

Hagen-Zanker, Alex; Jin, Ying

Conference Paper

Adaptive zoning and its effectiveness in spatial economic activity simulation

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Hagen-Zanker, Alex; Jin, Ying (2010) : Adaptive zoning and its effectiveness in spatial economic activity simulation, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/119095>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Adaptive zoning and its effectiveness in spatial economic activity simulation

Alex Hagen-Zanker and Ying Jin

Department of Architecture, University of Cambridge, 1-5 Scroope Terrace, Cambridge, CB2 1PX, United Kingdom

Email: ahh34@cam.ac.uk

Abstract

Spatial interaction models commonly partition the study area into a number of zones and model the interaction between all zone pairs. For many applications this implies that the computational demands, both in terms for processing time and memory allocation, are quadratic to the number of zones. These practical implications compel modellers to reduce the number of zones and trade accuracy for efficiency. This paper presents an approach to modelling spatial interaction to reduce the computational demands while maintaining accuracy. The main idea is to create for each zone a set of clustered zones that it interacts with. The degree of clustering of interacting is governed by the error caused by aggregation. Areas with strong spatial interaction are modelled at high detail whereas the precision for weak interactions is lower. For the same number of zones the interaction will be based on a considerably smaller number of zone-pairs. In a typical model where interaction diminishes with distance, each zone interacts with small zones close by and increasingly large aggregated zones at further distances. The new adaptive zoning approach is applied on a doubly constrained distribution model of commuting in England. The case highlights the model modifications necessary to work with the adaptive zoning systems as well as the trade-off in efficiency and accuracy that it entails. It appears that the number of zone pairs can be reduced by 95% whilst maintaining a fine granularity.

1 Introduction

In modelling spatial economic interactions such as embodied in the movements of people and goods, the study area is usually subdivided into geographic units, i.e. zones, to represent the origin, destination and any stop-over locations. In most cases, model precision tends to improve as the study area is subdivided into increasingly small zones. Smaller zones however

come at the cost of increased run-time. In practice the modeller often has to compromise on the resolution or the geographic coverage, which impinge upon model accuracy and applicability.

Current trends in spatial economic interactions are calling for larger study areas and finer spatial details: As transport and telecommunications improve, the realm of spatial interaction continues to expand. Furthermore, the rapid expansion in e-monitoring of the movements of goods and people is offering ever-increasing amount of spatial details. The combined impact is a strong need for spatial economic modelling to expand coverage whilst capturing local details.

In this paper we report the development of an innovative zoning methodology for spatial economic modelling. The basic idea is to use more detailed zoning where interaction is stronger, e.g. at shorter distance or between higher density areas. Instead of a single zonal division for the whole study area, the zoning scheme consists of one specific zonal division for each respective location (atomic zone) in the study area adapted to the interaction flows to and from that location. Incorporating this adaptive zoning scheme improves the scaling behaviour of spatial interaction and choice models, as the number of interactions per zone is no longer equal to the number of zones, but - depending on the precise nature of the interaction - logarithmic to it instead.

This paper first details the proposed methodology to arrive at the intended zoning system. The driving thrust is to minimize the expected error due to aggregation, subject to a limitation of computing resources. Subsequently, the new zoning system is applied on a common doubly constrained spatial interaction model that requires some modifications in order to function in combination with the new adaptive zoning system.

In theory, such a targeted application of computing power leads to higher precision at smaller cost. The adaptive zoning system however also introduces its own complexities that we explore by applying the model on an actual (albeit simplified) case of modelling the trip distribution for commuting in England on the basis of UK CENSUS 2001 data. In the discussion and conclusion we reflect on the merits of the zoning system for this particular model as well as the perspectives for further integration in spatial interaction models with a focus on transport modelling.

2 Methodology

2.1 Terminology

Zones at the finest level are *atomic zones*; these cannot be subdivided. *Clustered zones* are amalgamations of *atomic zones* and other *clustered zones*. The *top zone* consists of all zones and thereby covers the *model area*. The zones making up a clustered zone are called its *child zones* and the clustered zone is their *parent zone*. All zones, except the top zone, have exactly one *parent zone*. A *neighbourhood* is a configuration of atomic zones and clustered zones that is focused on one particular atomic zone. Each atomic zone has one neighbourhood.

The methods will be explained by means of an artificial example. The atomic zones in the example are arranged in a regular grid of 20 * 20 cells. For illustration purposes, two cases are considered, one in which the cells have a uniform population density, and one where the density varies such that there are two urban centres separated and surrounded by a ‘green belt’. Distances between zones are Euclidean.

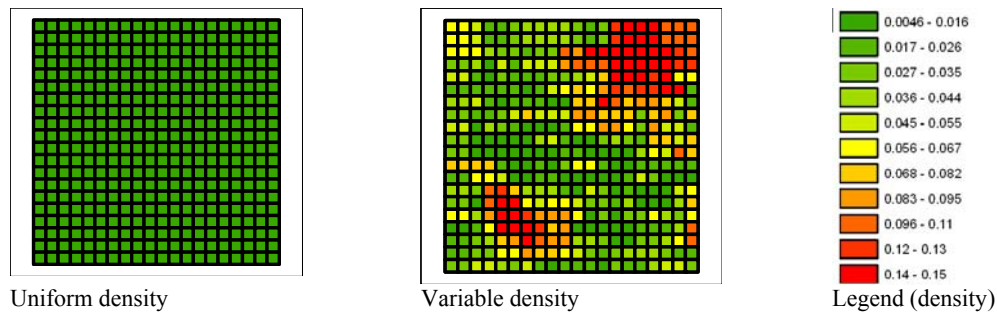


Figure 1. Atomic zoning systems used for illustration

2.2 Zonal hierarchy generation

There is a considerable literature on methods for grouping geographically detailed data into and making use of methods of clustering (Slater 1976) or optimization (Openshaw 1977).

Clustering methods are defined by their clustering algorithm. For instance, nearest neighbouring clustering iteratively merges region pairs that are the closest of all pairs.

Optimization methods are defined by their objective, which is to optimize particular characteristics. Duque et al. (2007) provide a review of clustering methods. Typically the methods aim to achieve high within-region homogeneity and high compactness. The work of Slater (Slater 1976; Slater 1987) details methods for hierarchical regionalization on the basis of links (primarily migration patterns) between zones.

For the first prototype a straightforward joining method is used. Starting point for this method is a distance matrix for all zone pairs and a vector of sizes (e.g. population) for each zone. Iteratively pairs of zones are joined on the basis of a criterion that combines distance and size. The criterion of distance follows from our overall objective to reduce the error caused by zonal aggregation and the consideration that this error is caused by the discarding of variations in distances within a zone. A zone with shorter internal distances will also have less internal variation in distances to other zones and hence cause less error. The criterion of zone size has another objective, in the following section we will explain how neighbourhoods of zones are created by selecting zones of different sizes as a function of their distance to the neighbourhood centre. The hierarchy of zones will over a more spatially balanced pallet of zone sizes if the aggregation aims to keep zone sizes equal. The combined criterion takes the following form:

$$c_{ij} = d_{ij} * (s_i + s_j) \quad (1)$$

Where, c_{ij} is the cluster criterion for zones i and j , d_{ij} is a measure of distance between zones i and j , s_i and s_j are the sizes of zones i respectively j

Upon joining, the distance matrix and size vector need to be adjusted; the two zones are taken out and a new *amalgamated* zone is introduced. The method for amalgamation that we applied for the distance matrix is the *weighted pair-group method using arithmetic averages* (WPGMA) (Sneath and Sokal 1973).

$$d_{ij} = \frac{\sum_{i \in I} s_i * d_{ij}}{\sum_{i \in I} s_i} \quad (2)$$

$$d_{jI} = \frac{\sum_{i \in I} s_i * d_{ij}}{\sum_{i \in I} s_i} \quad (3)$$

Where I is the amalgamated parent of zones i .

WPGMA does not cover intrazonal distance, i.e. the average distance from a location within a zone to another location within the zone, as it does not affect the cluster process. We will use self-distance later on however, in order to estimate aggregation errors. Therefore we use the following amalgamation rule for intrazonal distances

$$d_{II} = \frac{\sum_{i \in I} \sum_{i' \in I} s_i s_{i'} * (d_{ii'} + d_{i'i})}{2 \sum_{i \in I} \sum_{i' \in I} s_i s_{i'}} \quad (4)$$

Amalgamation of zone sizes is simple summation:

$$s_I = \sum_{i \in I} s_i \quad (5)$$

The process of finding the pair of zones with the lowest value for the criterion c_{ij} and amalgamating the corresponding zones is iterated until only one zone remains, which is the *top zone*.

When applying the aggregation algorithm on the illustration zones, a complication arises; because of the regular nature of the grid and the uniform density there are many pairs of zones scoring equally on the cluster criterion. Two different kinds of tiebreaking were used; random and north-west first. Even though this situation may not arise in reality, it indicates the role of path dependency in the formation of clusters. Figure 2 shows the results of the aggregation algorithm.

2.3 Neighbourhood generation

Neighbourhoods are derived with a view to minimize the error given the number of interactions. We try to achieve this by selecting interacting zones such that the error is distributed evenly. The rationale is to prevent spending computation time on levels of precision that are lost elsewhere. The algorithm for selecting interacting zones starts at clustered zone at the coarsest level of aggregation and incrementally breaks the zone with the smallest associated error into smaller zones, until the neighbourhood has the required number of zones. Note that even though this is a sensible heuristic approach, it is not guaranteed to arrive at the optimal solution.

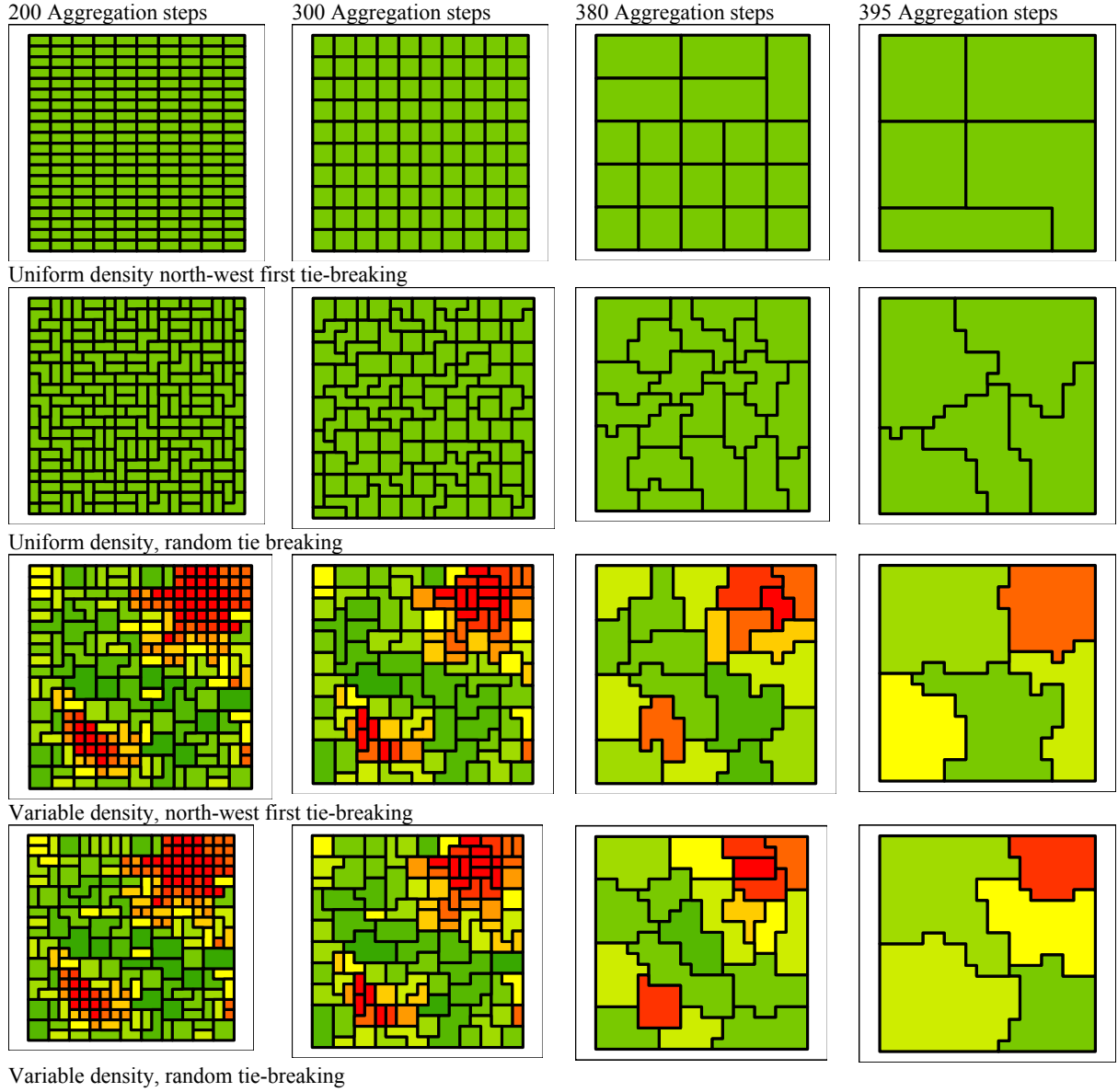


Figure 2. Aggregation results, for both demonstration cases, using two types of tie-breaking

The neighbourhood generation takes two inputs. The first is the hierarchy as derived in the hierarchy generation. The second input is the function estimating the aggregation error for interaction between two zones. The appropriate form of this function depends on the model that is being used. For now, we are assuming a typical gravity based spatial interaction model of the following form:

$$t_{ij} = s_i s_j e^{-\beta d_{ij}} \quad (6)$$

Furthermore, we are assuming that the error in d_{ij} is equal to the average intrazonal distance:

$$\Delta d_{ij} \approx \frac{d_{ii} + d_{jj}}{2} \quad (7)$$

where Δd_{ij} estimates the error in distance due to aggregation.

The derivative of the interaction with respect to distance is used to estimate the error:

$$\Delta t_{ij} = s_i s_j e^{-\beta d_{ij}} - s_i s_j e^{-\beta(d_{ij} + \Delta d_{ij})} = s_i s_j e^{-\beta d_{ij}} (1 - e^{-\beta \Delta d_{ij}}) \quad (8)$$

where Δt_{ij} estimates the error in trips due to aggregation. The interpretation of this equation is that the estimated error is a function of the relative error $(1 - e^{-\beta \Delta d_{ij}})$ and the strength of the interaction which is a function of distance and size.

2.4 Figure 3 shows the results of the neighbourhood algorithm for neighbourhoods of three atomic zones on the map. Note that there are two variables driving the size of the clustered zone in the neighbourhood; first there is distance: closer zones are smaller, second is the density: denser zones are smaller.

2.4 Crooked matrices and how to use them

Applying the neighbourhood for transport demand modelling will lead origin-destination (OD) matrices that have a non traditional form. Traditionally OD matrices tabulate interactions between all (atomic) zones in a system. In the proposed method, either the origins or the destinations are expressed as aggregated clusters. We call such matrices *crooked matrices* because they summarize the whole system of flows between origins and destinations, but not in the intuitive symmetrical form that we are used to. The matrices are sparse because every atomic zone only interacts with a limited set of clustered zones (its neighbourhood). Figure 4 illustrates the form of the new crooked matrices.

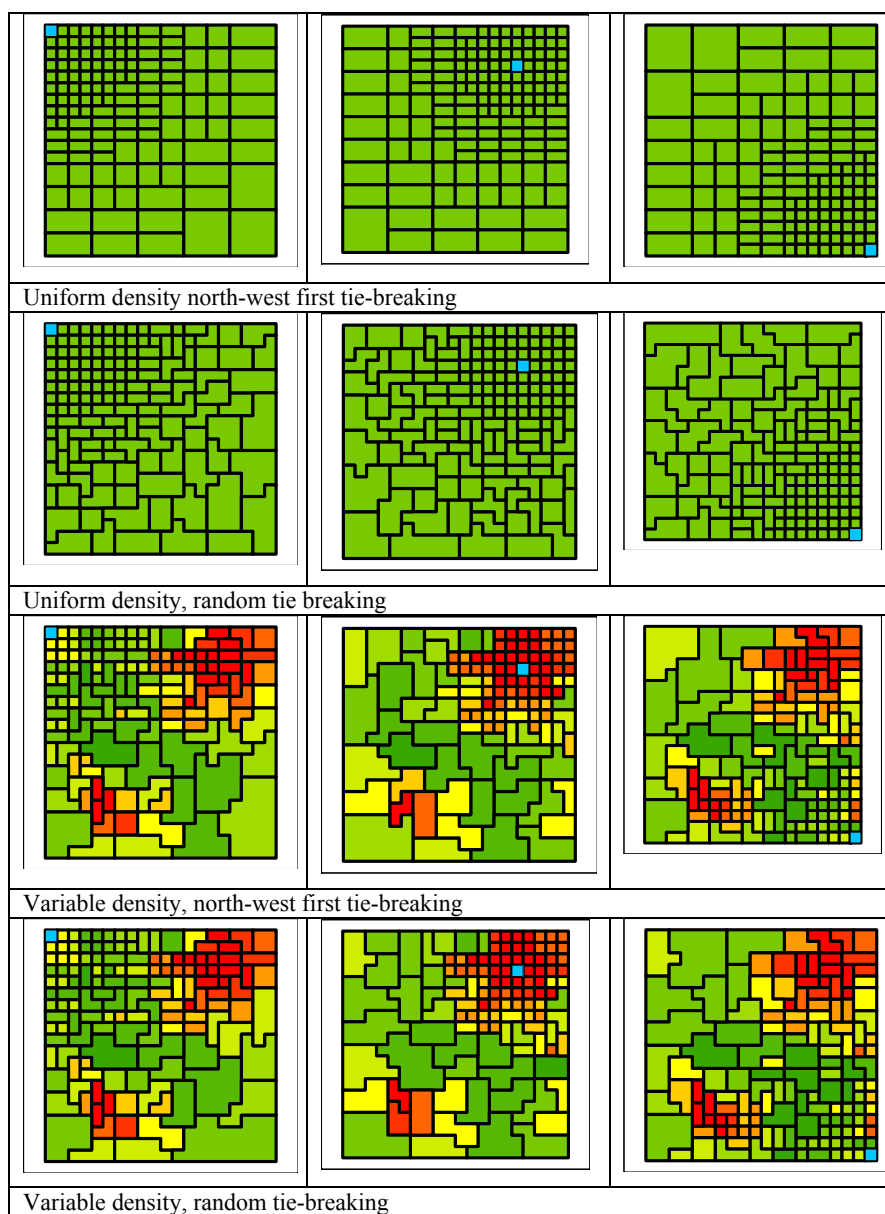


Figure 3. Results from the neighbourhood algorithm for three atomic zones (the atomic zone is coloured blue)

a. Traditional OD matrix

	Atomic Zones (destinations)
Atomic Zones (origins)	OD matrix for atomic zones (Full or Sparse)

b. Crooked matrix variant 1: Atomic origins

	Atomic Zones (destinations nearby)	Clustered Zones (farther destinations)
Atomic Zones (origins)	Sparse OD traditional matrix cells for atomic origin to nearby atomic destinations only	Sparse Crooked matrix

c. Crooked matrix variant 2: Atomic destinations

	Atomic Zones (destinations)
Atomic Zones (origins nearby)	Sparse OD traditional matrix cells for atomic origin to nearby atomic destinations only
Clustered Zones (farther origins)	Sparse Crooked matrix (variant 2)

Figure 4. Schematic representation of both traditional and proposed OD matrices

The sparse matrix of atomic zone – clustered zone interaction can pose model complications because of its underlying hierarchy and multi-scale character. Meaningful analysis will require some model adaptations. Without loss of generality, this paper uses clustered zones for destinations and atomic zones for origins (i.e. Figure 4(b)).

2.5 Adapting the doubly constrained spatial interaction model

2.5.1 Background

A commonly used spatial interaction model is the doubly constrained spatial interaction model. We incorporate adaptive zoning in this model because it presents a simple case for developing a methodology that can be applied to more complex spatial economic models. Inputs to this model are a matrix of prior distribution of interactions (probabilities) between zone pairs, a vector of constraints for interaction origins (row totals) and a vector of constraints for interaction destinations (column totals). The model finds *balancing factors*, one for each row and each column to satisfy the constraints.

The model takes the following form:

$$T_{ij} = a_i b_j P_{ij} \quad (9)$$

where T_{ij} is the interaction from i to j , P is the prior distribution and balancing factors a_i and b_j are chosen such that the following constraints are met:

$$\sum_j T_{ij} = R_i \quad (10)$$

$$\sum_i T_{ij} = C_j \quad (11)$$

where R_i is the constraint for the i -th row and C_j is the constraint for the j -th column.

Balancing factors are typically found by applying a Furness iteration that iteratively applies the following equations:

$$\begin{aligned} a_i &= \frac{R_i}{\sum_j b_j P_{ij}} \\ b_j &= \frac{C_j}{\sum_i a_i P_{ij}} \end{aligned} \quad (12)$$

In this paper the prior distribution will be based on the gravity model also applied in the step of neighbourhood generation (section 2.3). Thus:

$$P_{i,j} = s_i s_j e^{-\beta d_{ij}} \quad (13)$$

The challenge that arises in a system with zones aggregated to different degrees is twofold:

1. The values of balancing factors for clustered zones are not independent of balancing factors of their children, but instead are a function of those.
2. Interaction with clustered zones leaves undetermined how strong the interaction between atomic zones is. We worked out two mechanisms of dealing with this:
 - a. The extra degree of freedom relaxes the system and there will be a range of values for balancing factors that satisfies the constraints. In order to arrive at a unique solution an additional objective function will need to be formulated.
 - b. Alternatively, it is possible to disaggregate totals from clustered zones to aggregated zones and still apply constraints at the level of atomic zones. The advantage is that the system is fully determined again and has a unique solution.

2.5.2 *The parent-child relationship of balancing factors*

Our purpose of clustering zones is to trade precision for computation time. It therefore is imperative that the behaviour of a clustered zone is equivalent to the combined behaviour of its children. In statistical mechanics this would be called a mean field approximation.

This would require the following condition to hold:

$$T_{iJ} \approx \sum_{j \in J} T_{ij} \quad (14)$$

We are assuming that $d_{iJ} \approx d_{ij}$, which is sensible because proximity is one of the conditions for clustering.

Following through we get

$$\begin{aligned} a_i b_J s_i s_J e^{-\beta d_{iJ}} &= \sum_{j \in J} a_i b_j s_i s_j e^{-\beta d_{iJ}} \\ b_J s_J &= \sum_{j \in J} b_j s_j \\ b_J &= \frac{\sum_{j \in J} b_j s_j}{s_J} = \frac{\sum_{j \in J} b_j s_j}{\sum_{j \in J} s_j} \end{aligned} \quad (15)$$

In words, the balancing factor for a clustered zone is the size-weighted average of its children's balancing factors.

2.5.3 *Relaxed constraints and objective function*

By modelling interactions between atomic zones and clustered zones the nature of the constraints changes as well as the number of constraints. As we are clustering destinations and not origins, the constraints for origins remain the same:

$$\sum_j T_{ij} = R_i \quad (16)$$

For the clustered zones the constraints are different, because at any given clustering it is undetermined how much of the constraint is met at higher degrees of clustering. As a consequence the constraints for atomic zones become inequalities, rather than equalities:

$$\sum_i T_{i,j} \leq C_j \quad (17)$$

For clustered zones constraints follow from aggregation:

$$C_J = \sum_{j \in J} C_j \quad (18)$$

$$\sum_i \left(T_{iJ} + \sum_{j \in J} T_{ij} \right) \leq C_J, \text{ or} \quad (19)$$

$$b_J w_J + \sum_{j \in J} b_j w_j \leq C_J \quad (20)$$

where $w_j = \sum_i a_i P_{ij}$ to simplify the notation.

Only at the level of the whole model area is there a hard constraint:

$$C_T = \sum_i \sum_j T_{ij} = \sum_j b_j w_j \quad (21)$$

where T is the index of the top zone, corresponding to the whole model area.

The relaxation of constraints means that there no longer is a unique solution. In order to arrive at a well-determined system we add the objective to minimize the effect of the balancing factor:

$$obj = \sum_j ((b_j - 1) * w_j)^2 \quad (22)$$

2.5.4 *A system of equations*

The relations specified in the previous two sections (in particular equations 15, 20 and 21) can be summarized in a system of equations in the following form:

$$\begin{bmatrix}
w_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & w_2 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & w_m & 0 & 0 & \cdots & 0 \\
v_{1,m+1}w_1 & v_{2,m+1}w_2 & \cdots & v_{m,m+1}w_m & w_{m+1} & 0 & \cdots & 0 \\
v_{1,m+2}w_1 & v_{2,m+2}w_2 & \cdots & v_{m,m+2}w_m & 0 & w_{m+2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 \\
v_{1,n}w_1 & v_{2,n}w_2 & \cdots & v_{m,n}w_m & 0 & 0 & \cdots & w_n \\
w_1 & w_2 & \cdots & w_m & w_{m+1} & w_{m+2} & \cdots & w_n \\
v_{1,m+1}s_1 & v_{2,m+1}s_2 & \cdots & v_{m,m+1}s_m & -s_{m+1} & 0 & \cdots & 0 \\
v_{1,m+2}s_1 & v_{2,m+2}s_2 & \cdots & v_{m,m+2}s_m & 0 & -s_{m+2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
v_{1,n}s_1 & v_{2,n}s_2 & \cdots & v_{m,n}s_m & 0 & 0 & \cdots & -s_n
\end{bmatrix}
*
\begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_m \\
b_{m+1} \\
b_{m+2} \\
\vdots \\
b_n
\end{bmatrix}
\leq
\begin{bmatrix}
C_1 \\
C_2 \\
\vdots \\
C_m \\
C_{m+1} \\
C_{m+2} \\
\vdots \\
C_n \\
C_T \\
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
\quad (23)$$

where $v_{j,J}$ is introduced for convenience, it equals 1 if zone j is directly or indirectly a child of zone J . m is the number of atomic zones, and n is the total number of atomic and cluster zones .

The system of equations is linear, furthermore the objective function is quadratic (and convex), which means that the problem is a standard Quadratic Programming problem (Gould and Toint 2010). Conventionally the objective function is given in the following form:

$$f(x) = \frac{1}{2}x^T Qx + C^T x \quad (24)$$

Following that notation:

$$x = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, Q = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}, C = \begin{bmatrix} -w_1 \\ -w_2 \\ \vdots \\ -w_n \end{bmatrix} \quad (25)$$

There are several methods of solving this type of system that are offered as part of various software packages. (In this paper we used the library COPL_QP because it is well suited for large problems as well as sparse matrices (Zhang and Ye 1998) in the course of development we have also successfully used Excel Solver and QuadProg++, obtaining identical results for a simple test case.

2.5.5 *Alternative approach*

An alternative approach to relaxing the constraints is to assume that the arrivals in a clustered zone are distributed relative to size. This is equivalent to the assumption made before that

$d_{ij} = d_{iJ} \forall j \in J$. We then get

$$C_j = \frac{b_j s_j}{\sum_{j' \in J} b_{j'} s_{j'}} \sum_{j' \in J} b_{j'} w_{j'} = \frac{b_j s_j}{\sum_{j' \in J} b_{j'} s_{j'}} \sum_{j' \in J} \frac{\sum_{j'' \in J} b_{j''} s_{j''}}{s_{j'}} w_{j'} = b_j s_j \sum_{j' \in J} \frac{w_{j'}}{s_{j'}} \quad (26)$$

$$b_j = \frac{C_j}{s_j \sum_{j' \in J} \frac{w_{j'}}{s_{j'}}}$$

which brings the system back to a system of equalities, where the value of each b_j is independent of all others.

3 Test case

In order to evaluate the loss of accuracy due to the aggregations of the adaptive zoning system this paper will apply it on a simple test case, which nevertheless has a reasonably large number of zones. The doubly constrained distribution model is applied on commuting patterns in the area of England as measured by the UK Census of 2001 at the level of Standard Table Wards (ST Wards). The data used is Table W201 available from Centre for Interaction Data Estimation and Research (<http://cider.census.ac.uk>). ST Wards are the most detailed geography at which CENSUS commuting data is made available. ST Wards correspond to UK electoral wards, except that some small electoral wards have been aggregated to protect the anonymity of their inhabitants. In total there are 7932 ST Wards in England and the commuting matrix (commutes into and out from the England are conveniently ignored). The constraints for commutes into and out from each ST Ward are derived directly from the table. One complication in the data is the Small Cell Adjustment Method (SCAM) that is applied on the matrix before releasing it in order to prevent a breach of confidentiality. Commuting flows between ST Wards of exactly 1 or 2 persons have been reassigned to either 0 or 3 persons using an unspecified procedure by the UK's Office of National Statistics. As the current case is merely for testing and demonstration no further attention is paid to the effect of SCAM.

A second data source used is UK Borders (<http://edina.ac.uk/ukborders/>) , which disseminates digital UK boundary datasets. From this service we obtained the boundaries of all ST Ward in England and using ArcGIS 9.3 we calculated the coordinates of the (area weighted) centre of each ST Ward. Finally distances between ST Wards were calculated as the Euclidean distance between their centroids. Of course these are gross simplifications and today's interaction models apply more advanced methods of establishing the zone centre, for instance incorporating population patterns and road networks and tend to express the distance between zones in terms of generalized transport costs rather than Euclidean distance. For the purpose of the testing and demonstrating it is not opportune however to introduce further complexities.

4 Results

4.1 Calibrating β

The first step of the analysis is to find the β parameter by means of calibration. In a simple bracketing optimization β is found to optimize the correspondence between mean trip distance in reality and in the fully detailed distribution model. The resulting value for β is 0.1437 km^{-1} and is input to hierarchy and neighbourhood generation. The mean trip distance is 13 km.

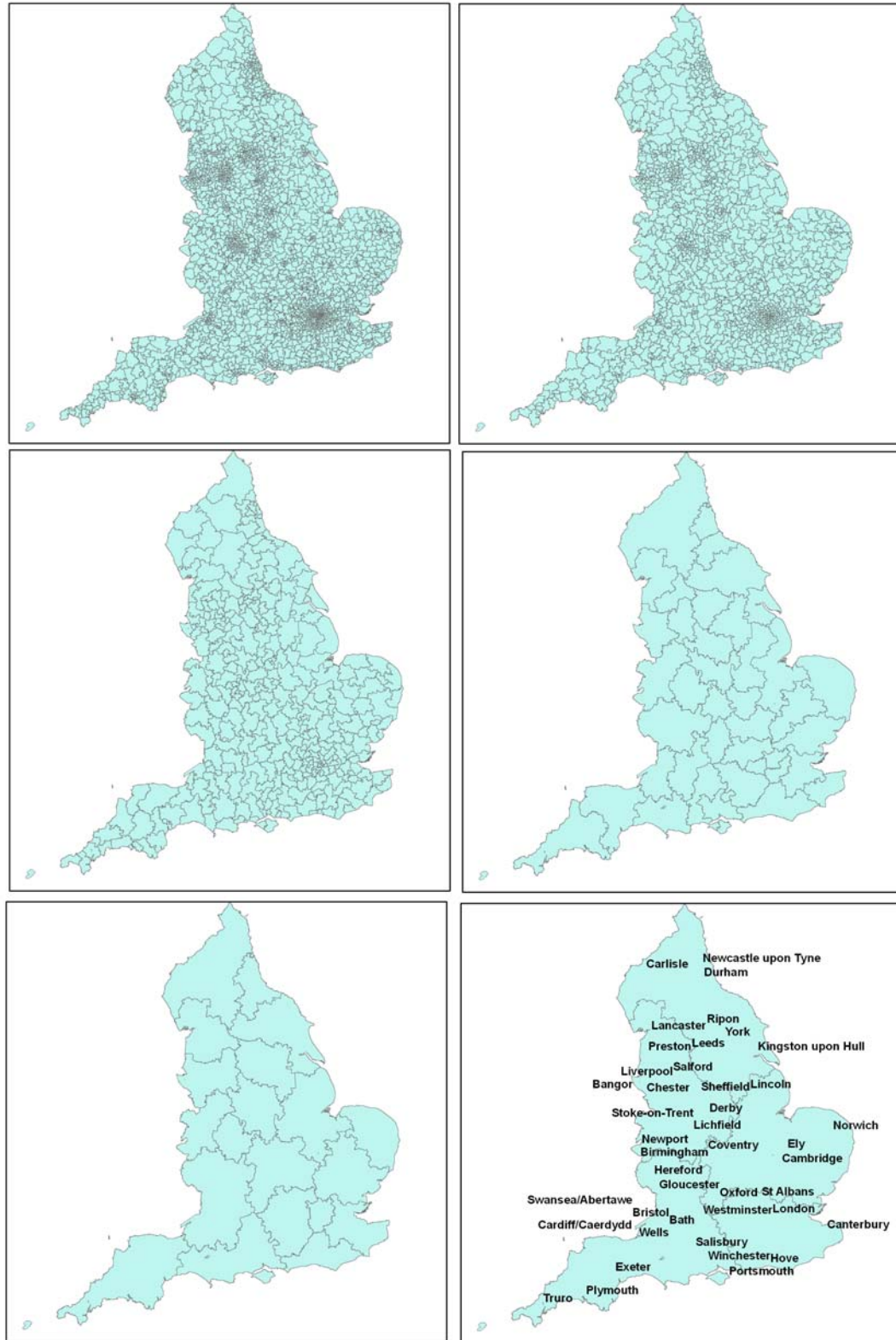


Figure 5. Six levels of aggregation in the hierarchy of zones. In each aggregation step two zones are amalgamated, therefore the number of levels of aggregation is identical to the number of atomic zone (7932).



Figure 6. Six neighbourhoods

4.2 Adaptive zoning

4.1Figure 5 shows a sample of the aggregated zones. The zone sizes are relatively even distributed, although areas with high job densities, i.e. cities and towns have smaller zones than others. The neighbourhoods are presented in 4.1Figure 6. it shows the neighbourhood for six different ST Wards, for reference the wards are named after close-by settlements. All neighbourhoods have the distinctive pattern of small zones close-by and larger zones further away. The gradient from small to large zones is more abrupt than in the synthetic case presented before, reflecting the steep decline in trip numbers with distance. The figure also illustrate that larger clustered zones make part of the neighbourhood of multiple atomic zones. The neighbourhood of Chichester is distinct from the other neighbourhood in having a large clustered zone close-by in its neighbourhood.

4.3 Distribution models

To make the results from the fully detailed distribution model and the adoptive zoning based distribution model comparable, the trips to the clustered zones are disaggregated to atomic zones on the basis of their respective sizes. Therefore three trip matrices are available for comparison, the table with observed trips (from CENSUS), the trip matrix from the fully detailed distribution model and the disaggregated trip matrix from the adapted zoning based model. The stopping criterion for the neighbourhood generation is such that each of the 7932 ST Wards has 333 neighbours, reducing the number of zone pairs by more than 95%.

Mean trip distance was calculated for all matrices (Table 1) and showed that the mean trip distance of the fully detailed model corresponds to the census data (because it was calibrated to that effect). The mean trip distance in the adaptive zoning based model is almost 30% longer.

Table 1. Mean trip distance for models and data

Trip matrix	Mean trip distance
CENSUS	13 km
Fully detailed	13 km
Adaptive zoning	18 km

Further analysis focuses on the distribution of trips over different distances. In order to compare the fully detailed distribution model and the adapted zoning based model the trip-distance distribution was calculated. From these results it appears that the adaptive zoning

model follows the fully detailed model closely, except for the very large distances. Moreover the discrepancy between the models and data is plotted against trip distance as well as the disagreement between the two models. From these results it appears that the difference between the two models is very small in relation to the disagreement with the data (Figure 7)

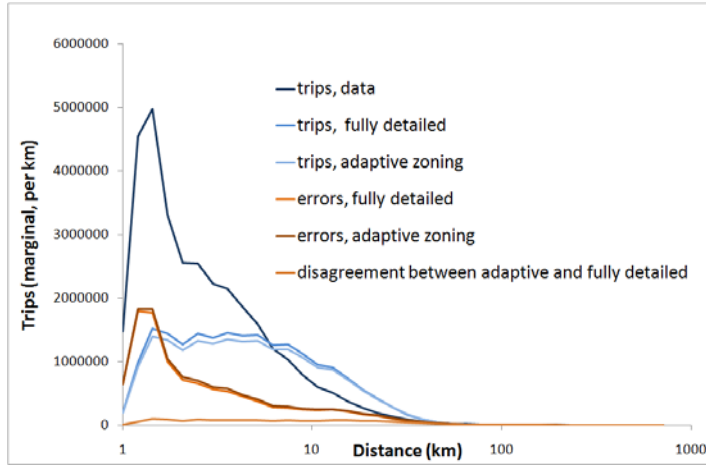


Figure 7. Trips and trip differences by distance

5 Discussion

The close agreement in trip distribution between the fully detailed model and the disaggregated model is promising. The results imply that the number of zone pairs can be drastically reduced without severely affecting the accuracy of the results. It must be noted though that at the further distances the match is not as strong and even though this concerns relatively small numbers of trips, these correspond to considerable distances travelled because of the long trip distances. This explains the strong disagreement over mean trip distance between the fully detailed and the adapted zoning based model, indicating a possible bias. A bias due to aggregation and the location of zone centroids could be expected. We placed the centroids in the geographical centre of zones and did not account for the fact that the trip-weighted centre is rather closer to the origin zone of the trip (as there are more trips at short distances than at longer distances). This bias would be to the contrary of our results, which show too many long distance trips, rather than too few. A better explanation probably is the disaggregation of trips in clustered destination zones to atomic destination zones that does not take into account the direction from which the trips are coming, in other words the far lying atomic zones are weighted equally to close-by atomic zones, leading to an overestimation of the far lying ones.

The layout of the neighbourhood of Chichester points to a weakness in the method. It appears that Chichester is positioned near the edge of two large aggregated clusters. The cluster to the East is disaggregated to a fine level of detail in the neighbourhood generation. The cluster to the West is not,. A probable explanation is that the centroid of the large cluster to the West is relatively far away from Chichester. The estimated error thus is relatively small. In reality Portsmouth, a major employment centre lies within the large aggregated zone close-by to Chichester, but its proximity is effectively not accounted for. It's an indication that the function for estimating error needs to be refined, for instance by including a term for either the variability or centrality of the activity pattern in the zone.

In order for the doubly constrained model to work with the adaptive zoning system it was necessary to modify the model. This is not a trivial task and we identified two places in the model where an aggregation or disaggregation step would suffice. By aggregating the constraints the demands on the model become relaxed, there is no longer a single solution and therefore the optimal solution within a solution space is sought instead. Although the system of equations that results is a standard QP problem, for which libraries are available it is rather more complex than Furness iteration that is used for traditional zoning systems. The alternative is to disaggregate the modelled trip ends in order to meet the constraints at the level of atomic zones. Although much simpler, this method has the disadvantage of rather slow convergence, compared to the model on the basis of the traditional zoning system. In effect, the new model outperforms the existing one in terms of memory use and storage, but not necessarily in terms of processing time. In the current implementation the model on the basis of clustered zoning requires longer computation time by a factor 4, but the data structures for the adaptive zoning based model have not been optimized yet, which means that the prototype implementation suffers the disadvantage of the overhead of clustered zones, without fully benefitting from the advantage of much sparser matrices.

The processes of hierarchy generation and neighbourhood generation are by large driven by a strong logic towards minimizing error. The methods are heuristic in nature however and it might be worthwhile to pursue a system of formal error minimization. Rather than finding a set of neighbourhoods that can reasonably expected to introduce only small aggregation errors it would be grand to have a rigid theoretic model. Such a system would be useful for the task of generating an adaptive zoning scheme, but also for many other model design questions. It may be a Holy Grail and warrant many years of research but the results of this paper do

suggest that estimations of modelling error can be an integration mechanism for complex, multi-criteria model design problems.

The model in this paper is static in nature and used under well-known conditions. In practice it is likely that conditions are now known a-priori. For instance, this paper estimated the β parameter on the basis of the fully detailed model. In a more realistic scenario the fully detailed model would not be available and β could only be estimated by means of the adaptive zoning system. At the same time, generating the adaptive zoning system requires the β parameter for estimating errors. As a solution, the modeller has to either accept a discrepancy between the β parameter underlying the model and the zoning system or come up with an iterative system.

A similar problem occurs in dynamic models, for instance migration and land use models. From one time-step to the next circumstance will be different as will the best matching zoning system. A further development may be towards adaptive *dynamic* zoning systems in which zoning systems are not developed offline in a pre-processing step, but instead as online part of the model; each change of circumstances can be cause for a newly generated adaptive zoning system. We already considered this dynamic application by making the neighbourhood generation a process of disaggregation, rather than aggregation. Distances, error estimates, etc. only need to be calculated down to the granularity required by the model and not at finer levels.

Distribution is only one step in classical transport model. Other steps include trip production and attraction, modal split and assignment. As the constraints for the distribution are given at the level of atomic zones, the process of production and attraction can remain the same.

Modal split can take place on the basis of interaction between atomic zones and their neighbour zones without difficulty. One complication of modelling model split is that the distance, or rather the disutility of travel in terms of generalized cost, between zones differs between modes generating one adaptive zoning system for all modes would be sub-optimal for the individual modes, but a separate adaptive zoning system for each mode would significantly complicate the modal split. The assignment step is the most computing intensive step of most transport models. In this step trips between zones are assigned to individual road segments taking into account the generalized cost of traversing these segments, Most transport models also incorporate the effect of transport intensity on travel speed. In order for the assignment models to benefit from the adaptive zoning system care must be taken to the

direction in which the transport is assigned to the road (i.e. building routes starting from the clustered zone or starting from the atomic zone). The nature of assignment models is such that the cost of building routes from one zone to a number of other zones is primarily a function of the distance to the *furthest* zone (and not as one might assume the number of zones).

Assigning transport starting from the atomic zone and building routes to all the neighbour zones would as a consequence not lead to significant efficiency gains. On the other hand starting at the neighbouring zone and building links to all zones of which it is a neighbour can be expected to make considerable efficiency gains. The crux is that large clustered zones are neighbours of atomic zones at large distances, whereas small clustered zones are only neighbours of atomic zones at small distances. There are many small clustered zones and only a few large clustered zones, so there are many low cost assignments to be made and only a few high cost assignments.

6 Conclusion

This paper has demonstrated how an adaptive zoning system can drastically reduce the number of interacting zone pairs in spatial interaction models while maintaining the original granularity. Thus the incorporation of this system can potentially reduce the running time of existing spatial interaction models, or offer an extended scope for modelling, e.g. a larger geographic area with a more extensive segmentation of spatial activities (such as required for energy demand analysis and accounting in mega-city regions), whilst maintaining the existing running time. Both could open avenues for systematic analysis, including calibration, validation, sensitivity analysis and scenario analysis.

Any advance however is subject to the adjustability of the model at hand. Modification is required, first of all such that it produces valid results using the more complex zoning structure and furthermore such that it actually reaps the potential benefits of the reduced number of zone pairs. The doubly constrained distribution model that served as a test case in this paper made clear that these are non-trivial matters. Essentially the adjusted model has a dual geography of atomic zones at the one hand and clustered zones at the other. Several steps may be identified where it is possible to ‘break into’ the original model and by means of aggregation of atomic zones or disaggregation of clustered zones reconcile the two geographies.

Finally, there is the famous 80-20 rule, which says that 80% of resources are spent on 20% of the problem. Substantial advances in computation time are only achieved when optimizations address the actual calculation bottlenecks. In the case of transport modelling ‘distribution’ - addressed in this paper - is a crucial step, but not the bottleneck. The true bottleneck is ‘assignment’, the process of attributing trips between zones to individual road segments in the transport network. This naturally is where we will devote our attention next.

Besides spatial interaction modelling of various kinds we see scope for several other applications. In particular, reducing the search space for complex optimization problems such as location-allocation modelling (Fotheringham, Densham et al. 1995), aggregating spatial interaction data at optimal granularity while protecting confidentiality (Martin 2000; Martin, Nolan et al. 2001; Young, Martin et al. 2009), as a means of geovisualization when the spatial distribution of zone sizes corresponds to the spatial distribution of the underlying patterns.

Acknowledgements

This research took place as part of the *Energy Efficient Cities initiative*, which is an EPSRC project at the University of Cambridge. Copyright for all CENSUS remains with the Crown. This work benefitted from fruitful discussions with Ian Williams of WSP, Cambridge.

References

- Duque, J. C., R. Ramos, et al. (2007). "Supervised regionalization methods: A survey." International Regional Science Review **30**(3): 195-220.
- Fotheringham, A. S., P. J. Densham, et al. (1995). "The zone definition problem in location-allocation modeling." Geographical Analysis **27**(1): 60-77.
- Gould, N. I. M. and P. L. Toint (2010). A quadratic programming bibliography, RAL Numerical Analysis Group Internal Report. **2000-1**: 140.
- Martin, D. (2000). "Towards the geographies of the 2001 UK Census of Population." Transactions of the Institute of British Geographers **25**(3): 321-332.

Martin, D., A. Nolan, et al. (2001). "The application of zone-design methodology in the 2001 UK Census." Environment and Planning A **33**(11): 1949-1962.

Openshaw, S. (1977). "Optimal zoning systems for spatial interaction models." Environment and Planning A **9**(2): 169-184.

Slater, P. B. (1976). "Hierarchical Regionalization of Japanese Prefectures Using 1972 Inter-Prefectural Migration Flows." Regional Studies **10**(1): 123-132.

Slater, P. B. (1987). "Algorithm 13. Strong component hierarchical-clustering." Environment and Planning A **19**(1): 117-125.

Sneath, P. H. A. and R. R. Sokal (1973). Numerical taxonomy; the principles and practice of numerical classification. San Francisco, W. H. Freeman.

Young, C., D. Martin, et al. (2009). "Geographically intelligent disclosure control for flexible aggregation of census data." International Journal of Geographical Information Science **23**(4): 457-482.

Zhang, X. and Y. Ye (1998). User's guide of COPL_QP. Computational Optimization Program Library: Convex Quadratic Programming. Iowa, Computational Optimization Laboratory, University of Iowa: 8.