The IMAGE Studio: a Tool for Comparative Internal Migration Analysis and Modelling

John Stillwell[1], Konstantinos Daras[1], Nik Lomax[1], Martin Bell[2] and Tom Wilson[2]

[1] Centre for Spatial Analysis and Policy, School of Geography, University of Leeds, Leeds, United Kingdom
[2] Queensland Centre for Population Research, School of Geography Planning and Environmental Management, University of Queensland, Queensland, Australia
email corresponding author: j.c.h.stillwell@leeds.ac.uk

Abstract

This paper presents a 'studio' that will facilitate the spatial analysis and modelling of internal migration in any country. The studio is being used in a research project trying to confront the MAUP challenge for comparative analysis of internal migration in different countries by reporting on the development of a tool that generates a series of indicators relating to spatial patterns of migration for a set of Basic Spatial Units (BSUs) and aggregations thereof. More specifically, the paper reports on the framework and components of the studio, its user interface and some initial experiments that explore how the frictional effect of distance on migration changes as BSUs are aggregated into larger regions in a stepwise manner, using data for the United Kingdom.

1. Introduction

Internal migration is an important and ubiquitous global phenomenon. There is on-going discussion about the definition of internal migration *vis à vis* residential mobility with the former generally taking place over longer distances and across administrative boundaries and the latter involving shorter distance movements within administrative areas. The approach of national statistical agencies in the United Kingdom (UK) when conducting censuses is to measure internal migration as anyone moving from one usual residence to another in the 12 months before the census, whatever their motivation or the distance involved in their move. However, when it comes to publishing census results, migration flow data are available for a limited number of different census geographies. In the case of the UK, migration taking place within the smallest spatial units is provided, but this is not the case in many countries.

The comparison of internal migration propensities and geographical flow patterns in different countries is seldom attempted because of the different systems of spatial units that are used by organisation tasked to collect, analyse and disseminate migration data for research or planning purposes. Whilst it is possible to use data on total migration to compute national propensities and age-sex migration schedules for individual countries which can be compared legitimately with other countries (e.g. Rogers and Castro, 1978), any comparison of sub-national movements between (and within) geographical areas is obfuscated by the different shape, size and number of census or administrative spatial units that are used for counting migration flows. This problem is a variant of the Modifiable Area Unit Problem (MAUP) described by Openshaw (1984), whose components include the *scale effect* or the variation in results obtained when data for one set of Basic Spatial Units (BSUs) is aggregated into larger aggregate spatial regions (ASRs, i.e. where the number of regions changes), and the *aggregation effect* or the variation in results obtained from different ways of subdividing geographical space at the same scale (i.e. where the number of regions remains the same but are configured differently).

Thus, the development of the IMAGE studio is partly to accommodate a methodological response to the MAUP challenge for comparative analysis of internal migration by reporting on the development of a tool that generates a series of indicators that relate to spatial patterns of migration patterns for a set of BSUs and aggregations thereof into ASRs. But it is also driven by the need for a computer-based system to facilitate the computation of a range of internal migration indicators and spatial interaction models *per se*. The presentation aims to present a ‘studio’ that will facilitate the analysis and modelling of internal migration in any country, dependent upon the provision of an origin-destination area matrix of flows between BSUs, a vector of area populations and a set of digital boundaries of the corresponding BSUs. More specifically, our aim is to use this tool to explore the sensitivity of the distance decay parameter of a doubly constrained spatial interaction model to changes in geography when we aggregate BSUs into larger regions in a stepwise manner and when we fit the same model to migration flows for different configurations of the same number of aggregated regions.

Thus, the objectives of the chapter are as follows: (i) to briefly distinguish the sources and types of internal migration data that can be used in the system; (ii) to explain the purpose and the general structure of the computational system; (iii) to outline two alternative spatial aggregation routines; (iv) to introduce the migration indicators, (v) to explain the spatial interaction modelling component; and (vi) to use an example of the UK to illustrate model results from the system. The remainder of the paper will consider each of these objectives in separate sections, presenting the results and some discussion of two initial modelling experiments and finishing with a short conclusion.

2. Sources of internal migration data

Internal migration data are collected in countries around the world using various different collection instruments that fall into three main categories: censuses, surveys and administrative sources (or what are often referred to as registers). Some countries collect migration data using more than one type of instrument; in England & Wales, for example, the national statistical agency – the Office for National Statistics (ONS) – retains a migration question in its decadal census but estimates migration by comparing the addresses of National Health Service (NHS) patient registers from one year to the next, and also draws on the Labour Force Survey (LFS) for samples of data on migrants whose behaviour is linked to the labour market.

Moreover, the concept of migration varies considerably between sources in different countries and between censuses across the world depending upon the time period within which the flows are recorded. Thus, we can distinguish lifetime migration (where only birthplace is captured in the census along with place of usual residence at the census) from migration in a prescribed period (place of usual residence 1 or 5 years before the census is recorded) or last migration (place of residence prior to the latest move, regardless of when it took place). The IMAGE inventory of global migration data has been created as part of the *Internal Migration Around the GlobE* (IMAGE) project[1] and a discussion of the methods used to collect internal migration data, the types of data collected, the intervals over

[1] http://www.gpem.uq.edu.au/image

which migration is measured and the spatial frameworks employed to collect internal migration data is found in Bell *et al.* (submitted).

In this paper, we use three sets of migration flows for the UK to illustrate results from the studio. The first is a matrix of the flows between 406 local authority districts (LADs) in the UK for the 12 month period prior to the 2001 Census; the second and third data sets are matrices containing flows for the 12 month period from mid-year 2001-02 and mid-year 2009-10 respectively, which have been extracted from a time series of estimated migration flows. Lomax *et al.* (2012) explain how the time series of matrices of migration flows in the UK has been estimated using data from administrative sources in each of the home countries. There are three national statistical agencies in the UK – one for England & Wales, one for Scotland and one for Northern Ireland – each of which undertakes an independent but partially harmonized census and each of which estimates migration within its respective country for inter-censal years. One consequence of this division of labour is that no single agency compiles a full set of sub-national migration flows between LADs in the UK. Thus, whilst administrative sources provide reasonably reliable data on internal flows between LADs in their respective countries, migration flows between LADs that cross the borders of England & Wales, Scotland and Northern Ireland are missing and need to be estimated from data on 'internal international' flows within the UK in order to generate a full matrix of internal migration in the UK equivalent to that available from the census. The LADs can be regarded as the BSUs that are input to the aggregation and the modelling and analysis system at the outset.

3. System framework

Whilst gathering internal migration datasets for each country all over the world is a difficult and time-consuming process in itself, it is essential to identify and select a methodological approach for analysing the datasets that have been collected in the IMAGE repository. To achieve a robust and flexible environment, the implementation of a unified framework is considered essential. Thus, the IMAGE studio has been designed to be used with data for each country, targeting special data characteristics and providing required tasks of data analysis and normalisation, the latter referring to the efficient organisation of data by eliminating redundancy and ensuring data dependencies. Both goals reduce the amount of space the data consume and ensure that data are stored logically.

The IMAGE studio is organized as a set of linked systems (Figure 1) associated with: (i) data preparation, (ii) spatial aggregation, (iii) internal migration indicators, and (iv) spatial interaction modelling. Each system is autonomous, supporting standardised input and output data in addition to the required tasks.

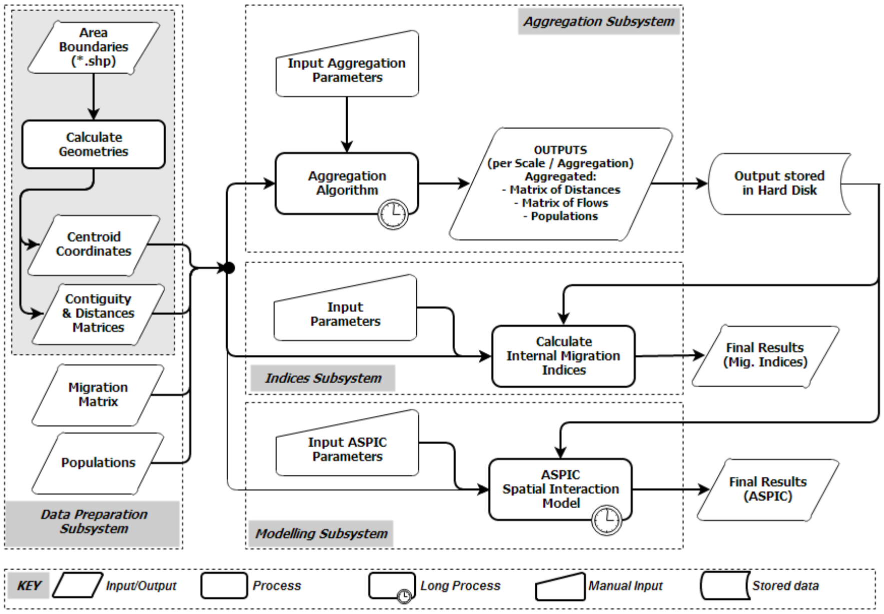


Fig. 1. System diagram of the IMAGE studio

The IMAGE studio is currently designed to prepare, aggregate and analyse data relating to one country at a time. The initial system is responsible for data preparation. It is necessary that the raw data for the country selected, such as the BSU boundaries, the migration matrices and the populations are transformed into normalized data sets for feeding the other two systems. The raw data input to the IMAGE studio includes geographic and tabular data. The geographic boundary data are usually either in the WGS84 projection system (geodetic projection) or in a national projection system (planar projection) of the country concerned whilst the tabular migration data are comma delimited origin-destination matrices or pairs of migration flows and vectors of populations.

In order to use the IMAGE studio for spatial aggregation, the construction of contiguity data deriving from the BSUs is required. The system uses the boundaries of BSUs to identify adjacencies and creates a graph representation of all BSUs, where a node refers to a BSU and an edge refers to the existence of adjacency between two BSUs. This process is performed automatically producing a pairwise output file. However, there are cases such as islands, e.g. Isle of Wight, where the adjacency is not available between the BSUs. These types of problems need to be tackled for a complete graph representation of BSUs by adding (manually) pair entries in the output file.

The second system shown in Figure 1 constructs the spatial aggregations at different scales and with various compositions of BSUs in a stepwise manner. It involves the implementation of an aggregation algorithm that is fed with normalised data from the data preparation system and produces aggregated information such as contiguities, flow matrices and populations for each newly created aggregation. Two aggregation routines are available as indicated in section 5. The third system computes global (systemwide) internal migration indicators for every spatial aggregation and local (area-specific) indicators for the set of BSUs. The indicators include those suggested by Bell *et al.* (2002) as being suitable for comparing migration in different countries and are outlined in more detail in section 6.

Finally, the fourth system calibrates a doubly constrained spatial interaction model (SIM) either for the migration flows for the initial set of BSUs or for the migration flows for each set of ASRs. The system makes use of a modelling package called ASPIC (ARC SPatial Interaction Collection) which has been written in FORTRAN which it provides with a configuration file with all the relevant information about the source of the data files in the hard disk and allows the user to set the required parameters for executing the SIM model. The system uses output data from the spatial aggregation process and for each aggregation produces a document with the results of each SIM analysis as well as averaged model statistics and goodness of fit measures. The SIM is outlined in section 7.

In general, all the spatial operations (such as adjacency and retrieval of polygon centroids) are delivered by making use of the SharpMap and Net Topology Suite (NTS) libraries. The NTS provides a group of methods that deliver topological functionality in geographical data while the SharpMap library handles the user interface. Both libraries are developed according to the simple feature specifications by Open Geospatial Consortium (OGC) and they are open source accessed.

4. Data preparation

Once the IMAGE studio is running the user will observe tabs along the top of the graphical user interface representing each system component. Figure 2 is a screenshot of the data preparation system interface. On the left side of the window, a user can load an ESRI shapefile and immediately on the right side the system draws the geographical boundaries of the shapefile, in this case the 406 LADs that constitute the UK. The studio automatically retrieves the projection system from the loaded geometries, informs the user what it is and subsequently uses it to calculate BSU areas and inter-centroid distances, measures that are crucial for calculating the migration indicators related to the distance and area factors as well as being used by the spatial interaction model to calculate the distance decay parameter.

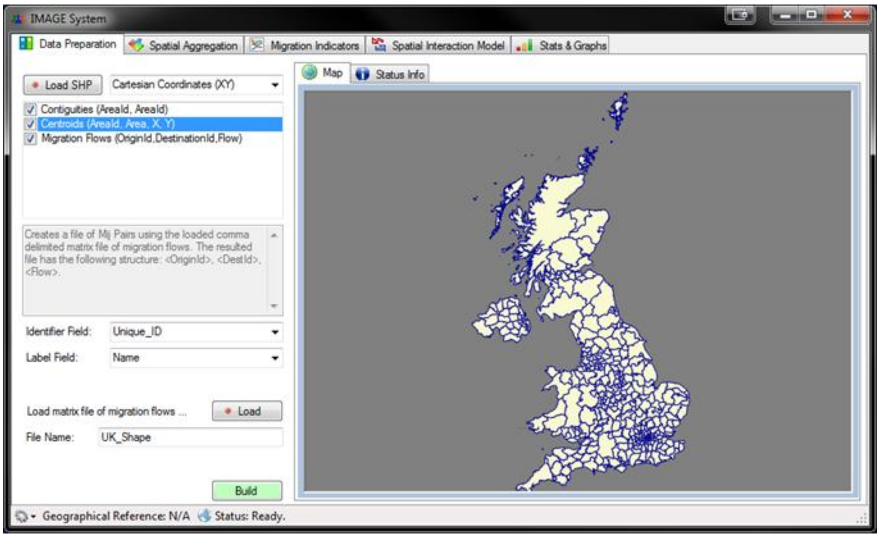


Fig. 2. The IMAGE data preparation interface

When the shapefile is loaded, three data output options are enabled: (i) contiguities, (ii) centroids, and (iii) pairwise migration flows. The contiguity option creates a pairwise file where pairs of BSUs (recorded as comma delimited text) represent the existing adjacencies of boundaries. The centroids option extracts the centroids and area from each BSU while the pairwise flows option converts the comma delimited flow matrix to a pairwise flow file. An important system parameter is the selection of the 'Identifier Field'. This field holds the unique number for each

BSU ensuring the correct association between the BSUs and the migration flows. The three output files are stored for subsequent reuse.

5. Spatial and attribute aggregation methods

One of the most important parts of any combinatorial optimisation method is the initial aggregation of BSUs. The IMAGE system contains two different aggregation algorithms for generating *m* contiguous aggregate statistical regions (ASRs) from *n* BSUs. These two approaches are the Initial Random Aggregation (IRA) and the IRA-wave algorithm. 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. In the IMAGE studio, the algorithm follows Openshaw's Fortran subroutine but it has been implemented with 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 algorithm for aggregating BSUs 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 *m* BSUs randomly and assign each one to an empty ASR. Using an iterative process until all the BSUs have been allocated to the *m* ASRs, the algorithm identifies the adjusted areas of each ASR targeting only the BSUs without an assigned ASR and adds them to each ASR 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 ASRs in comparison to the irregular shapes of the IRA algorithm. It is also important to note that the IRA-wave's randomness is limited only at the initial level where the algorithm randomly selects *m* BSUs and assigns one to each ASR. The IMAGE studio supports both algorithms for experimentation on different degrees of randomness and also allows the user the choice of modelling the initial system of flows or performing either single or multiple aggregations of the BSUs.

Figure 3 shows a screenshot of a multiple aggregation run. On the left side of the interface, the user loads the contiguity file and sets a series of aggregation parameters such as the type of initial random aggregation required (e.g. IRA-wave), the scale step (e.g. 10) and the number of iterations (e.g. 10) the system will execute at each step. The aggregation process always start at a scale of 2 ASRs and according to the scale step introduced by the user, increases in a stepwise manner until the number of ASRs becomes equal to or exceeds the number of BSUs. In addition, the user can change the first and last scales for targeting a specific range of scales. The selected IRA process is repeated for the required number of iterations per scale and the resulting aggregations are written to the storage device. Each scale is represented in the storage device as a directory and within each directory the system stores a series of files (equal to the number of iterations) that record the association of BSUs and ASRs. As shown on the right side of the interface in Figure 3, the system reports the archived progress as well as possible errors that occur and prevent completion.

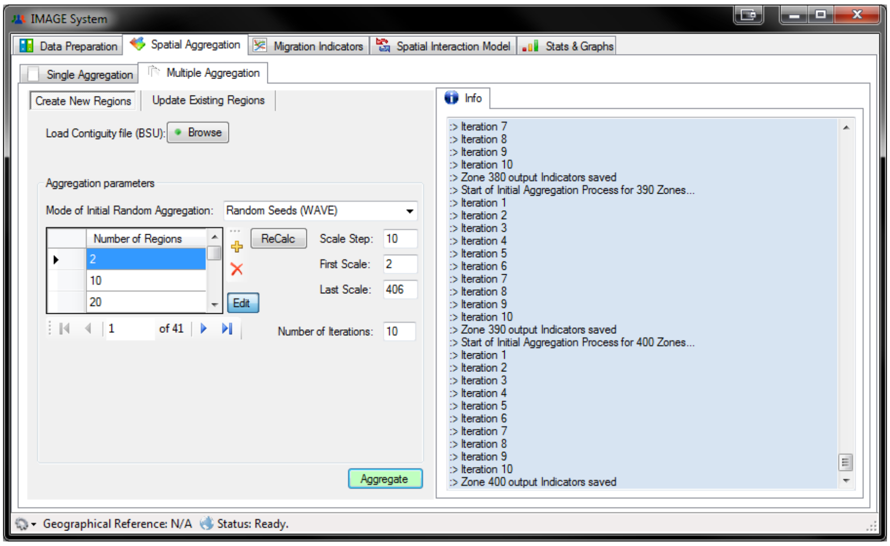


Fig. 3. The IMAGE data aggregation interface: Create new regions

The next step of the spatial aggregation process is to generate aggregated outputs of flows, distances, centroids/areas, and populations at the level of each aggregation by selecting the 'update existing regions' interface (Figure 4). The aggregated outputs are used as input data for the internal migration indicators and spatial interaction model systems.

The aggregated flows between the new ASRs are calculated by summarising the flows from the initial BSUs that constitute an origin ASR to the initial BSUs that comprise a destination ASR and these are calculated for all pairs of ASRs. Moreover, the flows between the BSUs within a new ASR are considered as an intra-region flow and are excluded from the analysis. In the case where the original BSUs include intra-BSU flows, then the system summarises the intra-BSU flows of BSUs occurring in the ASR and, at a second stage, summarises all the flows between the BSUs and within the ASR. The user has the choice to include or exclude intra-BSU and intra-ASR flows.

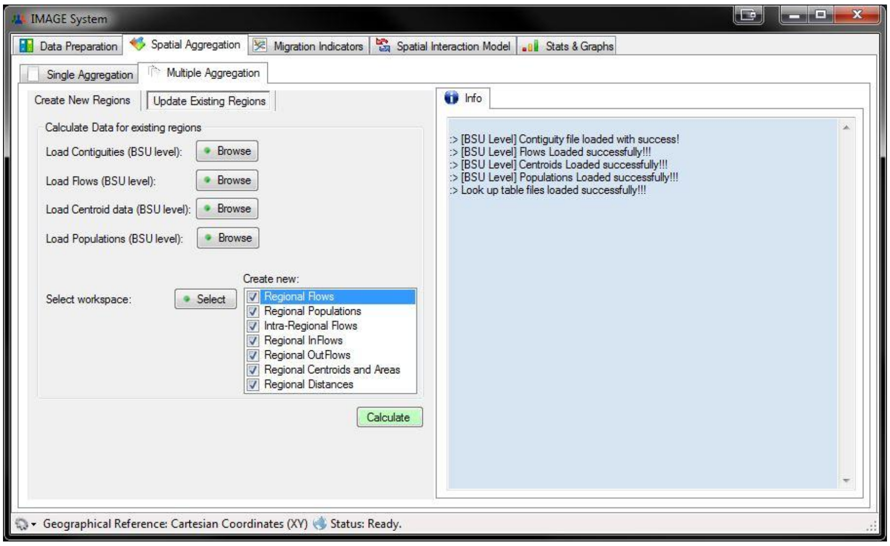


Fig. 4. The IMAGE data aggregation interface: Update existing regions

The distances between BSUs calculated by using the Pythagorean formula for Cartesian systems:

$$d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (1)$$

where *d* is the distance between the two points *i* and *j*, and x_i, x_j, y_i, y_j are the Cartesian coordinates of points *i* and *j* respectively, or by using the Haversine formula for geodetic systems:

$$d_{ij} = 2r\arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_j-\varphi_i}{2}\right) + \cos(\varphi_i)\cos(\varphi_j)\sin^2\left(\frac{\lambda_j-\lambda_i}{2}\right)}\right) \quad (2)$$

where *d* is the distance between the two points *i* and *j*, *r* is the radius of the Earth (treating the Earth as a sphere), φ_i is the latitude of point *i* and φ_j is the latitude of point *j*, and λ_i is the longitude of point *i* and λ_j is the longitude of point *j*.

The distances between ASRs that constitute each new aggregation are estimated on the basis of the initial distances between the BSUs. Each distance between a pair of regions is calculated as the mean of BSU distances between both ASRs. The formula for computing the distance d_{AB} between ASRs *A* and *B* is:

$$d_{AB} = \frac{\sum_{i \in A} \sum_{j \in B} d_{ij}}{mn} \quad (3)$$

where d_{AB} is the distance between the ASR *A* and ASR *B*, *i* is the BSU member of ASR *A*, *j* is the BSU member of ASR *B* and *n*, *m* are the number of BSUs in ASRs *A* and *B* respectively.

6. Internal migration indicators

The third system interface (Figure 5), enables the user to compute a selection of global or local migration or population indicators for either the system of BSUs or each of the systems of ASRs that are generated by the aggregation routine.

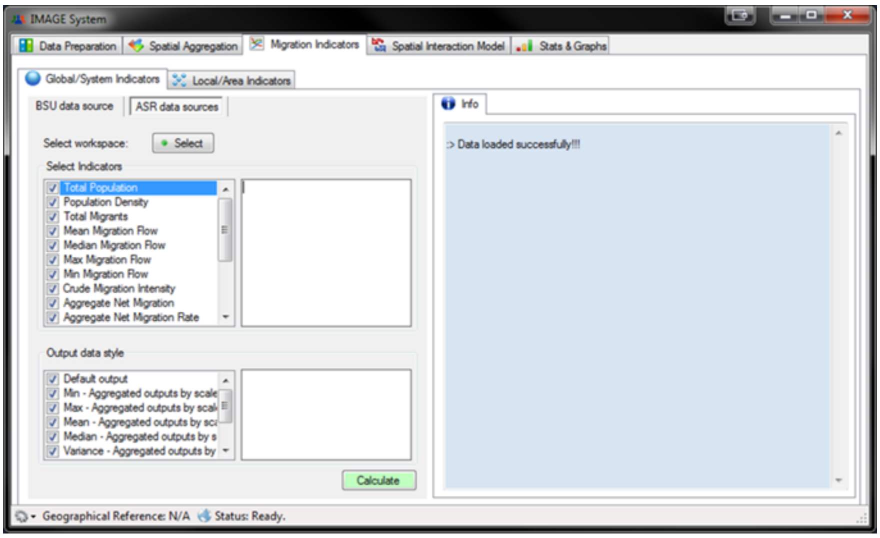


Fig. 5. The IMAGE internal migration indicators interface

The set of global indicators is listed in Table 1, together with the data that are required for their computation. Clearly the system-wide population and population density will remain the same regardless of whether the zone system is the BSUs or any one specific set of ASRs. However, the values of the migration indicators will change from the initial values for the BSUs as each new set of ASRs are generated. If the initial system contained 50 BSUs and the user decided to choose to aggregate in steps of 10 with 100 iterations at each step, then this would produce 500 values of each of the indicators. Indicators 3-7 include basic descriptive counts: total flows and the mean, median, maximum and minimum values in the cells of the matrix. The migration intensity is defined as a rate of migration by dividing the total number of migrants by the total population (at risk). The aggregate net migration is the sum of the absolute values of net migration across each set of spatial units and this is divided by the total migrants to give the aggregate net rate or by twice the total number of migrants to give the migration efficiency or effectiveness. The latter provides an indication of the importance of net migration in redistributing the population, as used by Stillwell *et al.* (2000) when comparing internal migration in Australia and in Britain.

Table 1. The global descriptive information and indicators supported by the IMAGE studio; values for system of UK BSUs, 2000-01

		Required source data			
	Global information or Indicator	**Flow data (Matrix)**	**Population data**	**Centroid/Area data**	**Indicator for UK 406 BSUs, 2000-01**
1	Total population		✓		58,836,694
2	Population density		✓	✓	230.7
3	Total migrants	✓			2,484,029
4	Mean migration flow	✓			24.8
5	Median migration flow	✓			8
6	Max migration flow	✓			4,225
7	Min migration flow	✓			0
8	Crude migration intensity	✓	✓		4.221
9	Aggregate net migration	✓			127,509
10	Aggregate net migration rate	✓	✓		0.217
11	Migration efficiency index	✓			5.133
12	Mean migration distance (km)	✓		✓	98.583
13	Median migration distance(km)	✓		✓	42.764
14	Coefficient of variation	✓			5.056
15	Index of connectivity	✓			0.609
16	Index of inequality	✓			0.456
17	Theil index	✓			2.998

Indicators 12 and 13 quantify how far migrants are travelling using the mean and median statistics respectively and the coefficient of variation provides information about the dispersion of values of migration flows around the mean. The index of connectivity is a simple measure of the proportion of spatial units that are connected by a migration flow involving one or more persons, whereas the global index of migration inequality is a measure of the difference between the observed flows in the migration matrix and the expected distribution that assumes all flows in the matrix are of the same magnitude. The Theil index is a measure of concentration and involves a comparison of each interregional flow (M_{ij}) with every other flow (M_{kl}) in a matrix of inter-regional migration (Plane and Mulligan,1997). Although the values of each indicator are stored in the system for each ASR set, average values iterations at each step will be used for analysis in order to reduce the volume of data.

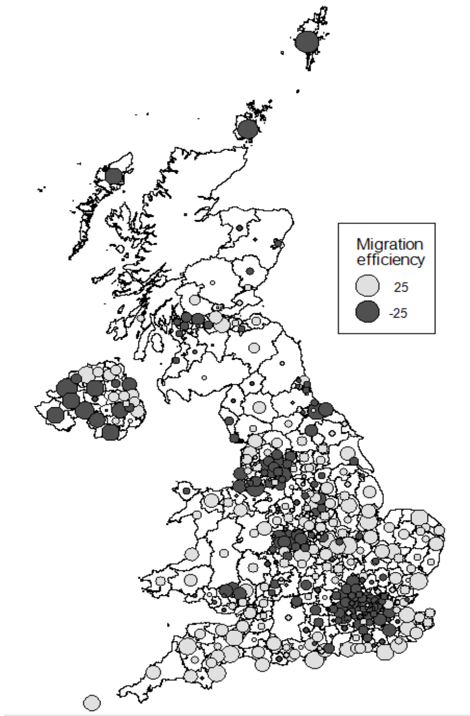


Fig. 6. Migration efficiency, UK districts, 2000-01

A set of local migration indicators are computed for each BSU; it is unlikely that this level of detail will be required for the sets of ASRs. The local indicators include those used for system-wide analysis extended to capture variation in out-migration and in-migration flows and distances, together with turnover (in-migration plus outmigration) plus churn (turnover plus intra-BSU migration). By way of example, Figure 6 illustrates the migration efficiency index for the 406 UK BSUs in 2000-01, indicating the process of metropolitan loss and non-metropolitan gain that has characterised the pattern of migration at this spatial scale for several decades.

Recognising that origin-destination migration flow data are not always available in some countries of the world and the paucity of directional flows disaggregated by demographic variables such as age, sex or ethnicity, the IMAGE studio provides the option for users to select some of the migration indicators using raw BSU inflow and outflow data, the marginal totals of the full migration matrix.

7. Spatial interaction modelling

One of the key indicators in the analysis of internal migration is the frictional effect of space or distance on flow magnitudes between origin and destination spatial units. Gravity theory applied to geospatial science (Zipf, 1946) tells us that whilst people move between places in proportion to the masses of the origin and destination spatial units, migration flows are inversely proportional to the distances between origins and destinations. Thus, following Tobler's 'first law of geography' (Tobler, 1970), more people travel shorter distances than longer distances and the negative relationship between migration and distance is measured through the calibration of distance decay parameters in gravity models where origin and destination mass was measured by population size. When constraints are introduced such that the outmigration flows from each origin to all destinations must sum to known out-migrant totals and in-migration 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 unconstrained gravity model becomes a doubly constrained spatial interaction model derived by Wilson (1970) from entropy-maximizing principles and can be written as follows:

$$M_{ij} = A_i O_i B_j D_j d_{ij}^{-\beta} \tag{4}$$

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. Both options are available in the IMAGE studio and the calibration method itself is explained in more detail in Stillwell (1990). Figure 6 is a screenshot of the spatial interaction model interface which contains windows on the left hand side that allow the user to enter some of the parameters required. An initial decay parameter value of 1 is chosen for the first run of the model and an optimum parameter is found automatically using a Newton Raphson procedure in which an increment value (0.01 in this case) is added to the initial β after the first model run and on alternate model runs. The optimum or best fit value of β is found when the mean migration distance calculated from the matrix of predicted flows is equal (or within close proximity) to the value of the mean migration distance computed from the observed migration flow matrix. Mean migration distance is therefore used as the convergence criterion in the spatial interaction model. The window on the right in Figure 7 illustrates model runs for consecutive sets of data from the spatial aggregation system.

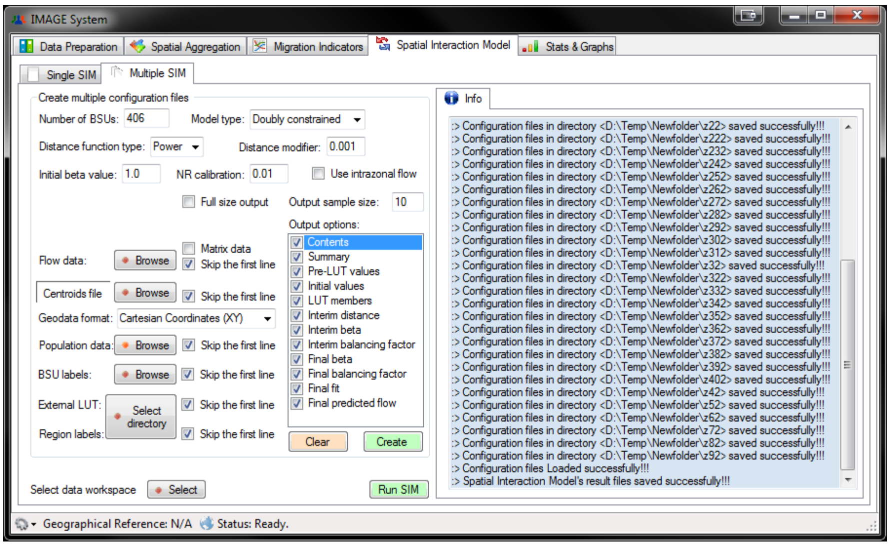


Fig. 7. The IMAGE spatial interaction modelling interface

8. Modelling experiments for the United Kingdom

This section reports on two experiments with IMAGE studio using data for a system of 406 LADs in the UK for the three data sets introduced earlier: the census data for 2000-01 and the data estimated from administrative sources for 2001-02 and 2009-10, each of which has the same set of BSUs. In the first experiment, we have selected to aggregate the BSUs in steps of 10 with 1,000 aggregation iterations generated from random seeds at each step using the IRA-wave option. No intra-BSU flows have been included so there is a steady decline in the number of migrants as the number of ASRs reduces down to 12 regions (Figure 8).

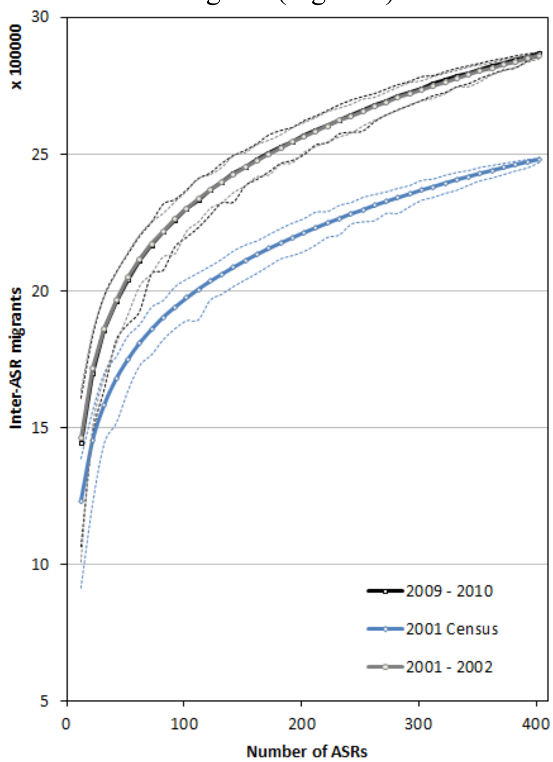
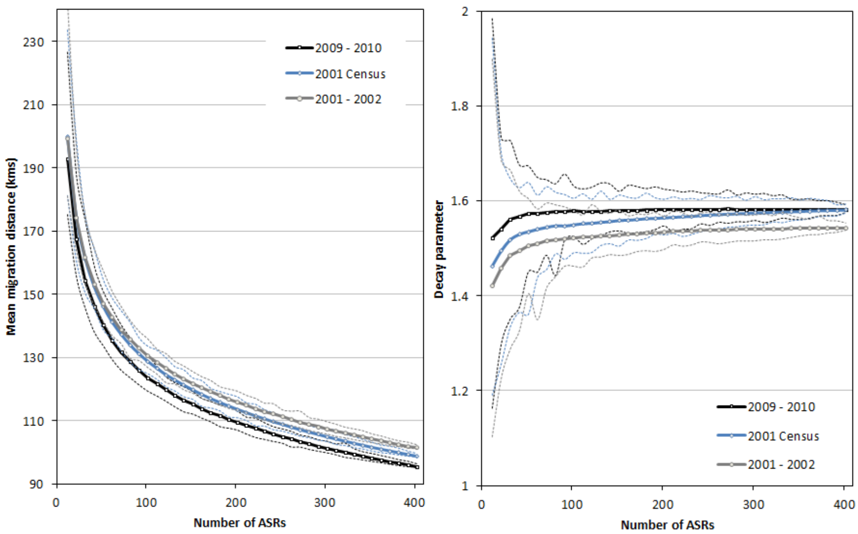


Fig. 8. Inter-ASR migration averages and ranges for 12-402 ASRs in the UK for three periods

It is clear from this graph that the number of migrants between the full set of BSUs that is recorded by the census (2.48 million in 2000-01) is significantly lower that the number of migrants estimated for 2001-02 or 2009-10 (approximately 2.87 million in each case). One of the reasons for this is the undercount in the 2001 Census caused by the number of migrants whose previous address was recorded as unstated. By the time that the BSUs have been aggregated up to 12 ASRs, the number of migrants being modelled has reduced to 1.23 million for the 2000-01 data, to approximately 1.45 million for 2001-02 and 2009-10 data. The dotted lines around each set of mean values are the minimum and maximum values that were generated from the 1,000 iterations at each step.

The mean distances of migration and the mean values of the model decay parameters (β) at each step are shown in Figure 9. The horizontal axes of both graphs have units that range from 12 to 402 ASRs in steps of 10, although models have been calibrated for the full set of 406 BSUs. The mean migration distance for all the BSUs is 98.5kms in 2000-01, 101.4kms in 2001-02 and 95.4kms in 2009-10. The distance decay values are very similar (1.58) for the original system of BSUs for the 2000-01 and 2009-10 data but the 2001-02 migration value is lower (1.54) indicating that distance has a lower frictional effect on migration in 2001-02. Thereafter, as the number of ASRs in the system decreases, there is a very gradual decline in the frictional effect of distance in 2000-01 until around 52 regions, after which the β value declines more rapidly and the frictional effect of distance on migration reduces whilst, at the same time the mean distance of migration increases considerably from 146kms with 52 regions to 200kms with 12 ASRs in 2000-01. Although the total number of migrants is much the same in 2001-02 and 2009-10, the decay parameters suggest that migrants in the most recent period were more influenced by the frictional effect of distance that those in 2001-02 and consequently moved on average over shorter distances.



(a) Mean migration distances (b) Distance decay parameters

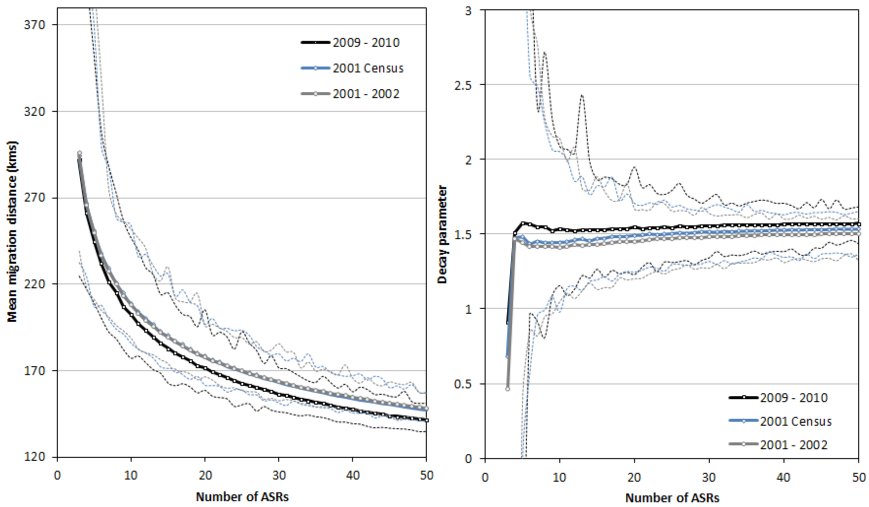
Fig. 9. Mean migration distances and distance decay parameters ranges for 12-402 ASRs in the UK for three periods

The range of β values associated with the 1,000 iterations at each step is also shown on the graph (dotted lines), indicating that as the number of ASRs 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. The range of values around the mean migration distances also increase as the number of ASRs reduces. In general, the decay parameters for all three periods show surprising consistency across the series of aggregations whilst the mean migration distances decline exponentially.

Given the relative stability of the decay parameter when the ASRs numbered above 52 with all three data sets, the second experiment involved using the IMAGE studio to calibrate the model for aggregations between 3 and 50 regions in steps of 1 with 1,000 aggregation iterations at each step. The volume of inter-ASR migration being modelled declines from 1.74 million between 50 ASRs to 0.62 million between 3 ASRs in 2000-01 and from 2.02 million to 0.71 million in 2001-02 and slightly lower figures in 2009-10.

The variation in the means and ranges of the decay parameter values and migration distances are shown in Figure 10. It is clear that whilst the distances increase in a non-linear manner as the number of ASRs get smaller such that migrants between 3 ASRs move, on average, around

293kms, the frictional effect of distance remains relatively consistent at around 1.5 throughout the series, although there is evidence of a linear decline from 50 to 6 ASRs, after which the parameter increases marginally and then drops to a value of 0.68 for migration in 2000-01 between only 3 ASRs and to 0.47 and 0.91 for the other two time periods. However, the range of values in 1,000 iterations of the model at each step is shown to expand significantly as the number of ASRs is reduced as was apparent in the previous experiment.



(a) Mean migration distances (b) Mean distance decay parameters

Fig. 10. Mean migration distances and distance decay parameters and for 3-50 ASRs in the UK for three periods

9. Conclusions

This paper has explained the structure and functionality of the IMAGE studio for analysing and modelling internal migration that incorporates spatial aggregation and interaction flow modelling facilities as well as the computation of migration indicators. In due course, it is envisaged that the studio will be used to facilitate comparative analysis of internal migration in different countries across the world.

The results of our two experiments using data for the UK exemplify how the system can be used with different types of data to examine variations in distance moved and distance decay at different levels of spatial aggregation. They 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 40-50 regions but varies more with lower numbers of regions. 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 ASRs 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 studio on data sets for different countries and for different demographic or socio-demographic groups.

Acknowledgements

This research is funded by Australian Research Council Discovery Project DP11010136 Comparing Internal Migration Around the World (2011-2014). Thanks Oliver Duke-Williams for sharing his code to compute migration indicators (MIGCONC) and for the upgrading and extension work that he has done on the spatial interaction model code (ASPIC).

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. (submitted) Internal migration around the world: Who collects what?, submitted.

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.

Lomax, N., Stillwell, J, Norman, P. and Rees, P. (2011) Data sources and methods of sub-national migration in the United Kingdom, *Working Paper 11/4,* School of Geography, University of Leeds, Leeds.

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.

Openshaw, S. (1984) *The Modifiable Areal Unit Problem*, Concepts and Techniques in Modern Geography, 38, GeoBooks, Norwich.

Plane, D. A. and Mulligan, G. F (1997) Measuring spatial focusing in a migration system, *Demography*, 34: 251-262.

Rogers, A. and Castro, L.J. (1978) Model migration schedules, *RR-81-30*, International Institute for Applied Systems Analysis, Laxenburg, Austria.

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.

Stillwell, J., Bell, M., Blake, M., Duke-Williams, O. and Rees, P. (2000) A comparison of net migration flows and migration effectiveness in Australia and Britain, *Journal of Population Research,* 17(1): 17-41.

Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region, *Economic Geography*, 46(2): 234-240.

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.