Introduction to the IMAGE Studio and the example data files

John Stillwell
University of Leeds

Presentation at the Workshop on
‘Analysing Migration Flows Using the IMAGE Studio’,
University of Leeds, 19-20 November, 2015
Aim of the IMAGE Studio

The IMAGE Studio has been developed to assist in the analysis of migration data sets with the following objectives in mind:

1. To address key methodological issues that currently impede international comparisons of internal migration

2. To develop a set of rigorous statistical indicators of internal migration that can be used to make comparisons between countries
MAUP

- Cross-national comparison of sub-national movements between (and within) geographical areas is problematic because of the different shapes and sizes of the spatial units that are used for counting migration flows

- This problem is a variant of the **Modifiable Area Unit Problem (MAUP)**

- Openshaw (1984) identified two MAUP components:
  - the *scale effect* or the variation in results obtained when data for one set of areal units is aggregated into larger spatial units (i.e. where the number of regions changes)
  - the *zonation or 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)
Four key functions/subsystems

- **Data Preparation**
  (Flow matrix, PARs, Shapefiles of Basic Spatial Units (BSUs))

- **Spatial Aggregation**
  (Randomly aggregates BSUs into Aggregate Spatial Regions (ASRs))

- **Migration Indicators**
  (Intensity, impact, connectivity)

- **Spatial Interaction Model**
  (Doubly constrained SIM)
Aggregation of BSUs to ASRs

- We start with an origin-destination matrix of migration flows between **Basic Spatial Units (BSUs)** - typically these refer to an administrative zonation used to collect migration flows data, such as local authority districts or municipalities.

- IMAGE studio contains algorithms that enable BSUs to be aggregated into larger regions that we call **Aggregated Spatial Regions (ASRs)**.

- The user is asked to choose between *single* and *multiple* aggregation, where the former involves the specification of single level of aggregation, whilst the latter provides progressively greater levels of aggregation with correspondingly fewer ASRs.

- At each level of aggregation (scale), the user can choose a number of configurations of BSUs.
Example aggregation for the UK

Start: n = 420 BSUs
M = 2.5 million
CMI = 4.2%

End: n = 10 ASRs
M = 1.1 million
CMI = 1.9%
Aggregation is part of series

Total number of migrants (M)
Total number of people (P)

Zero inter-ASR migrants

Scale (Number of ASRs)
**Link with Courgeau’s k**

Y axis: Inter-ASR migration volume transformed into the Crude Migration rate at scale n (CMI<sub>n</sub>)

X axis: Number of ASRs transformed into the Log of the mean ASR population at scale n (ln(P/n))

Courgeau’s K

\[ CMI_n = 12.08 - 0.66 \cdot \ln(P/n) \]
Migration indicators and SIM

- Algorithms are available in the Studio for the:
  (i) calculation of a range of global (systemwide) and local (zone-specific) indicators of migration intensity, connectivity, distance and impact
  (ii) calibration a doubly constrained spatial interaction model (SIM) that generates a distance decay parameter that quantifies the frictional effect of distance on migration
- The Studio is an environment for investigating what changes are apparent in the migration indicators/parameters at different spatial scales and what variation they show between different ASR configurations at any one scale
## Global and local Indicators

<table>
<thead>
<tr>
<th>Global information or Indicator</th>
<th>Local Information or Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Total population</td>
<td>1 Population</td>
</tr>
<tr>
<td>2 Area</td>
<td>2 Population density</td>
</tr>
<tr>
<td>3 Population density</td>
<td>3 Area</td>
</tr>
<tr>
<td>4 Total migrants</td>
<td>4 Intraregional flow</td>
</tr>
<tr>
<td>5 Mean migration flow</td>
<td>5 Intraregional rate</td>
</tr>
<tr>
<td>6 Median migration flow</td>
<td>6 Mean migration inflow</td>
</tr>
<tr>
<td>7 Max migration flow</td>
<td>7 Median migration inflow</td>
</tr>
<tr>
<td>8 Min migration flow</td>
<td>8 Max migration inflow</td>
</tr>
<tr>
<td>9 Crude migration intensity</td>
<td>9 Mean migration outflow</td>
</tr>
<tr>
<td>10 Aggregate net migration</td>
<td>10 Median migration outflow</td>
</tr>
<tr>
<td>11 Aggregate net migration rate</td>
<td>11 Max migration outflow</td>
</tr>
<tr>
<td>12 Migration effectiveness index</td>
<td>12 Net migration balance</td>
</tr>
<tr>
<td>13 Mean migration distance (between)</td>
<td>13 Net migration rate</td>
</tr>
<tr>
<td>14 Mean migration distance (within)</td>
<td>14 Turnover</td>
</tr>
<tr>
<td>15 Mean migration distance (All)</td>
<td>15 Turnover rate</td>
</tr>
<tr>
<td>16 Median migration distance (between)</td>
<td>16 Churn</td>
</tr>
<tr>
<td>17 Median migration distance (within)</td>
<td>17 Churn rate</td>
</tr>
</tbody>
</table>
| 18 Median migration distance (All) | 18 Migration effectiveness index |}

## Global Information or Indicator

<table>
<thead>
<tr>
<th>Local Information or Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Population</td>
</tr>
<tr>
<td>2 Population density</td>
</tr>
<tr>
<td>3 Area</td>
</tr>
<tr>
<td>4 Intraregional flow</td>
</tr>
<tr>
<td>5 Intraregional rate</td>
</tr>
<tr>
<td>6 Mean migration inflow</td>
</tr>
<tr>
<td>7 Median migration inflow</td>
</tr>
<tr>
<td>8 Max migration inflow</td>
</tr>
<tr>
<td>9 Mean migration outflow</td>
</tr>
<tr>
<td>10 Median migration outflow</td>
</tr>
<tr>
<td>11 Max migration outflow</td>
</tr>
<tr>
<td>12 Net migration balance</td>
</tr>
<tr>
<td>13 Net migration rate</td>
</tr>
<tr>
<td>14 Turnover</td>
</tr>
<tr>
<td>15 Turnover rate</td>
</tr>
<tr>
<td>16 Churn</td>
</tr>
<tr>
<td>17 Churn rate</td>
</tr>
<tr>
<td>18 Migration effectiveness index</td>
</tr>
<tr>
<td>19 Coefficient of variation</td>
</tr>
<tr>
<td>20 Index of migration inequality</td>
</tr>
<tr>
<td>21 Index of connectivity</td>
</tr>
<tr>
<td>22 Inflows</td>
</tr>
<tr>
<td>23 Inflow rates</td>
</tr>
<tr>
<td>24 Inflow mean migration distance</td>
</tr>
<tr>
<td>25 Inflow median migration distance</td>
</tr>
<tr>
<td>26 Outflows</td>
</tr>
<tr>
<td>27 Outflow rates</td>
</tr>
<tr>
<td>28 Outflow mean migration distance</td>
</tr>
<tr>
<td>29 Outflow median migration distance</td>
</tr>
</tbody>
</table>
Spatial Interaction Model

- Conventional modelling would typically involve calibrating a model for a selected set of Basic Spatial Units, e.g. fitting a doubly constrained SIM:

\[ M_{ij} = A_i O_i B_j D_j d_{ij}^{-\theta} \]

where \( O_i \) = the out-migration from zone \( i \) to all other zones
\( D_j \) = the in-migration to zone \( j \) from all other zones
\( A_i \) and \( B_j \) = balancing factors that ensure the constraints are satisfied
\( d_{ij}^{-\theta} = \) negative power distance decay function with parameter \( \theta \)

- Model uses an iterative search routine (Newton Raphson) to find the optimum \( \theta \) parameter when the mean distance migrated calculated from the modelled matrix is equal (or very close) to the mean migration distance from the observed matrix of flows.
What initial data are required?

For any country of interest:

I. an origin-destination matrix of flows between a set of BSUs
II. digital boundaries of the corresponding BSUs
III. populations at risk (PAR) of the respective BSUs
United Kingdom data files

- **Migration matrix** is from 2001 Census Special Migration Statistics Table MM01CUK - one-year transition data for all persons aged 1 and over between 404 local authority districts (LADs)
  - extracted via UK Data Service **WICID** interface
- **BSU boundaries** are for corresponding LAD polygons
  - Extracted via UK Data Service **Easy Download** interface
- **PAR** are 2011 Census (end-of-period) populations for 404 LADs
  - extracted via the UK Data Service **InFuse** Interface
Data collected by National Statistical Agencies

Northern Ireland Statistics and Research Agency (NISRA)

Office for National Statistics (ONS)

National Records of Scotland (NRS)
UK Data Service Census Support

Census Support

Who we are

We provide access to and user support for 1971, 1981, 1991, 2001 and 2011 UK census data.
Read more

LATEST NEWS

UK DATA SERVICE CENSUS SUPPORT

QUICK ACCESS TO

- UK Data Service
- Aggregate data
- Flow data
- Microdata
- Boundary data
Access tools: WICID, EasyDownload and InFuse

Census aggregate data

Aggregate data provide area counts, usually of individuals or households with particular characteristics. An example might be the number of males, aged 20-24, who are divorced and live in Devon.

We hold aggregate data from censuses in England, Wales and Scotland held in 1971, 1981, 1991, 2001 and 2011, as well as data for Northern Ireland from censuses in 1991, 2001 and 2011. These data are aggregated to a range of geographical levels (e.g. counties and districts) from output areas which are the smallest spatial units.

2011 Census data

InFuse now contains 2011 Census data. 2011 data are being released in phases. Find out more about the availability and plans for 2011 Census aggregate data.

- Guide to census aggregate data

ACCESS AGGREGATE CENSUS DATA

InFuse – access 2011 and 2001 Census aggregate data by selecting topics, categories within topics, and geographic areas.

Anyone can access InFuse.
Subsystems of the IMAGE Studio

Data Preparation Subsystem
- Area Boundaries (* .shp)
- Calculate Geometries
- Centroid Coordinates
- Contiguity & Distances Matrices
- Migration Matrix
- Populations

Indices Subsystem
- Input Parameters
- Calculate Internal Migration Indices
- Input ASPIC Parameters
- ASPIC Spatial Interaction Model

Modelling Subsystem
- Aggregation Subsystem
  - Input Aggregation Parameters
  - Aggregation Algorithm
  - Outputs (per Scale / Aggregation)
    - Aggregated:
      - Matrix of Distances
      - Matrix of Flows
      - Populations
  - Output stored in Hard Disk

KEY
- Input/Output
- Process
- Long Process
- Manual Input
- Stored data
IMAGE Interface

Tabs that represent the four different subsystems

Interface for loading the input data and setting the required configurations of each subsystem

Window used for presenting the results of analysis, error messages from the system and detailed status information

Popup button showing system

General status information about the
IMAGE Studio – Subsystem Interfaces

Data Preparation

Spatial Aggregation

Internal Migration Indicators

Spatial Interaction Modelling
Data Preparation Subsystem

**INPUTS**
- Polygon boundaries (*.shp file)
- Migration flows matrix (*.csv file)

**TASKS (C#)**
- Create centroid coordinates (x,y or lat,lon)
- Contiguities (matrix to pairs)
- Convert migration flows (matrix to pairs)

**OUTPUTS**
- Centroid coordinates (*.cen file)
- Contiguities (*.con file)
- Labels (*.lbl file)
- Migration flows (*.flw file)
Contiguities between all BSUs are required

User can check display and check contiguities
Aggregation Subsystem

**INPUTS**
- Output files from Data Preparation Subsystem

**TASKS (C#)**
- LOOP for given Scale Step
- LOOP for given Iterations
- Aggregate areas to larger regions (IRA)
- Calculate the data for the new regions

**OUTPUTS**
- Contiguities for each Aggregation
- Centroids/Area for each Aggregation
- Distances for each Aggregation
- Migration Flows (intra/in/out flows) for each Aggregation
- Populations for each Aggregation
Aggregation routines

• User can choose between two Initial Random Aggregation (IRA) algorithms:

• Original IRA algorithm developed by Openshaw provides a high degree of randomisation to ensure that the resulting aggregations are different during the iterations

• Second aggregation algorithm (IRA-wave) is a hybrid version of the original IRA algorithm with strong influences from the mechanics of the breadth-first search (BFS) algorithm
IRA wave algorithm

1) Select 2 random seeds
2) Select all neighbouring areas
3) Assign the selected areas to region

Final Aggregation to Aggregate Spatial Regions (2 ASRs)
Optimising routines

• Aggregation subsystem of the Studio also has facility to optimise the ASR configuration at each spatial scale using:
  - equal or similar ASR populations
  - equal or similar ASR area sizes
  - equal or similar ASR population densities
  - equal or similar ASR intra-area flows

• Difference between equality and similarity objective functions is that the former tries to minimise the variance between ASRs whilst the latter minimises the variance within the ASRs
Migration Indicators Subsystem

INPUTS
- Centroids/areas for each aggregation
- Migration flows for each aggregation
- Populations for each aggregation

TASKS (C#)
- LOOP for each aggregation
- Calculate internal migration indicators
- Write the results into file (*.txt)

OUTPUTS
- Internal migration indicators stored in a delimited file format for each aggregation
Modelling Subsystem

**INPUTS**
- Contiguities for each aggregation
- Centroids for each aggregation
- Migration flows for each aggregation
- Populations for each aggregation
- ASPIC configuration file for each aggregation

**TASKS (Fortran)**
- LOOP for each aggregation
- Doubly constrained SIM
- Write the SIM results into file (*.txt)

**OUTPUTS**
- SIM results stored in a simple text file format for each aggregation (summary comma delimited file also provided).
See results on screen in text editor

## Summary

The doubly constrained interaction model: \( m(i,j) = a(i) \cdot b(j) \cdot o(i) \cdot d(j) \cdot f(d(i,j)) \)

* Negative power distance decay model
* Initial beta value: 1.000
* Convergence criteria: 0.010
* Newton-Raphson increment value: 0.0100
* Constant: 1.00000000
* Sum of migration in system: 2484029.
* Max Mij value: 4225.
* Max Dij value: 1197.9
* Distance modifier: 0.001000

## Interim results

++ Round 1

* Balancing factor convergence iteration 4 the largest value difference is 0.005224
* Observed mean migration distance: 98.58
* Predicted mean migration distance: 147.05
* Difference: -48.47 (0.00 0.00)
* Beta value: 1.00 (0.00 0.00)
* Newton Raphson - beta changed to: 1.010000

Open file with default viewer
Use the IMAGE Manual
Questions

Contact details:

John Stillwell  j.c.h.stillwell@leeds.ac.uk