convergence of the outlined iterative bootstrapping procedure indicates not only that a stable solution can be found but also that executing the model with a suboptimal but reasonable adaptive zoning system is not necessarily problematic.

The acceptability of a suboptimal adaptive zoning system is an important consideration for dynamic models, too. From one time step to the next, circumstances will be different, as will the best-matching zoning system. A further development may be toward adaptive *dynamic* zoning systems in which the systems are developed not offline in a preprocessing step but rather as an

integral part of the model; then each change of circumstances can be cause for a newly generated adaptive zoning system. In this context, the neighborhood generation algorithm has the advantage of following a process of disaggregation rather than of aggregation. Variables such as interaction matrices, distances, and error estimates would need to be computed only for the geographic resolution required by the model and not at finer levels.

Management gurus like to quote the 80–20 rule, which says that 80% of resources are often devoted to 20% of a problem. Substantial advances in computation time are achieved only when optimizations address the actual calculation bottlenecks. In the case of transport modeling, "distribution" or spatial interaction—as addressed in this article—is a crucial step but not currently the major bottleneck. The true time-consumer is "traffic assignment," in particular, the process of attributing trips between zones to individual road segments in a transport network. We have already explored this issue with some success and hope to make it the subject of a subsequent article.

Besides spatial interaction modeling of various kinds, we see scope to investigate several other applications: to reduce the search space for complex optimization problems such as location-allocation modeling (Fotheringham, Densham, and Curtis 1995), to aggregate spatial interaction data to an optimal geographic resolution while protecting confidentiality (Martin 2000; Martin, Nolan, and Tranmer 2001; Young, Martin, and Skinner 2009), and finally to use geovisualization when the spatial distribution of zone sizes corresponds to the spatial distribution of an underlying pattern.

Acknowledgements

This research took place as part of the Energy Efficient Cities initiative, a project funded by the U.K. Engineering and Physical Sciences Research Council (EPSRC) at the University of Cambridge. This work benefited from encouragement by Professors Marcial Echenique and Koen Steemers, proofreading by Vassilis Zachariadis and Richard Mounce at the University of Cambridge, and extensive comments at an early stage by Ian Williams of WSP Policy & Research, Cambridge. We would also like to thank three anonymous referees and the journal editor for helpful and constructive comments.

Note

1 Note that there are always two aggregated matrices; by origin or by destination. The results are very similar, and for brevity only the worse result is given here.

References

Andersson, C., K. Lindgren, S. Rasmussen, and R. White. (2002). "Urban Growth Simulation from 'First Principles'." *Physical Review E* 66(2), Article 026204:1–9.

Andersson, C., S. Rasmussen, and R. White. (2002). "Urban Settlement Transitions." *Environment and Planning B: Planning and Design* 29(6), 841–65.

Arentze, T. A., and H. J. Timmermans. (2004). "A Learning-Based Transportation Oriented Simulation System." *Transportation Research Part B: Methodological* 38(7), 613–33.

Axhausen, K. W., and T. Gärling. (1992). "Activity-Based Approaches to Travel Analysis: Conceptual Frameworks, Models, and Research Problems." *Transport Reviews: A Transnational Transdisciplinary Journal* 12(4), 323–41.

Batty, M. (2009). "Urban Modelling." In *International Encyclopedia of Human Geography*, Vol. 12, 51–8, edited by R. Kitchin and N. Thrift. London: Elsevier.

Ben-Akiva, M. E., and S. R. Lerman. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge: MIT Press.

Berger, M. J., and P. Colella. (1989). "Local Adaptive Mesh Refinement for Shock Hydrodynamics." *Journal of Computational Physics* 82(1), 64–84.

Clarke, K. C., S. Hoppen, and L. Gaydos. (1997). "A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area." *Environment and Planning B: Planning and Design* 24(2), 247–62.

Crosetto, M., S. Tarantola, and A. Saltelli. (2000). "Sensitivity and Uncertainty Analysis in Spatial Modelling Based on GIS." *Agriculture, Ecosystems and Environment* 81(1), 71–9.

De La Barra, T. L., B. Pérez, and N. Vera. (1984). "TRANUS-J: Putting Large Models into Small Computers." *Environment and Planning B: Planning and Design* 11(1), 87–101.

De Vries, J. J., P. Nijkamp, and P. Rietveld. (2009). "Exponential or Power Distance-Decay for Commuting? An Alternative Specification." *Environment and Planning A* 41(2), 461–80.

Department for Transport. (2012). "Road Transport Forecasts 2011: Results from the Department for Transport's National Transport Model." Report. Available at http://assets.dft.gov.uk/publications/road-transport-forecasts-2011/road-transport-forecasts-2011-results.pdf (accessed on 20 July 2012).

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: Planning and Design* 25(4), 601–16.

Echenique, M. (2004). "Econometric Models of Land Use and Transportation." In *Handbook of Transport Geography and Spatial Systems*, 185–202, edited by D. A. Hensher, K. J. Button, K. E. Haynes, and P. R. Stopher. Amsterdam: Elsevier.

Fischer, M. M., and D. A. Griffith. (2008). "Modeling Spatial Autocorrelation in Spatial Interaction Data: An Application to Patent Citation Data in the European Union." *Journal of Regional Science* 48(5), 969–89.

Fotheringham, A. S., P. J. Densham, and A. Curtis. (1995). "The Zone Definition Problem in Location-Allocation Modeling." *Geographical Analysis* 27(1), 60–77.

Fratar, T. (1954). "Vehicular Trip Distribution by Successive Approximation." *Traffic Quarterly* 8(1), 53–65.

Gitlesen, J. P., G. Kleppe, I. Thorsen, and J. Ubøe. (2010). "An Empirically Based Implementation and Evaluation of a Hierarchical Model for Commuting Flows." *Geographical Analysis* 42(3), 267–87.

Goodchild, M. (1979). "The Aggregation Problem in Location-Allocation." *Geographical Analysis* 11(3), 240–55.

Gordon, I. (2010). "Entropy, Variety, Economics, and Spatial Interaction." *Geographical Analysis* 42(4), 446–71.

Guare, C. (1991). "Error, Precision, and Uncertainty." *Journal of Chemical Education* 68(8), 649–52.

Hagen-Zanker, A., and Y. Jin. (2011). "Reducing Aggregation Error in Spatial Interaction Models by Location Sampling." In *Proceedings of the 11th International Conference on GeoComputation*, London, July 2011.

Hagen-Zanker, A., and P. Martens. (2008). "Map Comparison Methods for Comprehensive Assessment of Geosimulation Models." In *Computational Science and Its Applications—ICCSA 2008*, part 2, 194–209, edited by O. Gervasi, B. Murgante, A. Laganà, D. Taniar, Y. Mun, and M. L. Gavrilova. Berlin: Springer.

Hillsman, E. L., and R. Rhoda. (1978). "Errors in Measuring Distances from Populations to Service Centers." *Annals of Regional Science* 12(3), 74–88.

Jin, Y., I. Williams, and M. Shahkarami. (2002). "A New Land Use and Transport Interaction Model for London and Its Surrounding Regions." In *Proceeding of the European Transport Conference*, Cambridge, September 2002.

Kaufman, L., and P. J. Rousseeuw. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken: Wiley.

Kocabas, V., and S. Dragicevic. (2006). "Assessing Cellular Automata Model Behaviour Using a Sensitivity Analysis Approach." *Computers, Environment and Urban Systems* 30(6), 921–53.

Kullback, S., and R. A. Leibler. (1951). "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22(1), 79–86.

LeSage, J. P., and R. K. Pace. (2008). "Spatial Econometric Modeling of Origin-Destination Flows." *Journal of Regional Science* 48(5), 941–67.

Martin, D. (2000). "Towards the Geographies of the 2001 UK Census of Population." *Transactions of the Institute of British Geographers* 25(3), 321–32.

Martin, D., A. Nolan, and M. Tranmer. (2001). "The Application of Zone-Design Methodology in the 2001 UK Census." *Environment and Planning A* 33(11), 1949–62.

Miller, S., A. Daly, J. Fox, and S. Kohli. (2007). "Destination Sampling in Forecasting: Application in the PRISM Model for the UK West Midlands Region." In *Proceedings of the European Transport Conference*, Noordwijkerhout, October 2007.

Openshaw, S. (1978). "An Empirical Study of Some Zone-Design Criteria." *Environment and Planning A* 10(7), 781–94.

Openshaw, S. (1984). "Ecological Fallacies and the Analysis of Areal Census-Data." *Environment and Planning A* 16(1), 17–31.

Parker, D. C., S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman. (2003). "Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review." *Annals of the Association of American Geographers* 93(2), 314–37.

Press, W. H. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge: Cambridge University Press.

Sancar, C. (2010). "Integration of Planning Models and GIS in Defining Urban Development Areas." *International Journal of Sustainable Development & World Ecology* 17(2), 133–41.

Schönfelder, S., and K. W. Axhausen. (2010). *Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel*. Farnham: Ashgate.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal* 27(4), 623–56.

Smith, D. P., and B. G. Hutchinson. (1981). "Goodness of Fit Statistics for Trip Distribution Models." *Transportation Research Part A: General* 15(4), 295–303.

Stillwell, J., and O. Duke-Williams. (2007). "Understanding the 2001 UK Census Migration and Commuting Data: The Effect of Small Cell Adjustment and Problems of Comparison with 1991." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170(2), 425–45.

Tobler, W. R. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46(Supplement: Proceedings. International Geographical Union Commission on Quantitative Methods), 234–40.

Tobler, W. R. (1989). "Frame Independent Spatial Analysis." In *Accuracy of Spatial Databases*, 115–22, edited by M. F. Goodchild and S. Gopal. London: Taylor & Francis.

Train, K. (2009). *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge: Cambridge University Press.

Van Vliet, J., R. White, and S. Dragicevic. (2009). "Modeling Urban Growth Using a Variable Grid Cellular Automaton." *Computers, Environment and Urban Systems* 33(1), 35–43.

Van Waveren, R. H., S. Groot, H. Scholten, F. V. Geer, H. Wösten, R. Koeze, and J. J. Noort. (1999). "*Good Modelling Practice Handbook*." STOWA report 99-05, STOWA / RIZA.

Waddell, P. (2000). "A Behavioral Simulation Model for Metropolitan Policy Analysis and Planning: Residential Location and Housing Market Components of UrbanSim." *Environment and Planning B: Planning and Design* 27(2), 247–63.

Waddell, P., G. Ulfarsson, J. Franklin, and J. Lobb. (2007). "Incorporating Land Use in Metropolitan Transportation Planning." *Transportation Research Part A: Policy and Practice* 41(5), 382–410.

Webster, C. J., and F. Wu. (1999). "Regulation, Land-Use Mix, and Urban Performance. Part 2: Simulation." *Environment and Planning A* 31(9), 1529–45.

Wegener, M. (2009). "From Macro to Micro—How Much Micro Is Too Much?" Paper presented at *the International Seminar on Transport Knowledge and Planning Practice, University of Amsterdam, 14–16 October*.

Wegener, M. (2011). "The IRPUD Model." *Working Paper 11/01*, Spiekerman & Wegener Urban and Regional Research consulting firm. Available at http://www.spiekermann-wegener.de/mod/pdf/AP_1101_IRPUD_Model.pdf (accessed on 20 July 2012).

White, R., G. Engelen, and I. Uljee. (1997). "The Use of Constrained Cellular Automata for High-Resolution Modelling of Urban Land-Use Dynamics." *Environment and Planning B: Planning and Design* 24(3), 323–43.

Wickramasuriya, R. C., A. K. Bregt, H. V. Delden, and A. Hagen-Zanker. (2009). "The Dynamics of Shifting Cultivation Captured in an Extended Constrained Cellular Automata Land Use Model." *Ecological Modelling* 220(18), 2302–9.

Williams, I., and C. Lindsay. (2002). "An Efficient Design for Very Large Transport Models on PCs." In *Proceeding of the European Transport Conference*, Cambridge, September 2002.

Young, C., D. Martin, and C. Skinner. (2009). "Geographically Intelligent Disclosure Control for Flexible Aggregation of Census Data." *International Journal of Geographical Information Science* 23(4), 457–82.