The effect of Initial Random Aggregation (IRA) on the analysis and modelling of internal migration

Konstantinos Daras[1], John Stillwell[1] and Martin Bell[2]

[1]School of Geography, University of Leeds, Leeds, LS2 9JT, United Kingdom
Tel. +44 113 34 30120, Fax +44 113 3433308
k.daras@leeds.ac.uk, j.c.h.stillwell@leeds.ac.uk, http://www.geog.leeds.ac.uk/
[2]School of Geography, Planning and Environmental Management, University of Queensland, Brisbane, Qld 4072, Australia
Tel. +61 7 336 57087, Fax +61 7 336 56899
martin.bell@uq.edu.au, http://www.gpem.uq.edu.au/

KEYWORDS: Initial Random Aggregation, Internal Migration, MAUP, Spatial Interaction Model

1. Introduction

This paper presents a comparison of two Initial Random Aggregation (IRA) algorithms implemented as part of an international project to analyse Internal Migration Around the GlobE (IMAGE). An inventory of internal migration data sources in different countries has been developed (Bell *et al.*, 2012) and a repository of internal migration data sets is in the process of being constructed, together with a virtual studio in which to analyse spatial patterns of migration using a range of indicators (following Bell *et al.*, 2002).

It is widely recognised that the intra-zonal and inter-zonal interactions in different countries are influenced by the different zoning systems that are used to collect migration flow data, i.e. the various administrative levels of each country. The IMAGE project is concerned with investigating the MAUP scale and zoning effects generated when calibrating spatial interaction models (SIMs) for different zone systems in the spatial hierarchy. More specifically, in this paper we report the results of calibrating doubly constrained SIMs for migration flows between sets of zones in the UK, starting with a set of Basic Spatial Units (BSUs) that are the local authority districts (LADs) and aggregating these spatial units into a series of aggregated 'regions' using two different IRA algorithms.

The objectives of the paper are as follows: (i) to measure and compare the performance of two different IRA algorithms; (ii) to illustrate the effects of both algorithms when applied to the aggregation of LADs in the UK and their associated migration flows; and (iii) to investigate the MAUP effects on the parameter measuring the frictional influence of distance on migration.

2. Data and Methods

We use a symmetric origin-destination matrix of migration flows for examining the speed and the effects of two IRA algorithms on the SIM parameter and the mean distance of migration. The matrix contains migration flows relating to the 12 month period prior to the 2001 Census between 406 LADs in the UK whose boundaries are those in existence at census date. The idea is that the BSUs are aggregated together in steps that can be defined by the user and that the user can determine the number of different iterations of the model at each step. Thus, for example, the user might choose to aggregate the BSUs in steps of 10 and generate 10 alternative spatial configurations at each step.

In this paper we compare the IRA and the IRA-wave algorithms for aggregating *'M'* BSUs into *'N'* contiguous regions. The original IRA algorithm developed by Openshaw (1977) provides a high degree of randomisation to ensure that the resulting aggregations are different during the iterations. Aggregation only takes place between contiguous zones and the algorithm is implemented following Openshaw's Fortran subroutine although it has been slightly improved by using object-oriented principles. The advantage of this approach is the use of objects instead of matrices which avoids the sustained sequential processes and results in much quicker random aggregation (Daras, 2006).

An alternative aggregation algorithm is the IRA-wave algorithm which is a hybrid version of the original IRA algorithm with strong influences from the mechanics of the breadth-first search (BFS) algorithm. The first step of the algorithm is to select '*N*' BSUs randomly from the initial set and assign each one to an empty *N* region. Using an iterative process until all the BSUs have been allocated to the *N* regions, the algorithm identifies the adjusted areas of each region targeting only the BSUs without an assigned region and adds them to each region respectively. One advantage of the IRA-wave algorithm versus the initial IRA algorithm is the swiftness for producing a large number of initial aggregations. Moreover, the IRA-wave provides well-shaped regions in comparison to the irregular shapes of the IRA algorithm. It is also worth mentioning that the IRA-wave's randomness is limited only at the initial level where the algorithm randomly selects *N* BSUs and assigns one to each region.

In order to investigate the speed at which the alternative algorithms produce solutions, we have experimented by selecting to aggregate a test tessellation containing 449 hexagonal BSUs in steps of 10 with 1,000 aggregation iterations generated from random seeds at each step. Figure 1 shows their respective time greediness (in milliseconds) and the time improvement rate between the IRA and the IRA-wave algorithms. What it is clear in the graph is that both algorithms spent the vast time on aggregations that involve small number of targeted regions. This is expected, as the number of BSUs that the algorithm allocates to each region increases considerably. However, the IRA-wave algorithm performs significantly faster, especially when the number of regions becomes very small or large. At the range of 300-442 regions, the algorithm performs twice the speed of the original IRA algorithm and from the 42 to 2 regions the IRA-wave algorithm executes nine times faster than the alternative IRA algorithm.

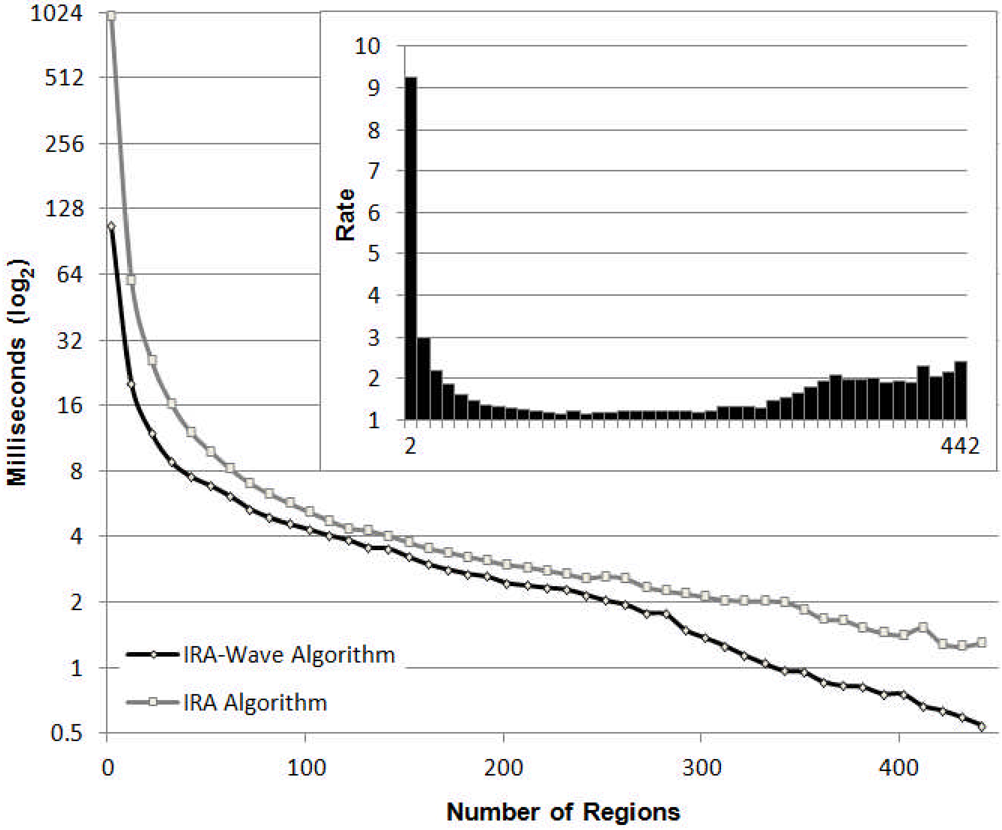


Figure 1. Speed comparison of IRA and IRA-wave algorithms using 2-442 artificial regions

An important indicator in the analysis of internal migration is the frictional effect of distance on flow magnitudes between origin and destination spatial units. Gravity theory in geospatial science (Zipf, 1946) tells us that whilst people move between places in proportion to the masses (or population sizes) of the origin and destination spatial units, migration flows are inversely proportional to the distances between origins and destinations. Thus, more people travel shorter distances than longer distances and the negative relationship between migration and distance is measured through the statistical or mathematical calibration of distance decay parameters in gravity models. When constraints are introduced such that the flows from each origin to all destinations must sum to known out-migrant totals and flows into each destination from all origins must sum to known destination in-migration totals, and the model is calibrated using mathematical rather than statistical calibration methods, the gravity model becomes a doubly constrained spatial interaction model derived by Wilson (1970) from entropy-maximizing principles and can be written as follows:

$$M_{ij} = A_i O_i B_j D_j d_{ij}^{-\beta}$$

where M_{ij} is the migration flow between spatial units i and j, O_i is the total out-migration from spatial unit i and D_j is the total in-migration into each destination spatial unit j, A_i and B_j are the respective balancing factors that ensure the out-migration and in-migration constraints are satisfied, and d_{ij}^{β} is the distance term expressed as a negative function to the power β where β is the distance decay parameter. In Wilson's derivation, the relationship between distance and the interaction variable is represented by an exponential rather than a linear function. The calibration method itself is explained in Stillwell (1990).

3. Modelling experiment using United Kingdom data

This section reports on the modelling experiment using data for a system of 406 LADs in the UK for the census migration flows for 2000-01. In this experiment, we have selected to aggregate the BSUs in steps of 10 (from 402-12) with 1,000 aggregation iterations generated from random seeds at each step using the original IRA and IRA-wave algorithms. No intra-region flows have been included so there is a steady decline in the number of migrants as the number of regions reduces down to 12 regions (Figure 2). The total number of migrants between the BSUs is around 2.5 million declining to around 1.2 million and 1.4 million for the IRA-wave and IRA algorithms respectively. The inter-regional migration rate between the two algorithms illustrates an exponential rise reaching to 1.12 as the number of regions becomes smaller. It is clear that the IRA-wave algorithm constructs regions with fewer inter-regional migrants.

The average values of the model parameters (β) and the mean distance of migration at each step are shown in Figure 3. The horizontal axes of all graphs have units that range from 12 to 402 regions in steps of 10. As the number of regions in the system decreases, there is a very gradual decline in the frictional effect of distance in both algorithms until around 150 regions, after which the β value for the IRA algorithm increases while the value for the IRA-wave algorithm constantly decreases. At the same time, the mean distance of migration increases in both algorithm favouring the IRA-wave when the number of regions are smaller. The increase rate of mean migration distances indicates that there is a linear increase of the rate by 5% between the IRA-Wave and IRA algorithms. The range of β values associated with the 10 iterations at each step is also shown on the graph (dotted lines), indicating that as the number of regions in the system gets smaller, the variation in the parameter value increases around the mean, suggesting much greater instability in the decay parameter when modelling smaller sets of regions.

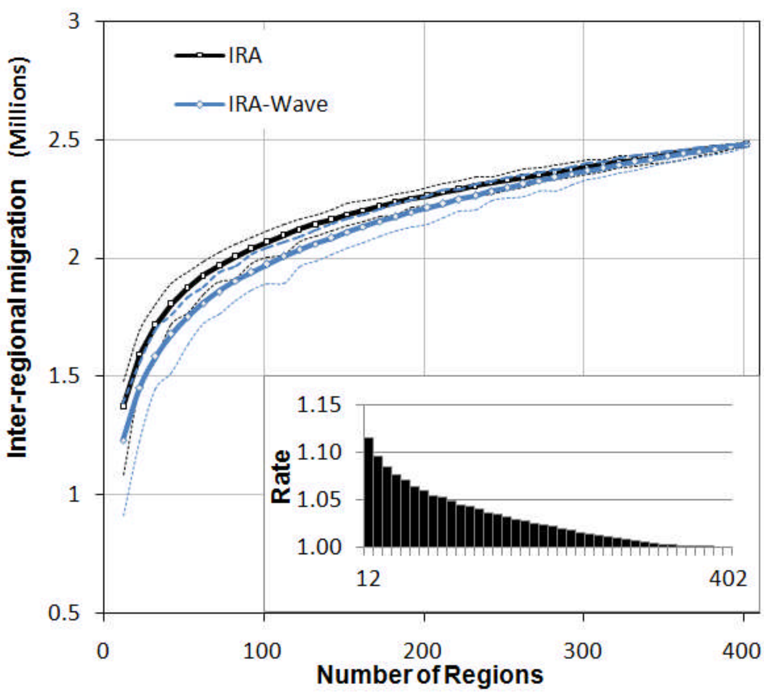
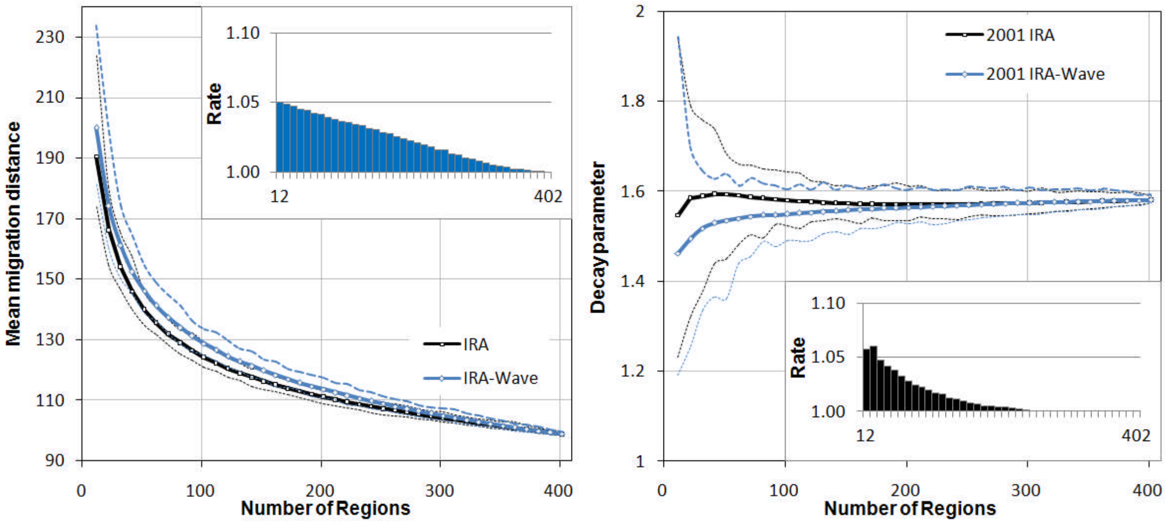


Figure 2. Average inter-regional migration totals and ranges for 12-402 regions based on IRA and IRA-Wave algorithms.



(a) Migration distances **(b)** Distance decay parameters

Figure 3. Average migration distances (and ranges) and average distance decay parameters (and ranges) for 12-402 regions based on IRA and IRA-Wave algorithms

4. Conclusions

This paper has explored the structure and functionality of two Initial Random Aggregation (IRA) algorithms for analysing and modelling internal migration. The results of our experiment using data for the UK exemplify how the IRA-wave algorithm performs in terms of processing performance compared to the original IRA algorithm. Furthermore, the results also illustrate the extent of the MAUP scale and aggregation effects when analysing internal migration in the UK. The results suggest that the scale effect of the friction of distance on migration is relatively small when the spatial system contains over 150 regions but varies more with lower numbers of regions. In contrast to the IRA-Wave algorithm, the IRA algorithm provides considerable higher decay parameters at the lower numbers of region suggesting that the irregularly shaped structures have a strong effect on the β values. Similarly, the aggregation effect is also more apparent when the spatial system contains relatively low numbers of regions, as indicated by the widening of the range around the mean values of β. On the other hand, there is a significant scale effect evident in the mean distance of migration which shows an exponential increase as the number of regions declines, but the aggregation effect is minimal throughout the series of steps. Further investigation using different step sizes and numbers of different aggregations at each step is required, together with testing the algorithms on data sets for different countries and for different demographic or socio-demographic groups.

5. Acknowledgements

This research is funded by Australian Research Council Discovery Project DP11010136 Comparing Internal Migration Around the World (2011-2014). Thanks to Oliver Duke-Williams for sharing his code and his extension work on ASPIC (ARC SPatial Interaction Collection).

References

Bell, M., Blake, M., Boyle, P., Duke-Williams, O, Rees, P., Stillwell, J. and Hugo, G. (2002) Cross-national comparison of internal migration: issues and measures, *Journal of the Royal Statistical Society A*, 165(2): 1-30.

Bell, M., Charles-Edwards, E., Kupiszewski, M., Kupiszewska, D., Stillwell, J. and Zhu, Y. (2012) Internal migration around the world: Towards a global inventory, submitted to *Population Research and Policy Review*.

Daras, K. (2006) An information statistics approach to zone design in the geography of health outcomes and provision, PhD Thesis, University of Newcastle upon Tyne, Newcastle upon Tyne.

Openshaw, S. (1977) Algorithm 3: A procedure to generate pseudo-random aggregations of N spatial units into M spatial units, where M is less than N, *Environment and Planning A*, 9: 1423-1428.

Stillwell, J. (1990) Spatial interaction models and the propensity to migrate over distance, in Stillwell, J. and Congdon, P. (eds.) *Migration Models: Macro and Micro Approaches*, Belhaven Press, London.

Wilson, A.G. (1970) *Entropy in Urban and Regional Modelling*, Pion, London.

Zipf, G.K. (1946) The P1P2/D hypothesis: on intercity movement of persons, *American Sociological Review*, 11: 677-686.

Biographies

Dr. Konstantinos Daras *is a human geographer with expertise in Quantitative Methods and Geographical Information Systems. His research interests include the way in which internal migration varies between countries around the world and the use of advanced techniques (Graph theory Visualisations, Geometric Algorithms and GPU programming) for developing innovative methods of regional aggregation.*

Professor John Stillwell *has research interests that include internal and international population migration, geographical information systems (GIS, and regional development and planning.*

Professor Martin Bell *is a demographer and population geographer with interests in population mobility and internal migration, demographic projections and forecasts, regional and local population change and population models and methods of analysis.*