Regional Studies, Regional Science

ISSN: (Print) 2168-1376 (Online) Journal homepage: http://www.tandfonline.com/loi/rsrs20

How far do internal migrants really move? Demonstrating a new method for the estimation of intra-zonal distance

John Stillwell & Michael Thomas

To cite this article: John Stillwell & Michael Thomas (2016) How far do internal migrants really move? Demonstrating a new method for the estimation of intra-zonal distance, Regional Studies, Regional Science, 3:1, 28-47, DOI: 10.1080/21681376.2015.1109473

To link to this article: http://dx.doi.org/10.1080/21681376.2015.1109473

Published online: 30 Nov 2015.

Regional Studies, Regional Science, 2016
Vol. 3, No. 1, 28–47, http://dx.doi.org/10.1080/21681376.2015.1109473

How far do internal migrants really move? Demonstrating a new method for the estimation of intra-zonal distance

John Stillwell[a*] and Michael Thomas[b]

[a]*School of Geography, University of Leeds, Leeds, UK;* [b]*Faculty of Spatial Sciences, University of Groningen, Groningen, the Netherlands*

(Received 29 April 2015; accepted 13 October 2015)

This paper makes use of data on origin to destination migration by postcode derived from a large biannual consumer survey, Acxiom's Research Opinion Poll, to measure distances of migration within England in the mid-2000s. These data provide an opportunity to evaluate conventional methods of centroid-based estimation of inter-zonal distance and area-based estimation of intra-zonal distance, and demonstrate how the latter in particular generates problematic estimates. A new regression-based approach for generating intra-zonal distance is presented, resulting in significantly improved goodness-of-fit statistics when used in doubly constrained spatial interaction modelling of migration flows for local authority districts in England from the 2001 and 2011 Censuses.

Keywords: internal migration; distance; postcodes; centroids; England; spatial interaction model

Introduction

The results of the 2011 Census, released as special migration statistics (SMS) in 2014, inform us that 11% of the population of England changed their place of usual residence in the 12 months before the day the census was taken. The figures also reveal that around two-fifths of those migrants relocated to another local authority district (LAD), whilst the majority moved to another address within the same district.[1] Whilst these statistics, like those from previous censuses, are indicative of a well-known axiom that migration declines with distance, the data available from the census do not provide a clear measure of how far people move when they relocate. Data on origin–destination flows are released at particular spatial scales (output areas, wards and local authority areas) and migration distances have to be computed from the zone polygons at each scale. The conventional means of measuring inter-zonal distance is to estimate the distance between zone centroids; the problem is more difficult for flows within zones and various alternatives have been suggested, derived from geometric properties of the zones concerned. Given the fundamental nature of this matter for the analysis and modelling of migration, the focus of this paper is to make use of an alternative source of survey data, Acxiom Ltd's Research Opinion Poll (ROP), in order to provide a more accurate measure of the mean distance that people travel in England when they move from a previous place of usual residence to a new location.

*Corresponding author. Email: j.c.h.stillwell@leeds.ac.uk

Thus, the aim of the paper is to employ the detailed origin and destination geo-identifiers held within the ROP in order to provide insights into the validity of conventional measurements of distance when modelling directional migration flows, including the vexed question of how to deal with intra-zonal distance. The first objective is to make comparisons of distance migrated by persons between (and within) zones using the postcode-to-postcode measures of distance derived from the Acxiom data with estimates of distance based on centroid-to-centroid (and area-based) distances for a series of zones at different spatial scales. The implicit assumption is that the former represent 'observed' data based on the precise locations of the origin and destination locations of the individuals migrating. The results of this analysis suggest that area-based estimates of intra-zonal distances are the most problematic, and therefore the second objective of the paper is to apply a new method of estimating intra-area distance based on the areas of the zone polygons concerned, and to test this method by assessing improvement in the goodness-of-fit of spatial interaction models calibrated using migration data from the two most recent censuses and distance measures estimated using the new methodology.

The paper is organized into five further sections, beginning with a review of previous work on the relationship between migration and distance by those attempting to model migration flows, the reasons why people move over different distances and ways in which distance has been measured for use in statistical or mathematical models. The third section introduces the survey data source used in the analysis as well as the systems of spatial units in England for which the sample of survey data on migration has been assembled. This section also identifies issues associated with the methodology adopted, for which results are presented in the fourth section, concentrating on a comparison of 'observed' postcode-based distances with centroid- and area-based distance estimates for various spatial systems of interest. The fifth section uses the relationship between mean postcode-based distance and mean zone area at different spatial scales to generate alternative estimates of intra-zonal distance for LADs in England. The resulting distance estimates are then used in combination with migration flows from each of the last two censuses to calibrate spatial interaction models. Various goodness-of-fit statistics are used to assess the extent to which the interaction models based on the new measures of distance generate better model fits to the observed flow data than those based on estimates of distance using existing approaches. Finally, some conclusions, and the more general implications of this work, are presented in the last section.

Previous studies of migration distance

Distance has been a primary and longstanding focus for analysis of a wide range of spatial interaction phenomena, including residential relocation through migration. People move from one 'place of usual residence' to another either as individuals or with partners, as families or in groups, over varying distances, for a particular reason or for a combination of different reasons. Life-course events, such as going to university or getting married or changing job, are frequently the triggers for changing address, but the gamut of migration determinants is very wide-ranging (Champion, Fotheringham, Rees, Boyle, & Stillwell, 1998) and distance from current place of residence is frequently one of the key considerations involved in the selection of a suitable destination. Indeed, one of the fundamental 'laws' of migration behaviour is that the volume or intensity of migration declines as distance increases. In the 19th century, Ravenstein (1885, p. 198), in his analysis of migration flows between counties in Great Britain and Ireland, concluded that 'the great body of our migrants only proceed a short distance'. The deterrence effect of distance on migration flows has been identified by many researchers in different parts of the world – including, for example, relatively early studies by Makower, Marschak, and Robinson (1938, 1939) in Great Britain, Olsson (1965) in Sweden, Sahota (1968) in Brazil, Beals, Levy, and Moses (1967) in Ghana, and Rose (1958), Gallaway (1967), Rogers (1967), and Schwartz (1968, 1973) in the USA Somewhat more recently, Long, Tucker, and Urton (1988a, 1988b) and Bell et al. (2002) have reported an international comparisons of migration distance.

Modelling migration and distance

The relationship between migration and distance has been incorporated into an array of formal models, commencing with the early gravity models, whose origin dates back to the 17th century when Isaac Newton formulated his Law of Universal Gravitation. Whilst Newton's theory was developed and articulated in the context of physics and astronomy, the principles have been used to underpin an ongoing stream of activity that has been termed 'social physics' (Stewart, 1950). Indeed, applications of gravity modelling can be found across a range of social and regional science disciplines, including population geography and spatial demography, where spatial interaction has been a predominant focus of research on many forms of behaviour, including that of population migration and mobility.

Beyond the very early theoretical contribution of Ravenstein (1885), and the more formal algebraic explanations of migratory distance decay by Zipf (1946) and Stewart (1948), one of the most important developments of the basic gravity model was that offered by Lowry (1966). Lowry's approach to the analysis of migration flows between Standard Metropolitan Statistical Areas (SMSA) in the USA involved the calibration of a model based on a function of vectors of explanatory variables which, when combined, aimed to describe the origin (push factors) and the destination (pull factors) as well as the basic measurement of gravity (distance). In the same year that Lowry's *Migration and Metropolitan Growth* (1966) was published, Lee published his theory of migration. The theory presented a scenario in which push and pull variables were conceived as sets of positive and negative factors associated with the areas of origin and destination, together with personal factors and 'intervening obstacles' – where the latter contained certain key 'impediments' which include the physical distance and the cost of transporting household goods (Lee, 1966, p. 51). Four years later, Tobler further extended the gravitational principle in the social sciences when he established the First Law of Geography in which 'everything is related to everything else, but near things are more related than distant things' (Tobler, 1970, p. 236). At the same time in the UK, Wilson began to produce a series of publications on the development of models akin to gravity models but derived using entropy-maximizing methods and called spatial interaction models (Wilson, 1970). Whilst spatial interaction models were subsequently applied in a range of contexts, including, for instance, shopping and transport, various mathematical formulations were developed in the field of migration by Stillwell (1978) and Fotheringham (1983), whilst statistical calibrations of spatial interaction models, using Poisson regression, were undertaken by Flowerdew and Aitken (1982), Congdon (1991), Flowerdew (1991, 2010), and, most recently, Champion, Bramley, Fotheringham, MacGill, and Rees (2003).

Reasons for distance decay

There are a number of reasons why most people move only short distances and why migration flows decline with distance from the point of origin. In many cases, migrants would prefer to remain in relative close proximity to family, friends and social contacts and not lose the benefits of community facilities and social capital they have built up, possibly over a considerable period of time. Yet, this is not always so, as with individuals reaching the age when they desire to leave the parental home for university, a life-course event in the UK that has traditionally encouraged long-distance migration (Duke-Williams, 2009). Alternative theories of distance decay in migration have been attributed to the decline from the origin in knowledge and information about potential destination locations (Ritchey, 1976) as well as the financial costs involved in the move itself, although the actual transport costs are likely to be relatively small vis-à-vis the transaction costs if properties are being bought and sold, and the other costs incurred when a removal company is considered.

As indicated by the example of students, the frictional effect of distance also varies according to the reasons for moving and therefore the propensities to move over distance are likely to vary according to the various demographic, economic and social characteristics of individual migrants as well as their households. For instance, Niedomysl (2011) used a large-scale survey in Sweden to explore how migration motives change over migration distance and these results confirm a well-known generalization dichotomy between those so-called 'home-based' migrants who move shorter distances for housing reasons and whose relocation does not involve a change of employment and those 'job-based' migrants who move longer distances and whose relocation involves a change of job as well as a change of residence. Whilst there is a temptation to classify the former as residential mobility and the latter as internal migration, there is no single cut-off point for these two types of mobility; short-distance migration will contain a small proportion of job-based migrants alongside a high proportion of home-based migrants with the proportions reversing as distance gets longer. Thus a comprehensive model of migration should include moves over all distances rather than just between administrative areas, particularly since the majority of movement takes place over relatively short distances within areas demarcated for local governance such as LADs in the UK.

Measuring migration distance

One of the key difficulties associated with the analysis of the relationships between migration, motivation and distance is the lack of data available in many countries. This applies to the collection of basic origin–destination migration statistics, let alone migration data cross-tabulated by reason for move and/or distance travelled. As identified by Bell et al. (2014) when constructing the IMAGE inventory of internal migration data across the world, in many countries the only data available are counts of flows between areal units at one or more spatial scales used for administration or statistical purposes.

Yet, researchers of internal migration in the UK, whilst not having access to the type of individual records available in a country like Sweden,[2] do have reasonably detailed statistics on origin–destination migration flows within and between spatial units at different spatial scales. Such data are derived from the decennial population census, which asks the question: Where was your place of usual residence 12 months ago? These data have been used for analyses of spatial patterns and intensities of migration over the last five decades, most recently by Dennett and Stillwell (2010), Champion and Coombes (2010) and Fielding (2012), for example, together with estimates of migration derived from the National Health Service Central Register (NHSCR) and the Patient Register Data Service (PRDS). One shortcoming of the data from the latter administrative sources is that the spatial scale at which data are released is that of health authority districts (Stillwell, Rees, & Boden, 1991), and LADs (Lomax, Stillwell, Norman, & Rees, 2014) more recently, both systems which involve relatively large spatial units and crucially lack estimates of intra-zonal flows. One of the most frustrating problems for migration researchers is the lack of detailed information about the exact points of origin and destination of migrants. Whilst the postcodes of each individual at the census and 12 months previously are collected by the national census offices, confidentiality constraints prevent the release of this detailed information (for 100 years). However, the origin–destination SMS data are released for a range of spatial scales including, since 2001, LADs, census wards and output areas. As such, the flows between and within these units can be aggregated to higher-level systems of spatial units, yet the abiding question remains as to how migration distance should be measured in these circumstances. Finding the average migration distance between zones can be far from straightforward.

The measurement of migration distance in these circumstances is tied closely to the size and shape of the areas since the simplest measure of distance is that between the centres of gravity or the centroids of the spatial units concerned. The zone centroids may be computed using a geographical information system (GIS) as geometric points or may be interpolated as population-weighted centroids. This approach has typified the use of distance in inter-zonal migration modelling with the centres of large/capital cities being used as centroids in some cases when the spatial units are large, with distances being measured typically as Euclidian distances between the origin and destination centroids. Whilst distances can be derived from origin and destination grid coordinates using Pythagoras' theorem in an automated manner, they are a huge simplification of the real distance involved, particularly if the two zones are separated by one or more physical barriers. More accurate distances between origin and destination centroids can be obtained using GIS tools and data on the road and rail transport infrastructure, taking into account physical obstacles such as mountains, rivers, estuaries and lakes. Airline distances were used by Lowry (1966) when modelling inter-state migration in the USA and alternatives such as social distance, time distance and travel cost measures have also been suggested, as well as Manhattan distances (the distance that would be travelled to get from the origin to the destination if a grid-like path is followed). A more radical measure of distance is that of intervening opportunities (Stouffer, 1940, 1960), whilst distance was conceptualized as intervening obstacles by Lee (1966) as discussed above. Weeden (1973) replaced the origin–destination distance term in his model with a contiguity dummy variable designed to pick up the shorter-distance 'home-based' flows that just happen to cross arbitrarily defined census boundaries.

Centroid-to-centroid distances, though less than perfect, can be used for inter-zonal flow modelling, but when intra-zonal flows are included, it becomes necessary to estimate intra-zonal distances. Several approaches for estimating intra-zonal distance exist such as half the drive time between the zone centroid and the centroids of all neighbouring zones (US Department of Commerce, 1965) or half the distance from the zone to the nearest centroid (Venigalla, Chatterjee, & Bronzini, 1999). Alternative options suggested by Batty (1976) and Fotheringham (1988) are based on the assumption that zones are circular in shape and that population is spread evenly across the area or varies systematically away from the centroid. In the simplest form, Batty (1976) suggests that the intra-zonal distance for zone i (d_{ii}) is computed as:

$$d_{ii} = \frac{r_i}{\sqrt{2}} \tag{1}$$

where r_i is the radius of a circle equivalent in area to zone i. The radius can be defined as $SQRT(A_i/\pi)$, where A_i is the area of zone i. We can regard Equation (1) as being the conventional measure of intra-zonal distance, although Fotheringham (1988) proposes an alternative formulation:

$$d_{ii} = \left(0.846\left(\frac{1.432}{0.846}\right)^{\frac{z}{r_i}}\right)r_i \tag{2}$$

where z is the distance from a destination point and the zone centroid; and the coefficients .846 and 1.432 are related to the potential minimum and maximum distances in a circular zone around the origin. Alternatively, Rogerson (1990) has suggested estimating migration distances in the USA from data on the proportion of migrants crossing regional boundaries following a procedure that makes use of Buffon's needle, a problem in geometrical probability from the 18th century. More recently, Kordi, Kaiser, and Fotheringham (2012) have developed an approach to the problem by scattering the origins and destinations of individuals either randomly or according to a given population density probability derived from a high-resolution density probability surface. They estimate the mean distance of travel using a fine regular grid and compute a weighted average between all possible pairs of grid points. This approach can be applied to either inter- or intra-zone distances, and Kordi et al. report the results of calibrating Poisson regression models for commuting flows in Lausanne using four datasets: (1) excluding the intra-zonal flows and using traditional centroid–centroid distances for inter-zonal flows; (2) using the circular-based intra-zone distances and inter-centroid distances; (3) using intra- and inter-zone distances based on randomly scattered points of origin and destination; and (4) using intra- and inter-zone distances based on density-based scattering of origin and destination points. Their results show a very good fit ($R^2 > .96$) against the observed data in each case, but the best fit is achieved with the density-based scattering model that includes intra-zonal flows and distances.

Not only is distance difficult to measure when statistical agencies only release data on migration for census or administrative areas, but also there is an accompanying major problem that arises due to the different size and shape of the spatial units at different scales with and between which movement is recorded. This issue, identified by Ravenstein (1885), is particularly problematic when we come to compare migration intensities or patterns in different places because the distances of moving within areas may vary considerably. The modifiable areal unit problem (MAUP) has two components: the scale component, which is associated with the number of spatial units into which the national territory is divided; and the zonation component, that is, the different boundary configurations of the areas at any one spatial scale. One response to the MAUP challenge is to compute measures of distance at a whole range of scales and area configurations and thus establish how the scale and zonation effects change as geographies change (Stillwell, Daras, Bell, & Lomax, 2014).

This review of previous studies has highlighted the fact that, whatever model calibration method is used, distance is a key variable in migration analysis and, as such, it often has to be measured in some way. Whilst the measurement of physical distance using zone centroids proves relatively unproblematic, at least computationally, for inter-zonal migration, it is more challenging for intra-zonal migration since there is only a single geographical centroid location for both the migration origin and destination. As noted above, ignoring the problem and calibrating the model using just inter-area flows reduces the sample size and does not capture internal migration behaviour in its entirety. It would also seem inappropriate to assume an intra-zone distance of zero, since the value is always going to be positive (Fotheringham, 1988) and may in fact be higher than some inter-centroid distances, depending on the size and shape of the area and the distribution of migrants within it. We will return to the discussion of intra-zonal distance after the next section in which the data and spatial units used in our analysis are introduced.

Data and methodology

Migration data from Acxiom Ltd

The source of data used for most of the analysis reported in this paper is the Research Opinion Poll (ROP) undertaken biannually until 2014 by Acxiom Ltd, a marketing technology and services company perhaps best known for its market segmentation product PersonicX. The ROP is a voluntary and principally paper-based postal lifestyle survey of individual household respondents aged 18 and over in Great Britain whose primary aim is to collect detailed and up-to-date information about consumer spending habits, preferences, socio-demographic information and the locations of respondents. The survey at large asks about 130 questions, allowing for over 1000 possible answers covering 26 broad topics such as groceries, shopping, local area, environment, outgoings, occupation, home, leisure, education and health. This information provides very useful insights into the spending patterns of different types of people and geographical areas on an annual basis. Details of the survey design, questionnaire structure, multi-channel collection methods and data processing are reviewed in Thompson, Stillwell, Clarke, and Bradbrook (2010). A number of address sources are used to ensure that the ROP generates a geographically representative response of georeferenced individuals. On average, Acxiom have received around 1 million responses from the ROP each year, making it the largest annual paper-based survey in Great Britain and the largest population study outside the Census of Population; in the 2009 ROP, for example, 'only .4% of all the Middle Super Output Areas (MSOAs) across Great Britain did not return a response' (Thompson et al., 2010, p. 13). In reviewing the ROP, Thompson et al. (2010) indicate that, when compared against government datasets including the 2001 Census, the Expenditure and Food Survey, the Labour Force Survey, the British Household Panel Survey, the General Household Survey and the Survey for English Housing, it scores favourably on dimensions such as sample size, geographic detail, consistency, and the quality and accuracy of the data.

A comprehensive description of the ROP as a source of data for internal migration analysis is available in Thomas, Gould, and Stillwell (2012), indicating that for the years 2005, 2006 and 2007, the ROP included the following two sponsored questions about previous migration: 'When did you move to this address? (month and year)' and 'Please tell us the house number and postcode of your previous address'. When the results of these questions are combined with the respondents' current address, the unique potential of this data source for analysis of migration becomes evident. Validation work by Thomas et al. (2012) has shown that although migrants are under-represented, descriptive benchmarking of the aggregated migrant flows from the ROP against the 2001 Census produced reassuringly similar results. Statistical comparison of in-migrants to LADs from the 2005 ROP and the 2001 Census yielded a Pearson correlation coefficient of .73, for example. In addition, at the micro-level, multivariate regression-based approaches have been used for validation (Thomas, 2014). Indeed, the results of a comparison of coefficients' effect size and direction (for various individual and household characteristics) on movement propensities, in weighted and unweighted models based on ROP micro-data as well as like-for-like comparisons with models using the 2001 Census Individual Licensed Sample of Anonymised Records (I-SAR, 3% sample), are particularly encouraging in showing the relative robustness of model-based results derived from the ROP for micro-level migration analysis (Thomas, Stillwell, & Gould, 2014).

The analysis presented here makes use of a subsample of the ROP data for England, derived from three tranches covering the period January 2005 to January 2007. The full initial ROP sample for Great Britain for January 2005, January 2006 and January 2007 included 950,658 records (Thomas et al., 2014) of individuals who either moved or stayed at the same address. Cleaning procedures for this initial sample are reported in Thomas et al. (2012), where strenuous efforts were made to validate and correct previous postcode addresses through cross-referencing with the Office for National Statistics (ONS) Postcode Directory. For this analysis a series of specific data conditions were required in order to identify migrants from non-migrants; these conditions led to the removal of a considerable number of records from the initial full sample. The removal included 53,919 records of individuals who indicated that they had moved, but for whom the precise origin location and distance moved could not be identified. For example, some respondents provided the same full postcode at origin and destination. There is a possibility that these individuals are moving within the same full postcode; however, the 2011 Census for England and Wales recorded an average of just 18 households per postcode (ONS, 2015), which suggests the high number of recorded intra-postcode moves in our sample is likely to be a reflection of the misreporting of postcodes rather than genuinely high rates of within-postcode spatial mobility. Furthermore, individuals who provided no previous postcode address, or an incomplete postcode address, were also removed due to the impossibility of calculating their distances moved. In addition to the individual response issues, records with an origin and/or destination in Scotland or Wales are removed due to sample scarcity – the ROP provided very few responses in particular regions of the two nations which, if not removed, would result in serious small number issues that can lead to wholly unreliable estimates of average migration distances, particularly at the smaller geographical scales. Beyond this, those who provide no prior address or year of move are considered to be non-migrants and therefore removed.

The final analytical sample is designed purely for the measurement of distance separating migrant origins from destinations. Thus, the measure of distance is based on the current and previous address location, regardless of individual/household characteristics and how long ago they moved. Following the selection conditions just described, the resulting sample of individuals who migrated is 125,494. The benefits of having detailed origin and destination postcode identifiers are twofold. Firstly, we can define origin and destination areas with far more precision than is allowed for in alternative sources such as the Census I-SAR (where only a regional geography is provided at the origin) or the Census SMS, where we only have counts of aggregate flows between spatial units such as LADs, wards and output areas, and would need to use the origin and destination centroids of these zones as the points of departure and arrival. Secondly, the ROP data allow us to measure continuous distance directly between the northing and easting coordinates (grid references) of the origin and destination of each migrant journey.

Methods

The first analysis seeks to evaluate the relative accuracy of conventional estimates of distance, as compared with real distances measured using postcode grid references from the ROP, for different zonal aggregations. To this end, postcodes are converted into national grid coordinates to give six digit grid references for both the origin and destination. The Euclidian distance between each pair of points is the hypotenuse of a right-angled triangle whose value is the square root of the sum of the squares of the other two sides. We can use Pythagoras' method to compute distances for each individual migration.

Five different sets of 'zones', a generic term used to refer to a spatial unit at any scale, are used in our analysis, where individual migrants are aggregated into flows within and between zones at each scale. The zones covering England include: (1) nine regions; (2) 27

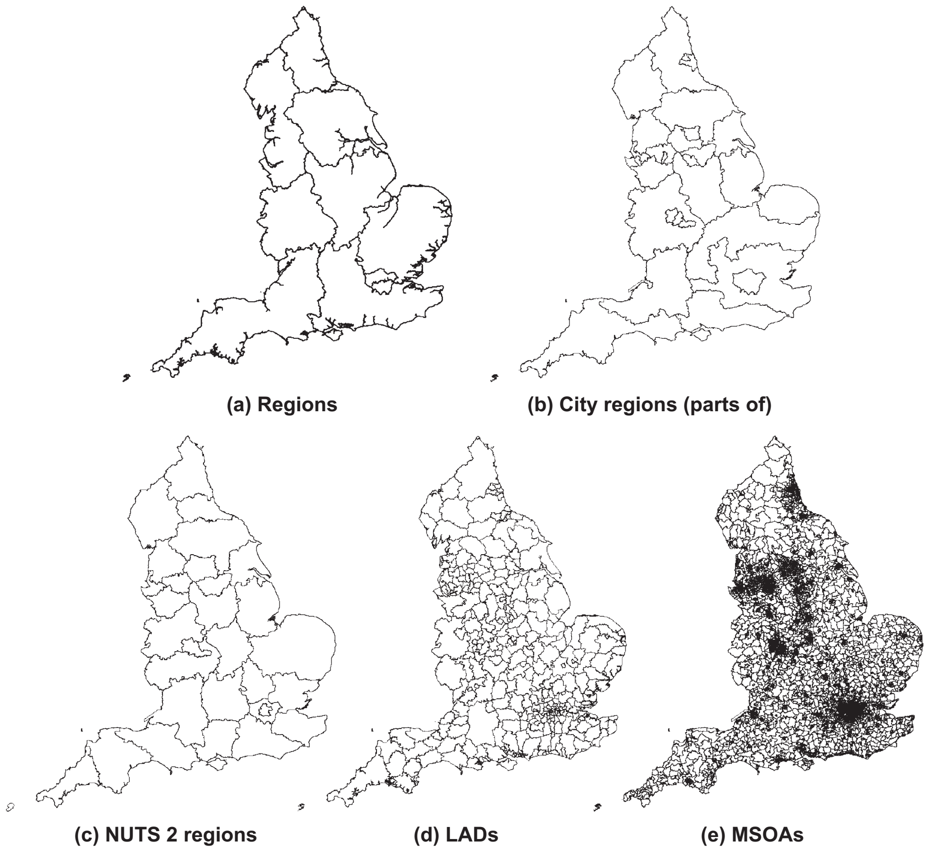

Figure 1. Zone systems in England used for reporting results.

city-region component areas; (3) 30 NUTS-2 regions; (4) 324 LADs; and (5) 6793 middle layer super output areas (MSOAs) (Figure 1). More detailed spatial units, below the MSOA level, are available from the UK Census (e.g. output areas (n = 171,372) and lower layer super output areas (n = 32,844)); however, their considerable detail would lead to serious sample sparsity and associated small number problems.

The mean zone populations and flows between and within zones at each spatial scale are shown in Table 1 and all five geographies are used in the comparison of distance measures. LADs are used use for spatial interaction modelling in the final section of the results.

The analysis of aggregated data for different zone systems involves comparison between the postcode-to-postcode distances with distances computed between and within zones using conventional methods: inter-zonal distances based on inter-centroid computations and intra-zonal distances using the radius of a circle of equivalent area divided by the square root of two, following Batty (1976). Once distances for individual migrations have been computed, the following statistical measures of average and variation are computed: (1) mean and standard deviation (SD); and (2) median and mean average deviation (MAD).

The mean migration distance (MMD) for all migrants in the system is computed as:

$$MMD = \frac{\sum_i \sum_j M_{ij} d_{ij}}{\sum_i \sum_j M_{ij}} \tag{3}$$

Table 1. Mean zone population size and migrants within and between zones at different spatial scales.

Set of zones	Mean zone population[a]	Intra-zone flows	Intra-zone %[b]	Inter-zone flows	Inter-zone %[b]
a. Regions	89,124.7	108,270	86	17,224	14
b. City-regions	29,708.2	101,008	80	24,486	20
c. NUTS-2 regions	26,737.4	100,951	80	24,543	20
d. LADs	22,65.91	83,220	66	42,274	34
e. MSOAs	118.3	30,564	24	94,930	76

[a]Based on a total Research Opinion Poll (ROP) sample for England (movers and non-movers).
[b]Percentage of total (intra- and inter-) moves.

where M_{ij} is the count of migrants from zone i to zone j; and d_{ij} is the distance between zone i and zone j. Measures of mean in- and out-migration distance can be computed. The *SD* associated with the mean is defined as:

$$SD = SQRT\left(\frac{\left(\sum_i \sum_j \left(M_{ij}d_{ij} - MMD\right)^2\right)}{\sum_i \sum_j M_{ij}d_{ij} - 1}\right) \tag{4}$$

The median distance moved is a simpler summary indicator for comparing the distance dimension of internal migration and is often considered preferable to the mean as the distribution of distances is negatively skewed, reflecting the strong distance-decay effect which consistently occurs with migration. The median can be calculated simply by locating the midpoint of a cumulative frequency distribution of individuals, sorted according to the distance moved and the MAD statistic is the median of the absolute deviation from the median value.

The approach in the next section is to compare the means and medians based on postcode–postcode distance of migration against centroid–centroid distances at different spatial scales defined above. Thereafter, ordinary least squares (OLS) bivariate regression analysis is used to quantify the relationship between mean migration distance and mean zone size (area) at different spatial scales; we derive linear and non-linear equations and use the regression relationships to estimate intra-zone (and inter-zone) distances given the areas of the districts. Finally, in the fifth section, we report the goodness-of-fit of doubly constrained spatial interaction models that are calibrated with different measurements of intra-zonal distance using three LAD migration matrices. The general form of the doubly constrained model for the flow of migrants between zone i and zone j, M_{ij}, is:

$$M_{ij} = A_i B_j O_i D_j f\left(d_{ij}\right) \tag{5}$$

where A_i is a balancing factor that ensures the sum of flows from origin i is equivalent to the sum of flows into all j zones from zone i; B_j is a balancing factor that ensures the sum of flows to destination zone j is equivalent to the sum of flows out of all zones to zone j; O_i is the total out-migration flows from zone i; D_j is the total in-migration flows to zone j; and $f\left(d_{ij}\right)$ is a distance-decay function that can be expressed in either linear or exponential form.

Given that a symmetric matrix of migration flows is available, since the number and order of origins and destinations are the same, the model can be calibrated either including or excluding intra-zonal flows.

A comparison of actual distances and centroid–centroid distances

National comparison

The aim of the analysis reported in this section is to compare the average inter-zonal distances of migration as measured by the postcode-to-postcode distances derived from the ROP with (1) the centroid-to-centroid distances for inter-zonal flows and (2) the intra-area distances measured using the conventional method as indicated in Equation (1). This comparison is reported at each of the five scales for which the numbers of residential relocations between and within zones are shown in Table 1. At the region scale, only 14% of migrants cross a zone boundary, whereas at the MSOA end of the scale, three-quarters of all migrants move between zones. The NUTS-2 and city-region systems are roughly equal in terms of numbers of zones but divide up the national territory in different ways and therefore provide a useful comparison.

Average migration distances, together with their measures of variation, are reported for the five spatial scales in Table 2, which is divided into three panels, each of which has a row indicating the mean and median for the actual postcode-to-postcode distance (Postcode) and a second row for the distance estimated from the area centroids (Centroid) based on Equation (1). Corresponding SD and MAD statistics are included in parentheses to indicate the variation around each measure of central tendency.

The top panel shows the results for inter-zonal flows, illustrating that differences in the mean and median distances between the two calculation methods are relatively small when large numbers of zones are involved but gradually increase as the number of zones gets smaller. At the region scale, the difference in the two mean values is over 22 km whilst the difference in the medians is over 50 km, and the median actually exceeds the mean when migration distances are much longer. The variation around the mean and median decreases significantly as the zone size gets smaller so that at the MSOA scale, the inter-centroid mean is less than 0.5 km more than the postcode mean. Although there are three more NUTS-2 spatial units than city-region areas, both the mean and median distance values are slightly higher for the former than the latter. The variation around the mean and median values is of a similar order of magnitude at each spatial scale.

The middle panel in Table 2 contains the same statistical measures but in this case for distances within zones at each spatial scale. As expected both means and medians increase as the number of zones gets smaller, but the differences in the two methods of calculation result in much higher values computed using the conventional formula than the actual distances measured between postcodes of origin and destination. Even at MSOA level, the mean intra-zonal distance computed using Equation (1) is twice the postcode-to-postcode distance and, at the region scale, the inter-centroid mean is over 47 km compared with the postcode–postcode mean which is only 7 km. Differences in distances measured using the two methods for all moves, referred to as the combined distance, are shown in the bottom panel of Table 2. In this case, the postcode distance remains the same (mean of 26 km, median of 3 km) at all spatial scales, while the centroid distance increases from a mean of 26 km at MSOA level to 63 km for regions. These results point to the inadequacy of the calculation of intra-zonal distances using the conventional formula indicated in Equation (1).

Table 2. Comparisons of inter-zonal, intra-zonal and combined migration distance at five spatial scales.

	MSOA		LAD		NUTS-2		City-region (parts of)		Region	
	Mean (SD)	Median (MAD)	Mean (SD)	Median (MAD)	Mean (SD)	Median (MAD)	Mean (SD)	Median (MAD)	Mean (SD)	Median (MAD)
	Inter-zonal distance (km)									
Postcode	33.81 (71.65)	5.00 (5.67)	71.63 (94.52)	23.93 (29.09)	113.10 (105.52)	83.14 (98.39)	110.30 (106.64)	77.32 (94.75)	142.5 (109.85)	123.20 (118.89)
Centroid	34.26 (71.46)	5.67 (5.94)	74.39 (92.41)	28.61 (28.50)	122.5 (96.10)	83.87 (56.05)	111.30 (103.91)	68.82 (80.13)	164.00 (92.06)	175.00 (95.41)
	Intra-zonal distance (km)									
Postcode	0.78 (1.13)	0.46 (0.42)	2.47 (3.14)	1.39 (1.50)	4.53 (8.04)	1.88 (2.18)	5.29 (12.95)	1.89 (2.19)	7.19 (17.34)	2.14 (2.54)
Centroid	1.47 (1.44)	0.86 (0.50)	6.54 (3.51)	5.70 (3.16)	24.64 (8.94)	24.94 (9.61)	32.73 (16.98)	34.18 (20.07)	47.29 (11.85)	49.53 (7.45)
	Combined distance (km)									
Postcode	25.77 (63.91)	2.89 (3.65)	25.77 (63.91)	2.89 (3.65)	25.77 (63.91)	2.89 (3.65)	25.77 (63.91)	2.89 (3.65)	25.77 (63.91)	2.89 (3.65)
Centroid	26.27 (63.73)	3.60 (4.00)	29.40 (62.55)	7.92 (6.03)	43.78 (58.12)	27.98 (12.48)	48.07 (57.52)	35.96 (29.70)	63.30 (53.81)	49.87 (6.94)

Cumulative frequencies

A second comparison of the two measures of distance involves plotting the cumulative frequency of migrants according to distance at the different spatial scales. This visual approach is a useful way of assessing how good the centroid based measures are for estimating inter- and intra-zonal distances; where postcode distance is taken as the 'observed' or real distance. In Figure 2, cumulative frequencies at four of the five scales are presented for inter-zonal migration. Results for the NUTS-2 scale are excluded because they bear strong similarity with those for the city-regions (parts of). Whilst there is a significant difference in the two distributions at region level, the two measures generate distributions that become increasingly similar as scale decreases and the volume of migration increases, that is, where more and more migrants move inter-zonally as the number of zones increases. These results suggest that use of centroids to measure inter-zonal migration distance is certainly legitimate at MSOA level, providing a pretty good fit against the actual distance measured between postcodes at MSOA level. For larger zones, however, the cumulative frequencies for the two measures diverge; whilst the postcode distance frequency graph remains a smooth curve, the centroid distance frequencies follow a stepped function. Thus, for example, at the regional scale the centroid-distance distribution does not capture any inter-zonal migrants below 36 km and then includes 2451 travelling between 37 and 94 km, at which point the curve increases to 4615 migrants and remains at this level until 98 km. The cumulative effect of this stepped distribution is an underestimate of distance for inter-zonal migration up until the distributions align at around 400 km. This stepped distribution comes from the fact that, when measured using centroids, all migrants within the specific origin–destination flows are recorded to move the exact same distance.

The stepped function is much more pronounced when the cumulative frequency distributions of intra-zonal migration distance are plotted (Figure 3). In comparison with the postcode-distance measure at the region scale, which suggests 52,165 migrants move 0–2 km and 97,160 move up to 16 km, the centroid-based measure captures zero migrants

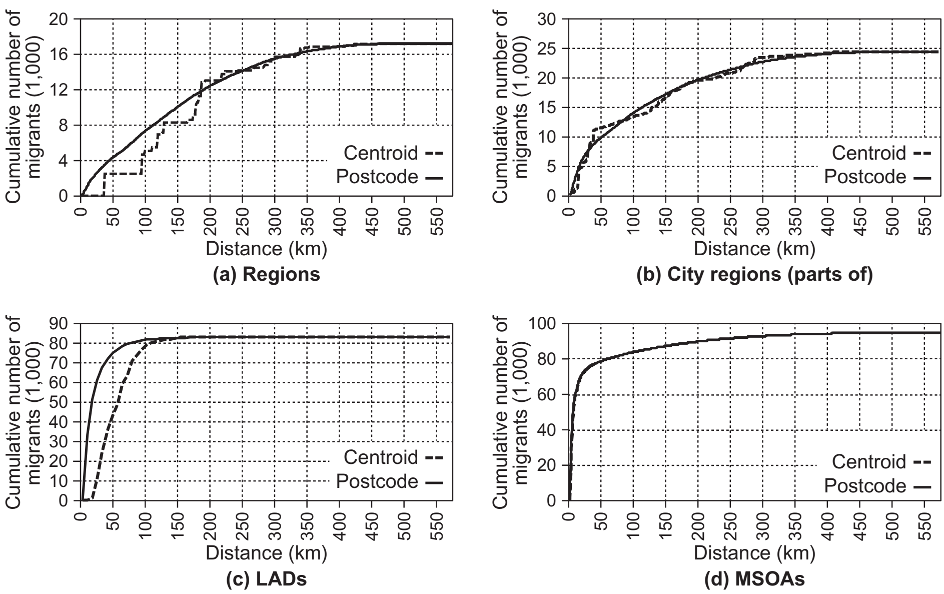

Figure 2. Cumulative frequency of the number of inter-zone migrants by scale.

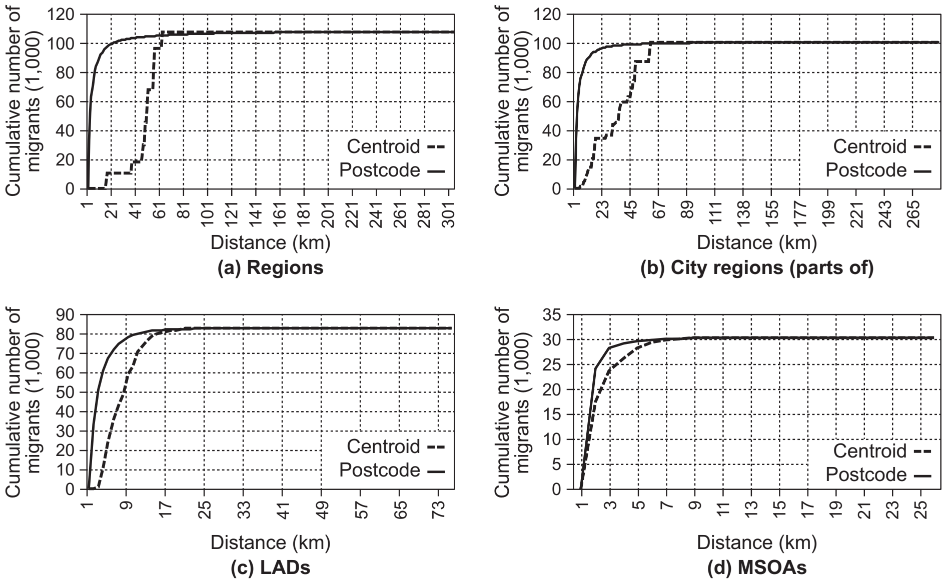

Figure 3. Cumulative frequency of the number of intra-zone migrants by scale.

up to just under 16 km and then just 10,405 migrants moving up to a threshold of 36 km. Again, the stepped function of the centroid-based distribution relates to the fact that the first group of recorded migrants are migrants moving within the smallest spatial region of England, London (where intra-moves are collectively estimated to travel 15.8 km), then at 37 km the distribution jumps to 18,375 cumulative individuals, when the intra-zone migrants in the next smallest region, the North East (where intra-movers are estimated to travel 36.98 km), are added. This same stepped pattern occurs right up to the final estimate for the largest region, the South West (where intra-regional movers are estimated to have travelled 61.6 km), at which point the full cumulative frequency of migrants (n = 108,270) is recorded, and an alignment with the postcode cumulative frequency curve is achieved. As with the inter-zone distributions, the discrepancies between the two intra-zone distributions reduces as the spatial scale gets smaller and the number of zones gets larger.

Taken together, these results indicate that the use of centroids for the measurement of distance, whilst proving suitable for inter-zone migration at the LAD scale and for smaller areas, is much less appropriate for intra-zone migration. This conclusion begs the question of whether it is possible to derive an alternative and indeed more accurate measure of intra-zone distance, particularly for larger areas.

Estimating intra-zonal distance for migration modelling: a new approach

This section proposes a new method of computing intra-zone distance that makes use of area size. Once defined, the method is subsequently used to estimate distances within LADs in England. In the final subsection, a series of doubly constrained spatial interaction models are calibrated with alternative distance functions and measures to evaluate goodness-of-fit.

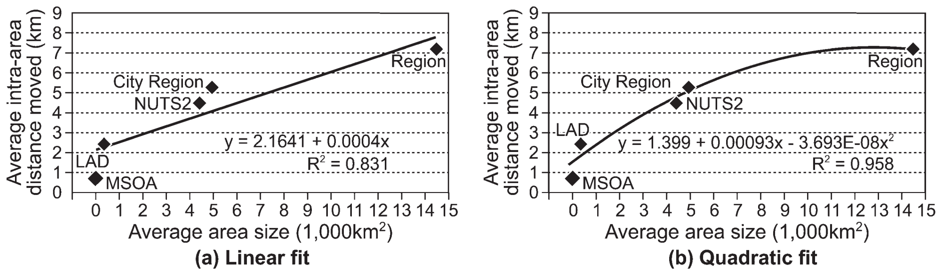

Figure 4. Linear and quadratic regression fits.

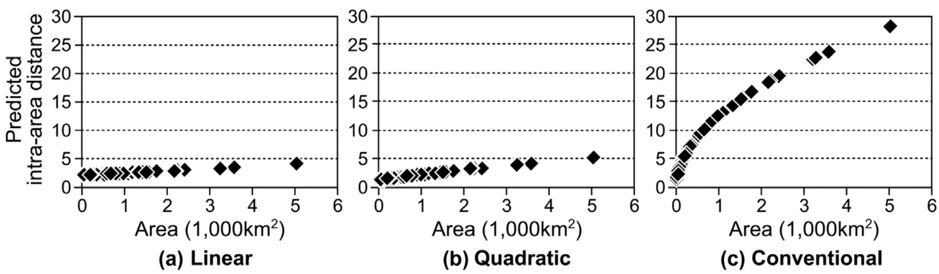

Figure 5. Comparison of predicted distance estimates using the regression and conventional methods for the 324 local authority districts (LADs) in England.

Comparison of within zone distance and area size

The assumption underpinning the proposed method of calculating intra-zone distance is that this distance is based on the size of an area (measured in km^2). The relationship between the two variables was computed from values of the mean intra-zonal distance and area at each of the five spatial scales. The regression parameters and fit statistics of linear and quadratic regression models are reported in Table 3 and the lines of best fit are illustrated in Figure 4. The results indicate that the quadratic model gives a better fit than the linear model.

Using the regression equations, intra-zone distance can be estimated for each LAD given its areal size. The graphs in Figure 5 illustrate the predicted mean intra-LAD distance moved against the corresponding LAD area in square kilometres for all LADs in England. Whilst the quadratic equation (Figure 5b) results in slightly greater variation between the shortest and longest predicted intra-LAD distances than the linear model (Figure 5a), both equations result in very similar estimates when compared with those derived from the conventional method (Figure 5c), where average intra-LAD distances are estimated on the basis of Equation (1) and predicted to be as far as 30 km.

Evaluation using spatial interaction models

In order to evaluate the impact of this approach to intra-zonal distance measurement, we present the results of calibrating doubly constrained models of migration at the LAD scale that use different measures of distance as well as either including or excluding intra-LAD flows. Two sets of migration data are used: (1) a matrix of flows from the recent 2011

Table 3. Regression parameters and fit statistics for relationships between mean inter-zonal distance and area size using linear and quadratic models.

	Linear		Quadratic	
	Estimate	Standard error	Estimate	Standard error
Intercept	2.1641	0.7220	1.399	0.7220
Area (km^2)	0.0003902*	0.0001016	0.0009342*	0.0001016
Area (km^2)2		−3.693e–08		1.51e–08
Adjusted R^2	0.775		0.915	

*Significant at the 95% level.

Table 4. Mean migration distances and decay parameters, 2000–01 and 2010–11.

Method of intra-zonal distance estimation	2000–01		2010–11	
	MMD	β	MMD	β
Conventional	37.03	2.0166	34.47	2.0358
Linear	34.29	1.5897	31.88	1.6079
Quadratic	33.97	1.5055	31.56	1.5206

Census for 324 LADs for 2010–11; and (2) a matrix of flows from the 2001 Census for an equivalent set of LADs for 2000–01. In each case, both inter- and intra-LAD migration flows are available. Three sets of distance estimates are used: (1) centroid–centroid distances for inter-LAD cells and conventional area-based distance measures for intra-district cells (diagonal), i.e. based on Equation (1); (2) centroid–centroid distances for inter-district cells and estimates based on the linear regression model for intra-LAD cells; and (3) centroid–centroid distances for inter-LAD cells and estimates based on the quadratic equation for intra-LAD cells.

A doubly constrained spatial interaction model (Equation 5) with a power-distance function was calibrated with the alternative matrices of distance estimates for the two single-year periods using the modelling subsystem of the IMAGE studio (Stillwell et al., 2014). The results presented in Table 4 indicate that the mean migration distance with the conventional method of measuring intra-zonal distance is 37 km in 2000–01 and drops to 34.5 km in 2010–11, with the decay parameter (β) that measures the frictional effect of distance on migration increasing marginally from 2.02 to 2.04 as expected. The mean migration distance decreases when the linear estimate of intra-zonal migration is included in the models for each time period and decreases further still when the quadratic estimates are used, with the decay parameters being lower in each case respectively.

Three measures of goodness-of-fit are reported in Table 5: the coefficient of determination (R^2), the index of dissimilarity (IoD) and the mean absolute percentage deviation (MAPD). In each case, the models that employ the estimates of intra-zonal distance with the linear model increase the R^2 and reduce the IoD and MAPD statistics significantly, whilst the use of the quadratic improves the model fit still further, resulting in R^2 values that approach .98 for both time periods.

Table 5. Goodness-of-fit statistics, 2000–01 and 2010–11.

Distance measure: inter-centroid distances plus	2000–01			2010–11		
	R^2	IoD	MAPD	R^2	IoD	MAPD
Conventional estimate for intra	0.9088	27.22	55.7	0.9212	25.66	51.5
Linear estimate for intra	0.9631	19.53	39.2	0.9608	18.64	37.5
Quadratic estimate for intra	0.9783	17.21	34.6	0.9777	16.58	33.3

Conclusions

Distance is one of the key variables influencing where people decide to move to. It is a variable that has been measured and incorporated into migration models in different ways but its frictional effect has been proven time after time in studies of internal migration in different countries, at different scales and for different time periods. However, in spite of its importance, the availability of precise data on how far individuals or households actually move remains a limitation for empirical analysis. It is, therefore, encouraging to discover a new source of survey data large enough to provide a representative sample of migrants between relatively small geographical areas which provides attributes relating to individuals or households that extend beyond those available from census micro-datasets. Moreover, availability of the postcode locations of both the origin and destination of the individual's migration enables a more accurate quantification of distance moved, particularly over shorter distances. This is not to say that Acxiom's ROP is not without certain limitations, for example: no data are available for those aged 17 and under; there are biases relating to certain socio-demographic characteristics; and, for a portion of the sample, the precise geographical details of the migrants' origin locations are missing. However, despite these issues, the ROP does present itself as a valuable alternative source of detailed geo-referenced data.

In this paper we have aggregated the ROP data for individuals into flows between zones for five different spatial systems. As such, this has enabled us to fulfil our first objective of comparing migration distances derived from individual postcodes of origin and destination with distances estimated between zone centroids for inter-zonal flows and using a conventional method for distances associated with intra-zonal movement. The headline result is that scale matters; as the number of zones gets smaller and the size of the zones gets larger, the difference between the two measures of distance increases, particularly for intra-zonal migration but also for inter-zonal migration. Whereas the inter-centroid distance measurement is almost identical to that of the postcode measurement at the MSOA level across the cumulative frequency distribution, the frequency distributions for intra-zonal migration at this scale show deviation and this inequality only increases as the zones increase in size.

With the results suggesting that intra-zonal distances measured in the conventional way (using the radius of a circle equivalent in area to zone i divided by the square root of two) are the most problematic, we have proposed a new method of estimating intra-zonal distance based on the regression of the mean distance of migration against the mean area of the zone polygons, using data from each of the five spatial scales. The regression analysis suggests that a quadratic model gives a better fit than a linear model. In order to test the validity of intra-zonal distance estimation using this approach, both equations have been used to generate intra-zonal estimates of distance within LADs in England and spatial interaction models have been calibrated with data from the last two censuses. The results demonstrate considerable improvements in the goodness-of-fit of the spatial interaction model to the observed data of flows between LADs from the last two censuses, particularly when the quadratic estimates of intra-LAD distance are used.

The more general implication of this work for migration modelling, when no precise information about the points of migration departure and arrival is available, other the origin and destination zones, is to consider aggregating the flow matrix and the spatial units to different levels of spatial resolution and using the relationship between migration distance and zone area at different scales for predicting the intra-zonal distances associated with the original flow matrix.

Acknowledgements

Both authors are grateful to Acxiom Ltd. for providing the ROP data and acknowledge the UK Data Service-Census Support as the source of census migration and boundary data sets.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The first author is grateful for funding support from the Economic and Social Research Council [ESRC grant ES/J02337X/1: UK Data Service-Census Support] and the Australian Research Council [ARC Discovery project DP110101363]. The second author is grateful for PhD funding from the ESRC [Grant RES-576-25-0039: TALISMAN].

Notes

1. 2011 Census data were downloaded from InFuse and WICID (UK Data Service-Census Support).
2. Where, for instance, population registers capture changes in employment alongside changes in address.

References

Batty, M. (1976). *Urban modelling: Algorithms, calibrations, predictions*. London: Cambridge University Press.

Beals, R. E., Levy, M. B., & Moses, L. N. (1967). Rationality and migration in Ghana. *The Review of Economics and Statistics, 49*, 480–486.

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

Champion, A., Bramley, G., Fotheringham, A. S., MacGill, J., & Rees, P. (2003). A migration modelling system to support government decision-making. In S. Geertman, & J. Stillwell, J. (Eds.), *Planning Support Systems in Practice* (pp. 257–278). Berlin: Springer Verlag.

Champion, A., & Coombes, M. (2010). Migration and socio-economic polarisation between British city regions. In J. Stillwell, O. Duke-Williams, & A. Dennett (Eds.), *Technologies for Migration and Commuting Analysis* (pp. 196–211). Hershey: IGI Global.

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*. Newcastle upon Tyne: Department of Geography, University of Newcastle upon Tyne.

Congdon, P. (1991). An application of general linear modelling to migration in London and the South East. In J. Stillwell, & P. Congdon (Eds.), *Migration models: macro and micro perspectives* (pp. 113–136). London: Belhaven Press.

Dennett, A., & Stillwell, J. (2010). Internal migration by age and sex at the start of the 21st century. In J. Stillwell, O. Duke-Williams, & A. Dennett (Eds.), *Technologies for migration and commuting analysis* (pp. 153–174). Hershey, PA: GI Global.
Duke-Williams, O. (2009). The geographies of student migration in the UK. *Environment and Planning A, 41*, 1826–1848.
Fielding, A. (2012). *Migration in Britain paradoxes of the present, prospects for the future*. Cheltenham: Edward Elgar.
Flowerdew, R. (1991). Poisson regression modelling of migration. In J. Stillwell & P. Congdon (Eds.), *Migration models: Macro and micro perspectives* (pp. 92–112). London: Belhaven Press.
Flowerdew, R. (2010). Modelling migration with Poisson regression. In J. Stillwell, O. Duke-Williams, & Dennett, A. (Eds.), *Technologies for migration and commuting analysis* (pp. 261–279). Hershey, PA: IGI Global.
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.
Fotheringham, A. S. (1988). Market share analysis techniques: A review and illustration of current U.S. practice. In N. Wrigley (Ed.), *Store choice, store location and market analysis*. (pp 120–159). London: Routledge.
Gallaway, L. E. (1967). Industry variations in geographic labor mobility patterns. *The Journal of Human Resources, 2*, 461–474.
Kordi, M., Kaiser, C., & Fotheringham, A. S. (2012). A possible solution for the centroid-to-centroid and intra-zonal trip length problems. In J. Gensel, D. Josselin, & D. Vandenbrouke (Eds.), *Proceedings of the AGILE'2012 International Conference on Geographic Information Science* (pp. 147–152). AGILE: Avignon.
Lee, E. S. (1966). A theory of migration. *Demography, 3*, 47–57.
Lomax, N., Stillwell, J., Norman, P., & Rees, P. (2014). Internal migration in the United Kingdom: Analysis of an estimated inter-district time series, 2001–2011. *Applied Spatial Analysis and Policy, 7*, 25–45.
Long, L. H., Tucker, C. J., & Urton, W. L. (1988a). Migration distances: An international comparison. *Demography, 25*, 633–640.
Long, L. H., Tucker, C. J., & Urton, W. L. (1988b). Measuring migration distances: Self-reporting and indirect methods. *Journal of the American Statistical Association, 83*, 674–678.
Lowry, I. S. (1966). *Migration and metropolitan growth: Two analytical models*. San Francisco: Chandler Publishing Company.
Makower, H., Marschak, J., & Robinson, H. W. (1938). Studies in mobility of labour: A tentative statistical measure. *Oxford Economic Papers, 1*, 83–123.
Makower, H., Marschak, J., & Robinson, H. W. (1939). Studies in mobility of labour: Analysis for Great Britain, part I. *Oxford Economic Papers, 2*, 70–93.
Niedomysl, T. (2011). How migration motivations change over migration distance: Evidence on variations across socioeconomic and demographic groups. *Regional Studies, 45*, 843–855.
Olsson, G. (1965). *Distance and human interaction: A review and bibliography*. Philadelphia, PA: Regional Science Research Institute.
ONS. (2015) FOI request: Households per postcode. Retrieved from http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/previous-foi-requests/people–population-and-community/households-per-postcode/index.html
Ravenstein, E. (1885). The laws of migration. *Journal of the Statistical Society, 46*, 167–235.
Ritchey, P. N. (1976). Explanations of migration. *Annual Review of Sociology, 2*, 363–404.
Rogers, A. (1967). A regression analysis of inter-regional migration in California. *The Review of Economics and Statistics, 49*, 262–267.
Rogerson, P. A. (1990). Buffon's needle and the estimation of migration distance. *Mathematical Population Studies, 2*, 229–238.
Rose, A. M. (1958). Distance of migration and socioeconomic status of migrants. *American Sociological Review, 23*, 420–423.
Sahota, G. S. (1968). An economic analysis of internal migration in Brazil. *Journal of Political Economy, 76*, 218–245.
Schwartz, A. (1968). *Migration and earning in the U.S.* (PhD Dissertation). University of Chicago.

Schwarz, A. (1973). Interpreting the effect of distance on migration. *Journal of Political Economy, 8*, 1153–1169.
Stewart, J. Q. (1948). Demographic gravitation: Evidence and applications. *Sociometry, 11*, 31–58.
Stewart, J. Q. (1950). The development of social physics. *American Journal of Physics, 18*, 239–253.
Stillwell, J. (1978). Inter-zonal migration: Some historical tests of spatial interaction models. *Environment and Planning A, 10*, 1187–1200.
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*, 5–23.
Stillwell, J., Rees, P., & Boden, P. (Eds.). (1991). *Migration processes and patterns volume 2 population redistribution in the United Kingdom*. London: Belhaven Press.
Stouffer, S. A. (1940). Intervening opportunities: A theory relating mobility and distance. *American Sociological Review, 5*, 845–867.
Stouffer, S. A. (1960). Intervening opportunities and competing migrants. *Journal of Regional Science, 2*, 1–26.
Thomas, M. (2014). *Modelling individual and place variations in residential moves using commercial data and official statistics* (Unpublished doctoral dissertation). Leeds: University of Leeds.
Thomas, M., Gould, M., & Stillwell, J. (2012). *Exploring the potential of microdata from a large commercial survey for the analysis of demographic and lifestyle characteristics of internal migration in Great Britain* (Working Paper No. 12/03). Leeds: School of Geography, Leeds University.
Thomas, M., Stillwell, J., & Gould, M. (2014). Exploring and validating a commercial lifestyle survey for its use in the analysis of population migration. *Applied Spatial Analysis and Policy, 7*, 71–95.
Thompson, C., Stillwell, J., Clarke, M., & Bradbrook, C. (2010). *Understanding and validating Acxiom's research opinion poll data* (Working Paper No. 10/6). Leeds: School of Geography, University of Leeds.
Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–240.
US Department of Commerce. (1965). *Calibrating and treating a gravity model for any size urban area*. Washington, DC: US Government Printing Office.
Venigalla, M. M., Chatterjee, A., & Bronzini, M. S. (1999). A specialized equilibrium assignment algorithm for air quality modelling. *Transportation Research Part D: Transport and Environment, 4*, 29–44.
Weeden, R. (1973). Inter-regional migration models and their application to Great Britain. *National Institute of Economic and Social Research Paper 2*. London: Cambridge University Press.
Wilson, A. G. (1970). *Entropy in urban and regional modelling*. London: Pion.
Zipf, G. K. (1946). The P1 P2/D hypothesis: On the intercity movement of persons. *American Sociological Review, 11*, 677–686.