

The IMAGE Studio: A Tool for Internal Migration Analysis and Modelling

John Stillwell · Konstantinos Daras · Martin Bell ·
Nik Lomax

Received: 28 October 2013 / Accepted: 13 January 2014 /
Published online: 13 February 2014
© Springer Science+Business Media Dordrecht 2014

Abstract This paper presents a ‘studio’ that facilitates the spatial analysis and modelling of internal migration in any country, dependent upon the provision of an origin–destination matrix of migration flows between a set of basic spatial units plus corresponding populations at risk and digital boundaries. Migration analysts undertaking comparative analysis of internal migration in different countries are confronted with problem that each country has a unique set of sub-national regions. In this paper, we outline the various features of the studio and show how its aggregation and modelling functions have been used to produce substantive results that confront the MAUP challenge. Results show varying systematic trends in the scale and zonation effects of the frictional effect of distance on migration and on the mean distance moved in a selection of northern European countries.

Keywords Internal migration · Area aggregation · MAUP · Migration indicators · Modelling

Introduction

Population redistribution through internal migration is an important and ubiquitous global phenomenon (Todaro 1980; Greenwood 1993; Rees et al. 1996). It is a process that may involve large numbers of people and generate significant changes in the demographic profiles of both origin and receiving regions. In most countries of the world, this movement takes place on a voluntary basis as individuals and families seek new residential locations to suit their specific requirements. In the United Kingdom (UK), for example, over 6.2 million people or 10 % of the population changed their place of usual residence in the 12 months before the 2001 Census (Stillwell et al. 2010).

J. Stillwell (✉) · K. Daras · N. Lomax
Centre for Spatial Analysis and Policy, School of Geography, University of Leeds, Leeds LS2 9JT, UK
e-mail: j.c.h.stillwell@leeds.ac.uk

M. Bell
Queensland Centre for Population Research, School of Geography Planning and Environmental
Management, University of Queensland, Brisbane, QLD 4072, Australia

These movements are usually selective with respect to a range of demographic and socio-economic variables associated with the migrants themselves and characteristics of the origin and destination areas between which they move (Champion et al. 1998). Of course, in some parts of the world, torn apart by war, famine or political uncertainty, forced internal migration takes place creating populations of displaced persons usually requiring humanitarian assistance from external agencies (Hampton 1998; Norwegian Refugee Council 2007), sometimes on a massive scale.

Statistics on voluntary internal migration are collected using different instruments (censuses, surveys and administrative sources) in different countries (Nam et al. 1990) and there are important conceptual, definitional and measurement issues associated with internal migration (Rees 1977) that need to be understood before migration data can be analysed and interpreted. Unlike population stock data, migration data typically involve counts of flows between regions that have been defined for administrative or statistical reasons and supplied in matrix format or as pairwise flows. Whilst migration researchers use these data sets to compute a variety of analytical measures as reviewed by Stillwell et al. (2010), it is apparent that there is a lack of software designed specifically for the computation of the full range of migration indicators (such as rates and efficiencies), indices (such as connectivity or inequality) and distance (such as mean distance travelled or the frictional effect of distance). Migration indicators are valuable tools that research analysts can utilise to assist in understanding migration behaviour but policymakers may also find these measures useful. Those with remits for housing or service provision in local government, for example, and planning practitioners whose focus is on the future demographic development of their towns and cities or rural areas, will derive benefits from measures that can be used to monitor the intensities of out-migration from and in-migration to different localities within their jurisdictions as well as measures of the net impact of migration on the resident population. This paper introduces the IMAGE studio which has been designed for exactly this purpose – to enable a researcher or a policy maker to compute a series of local (regional) and global (aggregate) migration indicators based on a matrix of migration flows for a set of Basic Spatial Units (BSUs), the populations at risk (PAR) for these BSUs and a set of boundaries of the BSUs that correspond to the attribute data.

However, the studio has also been designed to enable the user to explore the effects of the Modifiable Areal Unit problem (MAUP), described in detail by Openshaw (1984), whose components include the *scale effect* or the variation in results obtained when data for one set of BSUs is aggregated into larger aggregate spatial regions (ASRs), and the *zonation effect* or the variation in results obtained from different ways of subdividing geographical space at the same scale. The scale effect is identified by observing the change in an indicator or model parameter when the number of regions changes, whereas the zonation effect is identified by observing the indicator change when the number of regions remains the same but the regions are configured differently. The MAUP is at the core of comparisons of internal migration propensities and geographical flow patterns in different countries because each country has its own hierarchy of spatial units used by governments or agencies 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; Bernard and Bell 2012), 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.

Thus, the IMAGE studio has been developed to accommodate a methodological response to the MAUP challenge for comparative analysis of internal migration; a tool has been developed that generates a series of indicators that relate to spatial patterns of migration patterns for a set of BSUs and aggregations thereof into ASRs. In this paper, our aim is to explain the structure of the studio and demonstrate how it has been used 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. So, the key research question that underlies the analysis is as follows: what happens to the mean distance of migration and distance decay parameter when we aggregate a set of BSUs in steps of x and fit the model to y configurations (zonations) of ASRs at each step (scale)? We identify scale and zonation effects using census and registration migration data for the United Kingdom (UK) and then make comparisons of the same effects in the UK with three other northern European countries.

The paper begins with a brief discussion of the sources and types of internal migration data that can be used in the system before explaining the structural framework of the studio and its four sub-systems. Different sections of the paper outline the data preparation requirements, the two alternative spatial aggregation routines, the sets of migration indicators, and the spatial interaction modelling component. Screenshots of the interface are used to exemplify the functionality of each sub-system. The remainder of the paper reports on how mean distances of migration and distance decay parameters for different data sets in the UK vary by spatial scale and zone configuration and how these indicators compare with similar sets of indicators for Germany, Sweden and Finland. A final section provides some conclusions and suggestions for future development of the studio.

Internal Migration Data: Definitions and Sources

What constitutes internal migration is a matter of some academic debate. There is ongoing 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 (Long 1988). Likewise, there may be instances when flows between countries, considered to be international by some, are regarded as internal by others. One example of the latter is migration within the United Kingdom (UK) which takes place between the four countries of England, Wales, Scotland and Northern Ireland and is captured by independent but harmonised censuses carried out by each of the three national statistics agencies: the Office for National Statistics (ONS) for England and Wales, the National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA). The censuses in each country all 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. The responsibility for providing a UK-wide set of ‘internal’ migration data lies with the ONS who publish origin–destination migration flow matrices as Special Migration Statistics (SMS) at three spatial scales: districts, wards and output areas.

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, ONS retains a migration question in its decadal census but estimates migration on an annual basis between censuses 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 (e.g. place of usual residence one or five 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) project¹ and a discussion of the methods used to collect internal migration data, the types of data collected, the intervals over which migration is measured and the spatial frameworks employed to collect internal migration data is found in Bell et al. (2014). An IMAGE repository has been constructed which contains sets of migration flows and related data collected wherever possible for countries across the world.

In this paper, we use sets of migration flows for the UK, Germany, Sweden and Finland to illustrate results from the studio as described in the next section. Three UK migration matrices are used as indicated in Table 1: 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 periods from mid-year 2001 to mid-year 2002 and mid-year 2009 to mid-year 2010 respectively, which have been extracted from a time series of migration flows for the UK estimated using data from administrative sources in each of the home countries (Lomax et al. 2013). The BSU configuration is exactly the same in each of the UK data sets. Since each of the three national statistical agencies 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.

¹ <http://www.gpem.uq.edu.au/image>

Table 1 Characteristics of selected data sets for the BSUs in four selected countries

Country	Year	Number of BSUs	Area (km ²)				Total Population	Pop. Density	Total Migrants
			Min	Max	Mean	SD			
UK	2001	406	3	26,484	612	1,567	58,836,694	236.7	2,484,029
UK Est	2002	406	3	26,484	612	1,567	59,319,100	238.64	2,863,644
UK Est	2010	406	3	26,484	612	1,567	62,262,500	250.48	2,870,440
Germany	2009	412	27	3,068	866	635	81,902,416	229.52	2,568,234
Sweden	2008	290	7	17,320	1,163	1,643	5,388,351	13.79	281,537
Finland	2011	336	6	20,510	1,507	2,646	9,219,706	21.09	352,913

The migration data that were used for the three other countries are all register-based and refer to annual periods at the end of the first decade of the twenty-first century. The data for Germany are flows between 412 kreise in 2009; the Swedish data are flows between 290 kommun during 2008 and the data for Finland are flows between 336 kunta in 2011.

IMAGE Studio: System Framework

Whilst gathering internal migration datasets for each country across over the world has been 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 process of normalisation introduced here is not related to the statistical normalisation of data values but to efficiently organising 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 subsystems (Fig. 1): (i) the data preparation subsystem, (ii) the spatial aggregation subsystem, (iii) the internal migration indicators subsystem, and (iv) the spatial interaction modelling subsystem. Each subsystem is autonomous, supporting standardised input and output data in addition to the required tasks.

The IMAGE studio is currently designed to prepare, aggregate and analyse data relating to one country at a time. The initial subsystem is responsible for data preparation. It is necessary that the raw data for the country selected, such as the migration matrices, the populations and the BSU boundaries, are transformed into normalized data sets for feeding the other subsystems. The raw data input to the IMAGE studio include geographic and tabular data. The geographic boundary data are usually available either in the WGS84 projection system (geodetic projection) or in a national

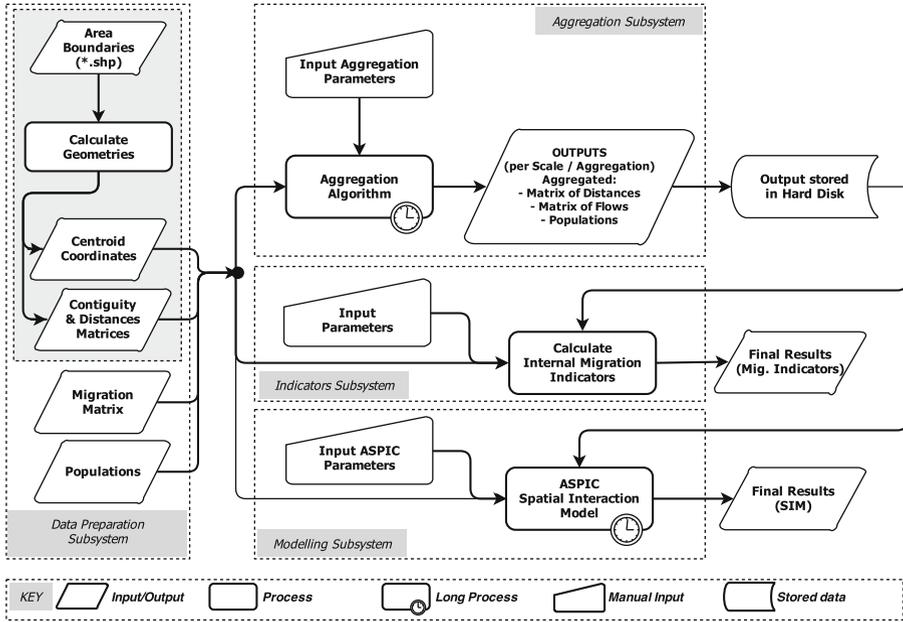


Fig. 1 System diagram of the IMAGE studio

projection system (planar projection) of the country concerned whilst the tabular migration data are comma delimited origin–destination migration matrices or pairwise migration flows and vectors of populations. The normalisation of these data sets is achieved by the system that provides the environment to load, convert and export the data.

In order to use the IMAGE studio for spatial aggregation, the construction of area contiguity data deriving from the BSUs is required. The system uses the boundaries of the 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. This approach is appropriate when the boundaries of BSUs are contiguous with one another. However, there are cases such as islands, e.g. Isle of Wight in England, where the adjacency is not available between the BSUs. This type of problem needs to be tackled for a complete graph representation of BSUs by adding (manually) adjacent pair entries in the output file. We have used ferry routes and nearest neighbours to establish the contiguities between mainland and island BSUs.

The second system shown in Fig. 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 subsystem and produces aggregated information such as contiguities, flow matrices and populations for each newly created aggregation. The third subsystem computes global (systemwide) and local (area-specific) internal migration indicators for every spatial aggregation and also calculates the descriptive statistics for each set of migration indicators with different ASR configurations. The indicators include those suggested by Bell et al. (2002) as being suitable for comparing migration in different

countries. Finally, the fourth subsystem enables the calibration of 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 subsystem makes use of a modelling code called ASPIC (ARC SPatial Interaction Collection) which has been written in FORTRAN (see acknowledgements) 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.

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) libraries². 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. Further details of each subsystem are now provided in the following four sections of the paper together with screenshots of the user interfaces to each subsystem.

Data Preparation

Once the IMAGE studio is running the user will observe tabs along the top of the graphical user interface representing each of the subsystem components. Figure 2 is a screenshot of the data preparation subsystem 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 the area of each BSU and distances between BSUs. These measures are crucial for calculating the migration indicators relating to BSU area and inter-BSU distances as well as being used by the spatial interaction model to calculate the distance decay parameter.

When the shapefile is loaded, three data output options are enabled: (i) the contiguities, (ii) the centroids and (iii) the pairwise migration flows. The contiguity option creates a pairwise file in which pairs of BSUs (recorded as comma delimited text) represent the existing adjacencies of boundaries. The option extracts the geometric centroids and areas from each BSU while the pairwise flows option converts the comma delimited migration flow matrix into a pairwise flow file. An important system parameter is the selection of the 'Identifier Field'. This field holds the unique number for each BSU and, by using this unique number, the correct association between the BSUs and the migration flow is secured. The three output files are vital inputs for the other subsystems of the IMAGE studio and therefore are stored for subsequent reuse.

² <http://sourceforge.net/projects/nts/>
<http://sharpmap.codeplex.com/releases/view/465>

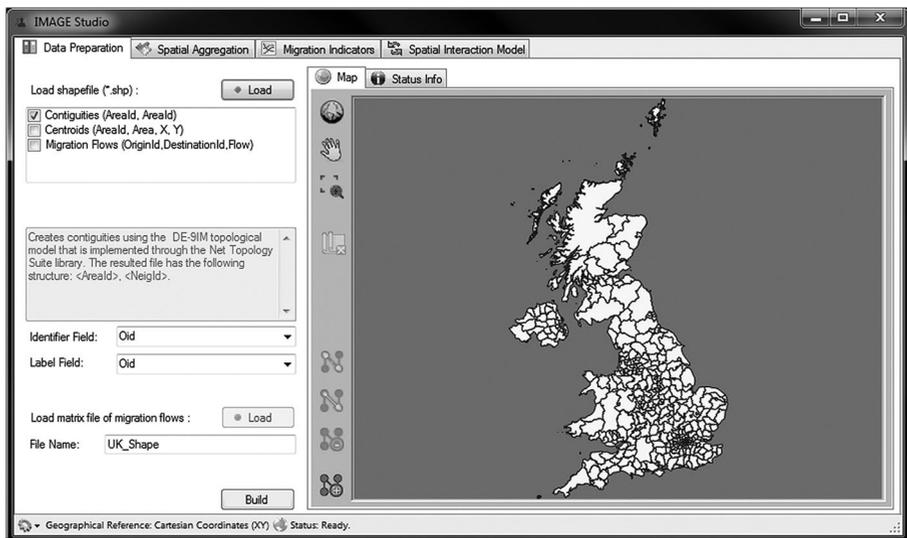


Fig. 2 The data preparation interface after loading the shapefile

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 ' N ' contiguous aggregate statistical regions (ASRs) from ' M ' 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). Detailed explanation of these methods are available elsewhere (Daras 2014).

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 ' N ' BSUs randomly and assign each one to an empty N ASR. Using an iterative process until all the BSUs have been allocated to the N 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. However, there is no objective function involved and therefore the ASRs can be of any size and population. It is also important to note 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 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 a single aggregation or multiple aggregations of the BSUs.

The single aggregation option simply requires the user to select one scale (number of ASRs) and to specify the number of configurations at that scale. 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. 100) that the system will execute at each step. The aggregation process always starts 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 become equal or exceed 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 Fig. 3, the system reports the archived progress as well as possible errors that occur and prevent completion. This process of directory and file creation is rather cumbersome, but it does mean that all the data created are stored and can be accessed so that any configuration of ASRs can be mapped.

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 (Fig. 4). The aggregated outputs are used for the internal migration indicators and spatial interaction model systems as input data. 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. The flows between the BSUs within a new ASR are considered as an intra-region flow and are excluded from the analysis so the volume of inter-ASR migration retained in the system decreases with each scale step as the ASRs reduce in size. A summary of the percentage

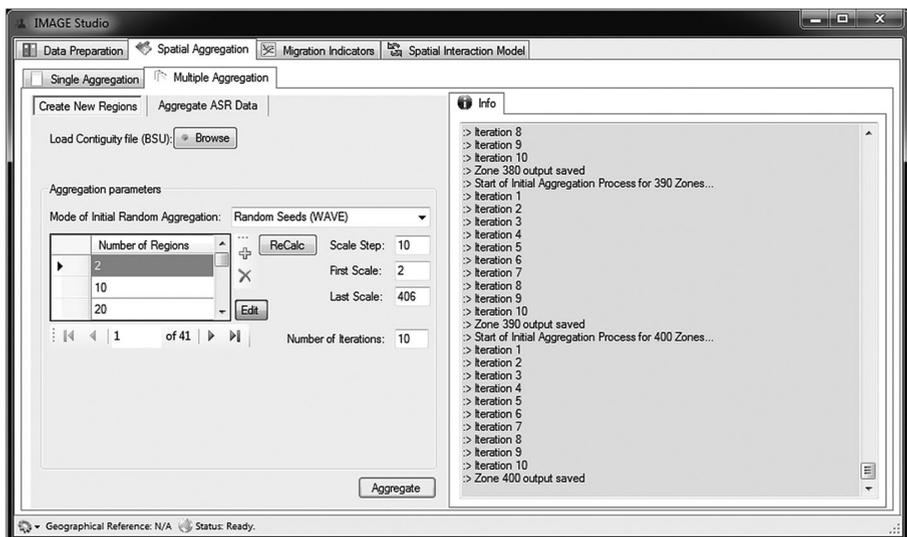


Fig. 3 The spatial aggregation interface: create new regions

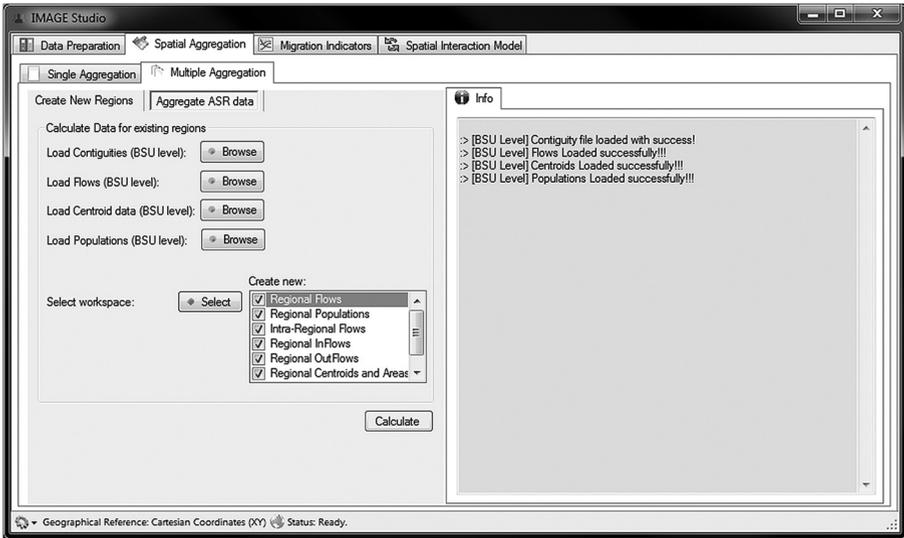


Fig. 4 The spatial aggregation interface: update existing regions

of flows which are internal will be provided. In the case where the original BSUs include intra-BSU flows, then the system summarises the intra-BSU flows for BSUs contained in the ASR and, at a second stage, summarises all the flows between the BSUs within the ASR. The user has the choice to include or exclude intra-BSU and intra-ASR flows.

The distances between BSUs are 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:

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

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.

Internal Migration Indicators

The third subsystem interface (Fig. 5) enables the user to compute a selection of 17 global or 29 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.

The global or system-wide population count and population density indicators 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 is 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. The set of global indicators includes basic descriptive counts: total flows and the mean, median, maximum and minimum values in the cells of the matrix. The global 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.

There are two indicators available to quantify how far migrants are travelling – the mean and median migration distance respectively – and the coefficient of variation provides information about the dispersion of values of migration flows around the mean.

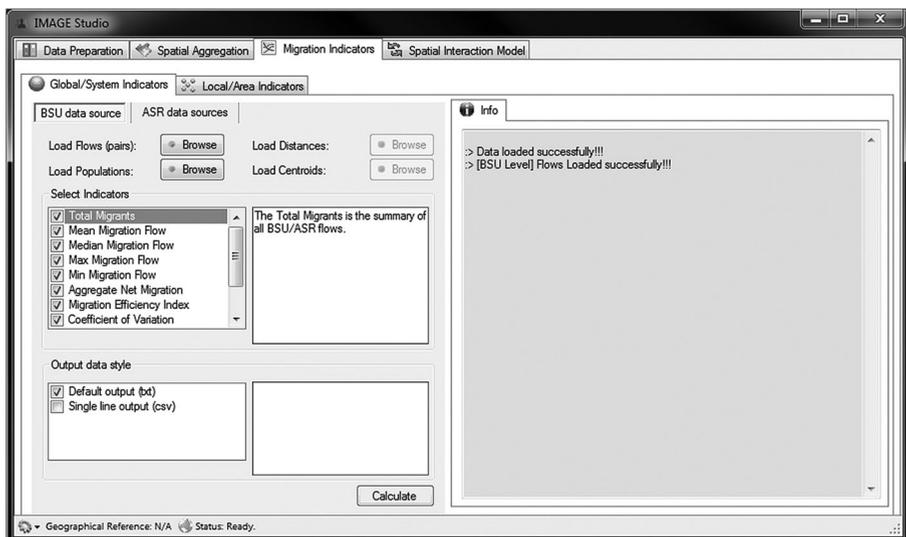


Fig. 5 The global internal migration indicators interface

The global 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. Finally, 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, an average value of all the iterations at each step will be used for analysis in order to reduce the volume of data.

The global indicators reported in Table 2 indicate some of the variation in the population and migration characteristics of the four selected countries. The population sizes range from nearly 82 million in Germany to almost 5.4 million in Finland whilst the population densities are over 200 persons per sq km in the UK and Germany but under 20 persons per sq km in Sweden and Finland. Whilst the total number of migrants also reflect the size of the populations, the migration intensities at the respective spatial scales defined by the BSUs suggests that migration rates are highest in Sweden and lowest in Germany although the global intensities range only from 3.1 to 5.2 per 100 persons.

Whilst the mean migration flow between origin and destination BSUs varies from 20 in Germany to 12 in Sweden, the skewed nature of the distribution of flows (small number of large flows, large number of small flows) means that the median flows are very small in each case. The difference between the mean and median is also reflected in the distances migrated, with median values less than half the mean values. It is not

Table 2 Global migration indicators

Global information		Indicator for			
		UK	Sweden	Finland	Germany
1	Total population	62,262,500	9,219,706	5,388,351	81,902,416
2	Population density	250.48	21.09	13.79	229.52
3	Total migrants	2,870,440	352,913	281,537	2,568,234
4	Mean migration flow	17	12	13	20
5	Median migration flow	1	2	2	4
6	Max migration flow	4,609	2,930	6,503	9,491
7	Min migration flow	0	0	0	0
8	Crude migration intensity	4.61	3.83	5.22	3.14
9	Aggregate net migration	116,759	17,451	10,999	134,424
10	Aggregate net migration rate	0.19	0.19	0.2	0.16
11	Migration efficiency index	4.07	4.94	3.91	5.23
12	Mean migration distance (km)	101.627	140.738	121.616	124.465
13	Median migration distance(km)	47.015	54.775	58.742	54.11
14	Coefficient of variation	4.98	8.96	17.02	6.69
15	Index of connectivity	0.83	0.35	0.19	0.75
16	Index of inequality	0.66	0.46	0.45	0.58
17	Theil index	1.91	3.78	4.42	2.99

surprising, given the size of the country, that people move on average over longest distance in Sweden, although median distance migrated is about the same in Germany as it is in Sweden. The aggregate net migration rate is slightly lower in Germany than in the other three countries but the efficiency of net migration in redistributing the population is slightly higher than the others. The dispersion around the mean migration flow is largest for Finland although the global index of inequality of its flows is the lowest and Finland also has the lowest level of connectivity between its BSUs.

The 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). 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 data on BSU inflows and outflows, the marginal totals of the full migration matrix.

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 masses are measured by origin and destination population size. There is a plethora of research and publications on internal migration modelling as summarized in Stillwell and Congdon (1991) and Stillwell (2008) with many studies using statistical calibration methods to quantify the significance of different explanatory variables on the decision to move and/or on migrant destination choice. A major study linking internal migration with policy variables in England and Wales using Poisson regression (MIGMOD) which was developed for the Office of the Deputy Prime Minister (ODPM 2002) and reported by Rees et al. (2004) and Fotheringham et al. (2004), emphasises the importance of the basic gravity variables. 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, Tobler’s unconstrained gravity model becomes a doubly constrained spatial interaction model (SIM) as derived by Wilson (1970) from entropy-maximizing principles and can be written as follows:

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

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 referred to as 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.

Whilst there is an extensive literature on determinants of migration, synthesised for England and Wales by Champion et al. (1998), the aims and objectives of the IMAGE project do not embrace the collection of explanatory variables for different countries of the world beyond population size and distance. This data collection exercise was considered beyond the scope of the project. As a consequence of assembling matrices of migration flows between BSUs in the IMAGE repository, a doubly constrained model calibration routine has been implemented in the IMAGE studio and both distance function options are available with a generalised decay parameter. The SIM calibration method itself is explained in more detail in Stillwell (1991) and it is intended that other modelling options, including singly constrained models and origin or destination-specific parameter models will be implemented in due course. Figure 6 is a screenshot of the SIM interface which contains windows on the left hand side that allow the user to enter some of the parameters required to run the model. An initial β value of 1 is chosen for the first run of the model with a power function 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 Fig. 6 is where the user observes model runs with sets of data from the spatial aggregation system.

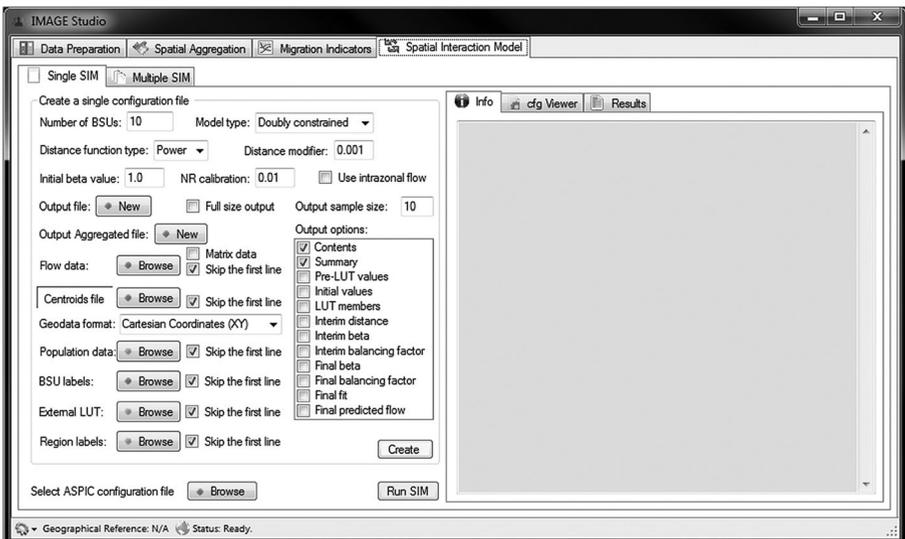


Fig. 6 The spatial interaction modelling interface

Modelling Results

This section reports on two comparative analyses of the scale and zonation effects of model indicators, the mean migration distance and the distance decay parameter. The first comparison is between three different data sets for a system of 406 local authority districts in the UK, each of which has the same set of BSUs. The second is between the 2009–10 data set for the UK and comparable data sets for the three other northern European countries. In the first comparative analysis, we 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. The number of migrants between the full set of BSUs that is recorded by the 2001 Census for 2000–01 (2.48 million) is significantly lower than 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 to 12 ASRs, the number of migrants being modelled has reduced to 1.23 million for the 2000–01 data, and to approximately 1.45 million for 2001–02 and 2009–10 data.

The median values of the 1,000 mean migration distances and model decay parameters (β) at each step are shown in Fig. 7, together with the inter-quartile ranges. Whilst the units on the horizontal axes of both graphs are labelled from 0 to 400, 40 values of the statistics are plotted from 12 to 402 ASRs in steps of 10. The mean migration distance for the original system of 406 BSUs is 99.3kms in 2000–01, 102.0kms in 2001–02 and 96.1kms in 2009–10. The decay parameter values are very similar (1.58) for 406 BSUs for the 2000–01 and 2009–10 periods but the 2001–02 value is lower (1.54) indicating that distance had a lower frictional effect on migration in 2001–02 than at the end of the decade. We observe in Fig. 7 that, 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 decay parameter value declines much

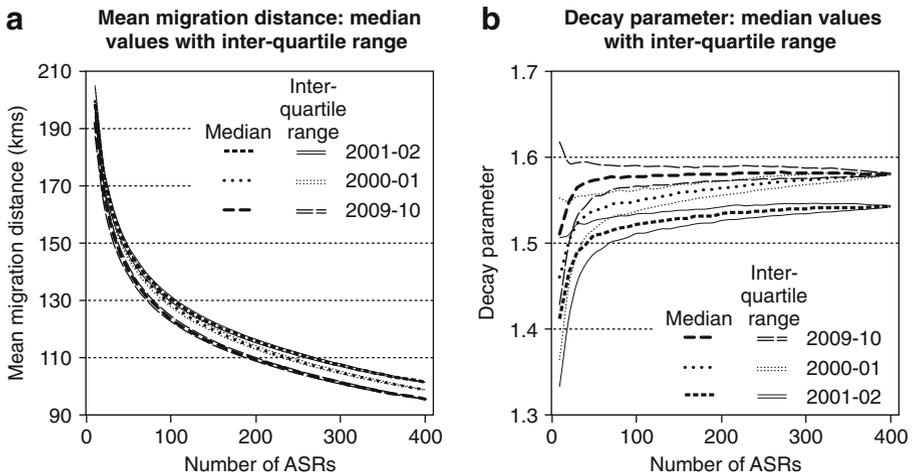


Fig. 7 Mean migration distances and decay parameters for 12–402 ASRs in the UK for three periods

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 than those in 2001–02 and consequently moved on average over shorter distances.

The variation of values around the mean migration distance at each step, as shown by the inter-quartile ranges, are very small and there is no obvious increase in variation as the number of ASRs reduces. It is worth noting that the mean and median values of both the mean migration distances and decay parameters are almost identical, suggesting that there is a normal distribution of values at each step. In general, the decay parameters for all three periods show surprising consistency across the series of aggregations. Variation in the decay parameter values associated with the iterations at each step are also shown in Fig. 7b, indicating that as the number of ASRs in the system gets smaller, the variation in the parameter value increases around the mean, suggesting much instability in the decay parameter when modelling smaller sets of regions.

The second comparison between the four northern European countries involves aggregation in steps of 10 with 100 configurations at each step (scale). The median values of the mean migration distances in each country increase exponentially as the number of ASRs gets smaller (moving from right to left on the x axis of Fig. 8a) with migrants in Sweden and Germany moving furthest on average at each spatial scale and migrants in the UK moving the shortest distances. This is evidence of a clear scale effect in each of the countries with variations between countries likely to depend on size and shape of the ASRs in each case. In terms of the zonation effect, it appears from the graphs showing inter-quartiles ranges that the variations in Finland in particular but also Sweden are larger than in the UK and Germany.

The sets of median distance decay parameters presented in Fig. 8b together with the inter-quartile range values at each spatial scale indicate that the frictional effect of distance is greater in Germany than in the UK but both are relatively insensitive to scale

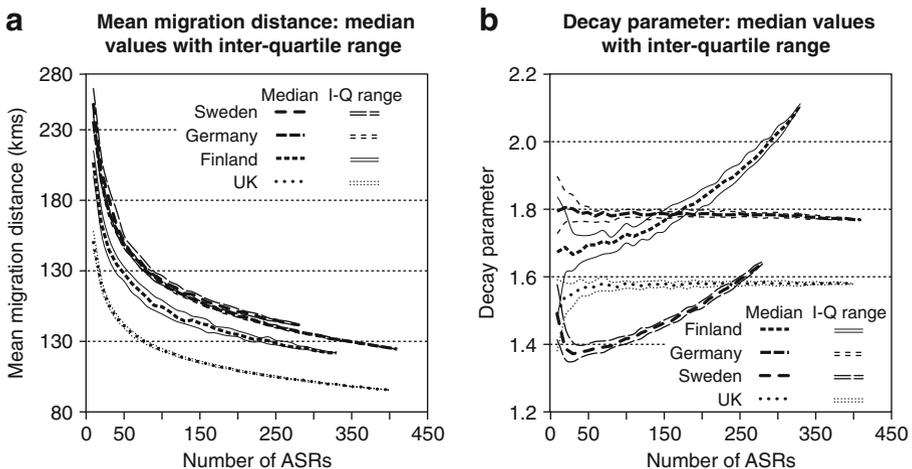


Fig. 8 Mean migration distances and decay parameters for ASRs for four countries

whereas migrants in Finland are more influenced by distance than those in Sweden but both Scandinavian countries experience a scale effect in which those travelling shorter distances between smaller ASRs in both countries are more affected by distance than those travelling longer distances between larger ASRs. Thus, migrants travelling shorter distances are more influenced by distance in Finland than those in Germany but those moving long distances in Finland are much less influenced by distance than those in Germany. Likewise, the frictional effect of distance on shorter distance migrants in Sweden is greater than for migrants travelling shorter distances in the UK. The results suggest a significant scale effect for the decay parameter in the two Scandinavian countries that is not apparent in the UK or Germany until the number of ASRs becomes less than 50. A zonation effect is most apparent in Finland at all spatial scales and increases in all countries as the number of ASRs gets very small.

Conclusions

This paper has explained the purpose, structure and functionality of the IMAGE studio for analysing internal migration. The computation of internal migration measures and the calibration of spatial interaction models represent a valuable toolkit for migration analysts to generate migration indicators that can be used to support policy making, whereas the spatial aggregation routines allow investigation of the scale and zonation effects of the MAUP on migration.

The results of our two selected analyses, using different types of data for the UK and data from four different countries respectively, exemplify how the studio can be used with different types of data to examine variations in distance decay and distance moved at different levels of spatial aggregation in one country or to make international comparisons. The results illustrate the extent of the MAUP scale and zonation effects when analysing internal migration in the UK and when comparing migration in the UK with migration in other northern European countries. In the case of the UK, the results suggest that the scale effect of the friction of distance on migration is very small when the spatial system contains over 50 regions but varies more with lower numbers of regions. Similarly, the zonation effect is also more apparent when the spatial system contains relatively low numbers of regions, as indicated by the widening of the inter-quartile range around the mean values of the decay parameter. 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 zonation effect is minimal throughout the series of steps.

The results of the international comparison suggest that migrants in Germany are more influenced by distance than those in the UK but like the UK, the scale effect on the frictional effect of distance is negligible until the ASRs become relatively large. On the other hand, whilst migrants in Finland are more influenced by distance than those in Sweden, there is a strong scale effect apparent in both the Scandinavian countries.

These findings stimulate the need for further investigation of scale and zonation effects in the internal migration patterns in other countries to ascertain whether there are regularities apparent in countries with different topographical or regional characteristics; this is one of the objectives of the IMAGE project. In addition, the IMAGE studio also provides the opportunity to undertake further experimental work using different

step sizes and numbers of different zonations at each step. Moreover, the studio can be used to compute internal migration indicators and calibrate spatial interaction models for migrant flows in one country disaggregated by demographic or socio-demographic variables such as age, sex, ethnicity, occupation or qualifications.

The studio itself would benefit from further work to develop an optimisation algorithm for producing ASRs with equal populations or to allow the user to choose to build ASRs based on the criterion of equality of any given variable (e.g. households, population density, number of migrations). There would also be value in developing a compactness algorithm using the mean coordinates of the ASR centroids which would be used in conjunction with the algorithm for producing ASRs using equal populated areas. Beyond this, it would be useful to extend the range of spatial interaction models and to automate some graphical facilities for summarizing and visualising the results of the analysis. The latter is important because, when being used in multiple aggregation mode with a large number of BSUs, the studio generates an enormous number of output files which require processing if only to extract the summary statistics.

Thus, in conclusion, it is envisaged that the studio will be used to facilitate comparative analysis of internal migration in different countries across the world and we hope that migration analysts will feel inspired to work with the IMAGE team using the studio with their own data or with data sets held in the IMAGE repository.

Acknowledgements This research is funded by Australian Research Council Discovery Project DP11010136 Comparing Internal Migration Around the World (2011–2014). Thanks to Tom Wilson for sharing his spatial aggregation code (MIRAGE) with us and to Oliver Duke-Williams for providing 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) first written by John Stillwell. We are grateful to Emma Lundholm for providing the Swedish data. Data for Germany were downloaded from the Federal Statistical Office (Destatis) and for Finland from Statistics Finland (StatFin). 2001 Census data for the UK were obtained from the Office for National Statistics, 2011 Census: Aggregate and Interaction data (UKs) [computer files] downloaded via the UK Data Service Census Support.

References

- Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., & 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. (2014) Internal migration around the world: Assessing contemporary practice, *Population, Space and Place* (forthcoming).
- Bernard, A., & Bell, M. (2012). *A comparison of internal migration age profile smoothing methods*, Working Paper 2012/01, Queensland Centre of Population Research. Brisbane: University of Queensland.
- Champion, A., Fotheringham, A. S., Rees, P., Boyle, P., & Stillwell, J. (1998). *The Determinants of Migration Flows in England: A Review of Existing Sources and Data*, Report for the Department of the Environment. Newcastle: Department of Geography, University of Newcastle.
- 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.
- Daras, K. (2014) IMAGE Studio 1.0 User Manual, School of Geography, University of Leeds, Leeds (available from the author).
- Fotheringham, A. S., Rees, P., Champion, T., Kalogirou, S., & Tremayne, A. R. (2004). The development of a migration model for England and Wales: overview and modelling out-migration. *Environment and Planning*, A36(9), 1633–1672.

- Greenwood, M. J. (1993). Internal migration in developed countries. In M. R. Rosenzweig & O. Stark (Eds.), *Population and family economics* (Vol. 1A, pp. 647–720). Amsterdam: Elsevier.
- Hampton, J. (Ed.). (1998). *Internally displaced people a global survey*. London: Earthscan.
- Lomax, N., Norman, P., Rees, P., & Stillwell, J. (2013). Sub-national migration in the United Kingdom: producing a consistent time series using a combination of available data and estimates. *Journal of Population Research*, 30, 265–288.
- Long, L. (1988). *Migration and residential mobility in the United States*. New York: Russell Sage Foundation.
- Nam, C. B., Scrow, W. J., & Sly, D. F. (1990). *International handbook on internal migration*. Westport: Greenwood Press.
- Norwegian Refugee Council. (2007). *Internal displacement global overview of trends and developments in 2006, internal displacement monitoring centre*. Geneva: Norwegian Refugee Council.
- ODPM. (2002). *The Development of a Migration Model, Research Report*. London: Office of the Deputy Prime Minister.
- 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* (Vol. 38). Norwich: GeoBooks.
- Plane, D. A., & Mulligan, G. F. (1997). Measuring spatial focusing in a migration system. *Demography*, 34, 251–262.
- Rees, P. H. (1977). The measurement of migration from census data and other sources" *Environment and Planning A*, 9(3), 247–272.
- Rees, P. H., Fotheringham, A. S., & Champion, A. G. (2004). Modelling migration for policy analysis. In G. Clarke & J. Stillwell (Eds.), *Applied GIS and Spatial Analysis* (pp. 259–296). Chichester: Wiley.
- Rees, P., Stillwell, J., Convey, A., & Kupiszewski, M. (Eds.). (1996). *Population Migration in the European Union*. Chichester: Wiley.
- Rogers, A., & Castro, L. J. (1978). *Model migration schedules, RR-81-30, International Institute for applied systems analysis*. Austria: Laxenburg.
- Stillwell, J., & Congdon, P. (Eds.). (1991). *Migration models: Macro and micro approaches*. London: Belhaven Press.
- Stillwell, J. (1991). Spatial interaction models and the propensity to migrate over distance. In J. Stillwell & P. Congdon (Eds.), *Migration models: Macro and micro approaches*. London: Belhaven Press.
- Stillwell, J. (2008). Inter-regional migration modelling: a review. In J. Poot, B. Waldorf, & L. van Wissen (Eds.), *Migration and human capital* (pp. 29–48). Cheltenham: Edward Elgar.
- Stillwell, J., Bell, M., Blake, M., Duke-Williams, O., & Rees, P. (2000). A comparison of net migration flows and migration effectiveness in Australia and Britain. *Journal of Population Research*, 17(1), 17–41.
- Stillwell, J., Duke-Williams, O., & Dennett, A. (Eds.). (2010). *Technologies for Migration and Commuting Analysis Spatial Interaction Data Applications*. Hershey: IGI Global.
- Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234–240.
- Todaro, M. (1980). Internal migration in developing countries: a survey. In R. A. Easterlin (Ed.), *Population and economic change in developing countries* (pp. 361–402). Chicago: University of Chicago Press.
- Wilson, A. G. (1970). *Entropy in urban and regional modelling*. London: Pion.
- Zipf, G. K. (1946). The P1P2/D hypothesis: On intercity movement of persons. *American Sociological Review*, 11, 677–686.