EUROPEAN TRANSPORT CONFERENCE 2020

SPATIAL MICROSIMULATION OF ANNUAL CAR KM TRAVELLED BY PURPOSE

Ian Philips
Institute for Transport Studies, University of Leeds
Claire Shadbolt
Leeds Institute for Data Analytics, University of Leeds

We investigated the potential for using a spatial microsimulation approach to generate easily accessible maps of small area estimates of car km travelled by purpose. Estimates of travel demand are often taken for granted in predict and provide approaches and are not visualised in a way which makes it easy to discuss demand reduction. We used two sources of data; a microsimulation based population projection (for the year 2017) and the English National Travel Survey anonymised individual records. We constructed the estimates of car km travelled for individuals within English LSOAs[1] using Monte-Carlo Sampling. To understand the potential to develop this approach further we chose to retain small area estimates of total car km travelled and estimates of commute distance by purpose from the 2011 census for validation purposes. The results are promising and show potential for further development. The project was carried out principally using R, the code is stored in an open access repository to aid reproducibility and further development.
https://github.com/DrIanPhilips/car_km_by_purpose_microsim

1. INTRODUCTION

Transport research must be driven by the need for rapid transport carbon reduction (Anable, 2019; Anderson et al., 2020; I Philips et al., 2020; Sorrell, 2015). Place-based understanding of travel demand is important for developing strategies to reduce the overall need for travel, as well as decarbonising the remaining demand (Creutzig, 2016; Marsden, 2019). With the exception of how people in small areas travel for commuting, which is gathered via the census, small area estimates of car travel demand by purpose are not frequently visualised and shared with policymakers, planners and other stakeholders. Travel by purpose is estimated in traffic models, but usually, they remain as matrices of data in transport the models, therefore they are not available beyond a technical modelling team. When travel demand is hidden in the model it is too easy to taken for granted. Visualising demand would aid discussions about demand reduction. It is important that demand is not considered as pre-determined – it can be influenced by policy.

This paper contributes to making small area estimates of current car use by purpose more easily available and more easily interpretable by putting information on planners and policymakers' desks in a digestible visual format. (e.g. maps). This then facilitates discussion about the desired level of demand and policies to achieve it.

In this paper we use spatial microsimulation (also known as population synthesis). It is a well established method (e.g. Beckman et al., 1996; Tanton and Edwards, 2013) of generating the individual or household populations of small areas by fusing anonymised individual survey records to small area aggregate data (usually a census). The most recent published National Travel Survey (NTS) data is approximately 2 years old. The most recent UK census was in 2011, the next is due in 2021 but it takes approximately 3 years before all outputs are available. This mismatch in the ages of the data sets is not ideal. Static spatial microsimulation – linking data sets such as the NTS and the latest census has been used in some cases to investigate social and environmental issues (Becker et al., 2019; Bonsall and Kelly, 2003; Lovelace and Philips, 2014; Philips et al., 2018). However, in this paper, to overcome some of the issues associated with fusing data of different ages, we test the applicability of dynamic spatial microsimulation based population projections to produce data-driven estimates of current and recent travel demand.

2. DATA

The data used is summarised in the table. The census data is openly available. National Travel Survey data is curated for research access by the UK Data Service.

Table 1 Data used in the simulation

Base synthetic population	**Data source**
2011 census tables Annual subnational population estimates	Nomis https://www.nomisweb.co.uk/census/2011 ONS https://www.ons.gov.uk/census/2011census
Car travel distance by purpose	
National Travel Survey (NTS) individual anonymised records 2015-17	Accessed via UKDS https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000037
Validation	
2011 census tables on travel to work by mode and purpose	Nomis https://www.nomisweb.co.uk/census/2011
2011 total private car miles travelled by LSOA	Data calculated from UK vehicle ownership and MOT (roadworthiness test) data (Cairns et al., 2014)
2014 total private car miles travelled by LSOA	Data calculated from UK vehicle ownership and MOT (roadworthiness test) data (Cairns et al., 2014)

3. METHODS

We first generated a base synthetic population for the year 2017. For this we applied SPENSER, a household level small area synthetic population estimation and projection model which uses dynamic microsimulation (Lomax and Smith, 2017). Data and code to generate populations using SPENSER is entirely open source[3].

Using the R language and RStudio we use a Monte-Carlo sampling approach to generate attributes of car km travelled per person per year for different purposes. Code is available at https://github.com/DrIanPhilips/car_km_by_purpose_microsim . An R markdown file is available in the repository and this can be used to call the various scripts which run the different stages of the simulation. The data comprises a mix of open source and safeguarded data. Table 1 shows the data sources. The stages of the method are shown in Table 2. This table follows the order of code execution.

TABLE 2: Spatial microsimulation of annual car km travelled by purpose for synthetic population of LSOAs in England. This table follows the order of code execution.

Section	**Subsection**	**Comment**
Getting started		
	1.1 Load packages	Load appropriate R packages
	1.2 Set date and time	Notes model run date and time for reference
Load data	2.1 NTS datasets	Read in data
	2.2 Constraint tables	Creates a probability constraint table, a cross tabulation of number of trips by purpose based on five constraints. In this model, the constraints are gender "Sex", age group "Age_B04", socio-economic classification "NSSEC", Region "GOR_new" and rural urban classification "ONSRuralUrban"
	2.3 ONS dataset	Load census geography and administrative area geography data
	2.4 SPENSER dataset	Load the base synthetic population for 2017, for every Local Authority District in England. This synthetic population was created by the programme SPENSER. Each Local Authority District is represented in a dataframe. It is cleaned and the Age and NSSEC variables are recoded to match the codes found in the NTS dataset. Next, each district dataframe is joined to the ONS dataset created in 2.3. This gives the synthetic population geographic identifiers: an LSOA[1], Rural/Urban Classification and a Region attribute too. District level dataframes are stored in a list called "dfList", and the name of each dataframe in the list is

		named after the Local Authority District Code for the dataframe.
	2.5 Split the data frames by Age and NSSEC socio-economic group	Each District Data Frame is split by Age and NSSEC group Each individual is then allocated a number of weekly trips based on a cross tabulation of mean weekly trips by population sub-group. After this step, recombine the subgroups to give a single dataframe for each district.
	2.6 Combine the Constraint Table with SPENSER	Join the constraint table generated in 2.2 to each district dataframe. For each District, this creates a master table of population and probability distribution of trip purpose based on constraints.
3 Monte Carlo Sampling	3.1 Trip Distance Constraint Table	Generate trip distance constraint tables from NTS data
	3.2 Count the mean trips for each age/NSSEC combination	Each individual is assigned a number of trips.
	3.3 Monte-Carlo Sampling	Monte Carlo Simulation, draw trip purpose and then trip distance bin.
4 Aggregation		
	4.1 Regional groups	Select individuals by region. The Following English Regions are used: "North_and_Yorkshire", "Midlands", "East_and_South_East", "London", "South_West"
	4.2 Convert the count into kms	Convert from bin count (3.3.) to kilometres.
	4.3 The total yearly and average kms travelled by LSOA and OA[2]	Aggregate individual travel distances to zones to calculate total travel by zone. Calculations take into account the proportion of people in the population who might not have access to a car
	4.4 The total yearly and average kms travelled per LSOA by Trip Purpose type	As 4.3 above but for each trip purpose.
5 Plot the datasets	5.1 Import aggregated data	Run data preparation code
	5.2 Plotting Functions	Map the initial outputs
6 Validation	6.1 MOT 2011 vs simulated total and average yearly distances travelled in a car per LSOA	Mark areas with highest residuals as not well simulated, remove these, and assess the correlation and bivariate regression of the remaining areas.
	6.2 MOT 2014: total and average yearly distances travelled in a car per OA	Mark areas with highest residuals as not well simulated, remove these, and assess the correlation and bivariate regression of the remaining areas.

	6.3 Commuting distance bins from the 2011 census data	Mark areas with highest residuals as not well simulated, remove these, and assess the correlation and bivariate regression of the remaining areas.

4. VALIDATION

We tested the simulated outputs against existing small area data sets. We compared the simulated data against the commuting distance bins recorded in the 2011 census. We also compared simulated outputs to 2011 total private car miles travelled by LSOA resolution and estimated 2014 total private car miles travelled by LSOA (see table 1 for data sources). We aggregated simulated individual variables to area totals to make the comparison.

We carried out Pearson's and Spearman rank correlation as well as bivariate OLS regression. We mapped the residuals to examine the spatial pattern. In spatial microsimulation, the algorithms and the input data may simulate most areas very well, but struggle to accurately simulate atypical types of areas (Smith et al., 2009). We visually inspected locations with the highest residuals. When comparing simulated and small area data on total car km travelled most of the highest residual areas included high levels of industrial and business premises. Examples include the Team Valley Industrial estate to the south of Newcastle. We then removed areas with the 10% highest residuals.

Table 3: Examples of areas with highest residuals

Area with high residuals	**Area Attributes**
Hunslet, Leeds	Warehousing and distribution centre.
North-West Leeds City Centre	University of Leeds.
Hillsborough, Sheffield	Industrial business use. Flats, which typically have the highest level of non-car ownership.
Sale	Retail, schools and places of worship.
Doncaster	Business parks right next to the motorway and the station. Retail area, places of worships and a little bit of terrace housing.
Leyland	Industrial estate.
South-East corner of York	University of York.

Figure 1: an extract from an interactive map, used to visually inspect the pattern of residuals – used as an indicator of difference between simulated and comparison small area data sources

4.1 Interpretation of validation

Examination of the residuals identified that uncertainty is greatest in areas with considerable industrial development. This is helpful in identifying how the modelling process may be improved. The considerable improvement in correlation and R squared when these high residual areas are removed, suggests a large number of residential areas are being simulated more effectively.

The comparison to 2011 commute data gave moderate correlations, but weak bivariate regression results. Two factors contributing to this are that firstly, the distance bins in the simulated population did not fully match the distance bins in the commute data. Secondly, and to a lesser extent, in the time since the census, the commute patterns may have changed in some areas.

Table 4. Comparison of simulated results against 2011 LSOA estimates of distance travelled to work.

Coefficient	England	North	Midlands	East and South East	London	South West
Spearman_bin1 (0 – 10km)	0.50	0.43	0.57	0.48	0.20	0.42
Pearsons_bin1	0.51	0.55	0.65	0.55	0.21	0.51
R_sqaured_bin1	0.26	0.30	0.43	0.31	0.04	0.26
Spearman_bin2 (10-30km)	0.41	0.32	0.48	0.34	0.20	0.32
Pearsons_bin2	0.39	0.33	0.52	0.35	0.21	0.34
R_sqaured_bin2	0.15	0.11	0.27	0.12	0.04	0.12
Spearman_bin3 (>30km)	0.62	0.38	0.55	0.33	0.11	0.40
Pearsons_bin3	0.61	0.54	0.60	0.43	0.10	0.40
R_sqaured_bin3	0.38	0.29	0.36	0.18	0.01	0.16

Comparison to total travel distance produced stronger correlations. The R Squared values from bivariate OLS regression are somewhat weaker than the correlation. The functional form of the relationship between simulated and validation data may not be completely linear when the largest residuals are removed and car ownership is accounted for. Even after accounting for car ownership and removing residuals, the estimations of car km travelled in London are poor. Other regions have strong correlations and R squared values. At LSOA resolution, the R squared value for England was 0.68. At OA resolution it was lower, 0.35. This is to be expected - as the number of individuals simulated within an area decreases, the level of simulation uncertainty increases. Validation at multiple spatial scales is useful to understand the finest scale at which spatial microsimulation can give useful small area estimates (Harland et al., 2012).

Table 5 Comparison of simulated results against 2011 LSOA estimates of total car km travelled

Total Yearly Distance per LSOA		England	North	Midlands	East & South East	London	South West
	Spearman_tot	0.7636	0.7169	0.7540	0.7183	0.2631	0.7704
	Pearsons_tot	0.8235	0.8002	0.8401	0.8062	0.3464	0.8506
	R_sqaured_tot	0.6782	0.6403	0.7057	0.6500	0.1200	0.7235

Table 6 Comparison of simulated results against 2014 OA estimates of total car km travelled

Total Yearly Distance per LSOA		England	North	Midlands	East & South East	London	South West
	Spearman_tot	0.5874	0.5374	0.5190	0.4627	0.3319	0.4900
	Pearsons_tot	0.5933	0.5497	0.5487	0.4982	0.3430	0.5291
	R_sqaured_tot	0.3520	0.3022	0.3011	0.2482	0.1176	0.2800

5. RESULTS

Figure 2 shows estimates of car use per annum for different purposes for the Whole of England. However note the discussion in the validation section, London in particular has not been simulated as effectively as other regions. Figure 3 shows The North of England. Larger versions of these figures and interactive maps are available at https://github.com/DrIanPhilips/car_km_by_purpose_microsim/upload/master/Plots_Purpose_Distance

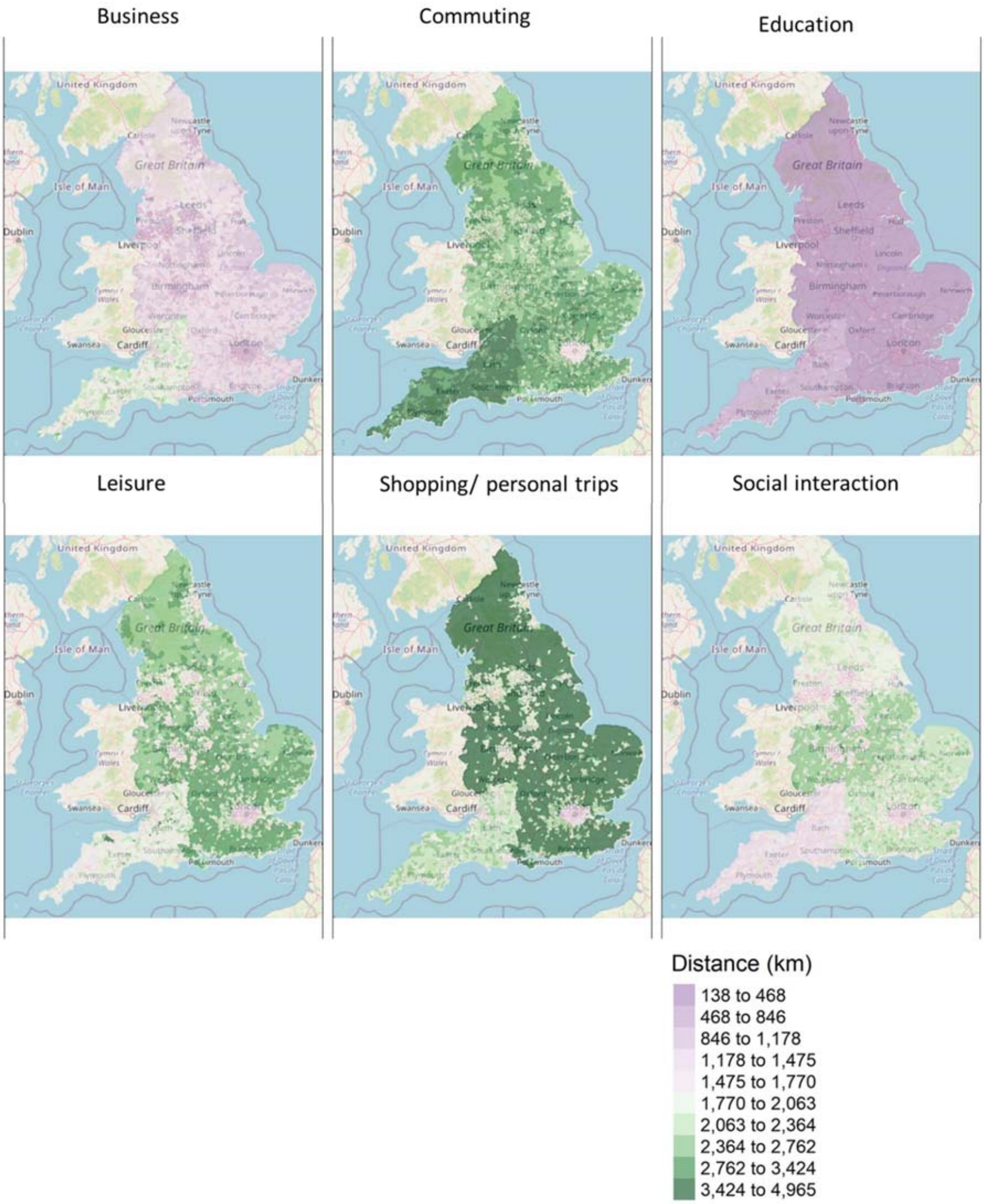

Figure 2 Simulated mean distance travelled by car per person per year by LSOA In England. This map uses a common scale, so distances travelled by different purposes are comparable.

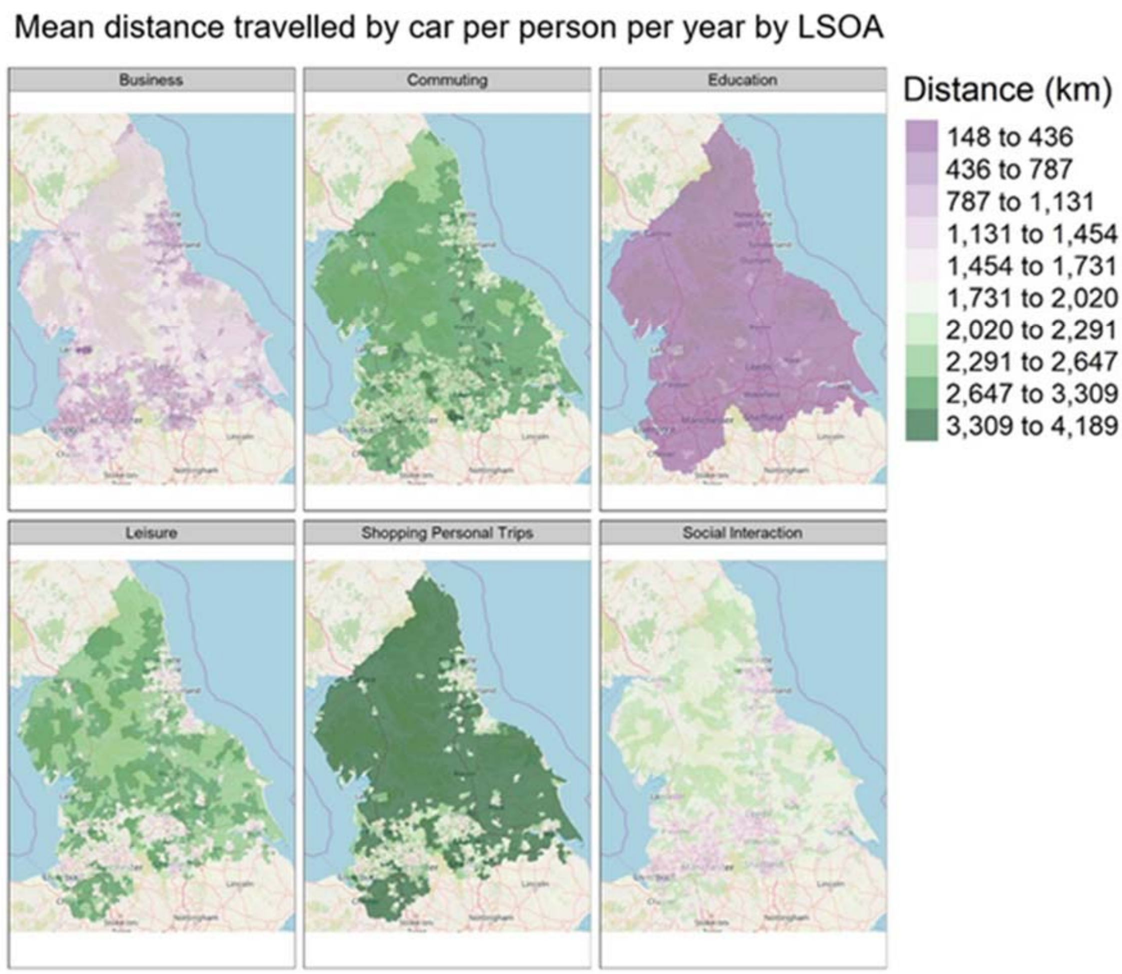

Figure 3 Simulated mean distance travelled by car per person per year by LSOA In Northern England. This map uses a common scale, so distances travelled by different purposes are comparable.

6. SUMMARY AND FURTHER WORK

This paper has produced a reproducible workflow and initial results, which forms a basis for further development of an open source and openly accessible method to visualise travel demand by mode and purpose at the small area level. It applies a dynamic spatial microsimulation based small area population for years more recent than the latest census. From this, small area estimates are made of the car km travelled for different purposes. The workflow goes as far as the production of openly available static and dynamic maps. The workflow described in this paper enables small area mapping of travel demand to be easily shared with policy makers, citizens and other stakeholders.

The potential for application is high. For example the need for place based decarbonisation of transport is established (Creutzig, 2016; Marsden, 2019). To facilitate decision making around policies to reduce travel demand and mode shift and thus reduce transport carbon emissions, such openly accessible tools are needed (Lovelace et al., 2020; Philips et al., 2020).

This initial model could be further developed in a number of ways.
The modelling process could be refined. For ease of implementation in this first version of the model, the mean number of trips per individual was estimated given age and socio-economic group and allocated to individuals. This could be refined to sample from the distribution and assign to individuals. In the present model, there is no geographical constraint on trip distribution such as region or rural / urban classification. Assignment of distance is based on bins. A refinement could be to include a sampling process to assign a distance within the bin (e.g. Philips, 2014; Philips et al., 2017). Multiple runs of the entire simulation would be good practice and would also help produce an estimate of the standard error of simulation. At present the model uses Monte-Carlo sampling from probability distributions. Because it is built using the R language – which is extensible because of the wide range of community contributed code packages, it may be possible to incorporate choice models to estimate trip distance by purpose using specifically developed packages (e.g. Hess and Palma, 2019).

Infrequent long distance journeys may be under represented in the National Travel survey. Further refinement of the trip distance distribution tables may better account for this. As the simulation is refined, a wider range of validation metrics may be used such as total absolute error and measures derived from it (e.g. Edwards and Tanton, 2012; Harland et al., 2012). The code could be modified to make these refinements.

Having estimated car km travelled, deriving other metrics such as cost of use and emissions would be possible. These would be valuable policy indicators. Potential application includes generating scenarios of travel demand policies, to reduce transport emissions, using a future population. This has considerable policy potential. Another useful facet of spatial microsimulation is that as individuals are generated with both travel demand profiles and socio-demographic attributes. This means that if we were to simulate car use and car emissions under a particular policy scenario, it may be possible to generate indicators of potential social impacts under said scenario. There is an established need to produce transport analyses which consider both social and environmental outcomes if we are to pursue equitable decarbonisation (Bonsall and Kelly, 2003; Castiglione et al., 2006; Lucas and Pangbourne, 2014).

BIBLIOGRAPHY

Anable, J., 2019. Rearranging elephants on the Titanic: the astonishing absence of the carbon reduction imperative in transport policy and research., in: UTSG 2019. Presented at the UTSG 2019, Centre for Research into Energy Demand solutions (CREDS), Leeds.

Anderson, K., Broderick, J.F., Stoddard, I., 2020. A factor of two: how the mitigation plans of 'climate progressive' nations fall far short of Paris-compliant pathways. Clim. Policy 1–15. https://doi.org/10.1080/14693062.2020.1728209
Axhausen, K., Horni, A., Nagel, K., 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity Press. https://doi.org/10.5334/baw
Becker, H., Balac, M., Ciari, F., Axhausen, K.W., 2019. Assessing the welfare impacts of Shared Mobility and Mobility as a Service (MaaS). Transp. Res. Part Policy Pract. S0965856418311212. https://doi.org/10.1016/j.tra.2019.09.027
Beckman, R.J., Baggerly, K.A., McKay, M.D., 1996. Creating synthetic baseline populations. Transp. Res. Part Policy Pract. 30, 415–429. https://doi.org/10.1016/0965-8564(96)00004-3
Bonsall, P., Kelly, C., 2003. Road user charging and social exclusion: the impact of a range of charging schemes on at risk groups. Association of European Transport.
Cairns, S., Wilson, R.E., Chatterton, T., Anable, J., Notley, S., McLeod, F., 2014. Using MOT Test Data to Analyse Travel Behaviour Change - Scoping Report.
Castiglione, J., Hiatt, R., Chang, T., Charlton, B., 2006. Application of Travel Demand Microsimulation Model for Equity Analysis. Transp. Res. Rec. 1977, 35–42. https://doi.org/10.1177/0361198106197700105
Creutzig, F., 2016. Evolving Narratives of Low-Carbon Futures in Transportation. Transp. Rev. 36, 341–360. https://doi.org/10.1080/01441647.2015.1079277
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.
Harland, K., Heppenstall, A., Smith, D., Birkin, M., 2012. Creating Realistic Synthetic Populations at Varying Spatial Scales: A Comparative Critique of Population Synthesis Techniques. J. Artif. Soc. Soc. Simul. 15, 1.
Hess, S., Palma, D., 2019. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. J. Choice Model. 32, 100170. https://doi.org/10.1016/j.jocm.2019.100170
Lomax, N.M., Smith, A.P., 2017. Microsimulation for demography. Aust. Popul. Stud. 1, 73–85.
Lovelace, R., Parkin, J., Cohen, T., 2020. Open access transport models: A leverage point in sustainable transport planning. Transp. Policy 97, 47–54. https://doi.org/10.1016/j.tranpol.2020.06.015
Lovelace, R., Philips, I., 2014. The 'oil vulnerability' of commuter patterns: A case study from Yorkshire and the Humber, UK. Geoforum 51, 169–182. https://doi.org/10.1016/j.geoforum.2013.11.005
Lucas, K., Pangbourne, K., 2014. Assessing the equity of carbon mitigation policies for transport in Scotland. Case Stud. Transp. Policy, Social Exclusion in the Countries with Advanced Transport Systems 2, 70–80. https://doi.org/10.1016/j.cstp.2014.05.003
Marsden, G., 2019. DecarboN8: a new approach to place-based decarbonisation [WWW Document]. DecarboN8. URL https://decarbon8.org.uk/decarbon8-a-new-approach-to-place-based-decarbonisation/ (accessed 8.11.20).

Philips, I., 2014. The potential role of walking and cycling to increase resilience of transport systems to future external shocks. University of Leeds, U.K., Leeds, UK.
Philips, I, Anable, J., Chatterton, T., 2020. E-bikes and their capability to reduce car CO2 emissions. Pre-Print Res. https://doi.org/10.13140/RG.2.2.27729.97120
Philips, I., Clarke, G.P., Watling, D., 2017. A Fine Grained Hybrid Spatial Microsimulation Technique for Generating Detailed Synthetic Individuals from Multiple Data Sources: An Application To Walking And Cycling. Int. J. Microsimulation 10, 167–200.
Philips, Ian, Walmsley, A., Anable, J., 2020. A scoping indicator identifying potential impacts of all-inclusive MaaS taxis on other modes in Manchester. Transp. Find. https://doi.org/10.32866/001c.11524
Philips, I., Watling, D., Timms, P., 2018. Estimating Individual Physical Capability (IPC) to make journeys by bicycle. Int. J. Sustain. Transp. 12. https://doi.org/10.1080/15568318.2017.1368748
Smith, D.M., Clarke, G.P., Harland, K., 2009. Improving the synthetic data generation process in spatial microsimulation models. Environ. Plan. A 41, 1251 – 1268. https://doi.org/10.1068/a4147
Sorrell, S., 2015. Reducing energy demand: A review of issues, challenges and approaches. Renew. Sustain. Energy Rev. 47, 74–82. https://doi.org/10.1016/j.rser.2015.03.002
Tanton, R., Edwards, K., 2013. Spatial Microsimulation: A Reference Guide for Users, Understanding Population Trends and Processes. Springer, Dordrecht.

Notes

1. An LSOA is a UK census geography. There are 32844 LSOAs in England each with between 1000 and 3000 residents. LSOAS nest inside Local Authority Districts. https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography
2. Output Areas are the smallest census geographies and nest inside LSOAs
3. SPENSER code https://github.com/nismod/household_microsynth and https://github.com/nismod/microsimulation