Article:
Stillwell, J, Bell, M, Ueffing, P et al. (4 more authors) (2016) Internal migration around the world: comparing distance travelled and its frictional effect. Environment and Planning A: Economy and Space, 48 (8). pp. 1657-1675. ISSN 0308-518X

https://doi.org/10.1177/0308518X16643963

Internal Migration Around the World: Comparing Distance Travelled and its Frictional Effect

Stillwell, J.[1], Bell, M.[2], Ueffing, P. [3], Daras, K.[4], Charles-Edwards, E. [2], Kupiszewski, M. [5] and Kupiszewska, D. [5]

[1] Centre for Spatial Analysis and Policy (CSAP), School of Geography, University of Leeds, Leeds,LS2 9JT, United Kingdom

[2] Queensland Centre for Population Research (QCPR), School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane Qld 4072, Australia

[3] Population Data Unit, Population Division, Department of Economic and Social Affairs, United Nations, New York, United States of America

[4] Department of Geography and Planning, The University of Liverpool, Roxby Building, Liverpool L69 7ZT, United Kingdom

[5] Department of Urban and Population Studies Institute of Geography and Spatial Organization Polish Academy of Sciences, Twarda 51/55, 00-818 Warsaw, Poland

Acknowledgements

The authors are grateful to the Australian Research Council for funding the Discovery Project DP11010136, Comparing Internal Migration around the World (2011-2015), to the project collaborators and national statistical agencies who assisted with data acquisition, and to the three anonymous referees for their very helpful comments on the initial version of this paper. The views expressed in this paper are those of the authors and not those of the United Nations.

Internal Migration Around the World: Comparing Distance Travelled and its Frictional Effect

Abstract

This paper examines how internal migration distance and its frictional effect vary between countries. Such comparisons are hampered by differences in the number and configuration of spatial units for which data are available − the modifiable area unit problem (MAUP). We use the flexible aggregation routines embedded in the IMAGE Studio, a bespoke software platform which incorporates a spatial interaction model, to elucidate these scale and pattern effects in a set of countries for which finely grained origin-destination matrices are available. We identify an exponential relationship between mean migration distance and mean area size but show that the frictional effect of distance remains remarkably stable across spatial scale, except where zones have small populations and are poorly connected. This stability allows robust comparisons between countries even though zonal systems differ. We find that mean migration distances vary widely, being highest in large, low density countries and positively associated with urbanisation, HDI and GDP per capita. This suggests a positive link between development and migration distance, paralleling that between development and migration intensity. We find less variation in the beta parameter that measures distance friction but identify clear spatial divisions between more developed countries, with lower friction in larger, less dense countries undergoing rapid population growth.

1. Introduction

Migration can be defined as changing residence from one geographical location to another. Whether this involves a permanent or a temporary relocation, travel occurs over a specific distance. As with many other forms of spatial interaction, migration conforms with the axiom following from Ravenstein's proposition in the nineteenth century that *"the great body of our migrants only proceed a short distance"* (Ravenstein, 1885, p. 198). Implicit in this statement is that fewer migrants travel longer distances and that distance therefore exerts a frictional effect on migration behaviour.

Bell et al. (2002) identified distance, along with intensity, connectivity and impact as the dimensions of internal migration that are important to consider when making cross-national comparisons. Migration intensity is the propensity to change one's place of usual residence, measured in the form of a rate or probability, whilst connectivity refers to the extent to which regions are linked by migration flows and can be measured using a simple index such as the proportion of the total flows between regions that are non-zero. Migration impact, on the other hand, indicates the extent to which migration transforms the pattern of population settlement and can be measured using a number of indicators such as migration effectiveness or aggregate net migration. Elsewhere we have examined the data available for making such comparisons (Bell et al., 2014b), developed software to compute comparative indicators and address key methodological issues (Stillwell et al., 2014), and assessed how countries differ with respect to overall migration intensities (Bell et al., 2015). In this paper, we turn to the distance dimension in order to examine how far people move and the frictional effect of distance on internal migration in countries around the world. One impediment we face is the lack of information about the exact origins and destinations of the moves individuals make and consequently the difficulty in measuring distance precisely. In almost all countries where internal migration is recorded, the published data refer only to flows between areas whose boundaries have been defined for administrative or statistical purposes. This lack of precision is exacerbated by inconsistencies in the size and shape of these sub-national areas, which is widely recognised as the Modifiable Areal Unit Problem (MAUP), comprising a scale component relating to the different number of regions and a zonation component relating to how the boundaries of zones have been defined. We confront the issues of scale and zonation by using the IMAGE Studio, bespoke software developed as part of the IMAGE project[1], to compare how distance travelled and distance decay vary at different scales and for different zonal configurations in a set of countries for which a matrix of aggregate flows between 99 or more zones is available.

The structure of the paper is as follows: context and previous work are reviewed in section 2; data issues and reasons for our selection of countries are explained in section 3; the methods used to process the data in the IMAGE Studio are outlined in section 4; results

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

of the analysis are discussed in section 5; and some conclusions and suggestions for future work are presented in section 6.

2. Distance and distance decay

People migrate over varying distances for a range of reasons. Life course events are frequent triggers but the range of migration determinants is very broad (Champion et al*.,* 1998) and distance from current residence is one consideration. There are several reasons why migration flows decline with distance, including the social or psychic costs associated with movement away from the neighbourhood of origin, the decline in knowledge about more distant destinations and the financial costs involved in the move itself.

The deterrence effect of distance on migration has been identified widely across the world, including studies by Makower et al*.* (1938; 1939) in Great Britain, Olsson (1965) in Sweden, Sahota (1968) in Brazil, Beals et al*.* (1967) in Ghana, and Rose (1958), Gallaway (1967), Rogers (1967) and Schwarz (1968, 1973) in the USA. Moreover, the relationship between migration and distance has been incorporated into an array of statistical or mathematical models, commencing with the early gravity formulations, derived from Newton's Law of Universal Gravitation, which stipulated that the force between two bodies is directly proportional to the origin and destination masses (P1, P2) and inversely proportional to the distance (D) between them.

Whilst Zipf's P1P2/D hypothesis in 1946 was one of the first attempts to establish an algebraic expression for distance decay, an important application was that by Lowry (1966) in which inter-SMSA migration flows in the US were modelled as a function of explanatory variables describing the origin and destination as well as the basic gravity variables. Four years later, Tobler generalised the gravitational principle as the First Law of Geography which states that "*everything is related to everything else, but near things are more related than distant things"* (Tobler, 1970, p.236) and Wilson (1970) proposed the family of spatial interaction models derived using entropy maximizing methods. Constrained spatial interaction models were subsequently applied by Stillwell (1978), Flowerdew and Aitken (1982) and Fotheringham (1983). The state of the art in migration modelling at the end of the 1980s is reflected in Stillwell and Congdon (1991) and the determinants of internal migration have since been classified by a number of researchers (e.g. Greenwood, 1997; Champion et al*.,* 1998; Borjas, 2000; Anjomani, 2002; van der Gaag et al*.,* 2003).

Distance can be measured either by asking migrants how far they have moved or by estimating distance between the origin and destination. In many countries, there is a dearth of data on migration flows, and they are rarely cross-tabulated by distance. Precise geographical locations are either not collected or not released for reasons of confidentiality. Thus, in most countries the only data available are counts of flows between areal units at one or more spatial scales (Bell et al., 2015). When migration data refer to moves across administrative boundaries, the simplest measure is the Euclidian distance between origin and destination centroids. Alternatively, zone centroids may be computed as geometric points in a GIS or interpolated as population-weighted centroids. The latter are likely to provide more accurate measures, particularly as zones get larger (Niedomysl et al*.,* 2014), but this requires populations at a lower level in the spatial hierarchy which are not always available. Euclidian distances derived from area centroids are a simplification of the real distance involved and may be particularly inaccurate if the two zones are separated by one or more physical barriers. More accurate distances can be obtained using GIS tools and data on the transport network. Gravity model formulations have been used to estimate distance; Tobler et al. (1970), for example, generated estimates of distance from an inverse gravity model whilst Plane (1984) used a doubly constrained spatial interaction model to infer functional distances that would give rise to the magnitude of an observed system of flows. While distance measurement has therefore been a longstanding problem, there is also an accompanying problem when undertaking cross-national comparisons with the different size and shape of the spatial units at which movement is recorded in different countries. We return to the MAUP in section 4.

One of the few studies of migration distance in different countries is reported in Long et al*.* (1988) using data from the US Current Population Survey in 1975 and 1976, the 1981 Census in Great Britain (GB) and the Swedish population register in 1974. The study compares rates of movement in distance bands from less than 50km to over 300km and reveals higher migration intensities in the USA at all distances. Long et al. (1988) also use health survey data to estimate median migration distances for movers in the US during the 1977-1980 period and show that whilst those making one move have a median distance of around 10km, migration distance increases for those making two or more moves. Although modelling migration distance using spatial interaction models has become common (Pellegrini and Fotheringham, 2002), few cross-national comparisons have been attempted. One example is the study by Yano et al. (2003) which reports strong similarity in the extent to which spatial separation reduced the propensity to make long-distance moves in the early 1990s in both Great Britain, where counterurbanisation trends were predominant, and Japan, where urbanisation continued apace.

Three key problems are inherent in attempting such comparisons: the lack of inter-zonal migration data in many countries; the difficulties in measuring inter-zonal distances; and the variability in the size and shape of zones used to record internal migration. In the next section, we report on the migration data available from different countries and how we selected countries for analysis.

3. Data and countries selected

As reported in Bell et al. (2014b), the IMAGE Inventory provides a contemporary assessment of countries around the world which collect internal migration data and their sub-national flows at different spatial scales. We now know, for example, that almost 93% of the 193 United Nations member states collect some sort of migration data, that 82% collect these data using a census, whereas 26% have a register of some sort and 57% use a survey instrument, and that 56% capture data from more than one source (Bell et al., 2014b). We also know that lifetime migration is the most common form of statistic collected, that the one-year interval is used most in Europe and Africa, and the five-year interval is more popular in Latin America, Asia and Oceania. Surveys are also commonly used, especially in Africa, while population registration systems are widely used in Europe, particularly in the Scandinavian countries where administrative systems have been in operation for a long time.

In addition to the IMAGE Inventory, a Repository has been created containing data on migration flows, the associated populations at risk (PAR) and the geographic boundaries of the zones between which migration takes place (Bell et al., 2014a). The migration data in the IMAGE Repository are of three types: national counts of the total number of moves or movers between spatial units; marginal totals indicating total migration to and from each zone; and origin-destination migration matrices. In the case of distance analysis, origin-destination matrices are essential. The matrices in the IMAGE Repository are for aggregate migration, with no demographic or socio-economic disaggregation and, in the large majority of cases, intra-zonal migrants are not available.

Migration matrices have been collected at the finest spatial scale wherever possible and are available for 105 countries. 'Basic Spatial Units (BSUs)' is the generic term we give to the zone system that is used for analysis; BSUs will differ in size and shape within and between countries. Three main criteria determined our selection of countries for the analysis presented here:

(i) *Time periods:* inclusion of both five-year transition data from a census and one-year data from a census or registration system; the five-year/one-year distinction is important since the former include more long-distance moves (Bell et al., 2002). Although the time periods for the data for all countries in the sample are not precisely aligned, they all refer to data collected in the 2001 Census Round or more recently.

(ii) *Number of Basic Spatial Units*: a sufficiently fine level of spatial detail to enable the effects of scales on the measures of distance to be observed; we include countries with 99 or more BSUs, but have not imposed an upper limit although some large matrices may be sparsely populated.

(iii) *Other data:* digital geographical boundaries are available.

Application of these criteria restricts the sample to a total of 29 countries, but with at least one country from each world region. Five-year data are available for 19 countries: Ghana in Africa; Indonesia, Iran and Malaysia in Asia; Switzerland in Europe; Bolivia, Brazil, Chile, Cuba, Dominican Republic, Ecuador, El Salvador, Honduras, Mexico, Nicaragua and Peru in Latin America and the Caribbean; USA and Canada in North America; and Australia in Oceania. One-year data are available for 13 countries: Austria, Denmark, Belgium, Finland, Germany, Italy, Netherlands, Norway, Sweden and the UK from Europe; Australia from Oceania; and the USA and Canada from North America. There are three countries for which one-year census data are available and the remainder produce aggregate data from administrative or registration sources. Both five-year and one-year data are available for Australia, Canada and the USA.

The census dates, and therefore the migration periods, are not entirely consistent and only registration data are available for a number of the European countries, so data for the most recent periods are used. There is wide diversity in the size of the countries involved, both in terms of area and population, as well as in the magnitude of the migration flows taking place between zones and hence in their corresponding crude migration intensities. Since only data on aggregate migration have been collected, it is not possible to distinguish migrants by motivation other than through the presumption that shorter-distance moves will primarily be housing-driven and longer-distance moves will be due to other reasons, such as jobs. In excluding intra-BSU flows, we attempt to restrict our analysis to internal migration rather than residential mobility.

4. Methodology: Modelling and aggregation with the IMAGE Studio

4.1 Distance measurement and spatial interaction modelling

The selected countries vary widely in the number and size of geographical areas for which migration data are produced, and which we refer to henceforth as BSUs. The IMAGE Studio (Daras, 2014; Stillwell et al., 2014) comprises a series of sub-systems that provide for the: (i) preparation of data; (ii) aggregation of BSUs into large areas referred to as Aggregated Spatial Regions (ASRs); (iii) computation of a suite of migration indicators; and (iv) calibration of a doubly constrained spatial interaction model (SIM). The system allows the computation of indicators and the SIM calibration using data for either BSUs or ASRs.

Included amongst the migration indicators are measures of mean and median migration distance across the whole spatial system. The median (*MedMD*) is the simplest indicator, calculated as the midpoint of a cumulative frequency distribution of migrants ranked according to how far they move, whilst the mean (*MMD*) for BSUs is defined when using the SIM as:

$$MMD = \sum_i \sum_j Mij \; dij \; / \sum_i \sum_j Mij \qquad (1)$$

where Mij is the number of migrants between origin BSU *i* and destination BSU *j*, where $i \neq j$ and both vary from 1 to m (the number of regions), and where dij is the distance between BSUs *i* and *j,* calculated using the Pythagorean formula for Cartesian systems.

The observed *MMD* is used as the convergence criterion for the doubly constrained spatial interaction model available in the IMAGE Studio which is defined as:

$$M'ij = Ai\ Bj\ Oi\ Dj\ dij^{-\beta} \qquad (2)$$

where $M'ij$ is the modelled flow between origin BSU *i* and destination BSU j, Oi and Dj are the observed totals of outmigration from BSU *i* and in-migration to BSU *j* respectively, Ai and Bj are balancing factors computed endogenously that enable the flows in cells across each row and each column of the modelled matrix to sum to the known totals Oi and Dj respectively, and β is the generalised distance decay parameter associated with a negative power function which is calibrated automatically using a Newton-Raphson iterative search routine. The best-fit distance decay parameter is identified when the *MMD* associated with the flows predicted by the model are equal to the *MMD* of the flows in the observed matrix, as explained in greater detail in Stillwell (1991).

4.2 MAUP and the IMAGE Studio aggregation method

As identified earlier, the problems of cross-national comparison of migration distance are underpinned by the MAUP, whose components are explained in detail by Openshaw (1984), and include the *scale effect* or the variation in results obtained when data for one set of BSUs are aggregated into larger ASRs, and the *zonation (or aggregation) 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 zones changes, whereas the zonation effect is identified by observing the indicator change when the number of zones remains the same but the zones are configured differently.

The IMAGE Studio has been developed to accommodate a methodological response to the MAUP challenge for analysis of internal migration indicators. As explained in Stillwell et al. (2014), the idea is to aggregate the BSUs to ASRs in a stepwise fashion in order to explore how indicators change at different scales but also to identify the sensitivity of indicators to alternative configurations at each scale. So, a key research question underpinning the analysis is: what happens to the *MMD* and the decay parameter when we aggregate the migration data for *m* BSUs to *n* ASRs with *x* zonations of ASRs at each scale?

The IMAGE Studio contains two different aggregation algorithms for generating contiguous ASRs from the set of BSUs. The original Initial Random Aggregation (IRA) algorithm, developed by Openshaw (1977), provides a high degree of randomisation to ensure that the resulting aggregations differ during each iteration. It has been implemented with object-oriented principles, thus avoiding the sustained sequential processes and resulting in much quicker random aggregation (Daras, 2006). However, an alternative algorithm for aggregating BSUs is the IRA-wave algorithm, a hybrid version of the original IRA algorithm and the breadth-first search (BFS) algorithm. The first step is to select BSUs randomly and assign each one to an empty ASR. Using an iterative process until all the BSUs have been allocated, the algorithm identifies the adjusted areas of each ASR, targeting only the BSUs without an assigned ASR, and adds them to each ASR. One advantage of the IRA-wave algorithm is its speed in producing a large number of aggregations. Moreover, the IRA-wave provides well-shaped ASRs in comparison to the irregular shapes of the IRA algorithm and for these reasons, we have chosen to use this approach.

The next step in the spatial aggregation process is to generate information on inter-ASR distances, flows, centroids, areas and populations at the level of each aggregation. Each distance between a pair of ASRs is calculated as the mean of inter-BSU distances between the two ASRs. The aggregated flows between the new ASRs are calculated by summing the flows from the BSUs that constitute an origin ASR to the 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 migration retained in the system decreases with each scale step as the ASRs increase in size. Thereafter, migration indicators can be computed and model parameters calibrated for each configuration at each spatial scale.

The results reported in the next section are based on computing values of the mean, maximum and minimum *MMD* and decay parameter for the two sets of countries using scale steps of 10 with 100 zonations at each scale. Changes in mean parameter values from one scale step to another indicate the scale effect whereas the zonation effect is captured by the difference between the maximum and minimum value of the parameters at each level of scale. Because of the time taken to run the aggregation where there are a large number of BSUs, a larger step size between scales is used for certain countries.

5. Results

5.1 Variations in migration distance and generalised parameter

There are marked differences between the *MMD* and *MedMD* in many countries; Figure 1 illustrates both measures for each set of countries ranked on the basis of the inter-BSU *MMD*.

Figure 1 Mean and median migration distances, five-year and one-year migration countries

As expected given the size of the countries involved, the *MMD* varies widely in both data sets; however, the relationship between the mean and median migration distances is not consistent. Whilst the *MMD* values for both North American countries are above 600km and the *MedMD* is between 120 and 145km, in Brazil, Iran, Bolivia, Ghana and Ecuador the *MedMD* values are much closer to the MMDs, indicating a more even distribution of flows between pairs of BSUs. In countries like Australia, Chile and Cuba, where the *MedMD* and *MMD* are further apart, the results suggest that the distribution of migrants is more concentrated on short-distance flows.

The decay parameters derived from modelling inter-BSU flows (Figure 2) for the five-year migration countries lie between 3.75 (Ecuador) and 0.87 (Cuba) with higher values suggesting that migrants are more affected by the frictional effect of distance. Amongst the one-year data, a very high parameter is calibrated for Belgium (19.8) (not included in Figure 2) whilst Italy (0.81) has the lowest value in the set. The unusually high frictional effect of distance on migrants in Belgium is very likely to be due to the cultural and linguistic division between the Dutch-speaking region of Flanders in the north and the French-speaking region of Wallonia in the south, with Brussels, the capital region, being a mostly French-speaking enclave within the Flemish region. Tensions between the Dutch and French-speaking communities due to different rates of economic development as well as language have meant that internal migration tends to be highly constrained to flows within the two communities and to and from Brussels. In fact, only 14.5% of all the cells in the full BSU matrix for Belgium contained non-zero values.

Figure 2 Beta values and spatial connectivity, one-year and five-year migration countries

Figure 2 also shows a connectivity index for each country measuring the proportion of all flows in the matrix that are non-zero and indicates that Italy, the country with the lowest beta parameter has the highest connectivity (95%). Very low values are observed in the USA for both one-year and five-year migration but, unlike Belgium, this is due to the size of this country rather than cultural division. Mexico also has very low levels of connectivity for five-year migration between BSUs whereas 70% of the matrix of five-year inter-BSU migration flows in Australia are non-zero in the five-year data and only 49% in the one-year data.

As might be expected, there is a negative relationship between the MMD and the distance decay parameter in each of our two sets of countries (where the extreme values have been removed), but the relationship is relatively weak for five-year data (R^2=0.33) and insignificant for one-year data (R^2=0.09). Whilst there is also a negative correlation between the beta parameter and the connectivity index in both data sets, with R^2 values of 0.26 and 0.14 respectively, spatial connectivity through migration is a separate dimension identified by Bell et al. (2002) and should not be regarded as an alternative to the decay parameter.

These indicators highlight the diversity of migration behaviour but the variation between countries can be traced at least in part to differences in geographic size, spatial structure and the zonal systems for which migration data are available. In the next sub-section, we report the results of our method for attempting to understand how scale and zonal configurations affect the measurement of mean migration distance and distance decay.

5.2 Distances migrated between zones: scale effects

The mean *MMD* is plotted against the number of ASRs in Figure 3, indicating the distances that people move at each spatial scale. Differences between countries are due in part to variations in the size of countries and their constituent regions. At one extreme, people in the USA migrate much further, on average, than people in the other countries at all spatial scales over both time periods. At the other end of the spectrum, domestic migrants in El Salvador move over the shortest distances in the five-year period whilst the shortest-distance movers over one year are those in Belgium and the Netherlands. The shapes of the schedules show the scale effects: that is, the extent to which migration distance increases in each country as the number of ASRs gets smaller (or the size of the ASRs gets bigger), moving from right to left on the graph. In some of the larger countries, including Canada and Australia, the scale effects are significant; in other cases, such as El Salvador and Switzerland, the distance moved changes very little as the number of zones reduces.

Figure 3 Mean inter-zonal distances migrated by scale, countries with five-year data and one-year

While these graphs reveal the scale effects for each country, the number of ASRs is a poor basis for comparison as ASRs differ between countries in terms of area and/or population. In order to make more robust comparisons, we use mean area size at each spatial scale to replace the number of ASRs on the horizontal axis. Mean area size is preferred to mean population size because distance is a physical phenomenon. When curves are fitted to the *MMD*-area relationship for each country using R, the best-fit is represented by a power function which can be written as:

$$MMD = a\ (A/n)^b \qquad (3)$$

where A/n is the mean ASR area size at scale n and a and b are parameters that define the function but are not directly interpretable. The exponent value (b) varies between 0.12 (Ecuador) and 0.42 (Malaysia) for the five-year migration countries whilst the a parameter varies between 218 (USA) and 6 (Malaysia) and is therefore more important in defining the function. This is also the case for one-year migration countries. In each case, the R^2 value measuring the goodness of fit of the model is above 0.94. The different effects of scale are evident from the mean *MMD* values predicted using the model for each country in Figure 4. In both graphs, the USA stands apart with migrants moving relatively long distances across the area size spectrum but with pronounced scale effects. The shape of the curves suggests that scale effects on *MMD* for five-year migrants are most apparent in Mexico, Iran, Brazil, and Canada and least evident in El Salvador and Switzerland, whereas the predicted *MMD* values for one-year migrants in Canada and Italy appear most influenced by scale and least affected are those in Belgium, Denmark and the Netherlands.

Figure 4 Modelled relationship between MMD and area size, five-year and one-year migration countries

These model schedules can be used to derive comparable values of *MMD* according to selected area size criteria. In effect, plotting *MMD* against mean area serves to adjust for differences between countries in the size of ASRs and therefore allows direct comparison of distance migrated. Figure 5 plots migration distances across the sample countries in rank order at two area sizes, 100 and 500 square kilometres. These league tables confirm that migrants in the USA move considerably further at both size thresholds than any other countries in our sample. Mexico, Brazil, Iran and Canada among the five-year migration countries also have MMD values above 200km when area size is 500 sq km while Italy and Canada take positions behind the USA for the one-year migration countries. At the other end of the five-year table appear several Central American countries including El Salvador and Honduras, as well as Switzerland and Malaysia, while Denmark, Netherlands and Belgium appear at the bottom of the *MMD* league table for one-year migration.

Figure 5 Mean migration distances for areas of 100 and 500 sq. kilometres, five-year and one-year migration countries

5.3 Variations in the decay parameter: scale effects

The beta parameter derived from the SIM provides a more general measure of the frictional effect of distance on inter-zonal migration. The mean beta values show remarkable stability when plotted against the mean population size for ASRs at each spatial scale for most countries (Figure 6). Mean population size is adopted as an alternative to mean area in this instance because it better captures the settlement framework within which people migrate, but it has a similar effect of standardising for differences between countries. Moving from left to right on each graph, the population size of ASRs increases but changes in the beta value are negligible. The graphs suggest that five-year migrants in Nicaragua and Switzerland are most influenced by distance at all spatial scales for areas with mean populations above 100,000, whilst those in Cuba experience the lowest frictional effects of distance. In the one-year countries, migrants in the Netherlands experience the highest distance effects while those in Italy are affected least.

Figure 6 Mean distance decay parameter by population size, five-year and one-year migration countries

The stability in the decay parameter as scale changes (the lack of a scale effect) is partly explained by the fact that as the number of ASRs gets smaller, the migrants between BSUs within the ASRs are removed, so progressively fewer short-distance migrants are included in the model. There are, however, some anomalies in the schedules depicted in Figure 7, where mean beta values increase significantly when the ASR populations are relatively small. This is particularly evident for Mexico, Ecuador and Peru among the five-year migration countries and Norway, Finland and Sweden among the one-year countries. As noted earlier in discussion of migration distance, these anomalies are explained by the low connectivity between ASRs when the mean population size drops below around 100,000. This is not a general feature; in other countries for which ASRs with small populations are available, such as Nicaragua, Switzerland and Bolivia, the tendency is for the beta value to fall marginally.

An important consequence of this stability is that the model decay parameter calibrated at one scale will be appropriate for use at another scale. For example, if we wanted to estimate flows for the USA for areas with populations of 200,000, we could use the parameter for the original set of BSUs. The corollary is that values of the distance decay parameter can be ranked and compared between countries, as indicated in Figure 7 at mean populations of 200,000 and 500,000. In the case of the five-year data, distance appears to have the greatest frictional effect in Nicaragua and Switzerland whereas Cuba is at the bottom of the league table. Between these outliers, however, there is marked uniformity across the distribution, and much less variation in the distance decay parameter than was apparent in *MMD*, with coefficients of variation from 0.2 to 0.35 compared with 0.7 to 0.95 for the *MMD*. At the 500,000 population level, parameters for Mexico, Peru, Ecuador, Bolivia and Brazil, together with Iran and Indonesia, lie within the 1.5 to 1.8 range, while for El Salvador, the Dominican Republic, Honduras and Chile, as well as the USA, Canada, Australia and Malaysia the friction of distance is lower, with parameters ranging from 0.9 to 1.3. The two clusters appear to broadly reflect differences in levels of economic development. For the countries which collect one-year data, distance friction appears to be highest in western Europe, but lower in North America, Australia and Norway.

Note: Values of beta for zones with mean populations of 200,000 are not available for Italy and Indonesia because migration data were not collected at a sufficiently fine scale

Figure 7 Beta values for area populations of 200,000 and 500,000, five-year and one-year migration countries

5.4 Links with development indicators

In the case of migration intensities, Bell et al. (2015) found that differences between countries were strongly correlated with a number of geographic, demographic, economic and social indicators, including the Human Development Index (HDI), GDP per capita, urbanisation and labour force participation. In the case of migration distance, the results in Table 1 provide a more mixed picture. Not surprisingly, geographic area is positively associated with *MMD* and inversely related to the friction of distance: people in larger countries tend to move longer distances and are less influenced by the friction of distance, though the latter relationship is much weaker and not significant for the sample of countries which measure migration over five years. High population densities tend to reduce migration distances and increase frictional effects, although beta parameters for the five-year countries show a contrary association. Level of urbanisation, HDI and GDP per capita are all positively associated with *MMD* for the range of countries in the five-year group, but the relationships are weaker, inverted and not significant among the one-year countries. It is notable that countries in the five-year group with higher overall levels of migration intensity also display longer migration distances. Together these results suggest a clear positive link between development and migration distance, paralleling that reported elsewhere between development and migration intensity (Bell et al., 2015). Results for the beta values are weaker and inconsistent across the two groups of countries. Most notable here is that among the one-year sample, distance friction appears to be low in large, low density countries with high mobility and high population growth rates. On the other hand, none of the selected variables account for the computed differences in distance friction across the sample of countries that collect five-year data.

Table 1 Correlation between migration distance indicators and selected variables

5.5 Distance migrated and decay parameter: zonation effects

Hitherto, it has been the effects of scale on the mean values of the migration indicators that have been reported. We have used the minimum and maximum values to identify variation around the mean *MMD* and beta values to indicate the zonation effects. In all countries, the zonation effects for *MMD* increase as the number of ASRs gets smaller or as the log of the mean area gets larger, as exemplified by the three countries selected from the five-year migration sample (Figure 8, upper panel). The zonation effects appear to be significant in large countries like the USA, Canada and Australia but also in Chile, Indonesia, Malaysia, Norway and Italy, due perhaps to the shapes of the BSU polygons and the unevenness of settlement across these countries. Zonation effects appear to be less evident in Latin American countries like Ecuador, El Salvador, Honduras, Nicaragua and Peru, as well as in Ghana, Switzerland and the Netherlands.

Figure 8 Zonation effects on MMD and beta for selected countries

The zonation effects for beta values in the lower panel of Figure 8 illustrate how the range diverges as mean population size increases and indicate a much greater effect in Nicaragua than in the USA, where the effect is negligible. In the five-year sample, the different spatial configurations appear to exert little effect in Brazil, Canada, Indonesia, Iran, Malaysia, Mexico and the USA whereas in Australia, Switzerland, Chile, Cuba and Ghana, the zonation effect is more marked. In Bolivia, Dominican Republic, Honduras and Nicaragua, different spatial systems result in quite wide variations in the parameter, particularly when mean populations are large. In the one-year migration countries, Austria, Finland, Norway and Sweden have larger zonation effects than the remaining countries but it is Denmark that stands out as having the most distinctive variation in beta at different spatial scales.

One characteristic of the zonation effect is that, for several countries, the difference between the minimum value of the MMD and the mean is lower and more stable across scales than the difference between the maximum value and the mean. It is likely that the shape of a country and its subnational areas plays a role in the aggregation procedure so that relatively large regions are sometimes created that produce longer MMD values and these may explain the higher deviation above the mean.

6. Conclusions

The research reported in this paper is part of the IMAGE project that is concerned with comparing internal migration in countries around the world. The focus on distance is particularly challenging because of the practical problems in obtaining matrices of migration flows and the difficulties created by the MAUP. Whilst many studies of migration distance have been undertaken, few have attempted to compare distance travelled and its frictional effect in different countries. We approached the task using the flexible aggregation routines embedded in the IMAGE Studio, which allows us to evaluate the sensitivity of migration distance to the two key dimensions of the MAUP – that is to changes in spatial scale and to variations in the configuration of sub-national areas at different scales. We focused on two discrete but inter-related measures: mean migration distance and the beta parameter from a spatial interaction model that captures the friction of distance. We have shown that plotting these indicators against mean area sizes and mean populations at various levels of spatial aggregation effectively standardises for some geographic differences and allows us to draw direct comparisons between countries.

The results reveal a positive relationship between mean migration distance and the mean area size that is consistent across all countries and is represented by power function. This consistency can be used to compare the scale and zonation effects of differing geographies on the migration distance apparent in different countries. The analysis also shows that scale effects on the distance decay parameter are minimal: the parameter is very stable in relation to mean population size except where subnational areas are relatively small and the migration matrix is sparsely populated. When connectivity becomes weak or non-existent, as in countries with large numbers of small areas or where there are constraints on movement (as in Belgium), the distance decay parameter increases sharply. Together, these results enable direct comparisons to be made between countries in terms of the distance migrated and the frictional effect of distance at selected values of area size or population respectively.

The highest migration distances across our sample of 29 countries are found in several large countries including the USA, Canada, Mexico, Brazil and Iran. This conclusion is intuitive given the size and settlement patterns of these countries which prompt longer distance moves. Distances migrated are substantially lower in many smaller countries, such as those in Central America, but also in Australia. Conversely, Italy has a mean migration distance above 200km, which is comparable to that for Canada. These differences persist when distances are computed between zones of equal area in each country, which suggests that migration distances are influenced by the shape and settlement pattern of countries, as well as by their aggregate geographic area. The influence on migration patterns exerted by the major cities was identified, for example, in Ravenstein's analysis of internal migration in the UK in the nineteenth century and incorporated into his 'law' that migrants travelling long distances generally choose one of the great centres of commerce of industry (Ravenstein, 1885). More recently, Henrie and Plane (2006) identify the influence of the spatial location of large metropolitan areas on migration patterns (and therefore distances) in the USA between 1950 and 2000. One potentially fruitful approach for further work would be to interpret the zone-specific balancing factors of the spatial interaction model as attractiveness factors for each regions within each country.

The friction of distance, as measured by the beta parameter of a spatial interaction model, shows much less variation. Among the 13 countries which measure migration over one year, it tends to be lowest among the large, New World countries of the USA, Canada and Australia, and in Norway, and somewhat higher in much of Western Europe, but with Italy (low) and Finland (high) being notable exceptions. Among the larger and more geographically diverse group of countries that collect data over five years, Cuba has the lowest distance decay parameter, and Nicaragua and Switzerland the highest. The remaining countries in the five-year group fall into two broad clusters, loosely distinguished by level of development, with those at higher levels of development generally displaying lower levels of distance friction. Simple correlations against a range of explanatory variables support this interpretation, showing longer-distance migrations in countries with higher levels of urbanisation, HDI, GDP per capita and aggregate migration intensity, at least for the broad and scattered sample of countries that collect migration data over a five-year interval. For the largely European sample of countries together with North America and Australia that collect data over one-year, lower levels of distance friction are more readily found in larger countries with higher population growth rates, high mobility and low population density. Further characteristics of the flow matrices in each country, such as the proportions of migration that are urban-urban, urban-rural, rural-urban and rural-rural, are required before any relationship between migration distance or its frictional effect can be linked with stages in Zelinsky's model of migration transition (Zelinsky, 1971).

On a methodological front, the IMAGE Studio provides an innovative approach to investigating scale and zonation effects in a systematic and automated manner and has potential for use in further investigations of migration behaviour. Whilst data assembly for the IMAGE project has been confined to matrices of total migrants between BSUs in different countries, there is a small number of countries for which intra-BSU flows are also available. The spatial interaction model within the IMAGE software has the option of including intra-zonal flows but one of the challenges in this context is the measurement of intra-zonal distance. The sensitivity of model results to alternative measures of intra-zonal distance merits exploration. Opportunities also exist to examine differences in the way migration distance, and the effects of scale and zonation, vary between migrants according to other demographic characteristics of interest. However, given the constraints on data availability in countries around the world that have become evident in assembling aggregate migration flows for this project, further research is realistically going to be confined to comparison between a relatively small number of countries or to analysis of a single country where origin-destination migration flows disaggregated by variables such as age, sex, ethnicity or occupation are available.

References

Anjomani A, 2002, "Regional growth and interstate migration" *Socio-Economic Planning Sciences* **36** 239-265

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 Series A* **165** 435-464

Bell M, Bernard A, Ueffing P, Charles-Edwards E, 2014a, "The IMAGE Repository: a user guide", Working Paper 2014/01, Queensland Centre for Population Research, University of Queensland, Brisbane

Bell M, Charles-Edwards E, Kupiszewska D, Kupiszewski M, Stillwell J, Zhu Y, 2014b, "Internal migration data around the world: assessing contemporary practice" *Population Space and Place* **21**(1) 1-17

Bell M, Charles-Edwards E, Ueffing P, Stillwell J, Kupiszewski M, Kupiszewska D, 2015, "Internal migration and development: comparing migration intensities around the world" *Population and Development Review*, **41**(1) 33-58

Beals R E, Levy M B, Moses L N, 1967, "Rationality and migration in Ghana" *Review of Economics and Statistics* **49**(4) 480-86

Borjas G J, 2000, Economics of migration, *International Encyclopedia of the Social & Behavioral Sciences*, Section No 3.4, Article No 38

Champion A, Fotheringham A S, Rees P, Boyle P, Stillwell J, 1998, *The Determinants of Migration Flows in England: A Review of Existing Data and Evidence* (Report prepared for the Department of Environment, Transport and the Regions, Department of Geography, University of Newcastle upon Tyne, Newcastle upon Tyne)

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 1.1 User Manual" School of Geography, University of Leeds, Leeds

Flowerdew R, Aitken M, 1982, "A method of fitting the gravity model based on the Poisson distribution" *Journal of Regional Science* **22** 191-202

Fotheringham A S, 1983, "A new set of spatial interaction models: the theory of competing destinations *Environment and Planning A* **15** 15-36

Gallaway L E, 1967, "Industry variations in geographic labor mobility patterns" *Journal of Human Resources* **2**(4) 461-74

Greenwood M J, 1997, Internal migration in developed countries, in *Handbook of Population and Family Economics Volume 1* Eds M R Rosenzweig, O Star, (Elsevier, London) pp. 647-720

Long L H, Tucker C J, Urton W L, 1988, "Migration distances, an international comparison" *Demography* **25** 633-640

Lowry I S, 1966, *Migration and Metropolitan Growth: Two Analytical Models* (Chandler Publishing Company, San Francisco)

Makower H, Marschak J, Robinson HW, 1938, "Studies in mobility of labour: a tentative statistical measure" *Oxford Economic Papers* **1** 83-123

Makower H, Marschak J, Robinson HW, 1939, “Studies in mobility of labour: analysis for Great Britain, Part 1” *Oxford Economic Papers* **2** 70-93

Niedomysl T, Fransson U, 2014, “On distance and the spatial dimension in the definition of internal migration” *Annals of the Association of American Geographers* **104**(2) 357-372

Olsson G, 1965, *Distance and Human Interaction: A Review and Bibliography* (Regional Science Research Institute, Philadephia)

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* (Geobooks, Norwich)

Pellegrini P A, Fotheringham A S, 2002, “Modelling spatial choice: a review and synthesis in a migration context” *Progress in Human Geography* **26** 487-510

Plane D A, 1984, “Migration space: doubly constrained gravity model mapping of relative interstate separation” *Annals of the Association of American Geographers* **74**(2) 244-256

Henrie C J, Plane D A, 2006, “Decentralization of the nation’s main street: new coastal-proximity-based portrayals of population distribution in the United States, 1950–2000” *Professional Geographer* **58** 448-459

Ravenstein E, 1885 “The laws of migration” *Journal of the Statistical Society* **46** 167-235

Rogers A, 1967, “A regression analysis of inter-regional migration in California” *Review of Economics and Statistics* **49** 262-67

Rose A M, 1958, “Distance of migration and socioeconomic status of migrants” *American Sociological Review* **23**(4) 420-23

Sahota G S, 1968, “An economic analysis of internal migration in Brazil” *Journal of Political Economy* **76**(2) 218-45

Schwarz A, 1968 Migration and Earning in the US PhD Dissertation, University of Chicago.

Schwarz A, 1973, “Interpreting the effect of distance on migration” *Journal of Political Economy* **81**(5) 1153-1169

Stillwell J, 1978, “Inter-zonal migration: some historical tests of spatial interaction models” *Environment and Planning A* **10** 1187-1200

Stillwell J, 1991, Spatial interaction models and the propensity to migrate over distance, in *Migration Models Macro and Micro Approaches* Eds J Stillwell, Congdon P, (Belhaven Press, London) pp. 34-56

Stillwell J, Congdon P, Eds, 1991, *Migration Models Macro and Micro Approaches* (Belhaven Press, London)

Stillwell J, Daras K, Bell M, Lomax N, 2014, “The IMAGE studio: a tool for internal migration analysis and modelling” *Applied Spatial Analysis and Policy* **7**(1) 5-23

Tobler W, 1970, “A computer movie simulating urban growth in the Detroit region” *Economic Geography* **46** 234-40

Tobler W R, Mielke H W, Detwyler T R, 1970, “Geobotanical distance between New Zealand and neighboring islands” *BioScience,* **20**(9) 537-542

Van der Gaag N, van Wissen L, Rees P, Stillwell J, Kupiszewski M, 2003, *Study of Past and Future Interregional Migration Trends and Patterns within European Union Countries in Search of a Generally Applicable Explanatory Mode,* Final Report on behalf of Eurostat, NIDI, The Hague

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

Yano K, Nakaya T, Fotheringham A S, Openshaw S, Ishikawa Y, 2003, “A comparison of migration behaviour in Japan and Britain using spatial interaction models” *International Journal of Population Geography* **9** 419-431

Zelinsky W, 1971, “The hypothesis of the mobility transition” *Geographical Review* **16**(2) 219-249

Zipf G K, 1946, “The P1 P2/D hypothesis: on the intercity movement of persons” *American Sociological Review* **11** 677-686

Table 1 Correlation between migration distance indicators and selected variables

Variable	Median migration distance				Beta parameter			
	One year		Five years		One year		Five years	
	100km²	500km²	100km²	500km²	200,000	500,000	200,000	500,000
Area	0.6477**	0.6979***	0.6398***	0.6596***	-0.7733***	-0.5578*	-0.1787	-0.1338
Population density	-0.3474	-0.3913	-0.3769	-0.4168*	0.7647***	0.5512*	-0.1409	-0.0854
Per cent urban	-0.2565	-0.2649	0.4463*	0.4648**	-0.1351	0.2006	0.0359	-0.1891
Population growth	0.2881	0.3162	-0.1875	-0.1852	-0.7727***	-0.5773**	-0.0201	-0.0238
Human Development Index	0.1412	0.1536	0.4608**	0.4764**	-0.343	-0.0414	-0.0353	-0.1167
GDP per capita, PPP	0.2123	0.2107	0.4159*	0.4313*	-0.3107	-0.1025	0.0357	-0.0423
Labor force participation rate	-0.2095	-0.1846	0.0327	0.0277	-0.0381	0.4941	0.0478	0.1129
Aggregate crude migration intensity	0.1369	0.1558	0.4039*	0.4297*	-0.5272*	-0.0146	0.0025	-0.1081

*p < 0.10; **p < 0.05; ***p < 0.01

Note: MMD values are estimated for mean zone sizes of 100km² and 500km²; beta parameters are estimated for zones of 2000,000 and 500,000 mean population size

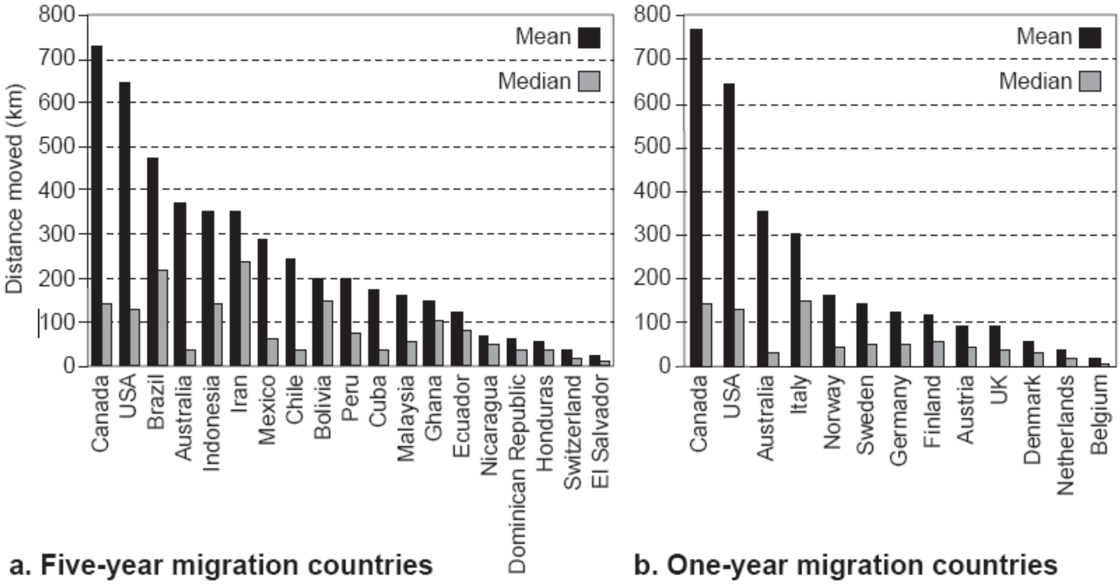


Figure 1 Mean and median migration distances, five-year and one-year migration countries

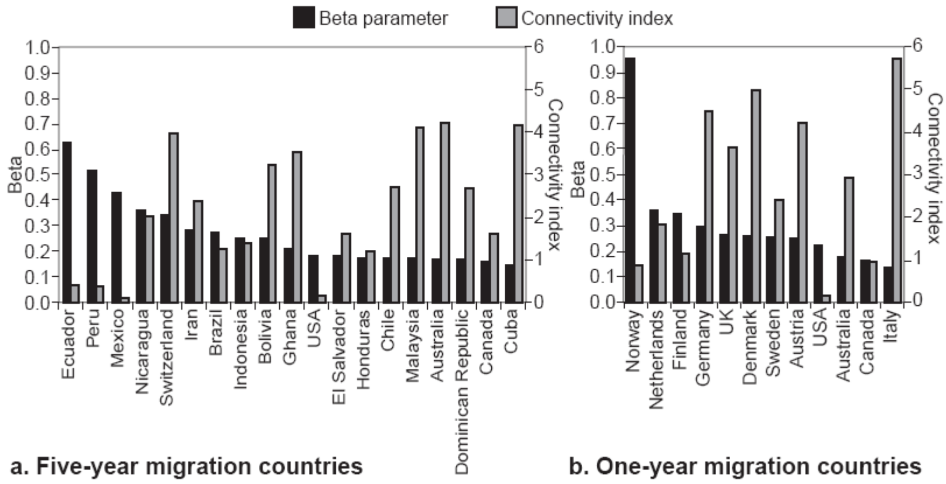


Figure 2 Beta values and spatial connectivity, one-year and five-year migration countries

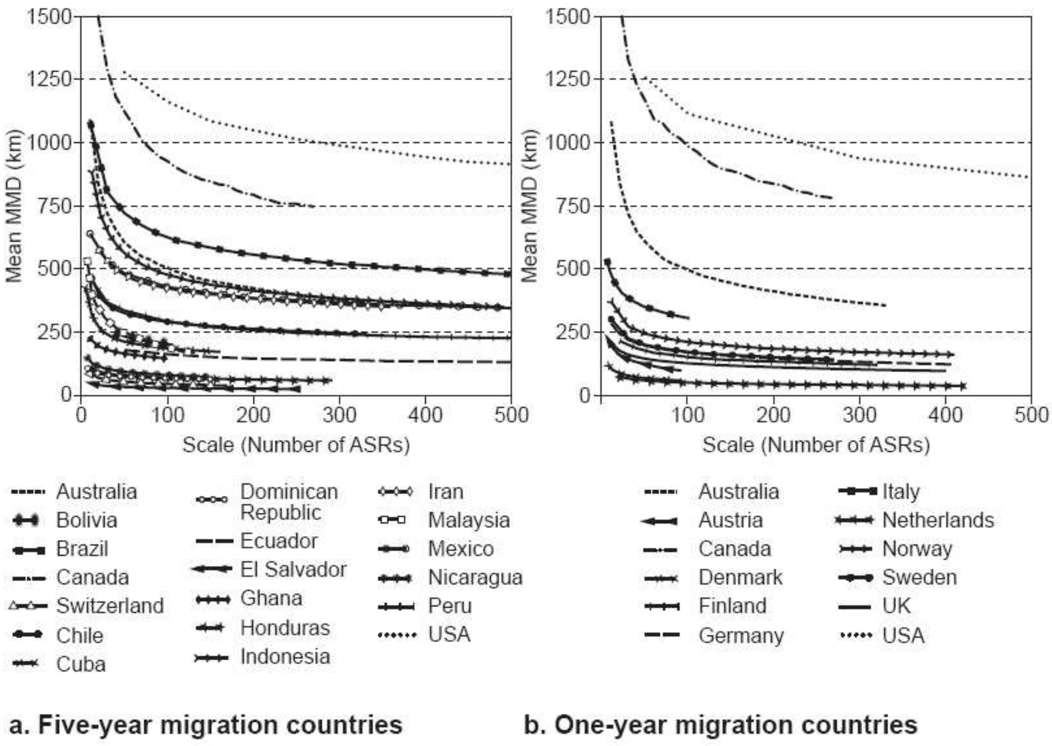


Figure 3 Mean inter-zonal distances migrated by scale, countries with five-year data and one-year

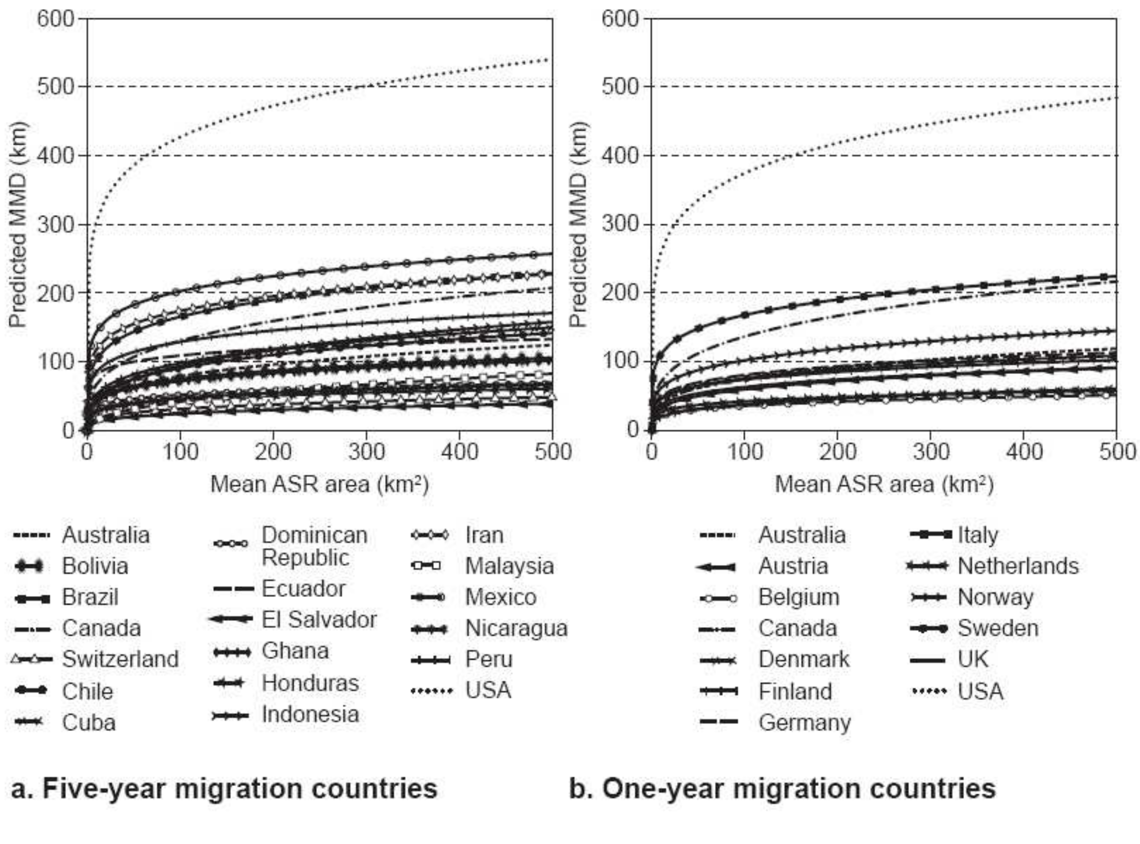


Figure 4 Modelled relationship between MMD and area size, five-year and one-year migration countries

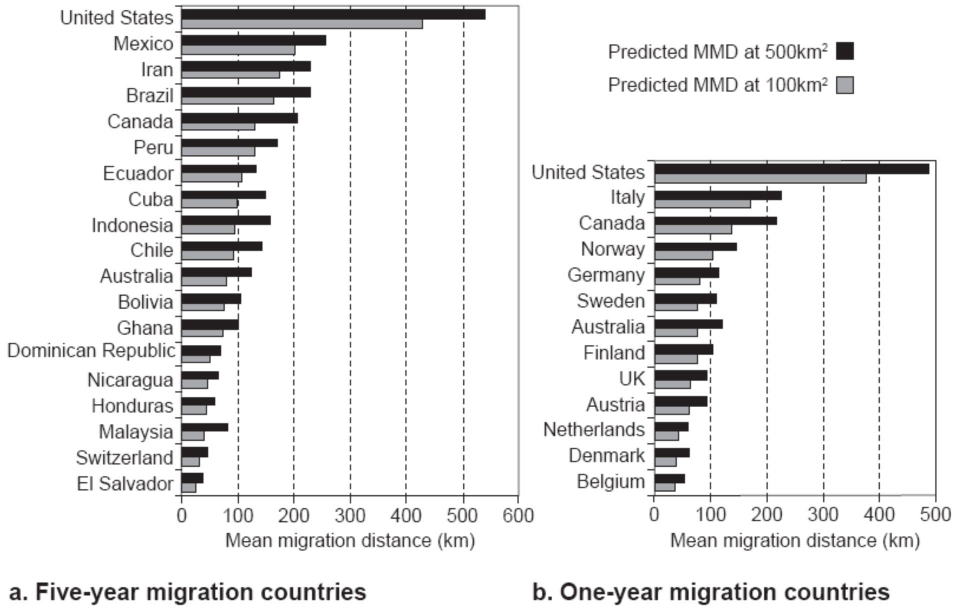


Figure 5 Mean migration distances for areas of 100 and 500 sq. kilometres, five-year and one-year migration countries

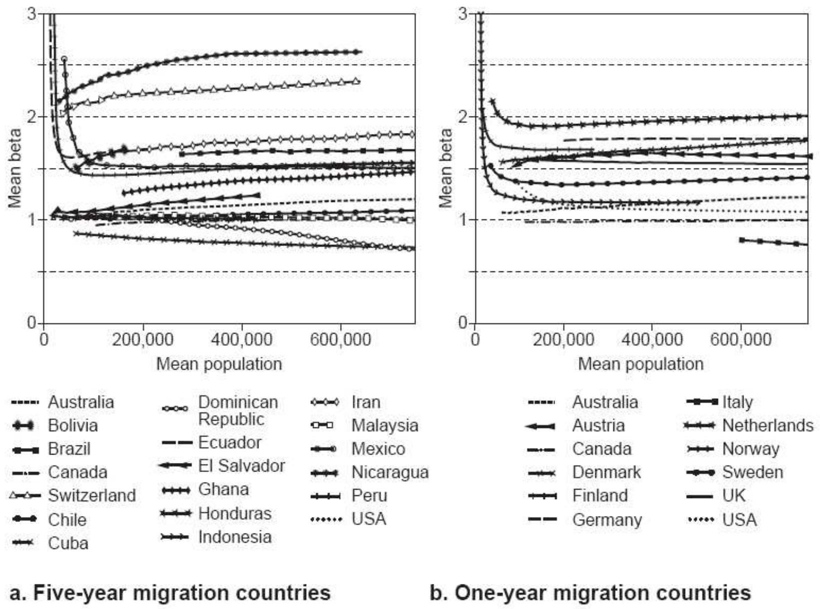
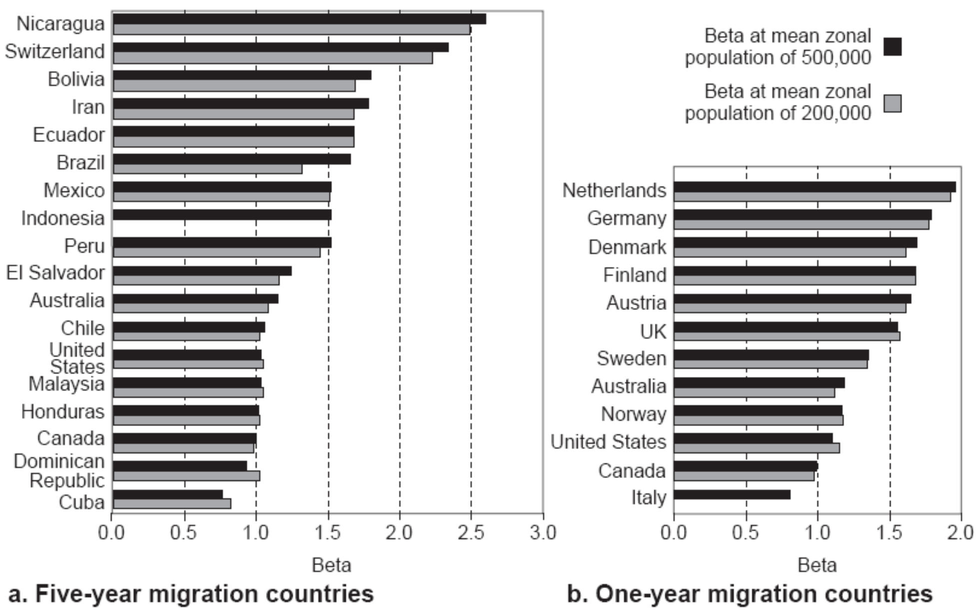


Figure 6 Mean distance decay parameter by population size, five-year and one-year migration countries



Note: Values of beta for zones with mean populations of 200,000 are not available for Italy and Indonesia because migration data were not collected at a sufficiently fine scale

Figure 7 Beta values for area populations of 200,000 and 500,000, five-year and one-year migration countries

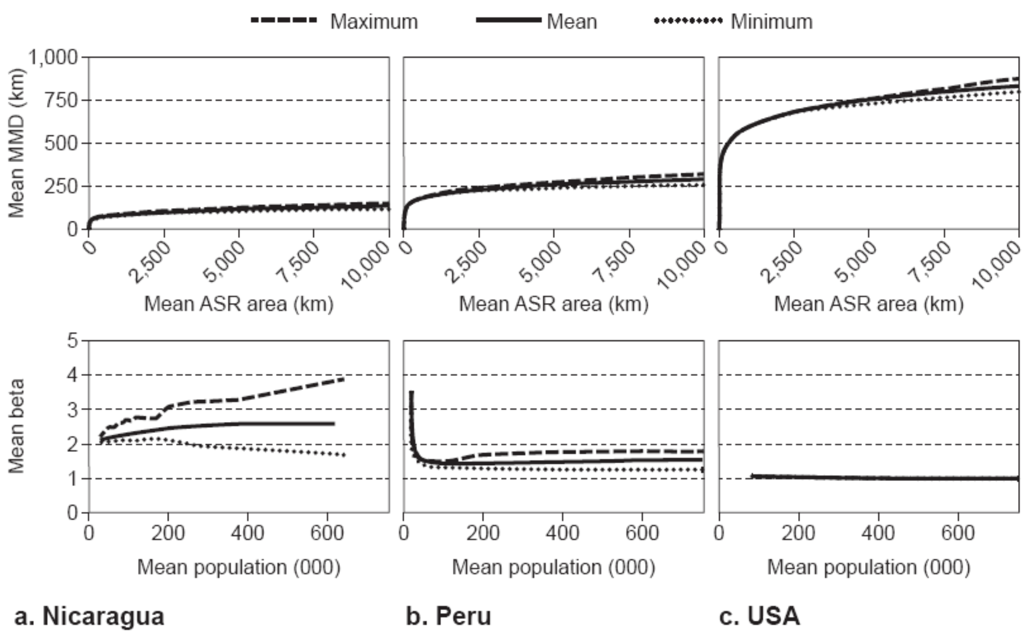


Figure 8 Zonation effects on MMD and beta for selected countries