Simulating the Pulse of the City: Using a Combined Approach for Modelling Commuter Patterns

K. Harland [a], M. Birkin [a] and D. Martin [b]

[a] School of Geography, University of Leeds, Leeds LS2 9JT, UK
Telephone: +44 (0 in UK)113 3433300
Fax: +44 (0 in UK)113 3433308

[b] Geography and Environment, University of Southampton, University Road, Southampton, SO17 1BJ, UK
Telephone: : +44 (0 in UK)238 0592216
Fax: +44 (0 in UK)238 059 3295

http://www.geog.leeds.ac.uk

1. Introduction

Aggregate data describing the demographic composition and social structure of small geographic areas is readily available in many countries. For example, the Census of Population and Households in the UK comprises counts for small geographical areas called output areas typically comprising around 125 households. Taken as a decennial snapshot, these data characterize the normal residential population and may be considered as providing a view of the regular night-time populations of different neighbourhoods. The regular night-time populations have been widely applied for planning the delivery of services such as education and health care, and for the strategic planning of land use or transportation.

The description of the night-time population is only half of a complicated story that makes up the daily movement patterns of the population with most people undertaking a regular commute to work or study. These regular 'workday' travel patterns provide a view into the dynamics of population movement patterns throughout the day. However, these travel patterns are further complicated by irregular journeys (such as participation in recreational activities, journeys to visit healthcare practitioners) and multipurpose journeys (for example picking a child up from school and calling to collect shopping on the way home from work). These journey patterns are much harder to capture with traditional datasets although the increased use of social media data, such as Twitter, does hold some promise to help estimate these travel behaviours. The representation of the population at any given point in time throughout a typical working day provides a significantly more useful view of the population for many purposes. With regard to planning for emergencies, whether natural or man-made in their origin, the actual likely location and distribution of people at a particular time of day is significantly more useful than knowing only where those people reside. A similar argument can be made for services such as healthcare or policing and for retail provision where the route taken maybe as valuable as the eventual destination.

2. Mapping the population

Work undertaken at the University of Southampton on the Population 24/7 project has developed a *mapping* approach that integrates a variety of directory resources to profile population change at user specified time intervals. For example, pupil roll data collated by the Department for Education are used to provide reliable counts of children during the school day. Work and education populations along with hospital patients are constructed in recursive layers that, when viewed in combination, provide a comprehensive picture of the day-time population at specified time points throughout the day. The Population 24/7 modelling method builds upon work undertaken by Martin (1989, 1996). It uses an adaptive kernel estimation algorithm to redistribute population counts associated with population-weighted centroids for small areas and results in conventional residential populations being distributed more realistically for a specific time of the day. The model represents each member of the population as being engaged in one of three activity classes; Residential (at home or engaged in an activity very close to home), Non-residential (currently defined as at work, in hospital or at school although other activities can be added to the simulation) or Transport (the individual is in transit between two activities).

The Population 24/7 model is a significant improvement over traditional residential datasets for estimating population distributions; however, the approach does have limitations. The method distributes the aggregate residential population to likely destinations given assumptions about the catchments of the work or study destination options. Information relating to the characteristics (other than broad age and economic activity) of the population travelling from one place to another is not tracked. Subsequently, the demographic and social structure of the resulting population at the destinations is limited. Likewise, interactions between individuals likely to drive multi-purpose journeys, such as parents collecting children from school, are not represented preventing the development of more complex travel patterns.

3. Tracking the population

An alternative method has been prototyped at the University of Leeds. This approach uses a blend of spatial modelling approaches to first create and move a population at an individual level (Harland and Birkin, 2013). The approach first creates a synthetic representation of small area residential populations at the individual-level with a static spatial microsimulation model using a Simulated Annealing algorithm (Harland, 2013). This process incorporates both small area statistics and anonymised individual records to generate synthetic individuals and households in accordance with the known aggregate characteristics of each neighbourhood. The top half of Figure 1 shows the process of combining small area statistics and survey records to produce a realistic synthetic population.

In this paper a commercial consumer survey provided by Acxiom was used to supply anonymised individual level records from which the synthetic population was constructed. The population generation process was constrained to UK Census of Population and Household 2001. These small area census statistics were used at the lowest level of administrative geography available in the UK, the output area, which typically consists of around 125 households. Population counts for the following demographic characteristics were used to constrain the synthetic population:

- Gender
- Age
- Ethnicity
- Economic activity
- Occupation
- Marital status

The movement patterns of these individuals – to schools, hospitals, workplaces – are then simulated in relation to the location of services and facilities using a series of spatial interaction models. The spatial interaction modelling architecture follows that suggested by Wilson (1971) in his work to more accurately represent different behaviour in commuting to work through the disaggregation of spatial interaction models to represent different modes of travel. The correct spatial interaction model from the series of models is probabilistically joined to the synthetic population, as shown in Figure 1, to provide the destination for regular movement patterns for work or school etc. The synthetically generated individuals are finally incorporated into the Agent-Based Modelling framework, MASON, developed at George Mason University where behavioural characteristics can be exploited, as has been demonstrated by Malleson *et al.* (2013), although currently the tool is used purely for visualisation purposes.

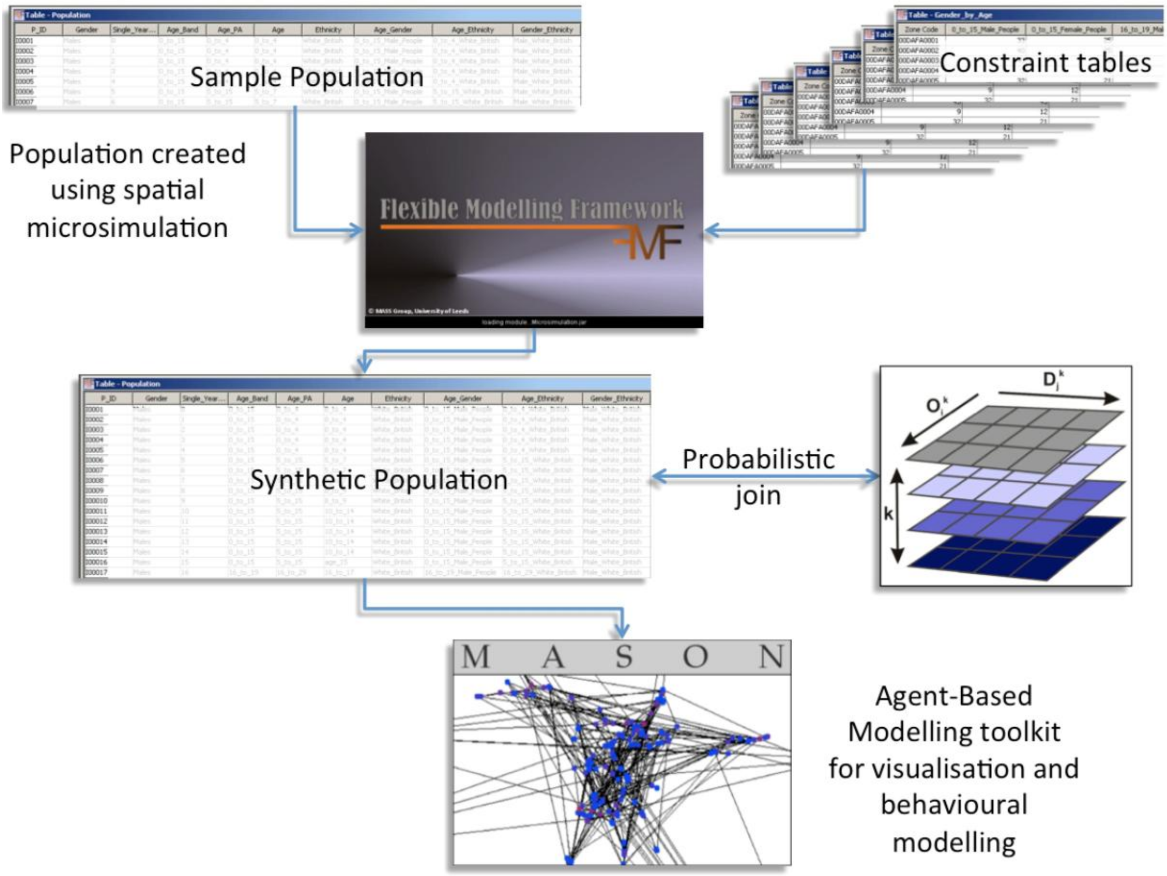

Figure 1: Diagrammatic representation of the population tracking approach

Figure 2 shows screen shots for three time slices of the model run for regular car commuters in the city of Leeds, UK. Each agent in the simulation darkens as the population density for the output area it is within increases. Each output area is represented as a square and space has been added between each one to ensure that each area can be clearly observed. Although not geographically accurate, each output area is correctly scaled so that the relative geographical area of each output area is correct.

Figure 2 clearly shows the residential population ubiquitously distributed throughout the cities output areas, a characteristic of UK administrative zones that are designed to contain relatively consistent population counts. As the population begins to commute the density of the population begins to increase in particular output areas and decrease in others. Once all of the commuters have arrived at work a selection of small central output areas are extremely densely populated while suburban areas are far less populated. This modelling approach may be thought of as a means for synthetic *tracking* of individual movement patterns. It enables detailed (re-)construction of the demographic profiles of geographical areas throughout the day as demographic attributes for the synthetic individuals can be linked to each on wherever they move in the simulated environment.

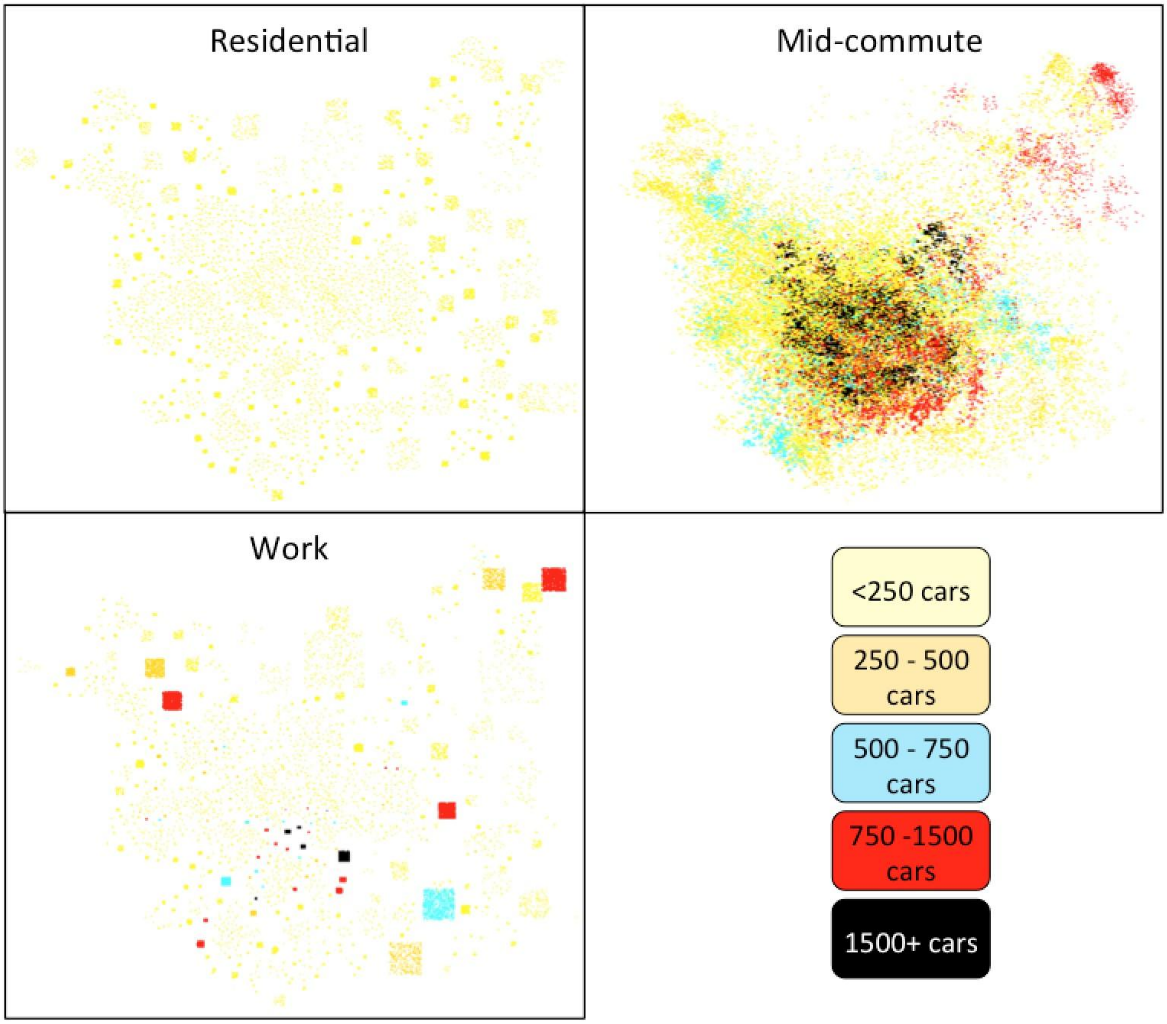

Figure 2: Screen shots from the visualisation of daily car commuters in Leeds

4. Tracking and mapping combined

The first steps towards the integration of the two modelling approaches to the problem of estimating day-time population distributions are reported here. This research demonstrates, for a case study of the city of Leeds, how a more complete picture emerges from the combination of these two approaches with far greater flexibility and more possibilities for both future development and areas of application. The tracking of both tangible and intangible commodities around the city throughout the day is demonstrated, such as the change in income distribution between the residential night-time population locations and day-time population locations shown in Figure 3 below.

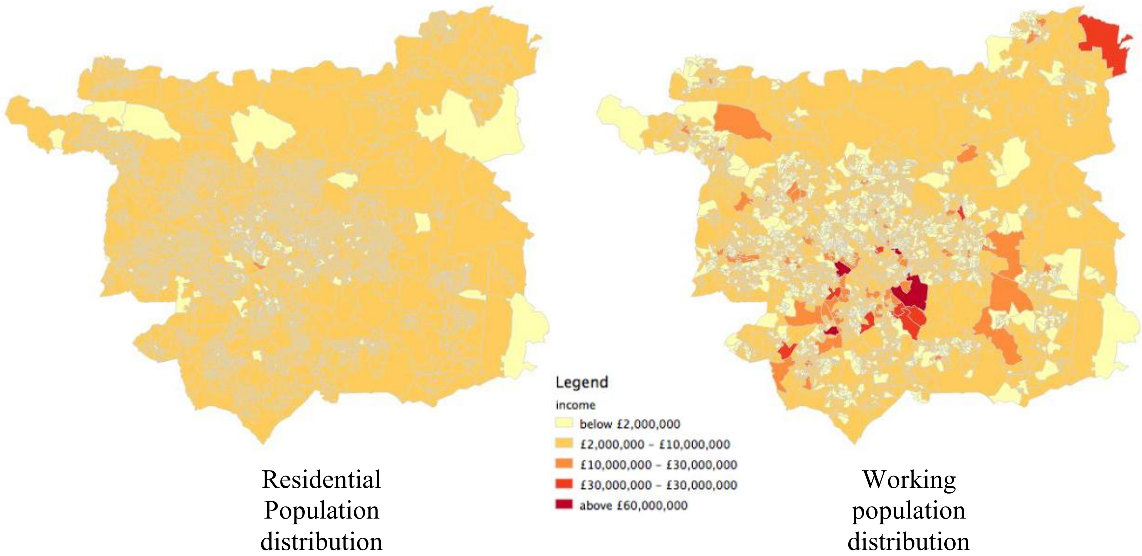

Figure 2: Simulated distribution of annual income.

The combined approach introduces the 24/7 populations as a detailed set of constraint vectors on the underlying interaction models with the time structures providing realistic timings for the underlying journeys start or end time. Alternatively this may be conceived of as a filtering process in which the day-time population counters are filled up with appropriate individuals from the surrounding areas at a particular time of the day. The sensitivity of the probabilistic elements of the combined modelling approach (the generation of the synthetic population and the assignment of the destinations derived from the Population 24/7 to the synthetic individuals) is presented and shows that these elements of the combined model structure have little impact on output distributions for regular daily commuting patterns.

Through this combined approach it is possible to arrive at a superior end product coupling the best features of both methods: realistic destination populations and time structures from the mapping approach and enhanced attribute capture at the destinations alongside the potential for further behavioural modelling from the tracking approach. Tracking the purpose and spatial realisation of individual flows facilitates the representation of individual movement patterns within cities. This opens up a range of further research opportunities in understanding the transmission of all kinds of tangible and intangible commodities around the city. The example briefly introduced above shows the change in distribution of income through the day and is achieved by using a commercial consumer survey for the population synthesis stage of the modelling process. However, this is only one commercially orientated example of the potential applications for this modelling approach with other uses such as emergency evacuation planning and increasing the realism of populations at risk of crime throughout the day also being discussed.

The results of the case study simulation have been compared not just to census data but also to patterns in the use of the Twitter social messaging service. Despite the limitations of Twitter data, such as the skewed age profile of users and the limited number of geographically encoded tweets, currently less than 3% of all tweets, interesting patterns can be teased from the data. These patterns can be used, albeit with care, to augment more traditional sources of information to help deliver a more complete picture of day-time travel patterns. Further work is still needed in order to provide a robust assessment of model performance, and indeed in understanding the contribution social media data can make, but early indications are positive.

Acknowledgements

Philip Cross presented this paper at GISRUK-2014, for and on behalf of the authors.

References

Harland, K. (2013) Microsimulation Model User Guide (Flexible Modelling Framework). NCRM Working Paper 06/13. NCRM.

Harland, K., Birkin, M. (2013). Using Synthetically Generated Populations in Agent-Based Models. Paper presented at the Annual Association of American Geographers Conference 2013. Los Angeles, USA. 9-13 April 2013.

Malleson, N., Heppenstall, A., See, L., Evans A. (2013). Using an agent-based crime simulation to predict the effects of urban regeneration on individual household burglary risk. *Environment and Planning B: Planning and Design*. vol 40. pp 405-426.

Martin, D. (1989). Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers NS*. vol 14. pp 90-97

Martin, D. (1996). An assessment of surface and zonal models of population. *International Journal of Geographical Information Systems.* vol 10. pp 973-989

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning* . vol 3. pp 1 – 32.