Development and application of a two stage hybrid spatial microsimulation technique to provide inputs to a model of capacity to walk and cycle.

Ian Philips[*]

Institute for Transport Studies
University of Leeds

November 7, 2014

Summary

This paper demonstrates the development and application of a two stage hybrid static spatial microsimulation technique. The first stage makes best use of Simulated Annealing with available micro-data, and the second uses Synthetic Reconstruction to add attributes not available in a single micro-data source. The new technique is applied to Leeds UK to generate a synthetic population which can be used as an input to a model of capacity to commute using only walking and cycling.

KEYWORDS:

Hybrid spatial microsimulation method, Simulated Annealing, Synthetic Reconstruction, walking and cycling, transport planning model.

1. Introduction

In transport applications spatial microsimulation is often used to generate a synthetic population of individuals as a start point for a transport planning model (e.g.Beckman et al., 1996; Frick and Axhausen, 2004; Müller and Axhausen, 2010). A transport planning model of individuals' capacity to make journeys by walking and cycling has been constructed (the details of this model is reported elsewhere and is not the focus of this paper). It required a synthetic population of individuals as inputs. This paper focuses on the production of that synthetic population.

Existing populations were not suitable as these populations did not contain all of the attributes required to estimate capacity to walk or cycle used in the model. Table 1 shows the attributes required by the model. In addition to this correlated constraint attributes were selected. The synthetic population had to represent the variation in the physical capacity of individuals to walk and cycle. For example, Parkin (2008) expressed the importance of considering human power output in models of cycling. It should also consider constraints on walking and cycling such as bicycle availability and the need to escort children as part of a commute.

The population had to be available at a fine spatial resolution (Output Areas for UK applications). This is because journey origins were based on zone centroids. Typical trip length for walking and cycling is short (generally under 8km). This means that using the centroid of a large zone would introduce considerable error into estimates of distances (Iacono et al., 2010). Additionally being able to report results at a fine spatial resolution allows aggregation to coarser resolutions within a hierarchy and can be used to demonstrate heterogeneity to decision makers. Using the results in this way reduces the dangers of making poor decisions affected by the ecological fallacy.

[*] Gy09ip@leeds.ac.uk

Table 1: Attributes required in the synthetic population

VO_{2max} a measure of fitness in terms of the body's ability to make use of oxygen for exercise Physical activity Body Mass Index Age Gender Weight	*These attributes are used to derive pedalling power and the model uses this to estimate an individual's cycling speed and a corresponding value for walking speed.*
Bicycle availability The need to escort children on the way to or from work Current commute distance	

There are several existing static spatial micro-simulation techniques (see Hermes and Poulsen, 2012 for an introduction). The suitability of existing spatial microsimulation techniques was examined. Simulated Annealing based combinatorial optimisation was reported as the best performing technique, particularly at Output Area resolution (Williamson, 2012; Harland et al., 2012). Synthetic Reconstruction (using Monte-Carlo sampling) is useful when a micro-data sample is not available (Barthelemy and Toint, 2012). The practical issues of constructing a population for Output Areas in the UK city of Leeds were examined. Not all of the required attributes were available in a single micro-data sample however the majority were available in the 2008 Health Survey for England.

The principal problems of using an existing technique are: Firstly, to use an existing technique would require sacrificing either spatial microsimulation performance (if Synthetic Reconstruction were used) or limiting the inputs to the indicator (if Simulated Annealing were used). Secondly, some attributes cannot be assigned to an individual until that individual has been allocated a location. For example, commute distance is strongly associated with location as well as individual socio-demographic attributes. Commute distance is collected in some micro-data surveys, but, it would not be appropriate to allocate this value out of its original spatial context. This is because individual survey data has geographical detail removed. This means that some individuals will not be allocated to areas to where they actually live. These problems led to development of a new hybrid technique.

2. A hybrid 2 stage spatial microsimulation technique

The two stage hybrid method works follows: In the first stage, a single synthetic population is constructed using Simulated Annealing In the application the open source FMF software was used (see Harland, 2013). The available micro-data is used as a sample population and constraint tables are taken from the census. In the second stage, Monte-Carlo sampling (Synthetic Reconstruction) is used to add attributes which are not available in the micro-data or which are geographically dependent. Monte-Carlo sampling can then be used to draw multiple synthetic populations.
This approach makes progress towards addressing the problems above: The performance benefits of Simulated Annealing are used with available data. Though Monte-Carlo sampling introduces increased computing time and data storage requirements, this is less of a drawback than it once was. A greater gain is made because it allows the full range of desired attributes to be modelled rather than having a model constrained by limited data sources. Introducing Monte-Carlo sampling also introduces a source of stochastic variation.
This is not a problem if a suitable number of draws is made; the standard error of the mean should not be excessive. Because only a minority of attributes are being added using Monte-Carlo sampling, the

stochastic variation between draws should be less than if the entire population was built using Synthetic Reconstruction. This will give an overall advantage in terms of performance. The process is outlined in Figure 1.

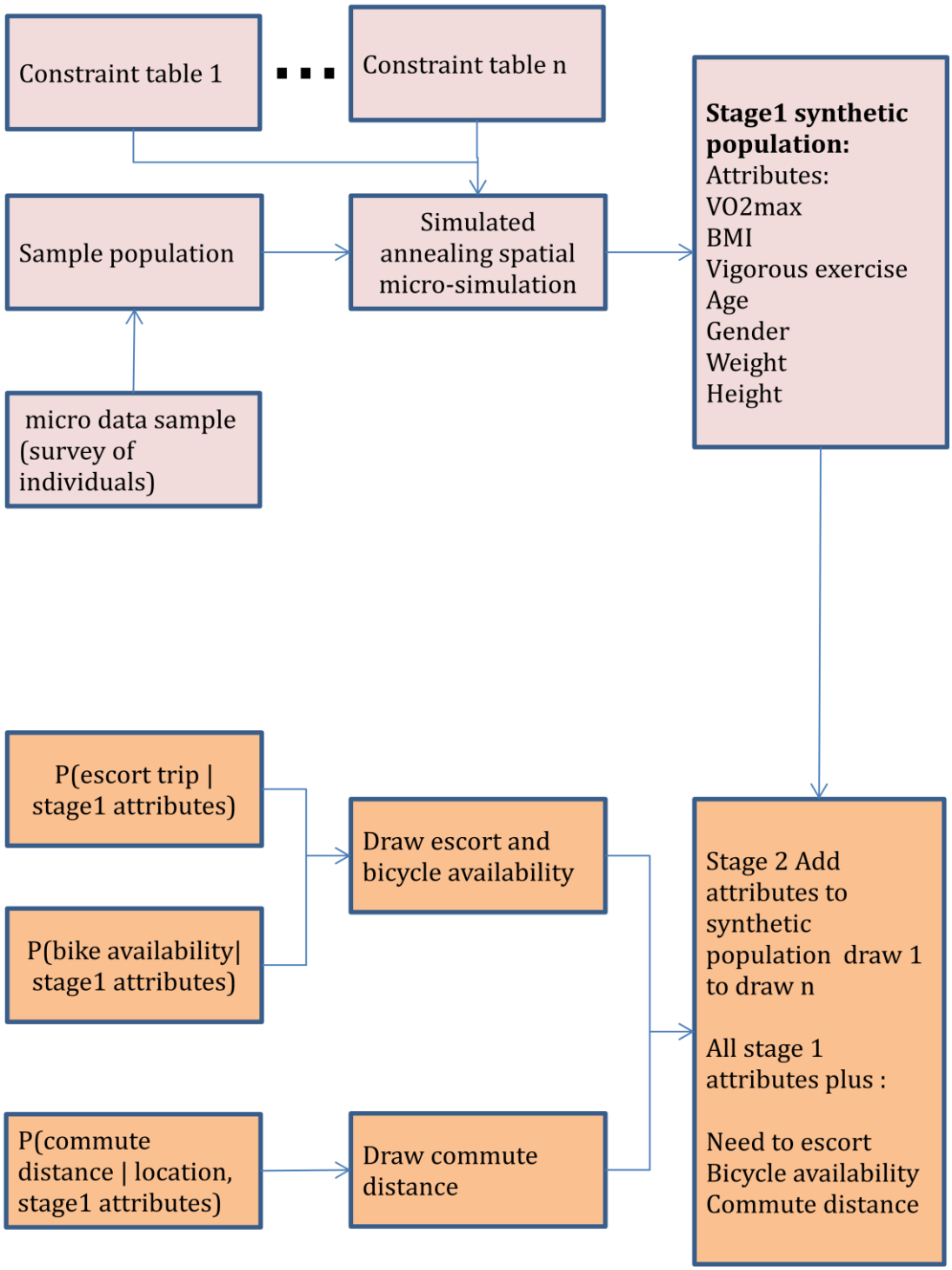

Figure 1: Outline diagram showing the two stage hybrid spatial microsimulation method.

3. Validation

Internal and external validation tests were performed using established validation techniques, TAE based measures and Z-scores (See Edwards and Tanton, 2012 for discussion of these techniques).
The sensitivity of the final model result resulting from the stochastic variation and the different construction of constraints was estimated as less than $\pm 5.1\%$ in 95% of Output Areas. This was acceptable for the application of the model.

4. Results

The spatial pattern of both the model output and the attributes contributing to it were mapped.
The distribution of contributing attributes was internally consistent. Mapping individual attributes aids analysis of where specific attributes exert influence over the model. For example age affects pedalling power. The age structure in Figure 2 clearly shows an influence on the spatial distribution of pedalling power in Figure 3.

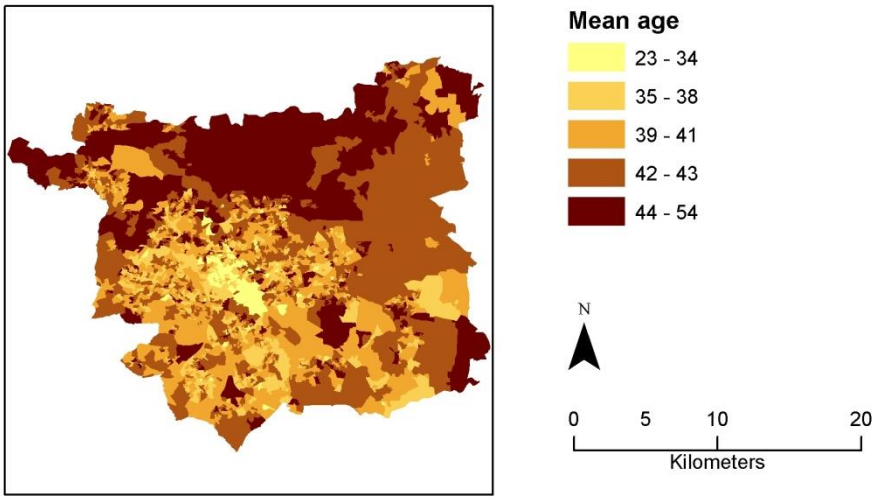

Figure 2 Mean age of working population in Leeds Output Areas

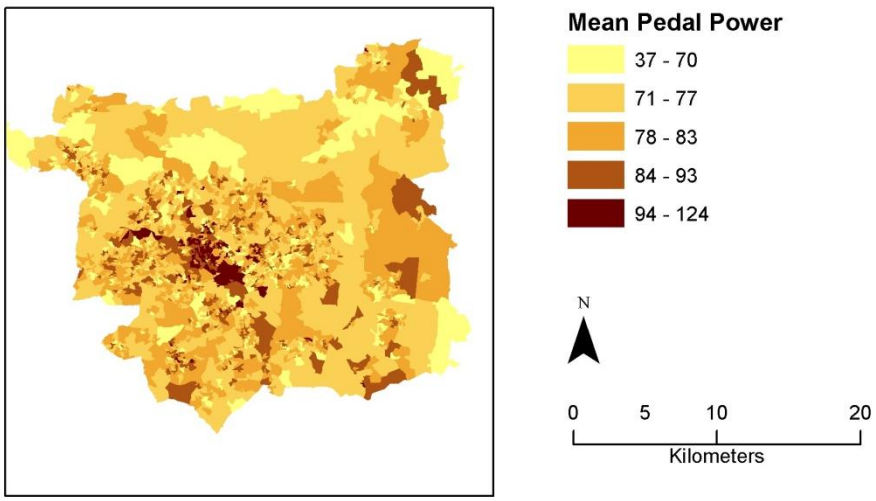

Figure 3: Mean Pedal Power of working population by Leeds Output Areas

Further analysis was conducted mapping other attributes output from the spatial microsimulation which contribute to the final model output shown in Figure 4.

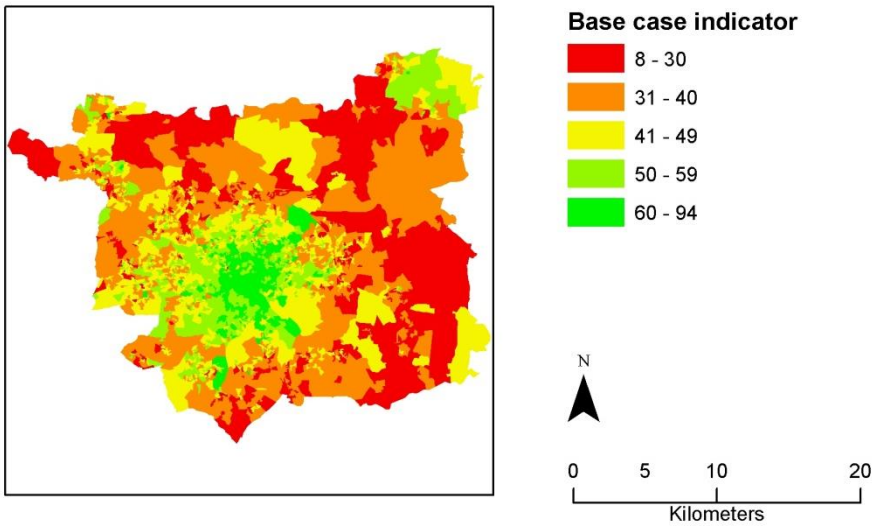

Figure 4. Model output based on the outputs of the hybrid spatial microsimulation method: Percentage of working population with capacity to walk or cycle to their current place of work in a network with no motorised traffic. Leeds Output Areas

5. Conclusion

The conclusion of the work is that this method can be usefully applied and is particularly useful where a synthetic population with many attributes is required as an input to a further modelling process.

6. References

Barthelemy, J., Toint, P.L., 2012. Synthetic Population Generation Without a Sample. Transportation Science. doi:10.1287/trsc.1120.0408

Beckman, R.J., Baggerly, K.A., McKay, M.D., 1996. Creating synthetic baseline populations. Transportation Research Part A: Policy and Practice 30, 415–429. doi:10.1016/0965-8564(96)00004-3

Edwards, K.L., Tanton, R., 2012. Validation of Spatial Microsimulation Models, in: Tanton, R., Edwards, K. (Eds.), Spatial Microsimulation: A Reference Guide for Users. Springer Netherlands, Dordrecht, pp. 249–258.

Frick, M., Axhausen, K.W., 2004. Generating Synthetic Populations using IPF and Monte Carlo Techniques: Some New Results. Presented at the Swiss Transport Research Conference, Ascona.

Harland, K., 2013. Microsimulation model user guide Flexible Modelling Framework.

Harland, K., Heppenstall, A., Smith, D., Birkin, M., 2012. Creating Realistic Synthetic Populations at Varying Spatial Scales: A Comparative Critique of Population Synthesis Techniques. JASSS 15, 1.

Hermes, K., Poulsen, M., 2012. A review of current methods to generate synthetic spatial microdata using reweighting and future directions. Computers, Environment and Urban Systems 36, 281–290. doi:10.1016/j.compenvurbsys.2012.03.005

Iacono, M., Krizek, K.J., El-Geneidy, A., 2010. Measuring non-motorized accessibility: issues, alternatives, and execution. Journal of Transport Geography 18, 133–140. doi:10.1016/j.jtrangeo.2009.02.002

Müller, K., Axhausen, K.W., 2010. Population synthesis for microsimulation: State of the art. ETH Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau (IVT).

Parkin, J., 2008. NECTAR Workshop Abstract The importance of human effort in planning networks. Presented at the NECTAR workshop.

Williamson, P., 2012. An Evaluation of Two Synthetic Small-Area Microdata Simulation

Methodologies: Synthetic Reconstruction and Combinatorial Optimisation, in: Tanton, R., Edwards, K. (Eds.), Spatial Microsimulation: A Reference Guide for Users. Springer Netherlands, Dordrecht, pp. 19–47.

7. Acknowledgements

This work was carried out as part of an ESPRC CASE PhD studentship partnered with Sustrans. I would like to acknowledge the help and support of my PhD supervisors David Watling and Paul Timms.

8. Biography

Ian Philips has recently completed his PhD on resilience to fuel shocks. This involved creating a spatially explicit indicator of who could get to work by walking and cycling if there was no fuel for motorised transport. Ian's interests include spatially explicit modelling to inform sustainable transport policy.