Article:

Lovelace, R, Ballas, D and Watson, M (2014) A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. Journal of Transport Geography, 34. 282 - 296. ISSN 0966-6923

https://doi.org/10.1016/j.jtrangeo.2013.07.008

Journal of Transport Geography xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels ☆

Robin Lovelace *, Dimitris Ballas, Matt Watson

Department of Geography, The University of Sheffield, Sheffield S10 2TN, United Kingdom

ARTICLE INFO

Keywords:
Spatial microsimulation
Commuting
Policy evaluation

ABSTRACT

The daily trip to work is ubiquitous, yet its characteristics differ widely from person to person and place to place. This is manifested in statistics on mode and distance of travel, which vary depending on a range of factors that operate at different scales. This heterogeneity is problematic for decision makers tasked with encouraging more sustainable commuter patterns. Numerical models, based on real commuting data, have great potential to aid the decision making process. However, we contend that new approaches are needed to advance knowledge about the social and geographical factors that relate to the diversity of commuter patterns, if policies targeted to specific individuals or places are to be effective. To this end, the paper presents a spatial microsimulation approach, which combines individual-level survey data with geographically aggregated census results to tackle the problem. This method overcomes the limitations imposed by the lack of available geocoded micro-data. Further, it allows a range of scales of analysis to be pursued in parallel and provides insights into both the types of *area* and *individual* that would benefit most from specific interventions.

1. Introduction

Commuting is a major reason for personal travel,[1] and a broad research area within transport geography. In many cases zonally aggregated census statistics—often the most reliable source of information about spatial variation in commuter patterns—form the basis of geographical commuting research (Horner and Murray, 2002; Titheridge and Hall, 2006). Recent advances in data availability and computational methods have, however, facilitated the analysis (Helminen and Ristimäki, 2007) and modelling (Buliung and Kanaroglou, 2002; Buliung and Kanaroglou, 2006) of commuting at *individual* and *household* levels. This trend—towards micro-level social and spatial analysis—has several potential benefits for decision makers, including:

- The ability to target specific *types* of commuters.
- The potential to model the impacts of small-scale interventions (e.g. a new bicycle path) on individuals living in the local area.
- Higher spatial resolution, allowing for realistic insight into the impacts of change on network usage (e.g. identify likely points of congestion).
- The results provide a foundation for agent-based and dynamic microsimulation models.

The shift towards micro-level analysis also has some potential disadvantages. These include greatly increased computational requirements for analysis, lack of available software or expertise, and the pitfalls of overcomplexity. As recent literature shows, new techniques for spatial microsimulation, which model individual characteristics and behaviour, can overcome the majority of these problems (see Section 2.3). A more fundamental barrier preventing the use of micro-level methods in many contexts is that accurate, geocoded microdata are simply unavailable. In the UK, for example, census-derived microdata are made available only as a Sample of Anonymised Records (SARs) at coarse geographical levels (Dale and Teague, 2002).[2] More specific surveys (such as the UK's National Travel Survey) can provide further insight into travel patterns at the individual level but these also omit high resolution geographical information to protect participants' anonymity.

We believe spatial microsimulation techniques, of the type described in this paper, hold great potential benefits for transport

* Corresponding author. Tel.: +44 07837788663.
E-mail address: rob00x@gmail.com (R. Lovelace).

[1] In the UK, for example, commuting accounts for 16% of trips and 20% of the total distance travelled by personal transport modes (DfT, 2011).

[2] The SARs are divided into two parts: the 2% SAR, which allocates each individual to a geographic region with a population size of at least 120,000 (narrowing-down the results to one or more Local Authorities), and the 1% sample, which allocates each individual to a country (Dale and Teague, 2002).

0966-6923/$ - see front matter © 2013 The Authors. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtrangeo.2013.07.008

planners and policy makers who lack access to official, geocoded microdata (individual-level data allocated to small areas). With such 'spatial microdata', new analysis options are created, including route choice between origin destination pairs, localised intervention evaluations and cross-tabulated contingency tables. These applications should also be of use in the rare (yet increasingly common) situations where official geolocated microdata are provided. In the UK, as in many other countries, spatial microdata must be simulated, as reliable secondary data sources are limited to (1) zonally aggregated census data, and (2) non-geographical, individual level microdata from national surveys. This paper builds on the pioneering theoretical work on spatial microsimulation and applies it to the issue of commuting.

The aim of the research presented in this paper is to bring micro-level analysis within reach for transport planners and researchers already acquainted with aggregated census data on commuting. Detailed non-geographical microdatasets on commuting already exist, but many analyses for evaluating the impact of commuting policies require *spatial* microdata. As indicated above, there are a number of reasons why such spatial microdata may be needed: planning for more sustainable commuting is a complex problem that operates on a range of scales, including that of individuals (Vega, 2012; Verhetsel and Vanelslander, 2010). In the words of Li et al. (2012, p. 313), "a more spatially disaggregated method is needed". To summarise the research problem, tools to aid the design and evaluation of policies affecting commuters are needed. These tools should be flexible, able to operate at a range of levels and shed light on various issues, from the potential of telecommuting (where internet access facilitates working from home, saving transport fuel) to levels of access to public transport, walkways and cycle paths.

The remainder of this paper is organised as follows: Section 2 reviews relevant literature on commuting, transport modelling and spatial microsimulation, highlighting the potential benefits of incorporating individual level socio-demographic data into transport studies. Section 3 outlines the data and methods required to fulfil this potential, and and shows how spatial microsimulation has been implemented in this paper. Section 4 presents some outputs from the spatial microsimulation model. The purpose is to illustrate the new types of analysis opened-up and policy relevance of distributional impacts. Finally, in Section 5, these results are discussed and placed in the context of current practise in transport planning and policy evaluation.

2. Literature review

2.1. Modelling commuter patterns

Commuting has been a topic of research for many decades, reflecting its role in relation to economy, to individual and household well-being and, increasingly, to environment. From this extensive literature, it is apparent that commuting should, in theory, be relatively easy to model. This is because journeys to work tend to be:

- Regular, occurring on a near-daily basis for most people and following predictable hourly, weekly and annual patterns (Akkerman, 2000).
- Non-discretionary—work trips, unlike trips made for socialising and holidays, are an essential part of daily working life. In other words, the demand for commuter travel is non-elastic, and responds slowly to changes in the cost of travel (DePalma and Arnott, 2012).
- Destination-constrained. It is often challenging to change one's work location (e.g. after moving house), as embodied in the common assumption of fixed workplaces (Vega and Reynolds-Feighan, 2009).

These characteristics mean that commuting flows should follow more regular patterns over space and time than travel for other purposes, such as holidays or shopping. In addition, commuting statistics are widely available from national censuses, which often contain a question on travel to work. This data availability and relative predictability has made commuting well-suited to academic research, and a number of methodological advances have been demonstrated using travel to work statistics.

This is well illustrated by comparing the methods of Ibeas et al. (2012) with those employed 16 years earlier by Forrest et al. (1996). In the former, four (increasingly complex) spatial econometric models were harnessed to investigate links between house prices and commuter accessibility. The latter used a single linear regression model to explore the house-price accessibility relationship with respect to a case study of Metrolink, a light rail scheme in Manchester. Increased range and complexity of methodologies can also be seen by comparing the descriptive methods used by Knowles (1996) with the statistical tests employed by Senior (2009) for exploring the transport impacts of the same scheme.[3]

The most recent major methodological advance to use commuting data is the radiation model (Simini et al., 2012). Based on census-derived inter-county commuter flow data across the USA, Simini et al. (2012) developed a probabilistic method of predicting the flows between any two zones, based only on knowledge of population and employment. If the claims stand up to further tests, this could represent a step forward in the modelling capabilities of transport geographers (Brockmann, 2012), for example by allowing individual trips to be predicted and by providing realistic estimates of commuter flows in areas where no flow data is available. In general, however, methods for investigating commuting have advanced gradually, in-line within the 'normal science' of transport geography. In addition, most modelling efforts have been constrained to the geographical scale at which data is made available.

2.2. Scales of analysis

Despite the advances outlined above many geographic approaches for analysing commuting patterns operate only at a single level of analysis. This is often the lowest geographical level for which the required data are available. Indeed, prior to the 21st century, personal transport models tended to be simplistic, assuming 'mono-centric' cities (see Fig. 3) and and taking little or no account of geographic factors beyond distance (Akkerman, 2000; Horner and Murray, 2002). This was problematic for practitioners aiming to evaluate interventions, the impacts of which may be geographically heterogeneous and highly localised (e.g. bicycle paths) or focused on specific socio-economic groups (e.g. telecommuting). Due to data, software and computing limitations, evaluations of the impacts of policies affecting personal transport have tended to be over-simplistic, considering only a single scale of analysis.[4] Ideally, however, macro (geographic) *and* micro (individual-level) factors would be included. The efforts towards such an approach "that integrates [spatial] demographic microsimulation with urban simulation and travel demand" are making progress and could signify a major step forward for personal transport models for policy evaluation (Ravulaparthy and Goulias, 2011, p. 4). Increasingly, newly available micro-level datasets are being incorporated into

[3] The former study harnessed descriptive statistics based on primary data and hand-crafted maps to investigate the transport impacts of Metrolink. The latter employed multiple regression and chi-squared tests of survey data to identify longer-term changes in behaviour attributed to the light rail system.

[4] See, for example, Lovelace et al. (2011) for a non-geograhical example of city-level aggregation, and Li et al. (2012) or Titheridge and Hall (2006) for analyses that use only a single geographical level of analysis.

Please cite this article in press as: Lovelace, R., et al. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. J. Transp. Geogr. (2013), http://dx.doi.org/10.1016/j.jtrangeo.2013.07.008

geographical analyses of personal travel and commuting in particular (the next section provides examples of this work).

Advances in software, data availability and computers have been central drivers of this methodological change, leading to a plethora of options for many types of analysis.

2.3. Incorporating the micro level

Modern computers facilitate the simulation of hundreds of thousands of simultaneous trips. A good recent example illustrating this is the work of Ferguson et al. (2012), who used microdata on company location in combination with the road network to produce traffic simulations at high spatial and temporal resolutions. A major advantage of such detail is the opportunity to test our understanding of transport systems directly, through prediction and corroboration. The close fit between simulated and independent observations made of commercial vehicles by Ferguson et al. (2012), in both space and time dimensions, illustrates the potential of combining microdata with geographical inputs for policy analysis. In the realm of public transport, Tribby and Zandbergen (2012) combined demographic data of small areas with bus and walking networks for Albuquerque, New Mexico. The results of this study (which is further discussed in Section 5) were used to evaluate the accessibility impacts of new bus routes. It was found that the impacts varied greatly between neighbourhoods and, crucially for social justice, that disadvantaged groups benefited *least* from the intervention. From a methodological perspective, Tribby and Zandbergen (2012) used the study to highlight the importance of geographical *and* socio-economic disaggregation of results. While the preceding literature is new, it is worth noting that the benefits of including spatial and non-spatial factors in personal travel analysis have been expounded since the 1970s (Horowitz, 1986). What is new is the widespread availability of data, computers and software to meet the challenge.

A couple of national-level studies serve here to illustrate the utility of analysing spatial microdata for the geographical investigation of commuting patterns. Helminen and Ristimäki (2007) investigate the relationship between distance from workplace and telecommuting in Finland. They used an individual-level geolocated database of all 2 million workers to calculate average trip distances and total annual distance travelled. As the authors note, "distance is a basic characteristic of the spatial pattern of commuting" Helminen and Ristimäki (2007, p. 333), yet it is difficult to calculate accurately in practise: Distance data are usually 'Euclidean' (provided as a straight line between home and work), yet the actual route distance travelled is almost always longer and invariably difficult to calculate. Network analysis methods have recently emerged to overcome this problem (Ehrgott et al., 2012; Levinson and El-Geneidy, 2009). However, these methods would be difficult to conduct at the national scale: Helminen and Ristimäki (2007) tackle this issue by explicitly using Euclidean distance and citing estimates of circuity (the ratio of route distance to Euclidean distance). The results illustrate the utility of geographically disaggregated microdata.[5]

Another recent application of individual-level geolocated census data to commuting policy was the investigation of the impact of location (relative to railway stations and bus stops) on sustainability of work travel in Flanders (Verhetsel and Vanelslander, 2010). As with Helminen and Ristimäki (2007), a problem encountered was the sheer size of the raw commuting database: 1.2 million individuals. This problem was overcome by aggregating the results into small areas (each containing around 130 people). The diversity of the data was tackled by classifying small areas into 5 groups, depending on the number of train stops made in each per day. Simplifying classifications may be an important way of interpreting complex spatial data, as will be seen in Section 3.

With the increasing availability of individual-level transport data geographical methods for analysing them, that are accessible to transport planners, have (in general) struggled to keep up. Notable exceptions include the work of Bhat et al. (2004), who presented an econometric microsimulation approach to modelling daily travel patterns, and Guo and Bhat (2007), who refined the iterative proportion fitting procedure of Beckman et al. (1996) to create accurate synthetic microdata for transport modelling applications. However, in neither case are methods for the *geographic* analysis of the microdata results presented. Buliung and Kanaroglou (2006) addressed this problem by developing bespoke extensions to ArcGIS software. Their toolkit facilitated the geographical analysis of the travel spaces of households based on a detailed travel-diary dataset. The research illustrates the potential for new software to pose relevant hypotheses and visualise travel patterns. The research agenda pursued by Buliung and Kanaroglou (2006) raises the following questions: Can the behaviour of *all* citizens in a study area be simulated (rather than just the survey respondents)? How can methods of individual-level transport analysis be presented and disseminated such that they are used by others?[6] Pibyl and Goulias (2005) presented an activity-based approach to the analysis of travel demand and travel schedules taking into account household characteristics.

Overall, micro-level transport models studies demonstrate the additional insight into transport patterns, and the effects of interventions, which geolocated individual-level data can provide. Unfortunately the geolocated microdatasets used by these studies are unavailable in many settings, or lack key socio-demographic variables. This is where spatial microsimulation comes in: there is a much experience in the field which, we argue, has great potential for transport researchers.

2.4. Spatial microsimulation and transport

In addition to micro-level transport models, there has been considerable progress in the development of micro-level models for analysing residential populations using spatial microsimulation. This body of work small area microdata (also often described as 'spatial microdata') by combining individual-level survey data with geographical data to simulate populations of individuals assigned to households, whose characteristics are as close to the real population as possible. The approach has added a geographical dimension to previous earlier work in which (non-spatial) microsimulation models were used to asses the impact of national government policies (e.g. Mitton et al., 2000; Redmond et al., 1998). The model outputs of spatial microsimulation include all the variables contained in the non-geographical survey data. These so-called 'target variables' can include policy relevant variables such as earned income, household type and socio-economic group, about which geographical data is unavailable (for recent reviews of spatial microsimulation methods see Harland et al., 2012; Hermes and Poulsen, 2012; Ballas et al., 2013; Birkin and Clarke, 2011).

Despite the methodological advances and growth in the number of applications over the past decade, there has been very little research applying the spatial microsimulation method to the

[5] In this case for calculating the transport impacts of telecommuting in Finland and identifying the characteristics of telecommuters (Helminen and Ristimäki, 2007).

[6] The ArcGIS-based methods of individual-level analysis and visualisation advocated by Buliung and Kanaroglou (2006) appear, based on the academic literature, not to have adopted by researchers using microdata. None of the 59 articles citing Buliung and Kanaroglou (2006) in Google Scholar (September 2012) reported using their software to investigate travel patterns using microdata, instead dealing with the broader concept of activity spaces. This is despite the efforts made to ensure the software was user friendly, with the addition of a graphical user interface (Buliung and Kanaroglou, 2006).

Please cite this article in press as: Lovelace, R., et al. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. J. Transp. Geogr. (2013), http://dx.doi.org/10.1016/j.jtrangeo.2013.07.008

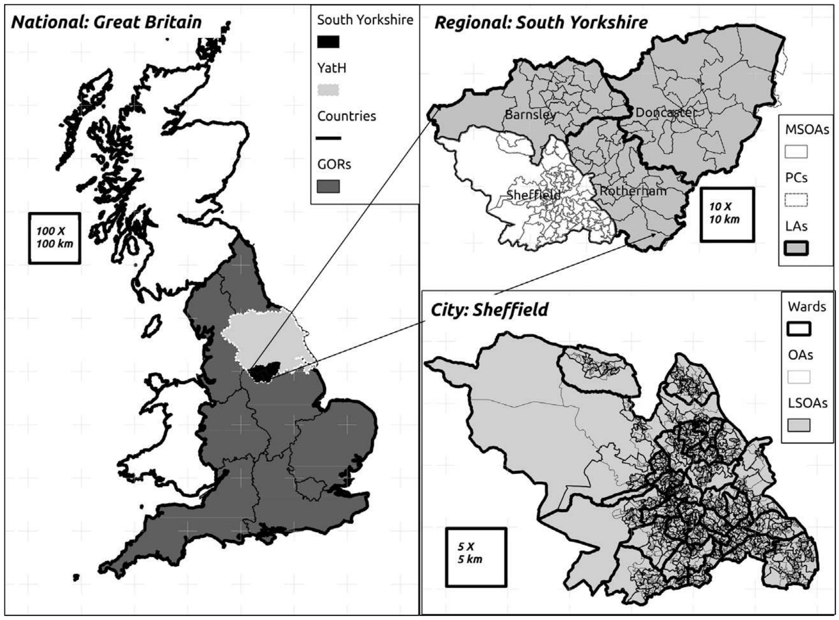

Fig. 1. National, regional and city-wide scales of analysis, as illustrated by a range of administrative boundaries. Yorkshire and the Humber (YatH, left), South Yorkshire (top right) and Sheffield (bottom right) are the study areas used for this paper. (See footNote 7 for definitions of acronyms.)

analysis of spatially variable transport patterns and policies. This is surprising, as there is great potential to enrich transport models such as the those presented in Simini et al. (2012) and Tribby and Zandbergen (2012) with additional policy-relevant attributes at the small area level using the methods and datasets typically used in population spatial microsimulation models. Based on this research problem, the following section outlines an approach that simulates individual commuters, to demonstrate the potential benefits of spatial microsimulation as a tool for policy evaluation in the absence of real geocoded micro-level data.

3. Methods and data

3.1. Selecting geographical data and scales of analysis

UK census datasets are available at a range of administrative levels, through the portal Casweb (Census Area Statistics Web) (Fig. 1). It is important to consider the range of options at the outset, because research findings can depend on the size and shapes of geographic zones, the 'areal units' of analysis (Horner and Murray, 2002; Openshaw, 1983). Selecting zones that are too small relative to the study area can lead to long processing times, messy maps and overcomplexity. Analyses based on overly large zones, on the other hand, can gloss over spatial variability by presenting space in extensive, homogeneous blocks. Regardless of the scale of analysis selected, it is important to remember that *all* analysis based on geographically aggregated data may be susceptible to the modifiable areal unit problem (MAUP) (Wong, 2009).

In contrast to research that uncritically uses only one scale of analysis, the methods described in this paper are designed to facilitate 'frame-independent' (scale independent) analysis (Horner and Murray, 2002). Spatial datasets related to commuting in the UK, and their scales, are outlined in Table 1.[7]

These datasets are all derived from the National Census, which takes place every 10 years and covers every individual living in the UK by law, so are highly reliable. However, they vary in terms of geographical scale and precision: Casweb data is the most reliable, as it is taken directly from the national Census (which happens every 10 years) and provides precise absolute count data at every level for which the data is provided. The Nomis data is also taken from the Census. Its main advantage is that it provides more cross-tabulation options for small (sub-ward level) areas than Casweb data. To deal with this, Nomis applies a randomisation algorithm to the count categories to avoid individuals being identified from the various cross-tabulations provided. In addition, following the 'confidentiality principle' of census data release (Rees and Martin, 2002), small numbers (3 or below) are allocated as either 0 or 3 in the Casweb data, meaning reduced accuracy for small counts. This makes cross-tabulated datasets of unusual categories such as 'cycles to work' unreliable at the smallest Output Areas (OA) level. Census data are the 'gold standard' in terms of accuracy and comprehensive geographical coverage (Rees et al.,

Table 1
Aggregate data related to commuting behaviour and the scales at which they are available for South Yorkshire. The abbreviations are as follows: OA refers to Output Areas, LS and MS referring to Lower and Medium Super Output Areas respectively; ST Wards refer to Statistical Wards, and LA refers to Local Authorities. Source: Casweb (Census Area Statistics Web), unless otherwise stated.

Variable	OA	LSOA	MSOA	ST Ward	LA
N. zones in South Yorkshire	4278	845	173	59	4
Average population	296	1450	7320	21,500	317,000
Mode of transport to work	Y[a]	Y	Y	Y	Y
Average distance	N	Y	Y	Y	Y
Distance categories	Y[a]	Y[c]	Y[c]	Y	Y
Car access[b]	Y	Y	Y	Y	Y

[a] Output area statistics are often unreliable because values less than 3 are randomly allocated the value of 0 or 3. This is problematic for sparsely populated categories such as those who travel 60 km or more to work.
[b] 'Car access' refers to the census dataset 'cars or vans' which provides counts for the number of houses with access to no cars, one car etc., and total number of cars in each area. This is for estimating reliance on public transport.
[c] Data provide by Nomis (obsolete acronym of the National Online Manpower Information System) government data portal, providing various cross-tabulation options (https://www.nomisweb.co.uk/Default.asp).

[7] The administrative acronyms OA, LSOA, MSOA, and LA refer to Output Areas (which contain ~300 people), Lower Super Output Areas (~1600 people), Medium Super Output Ares (~7000 people) and Local Authorities (more than 100,000 people) respectively. PCs are postal codes, used in some geographical classifications.

Table 2
The four constraint variables and their associated categories used as the aggregate-level inputs into the spatial microsimulation model. The category notation for numeric variables follows the International Organization for Standardization (ISO) 80000-2:2009: Square brackets indicate that the endpoint is not included in the set, curved brackets indicate that the endpoint is included.

Variable	N.	Categories/bin breaks	Comments
Age/sex	12	(16, 20] (20, 25] (25, 35] (35, 55] (55, 100]	Female and male categories, in employment (excludes full-time students)
Mode	11	mfh metro train bus moto car.d car.p taxi cycle walk other	Main mode of travel to work (no data on variability of mode choice)
Distance	8	(0, 2] (2, 5] (5, 10] (10, 20] (20, 30] (30, 40] (40, 60] (60, 250]	Euclidean distance between respondents' home postcode and their main place of work (does not capture multiple work destinations)
NS-SEC	9	NS-SEC 1.1, 1.2, 2, 3, 4, 5, 6, 7 and other	Classes range from higher managerial (NS-SEC 1.1) to routine occupations (NS-SEC 7) – see (Chandola and Jenkinson, 2000) and on the ONS website (http://www.ons.gov.uk)

2002, p. 4), but lack cross-tabulations and insights into individual-level variability within zones, details which are essential for the 'intelligent' analysis of commuter patterns including individual-level factors. For the purposes of this paper, we use the MSOA level of geographical aggregation (see top right, Fig. 1)) to constrain the spatial microsimulation model, a medium level of geographic detail. The four constraint variables used were age/sex (to capture the demographic profile of each area), mode and distance travelled to work (for insight into commuter patterns) and National Statistics Socio-Economic Classification (NS-SEC) to provide information about class (a proxy for income). These constraints are described in Table 2.

We selected these variables based on a consideration of possible correlations between them and the 'target variables' from the survey dataset that we are aiming to estimate—in this case travel behaviours and socio-economic profile, such as income. This to selecting constraint variables is recommended by Ballas et al. (2007) and Edwards and Clarke (2009), who used correlation and regression analysis to identify the appropriate constraints. In our case the selection process was more straightforward: mode and distance are the only commuting-related variables available at the small-area level; age/sex are needed to create a realistic population profile (age is also related to commuting behaviour and income); and NS-SEC is a good indicator of socio-economic background and disadvantage (Anderson, 2013; Ballas et al., 2007; Kavroudakis et al., 2012). The model presented in this paper is underpinned by the assumption that our target variables are associated with the geographical constraint variables.

3.2. Microdata

The aggregated census data described above form a solid foundation for analysing commuting patterns. However, they omit a number of relevant variables and mask intra-zonal variability. Table 3 illustrates some important individual level, transport related variables ('target variables', to be estimated at the local level) that are available through a single dataset: the Understanding Society dataset (USd).[8]

It should be noted that the USd variables described in Table 3 are *proxies* of the attributes assigned to them: therefore they should be interpreted with caution. The propensity of households to move (linked to commuting via job mobility), for example, does not just depend on the number of children:[9] it also depends on other factors such as the ownership status of the house, years left on mortgage, time spent at current location and satisfaction with the local community (Mellander et al., 2011). Some of this information is in fact provided by the USd (in variables 'hsownd' and 'mglife', at the household level and 'mvyr' and 'lkmove' in the individual questionnaire): Table 3 represents only a snapshot of the available variables.

Some individual-level data may be inferred based on aggregate-level data. The number of people who commute in the same direction, for example, can be derived from commuting flow data provided by Nomis. This additional level of detail, about *where* people travel to and from may well be relevant for decision makers. Simulation of the potential for trip sharing or for the increased uptake of non-motorised modes along frequently used home-work corridors, for example, could harness this data. However, for the individual-level investigation of *socio-economic* variables not included in census aggregates related to commuting behaviour, spatial microsimulation is needed. The geographical data presented in Table 1 can be combined with the non-geographical microdata presented in Table 3.

3.3. Spatial microsimulation

As noted in Section 2.4, spatial microsimulation models generate individual-level data allocated to administrative zones. The methods underpinning spatial microsimulation models range from synthetic reconstruction techniques to combinatorial optimisation and deterministic reweighting based on iterative proportional fitting (IPF). *'Synthetic reconstruction'* approaches are typically used in situations where the only available data are small area cross-tabulations (e.g. see Table 2) and suitable microdata (e.g. Table 3) are unavailable. The first demonstration of this technique was presented by Birkin and Clarke (1988, 1989) in their Synthetic Spatial Information System for Urban and Regional Analysis (SYNTHESIS). This system used Monte Carlo sampling in combination with a statistical method known as Iterative Proportional Fitting (IPF) (Deming and Stephan, 1940; Mosteller, 1968; Fienberg, 1970; Wong, 1992) to combine joint-probability distributions from small area census tables. The increasing availability of suitable social survey microdata sets from the 1990s onwards, combined with major advances in computer hardware and software technologies created the enabling environment for the *'combinatorial optimisation'* approach. The method work by searching for the optimal combination of individuals and households from survey microdata to match aggregated count data. The first application of combinatorial optimisation algorithms to produce spatial microdata was presented by (Williamson et al., 1998). 'Hill climbing', 'genetic algorithms' and 'simulated annealing' algorithms were tested by combing UK census data with the Sample of Anonymised Records (SAR). This work has led to further refinements and applications of the method (e.g. Voas and Williamson, 2001; Williamson et al., 2002; Ballas et al., 2006; Morrissey et al., 2008; Kavroudakis et al., 2012).

[8] Understanding Society replaces the British Household Panel Survey (BHPS) as the UK's largest national governmental survey (see www.understandingsociety.org.uk). The Department for Travel's National Travel Survey and the Living Costs and Food Survey provide additional options for individual-level variables related to commuting. The USd is the most comprehensive (with a longitudinal sample size of 50,000), so it is used here.

[9] To provide another example, the USd provides three categories of car engine size rather than describing the exact make and model, a substantial oversimplification from the perspective of energy use.

Please cite this article in press as: Lovelace, R., et al. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. J. Transp. Geogr. (2013), http://dx.doi.org/10.1016/j.jtrangeo.2013.07.008

Table 3
Selected individual-level variables related to commuting, available from the Understanding Society dataset.

Attribute	Variable	Measurement	Comment
Type of car	Household variable 146	Engine size of cars: <1.4, 1.4–1.9, or ⩾2*l*	Data on additional cars also available
Household income	Household variable 193	Net household income, £/month	Equivalised income must be calculated
Telecommuting potential	Individual-level variable 953	7 point scale from ''no access'' to ''everyday''	Must be linked with type of work
Ease of moving home	Household variable 171	Number of children (aged 15 or under) in household	One indication of how settled household is

It should be noted that synthetic reconstruction and combinatorial optimisation techniques rely on the use of random number generators and (partly as a result) tend to be computationally intensive (Pritchard and Miller, 2012). An alternative approach, which does not rely on random numbers (and can therefore be guaranteed to produce the same output with each run) is the so called '*deterministic reweighting*' approach (Ballas et al., 2005b). This, like synthetic reconstruction, uses IPF to weight individuals: individuals highly representative of a zone will receive a high weight. There have also been many applications and refinements to this technique (e.g. Anderson, 2013; Campbell and Ballas, 2013; Edwards and Clarke, 2009; Smith et al., 2009). A key advantage of IPF-based deterministic models over probabilistic combinatorial methods is that the results are the same with each model run. Other benefits include the robustness and reliability of the technique (Mosteller, 1968; Fienberg, 1970; Wong, 1992), as well as its speed and simplicity (Pritchard and Miller, 2012; Lovelace and Ballas, 2013). A major disadvantage for certain applications (e.g. agent based models) is that deterministic reweighting produces non-integer weights: researchers are not left with whole individuals to deal with, but many (in many cases tiny) fractions. However, this problem can be overcome using 'integerisation' (Ballas et al., 2005a; Edwards and Clarke, 2009; Lovelace and Ballas, 2013).

Each of the approaches discussed above result in the synthesis of spatial microdata: by combining small area census data with survey data, the individuals from the survey which most closely match the aggregated data are preferentially selected (or given high statistical weights). In other words, the models simulate virtual populations to match real aggregate data. Having considered all the methods that have been tried and tested to date and in particular the strong arguments regarding the robustness, simplicity and computational speed of IPF-based approaches, it was decided to adopt the deterministic reweighting technique presented by Ballas et al. (2005a) and the refinements and integerisation method presented by Lovelace and Ballas (2013).

There is an ongoing debate within the spatial microsimulation community about which methods and refinements are best suited to different situations (e.g. Clarke and Harding, 2013). The intention of this paper is not to intervene in the methodological debate, however, but to flag the alternative techniques available: this paper focusses on methods for *applying* spatial microdata to the analysis of commuting patterns. Suffice to say that the methods presented here would be compatible with spatial microdata generated through any means; deterministic reweighting via IPF is used here as it is widely considered to be well-established, fast and accurate (Wong, 1992; Pritchard and Miller, 2012; Lovelace and Ballas, 2013). The code used in this paper is written in R, is freely available and is described in more detail in a ''user manual'' for generating integer weights from the IPF procedure (Lovelace and Ballas, 2013, Supplementary Information).

As with any modelling method, spatial microsimulation simplifies real-world complexity. Its ability to create individual-level results should therefore be treated with caution, and interpreted as a modelling tool to aid understanding rather than 'real data'. It is important that practitioners are aware of its limitations before using it as a decision-making tool. These limitations, and strategies to overcome them, can be summarised as follows:

- The individual-level results are simulated, and are unlikely to be totally representative of the zones in question. We can have confidence in the contrained variables (although large bin sizes for continuous attributes such as age may not fully capture unusual distributions),[10] but the target variables are simply the result of their relationship with constraint variables at the national level. This can be tackled through validation methods (see Edwards and Clarke, 2009, and below) or, in the long run, through increased access to real spatially disaggregated microdata.[11] In fact, awareness of the policy insights offered to researchers by spatial microdata could encourage the release of real geographically disaggregated microdata (see Lee, 2009).
- Lack of accurate distance travelled estimates in the main model (currently broad distance categories are used). This could be overcome by creating more accurate origin–destination pairs for individuals. Lower-level commuter flow data (compared with the data presented in Fig. 5) is available to do this.[12] Also, undertaking network analysis of roads, railways, and walkways (see Fig. 6 for an example) for all individuals could allow more accurate estimates of route distance. However, this is computationally challenging, although increasing feasible (Gao et al., 2010).
- Omission of explanatory variables such as car parks, the quality of paths, and even the provision of showers for cyclists at work destinations. These variables can be included by appropriate survey questions (Buehler, 2012) or analysis of environmental variables (Rietveld, 2004).

Each of these issue presents a major methodological challenge, but none of them invalidates spatial microsimulation as a modelling tool to better understand travel behaviour. These issues are partly tackled in Section 3.6 and their implications discussed in the final section.

3.4. Model implementation

The method requires both aggregate and individual level datasets described in Sections 3.1 and 3.2 to share at least one 'linking variable'. These linking (or constraint) variables, described in Table 2, preferentially sampled representative individuals, in this case via the deterministic reweighting technique of IPF (Wong, 1992;

[10] The distance bins presented in Table 2, for example, are quite widely spaced. In a situation where many people travelled a distance close to the edges of one of these bins—for example due to a factory located 11 km from an employment centre—the results, which would represent an even distribution of all individuals in the sample who 10–20 km to work, would be inaccurate.

[11] For example, a dataset of geo-coded individuals and their workplaces provided by Finnish government allows destination/origin analysis and insights into the directions of flow (Helminen and Ristimäki, 2007)

[12] Commuter flow datasets of the type presented in Fig. 5 are available at the much smaller Output Area level (from http://www.ons.gov.uk/ons/guide-method/census/census-2001/data-and-products/data-and-product-catalogue/origin—destination-statistics/output-areas/index.html the Office of National Statistics). However, the data are available only on a DVD, with the following proviso: ''analysis [of the Output Area commuter flow data] requires the use of specialist software, which is not supplied with the product, but which is available from intermediary organisations (for more information contact Census Customer Services).''

Please cite this article in press as: Lovelace, R., et al. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. J. Transp. Geogr. (2013), http://dx.doi.org/10.1016/j.jtrangeo.2013.07.008

Table 4
Sample of linking variables at the individual level (USd).

ID	Age/sex	Mode	Distance (km)	NS-SEC
1	Male, 59	Car driver	3	Lower management
2	Female, 51	Car driver	9	Higher professional
3	Male, 31	Car driver	2	Other
4	Female, 24	Walk	1	Lower management

Table 5
Sample of linking variable values for zones. The population of the most populous category is presented for each variable.

Variable ⇒	Age/sex	Mode	Distance (km)	NS-SEC
Area code	Males, 35–54	Car drivers	5–10 km	Lower management
E02001509	116	1616	914	499
E02001510	94	1430	665	402
E02001511	82	1467	848	340
E02001512	152	2280	573	791

Pritchard and Miller, 2012). The target variables (Table 3) are thus simulated.

The mathematics (Fienberg, 1970) and code (Lovelace and Ballas, 2013, Supplementary Information) used to implement IPF are described in detail elsewhere; here we illustrate in more intuitive terms how the model works. Tables 4 and 5 provide samples of the raw constraint data, on aggregate and individual levels respectively. The spatial microsimulation model works by adjusting a large array of weights—rows corresponding to individuals and columns corresponding to the geographic zones under investigation—iteratively, to maximise the fit between simulated and census data. Assuming temporarily that only the four individuals represented in Table 4 were used, constraining by the distance variable in Table 5 would lead the individual with an ID of 2 to be allocated a weight of 914 for zone 1, 665 for zone 2 etc., as they are the only person who fits into that category. Clearly, many other individuals, with other characteristics would fit into the 5–10 km distance category in the entire microdataset, and this diversity is what allows the weights to converge towards a single result for each individual-zone combination (Fig. 2).

To ensure the model is working, the simulated micro-data are aggregated and then compared with census data. Total absolute error (TAE), a simple and effective goodness-of-fit metric (Williamson et al., 1998; Voas and Williamson, 2001), was calculated after constraining for linking variable and after each complete iteration (Fig. 2). Further validation tests are described in Section 3.6.

The weighted data provided by IPF-based spatial microsimulation is bulky (containing rows even for individuals who contribute very little: whose weight is close to zero) and prevents certain types of analysis. To tackle this problem, and provide a single dataset for analysis using various techniques (e.g. individual-level, geographic, or agent-based methods), the ‘truncate, replicate, sample’ method of integerisation was used Lovelace and Ballas (2013). Still, the final output dataset contained 532,130 rows, representing every commuter in South Yorkshire.

3.5. Assigning work location

The spatial microsimulation model results in a large dataset containing hundreds of individuals for each zone under investigation. For micro-level spatial analysis, origin–destination pairs are needed: simulated places of home and work need to be geotagged. The simplest solution to this problem is to allocate all individuals in each zone home coordinates corresponding to the zone’s population-weighted centroid. Likewise, work coordinates can be set to the nearest employment centre. This method allows for simple analyses such as the proxy for geographic isolation presented in Fig. 3.

Rather than assuming that work centres are always located in the city centre, a more realistic approach is to acknowledge that a variety of employment centres exist, and that the relative importance of each varies from place to place. This is illustrated in Fig. 5,, a ward-level flow diagram of the work locations of commuters based on the outskirts of Sheffield. Although Barnsley is the closest city centre to Stocksbridge (see Fig. 3), this analysis makes it clear that Sheffield is the primary non-home workplace.

At an even finer geographical level, it is possible to discern the localities within each city and ward where people are most likely

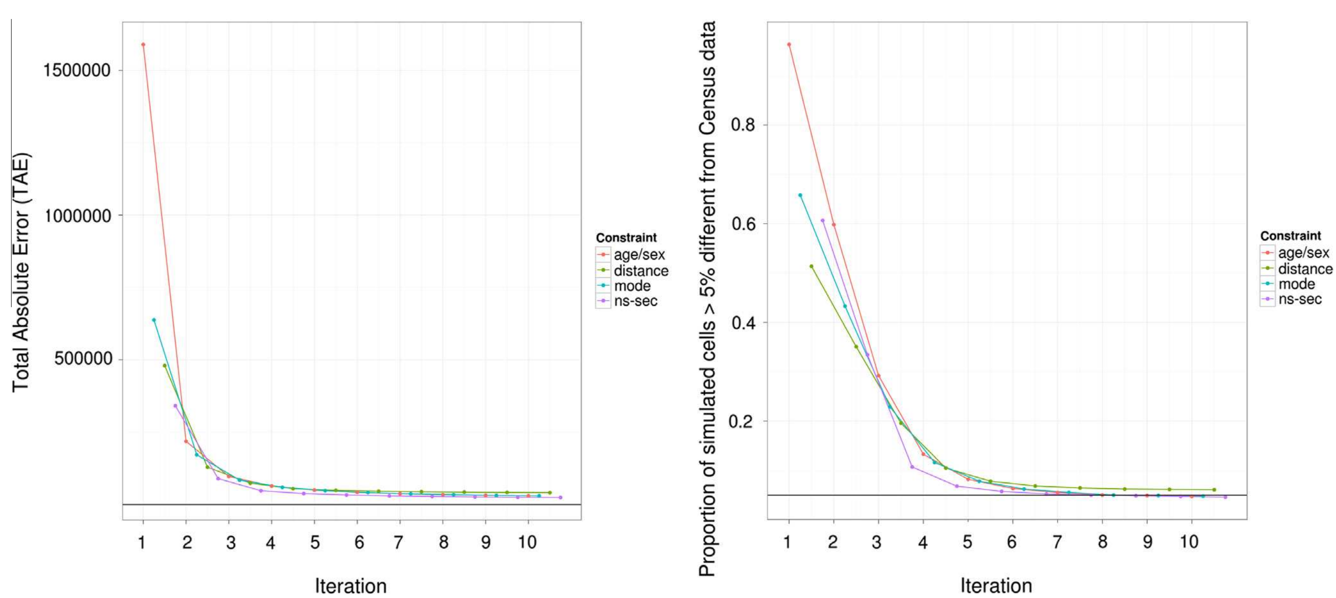

Fig. 2. Improving fit between simulated and census data across all 4 constraint variables outlined in Table 4, as illustrated by decreasing values of the total absolute error (TAE) (left) and decreases in the proportion of simulated aggregate cell values that differ from census data by more than 5% (right) after each constraint and iteration. The horizontal black lines represent 0 error and 5% of cell values, respectively.

Please cite this article in press as: Lovelace, R., et al. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. J. Transp. Geogr. (2013), http://dx.doi.org/10.1016/j.jtrangeo.2013.07.008

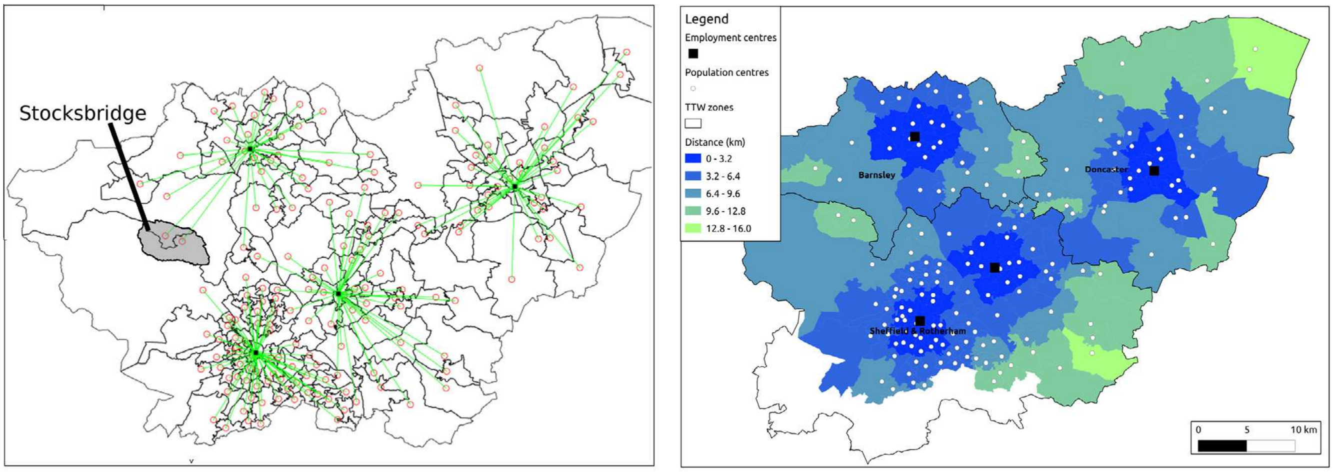

Fig. 3. Average distance to employment centre in South Yorkshire. The left-hand map illustrates how distance was calculated (using the command nncross () in the R package 'spatstat'). The right-hand map illustrates the results—Sheffield and Rotherham are grouped together in the same travel to work zone.

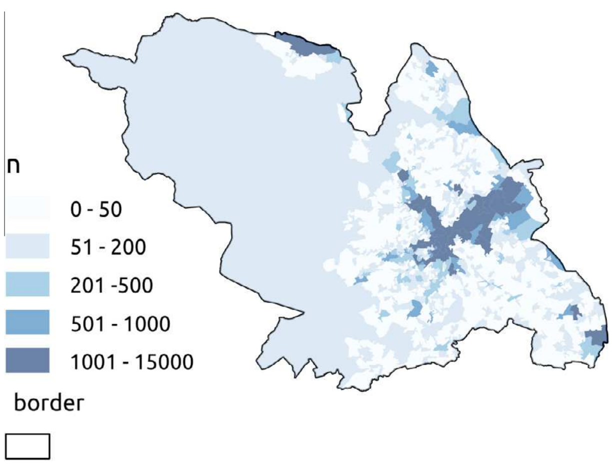

Fig. 4. Employment density at the local level in Sheffield (*n* is the number of employees registered to each zone). These results were generated by summing all incoming flows to all of Sheffield's 1744 Output Area (OA) administrative zones. Data provided on a CD, on request from http://www.nomisweb.co.uk/.

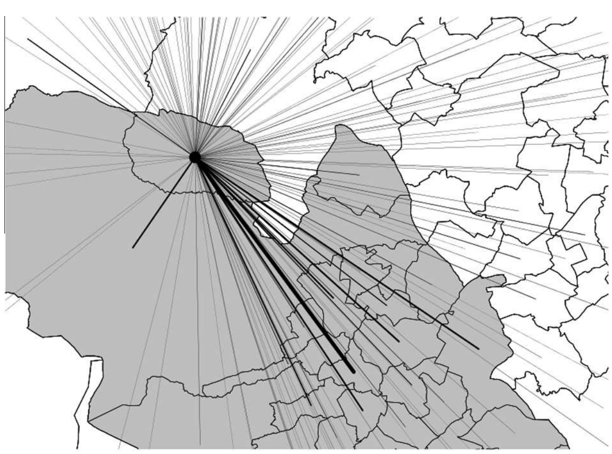

Fig. 5. Flow diagram illustrating popular commuter destinations for citizens of Stocksbridge. The thickness of the lines is proportional to the number of people who travel there (for reference, 661 people travel to the centre of Sheffield—illustrated by the thickest line—and 2036 people work in Stocksbridge—illustrated by the dot from which all lines radiate. $n = 6338$).

to work based on UK census data. This is illustrated in Fig. 4. Although this level of geographic detail was not used in the final results due to aggregation issues,[13] it demonstrates the potential for highly localised work allocation based on census-derived flow data.

The analyses presented in both Figs. 3 and 5 both greatly oversimplify trip routes. The straight lines underestimate travel distance, completely ignoring the transport network. A more realistic method is to randomly allocate each individual to a unique home location based on population density (or, potentially, local area classification) and estimate the route taken using shortest trip algorithms dependent on the mode of transport used (Fig. 6). This latter method allows for the calculation of route distances by mode, but is more complex and difficult to implement over large areas.

These methods of spatial analysis provide great insight into the meaning of aggregate statistics for groups of individuals at the city level of policy intervention. However, to gain insight into the impacts of schemes on individuals and local communities, agent based models may be needed. In particular, there is great potential to link the work presented here with relevant agent-based simulation work in the social sciences (e.g. Gilbert and Troitzsch, 2005; Gilbert, 2007) and attempts to add a geographical dimension to this work (see Wu et al., 2008).

To this end Fig. 6 presents the simulated route choice of the 18 commuters selected from the spatial microsimulation model, and contains both socio-demographic and geographic detail.[14] The distances travelled along the transport network are clearly substantially further than represented by simple straight lines. This concept can be defined formally as *circuity*, the ratio of straight-line distance to route distance (Ballou et al., 2002). Fig. 7 illustrates the impact of the road network on distance travelled. Overall, the route distance represented in Fig. 6 is 223 km, 24% further than the straight-line distance (179 km) for the 17 commutes. As in previous studies, circuity tends to decrease approximately logarithmically as a function of distance (Levinson and El-Geneidy, 2009). The spatial microsimulation method holds great potential for investigating the impact of

[13] The Output Area flow data presented in Fig. 4 is difficult to work with for individuals allocated to specific zones, because any number between 1 and 4 is randomly set as either 0 or 3. This makes the flow data essentially probabilistic for single Output Area pairs, hence our limitation to aggregate-level analysis of this dataset here.

[14] For example, the simulated car passenger who commutes to central Sheffield in Fig. 6 is 16 years old, is classified as class 'other', and lives in a family that has access to 5 cars. These, and further simulated details such as income, could, once validated, contribute towards transport interventions targeting specific commuter groups.

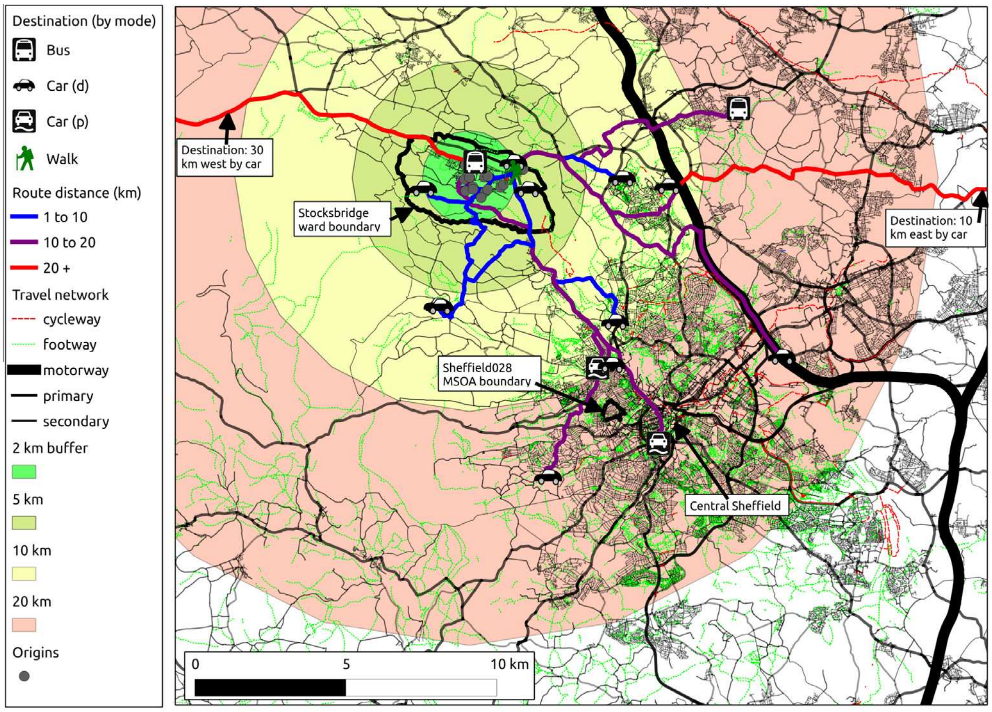

Fig. 6. Simulated route choice for 20 randomly selected individuals from the spatial simulation model. Destinations were determined by (1) subsetting destination wards by distance from Stocksbridge centre, (2) assigning probabilities of working in each ward for each distance band (based on flow data presented in Fig. 5) and (3) randomly selecting points within the resulting destination wards. (Workplaces of 3 people who work from home are not mapped).

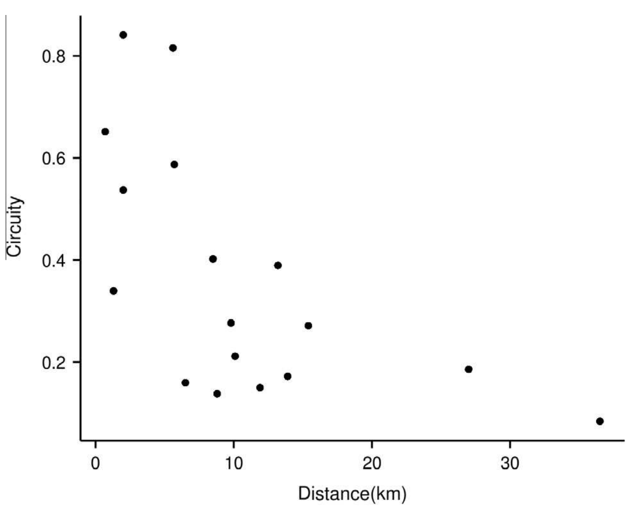

Fig. 7. The circuity of the route distance as a function of the straight-line distance for 17 commuter trips modelled in Stocksbridge.

the travel network, especially when combined with new tools for batch-processing of shortest-route algorithms.[15]

3.6. Model validation

Due the dangers of using incorrect model data to inform policy, the importance of validation has been emphasised repeatedly in the spatial microsimulation literature (Clarke and Holm, 1987; Chin and Harding, 2006; Smith et al., 2009; Edwards et al., 2010; Ballas et al., 2013). Because the outputs of spatial microsimulation are by nature detailed and provided at the individual-level, validation is challenging: "such detailed information is virtually never available at the disaggregate level for an entire region" (Ravulaparthy and Goulias, 2011, p. 37). In fact, one could argue that if individual microdata were made available at the small area level, spatial microsimulation would be obsolete.

Researchers using spatial microsimulation have been innovative at overcoming this 'catch 22' situation, using a variety of methods. In broad terms, there are two types of strategy available: internal and external validation (Edwards and Tanton, 2013). The first of these is relatively straightforward: the aggregated constraint variables are compared with the aggregated results of the spatial microsimulation model for the same variables. In our model, the results of this test were reassuring: the correlation between the aggregate counts from the census and those generated in our spatial microsimulation were 0.9989 overall for all 6920 data points (40 categories by 173 zones). However, the quality of the fit was better for some constraint variables than for others: the r^2 values for the distance and mode variables were 0.9993 and 0.9983, primarily due to the inaccuracy or our estimates of individuals who work mainly from home (mfh) (Fig. 8).

This internal validation result is less impressive when one considers that IPF always converges towards the optimal result for known constraint variables: it is the unknown cross-tabulations and target variables that we are trying to simulate with spatial microsimulation, so external should, in most cases, be the focus (Morrissey et al., 2008; Edwards and Tanton, 2013). Four methods of corroborating spatial microsimulation results with external data were identified:

[15] The analysis conducted one trip at a time, using the QGIS plugin "Road Path" for a simple solution with a user-friendly interface. (http://docs.qgis.org/2.0/html/en/docs/user_manual/plugins/plugins_road_graph.html http://plugins.qgis.org/). To automate the process, Routino (http://www.routino.org/), PGRouting (http://pgrouting.org/) or the recently released R package osmar (http://cran.r-project.org/web/packages/osm) could be used. The rapid evolution of transport network data and software provides avenues for methodological advance.

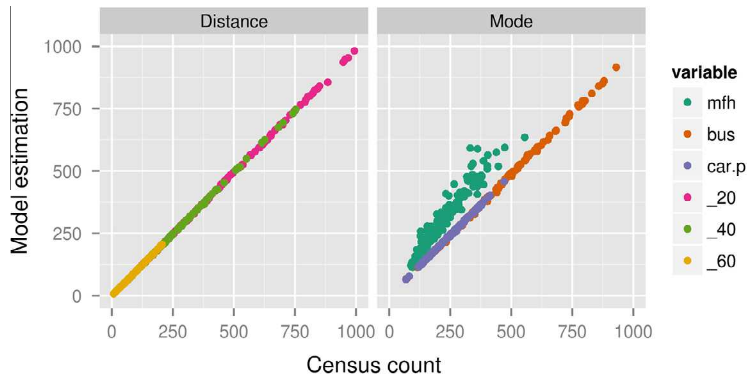

Fig. 8. Comparison of census and simulated results at the aggregate level for a selection of six categories from the mode and distance constraints. The "_20" category, for example, refers to the number of people travelling 10–20 km to work.

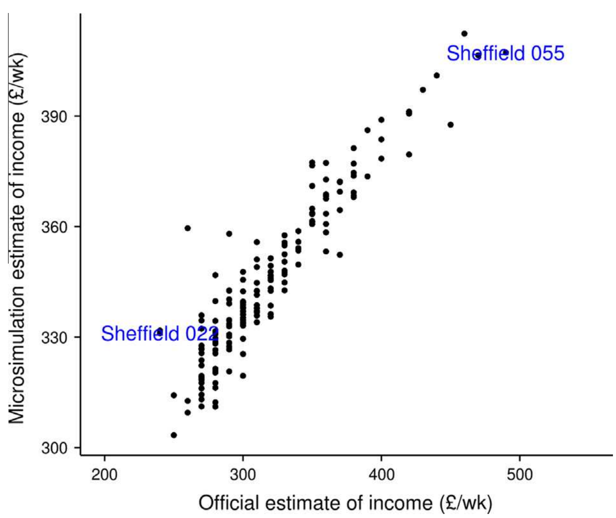

Fig. 9. Scatter graph of mean equivalised household income produced as an output from the spatial microsimulation model (y axis) and official estimates from the Office of National Statistics for the 173 Medium Super Output Areas of South Yorkshire. (Max. and min. official estimates labelled in blue.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Contingency table illustrating the link between 2nd most common mode of TTW in an area and average values for other variables.

2nd mode	N. zones	Total (%)	$\bar{D}$ (km)	$\bar{P}car$ (%)	$\bar{D}ens$ (People/km^2)
MFH	18	10	17.0	68	31
Tram	4	2	10.8	53	179
Bus	95	55	11.2	54	106
Car (p)	10	6	13.5	63	40
Foot	46	27	13.2	53	112

- Compare simulation results with real spatial microdata.[16]
- Collect primary data from specific areas against which the simulated results can be tested.[17]
- Compare simulation results at the aggregate level with estimates from a dataset external to the model (Morrissey and O'Donoghue, 2013).

[16] Income, for example, is collected by the Census, but is not disseminated at aggregate levels, let alone the individual-level geocoded data required to validate the individual-level results of the spatial microsimulation model. Access to such sensitive real microdata limits the applicability of this method.

[17] In some cases (e.g. environmental attitudes) this may be the only reliable validation option, as the data is simply not collected in geo-coded surveys.

Table 7
Summary statistics of the commuting behaviour of individuals in South Yorkshire disaggregated by mode. (Motorbike, taxi, metro and 'other' modes have been removed for brevity).

Mode	N.	%	% National	Age	Distance (km)	Ncars
Bus	31,486	7.2	7.4	38.3	7.5	0.5
Car (d)	268,496	61.1	54.6	40.1	14.3	1.9
Car (p)	38,233	8.7	5.9	33.5	14.5	1.5
Cyc	4498	1.0	2.6	38.3	5.0	1.1
MFH	45,326	10.3	9.3	40.0	0.0	1.9
Train	5709	1.3	4.6	36.9	24.6	1.2
Walk	38,406	8.7	9.7	36.6	3.1	0.8
Average	–	–	–	39.0	11.3	1.6

- Aggregate-up the small area estimates provided by spatial microsimulation to compare the results with real data that *is* provided at higher geographies (Edwards and Clarke, 2009).

Each of these options was considered for our case study, but data constraints meant that only one, comparison of aggregate data on a target variable with a reliable external dataset, was deemed viable. The target variable chosen for this was income; Neighbourhood Statistics provides estimates of this at the MSOA level, allowing for direct comparison with our results (Fig. 9). The results show high levels of correlation ($r^2 = 0.93$) between simulated incomes and official estimates, although the spread of the values resulting from spatial microsimulation underestimated the true level of inter-zone variation in average incomes (see Table 6).

4. Results

4.1. Aggregate-level results

Our results show that, at the aggregate level, South Yorkshire's commuting behaviour is comparable to the national average. Nevertheless, the microdata illustrate inter- and intra- zone variability. Table 7 illustrates the cross-tabulations (contingency tables) that are made possible when spatial microdata are used. Univariate statistics are available on mode of transport, age and number of cars but the interaction between these variables remains hidden in aggregated Census data.

Beyond illustrating the capability of spatial micrsimulation to provide estimated cross-tabulations of aggregate level data, Table 7 also provides substantive information about commuting patterns that could be applied to transport policy:

- Cars dominate travel to work in South Yorkshire, to an even greater extent than in England as a whole.

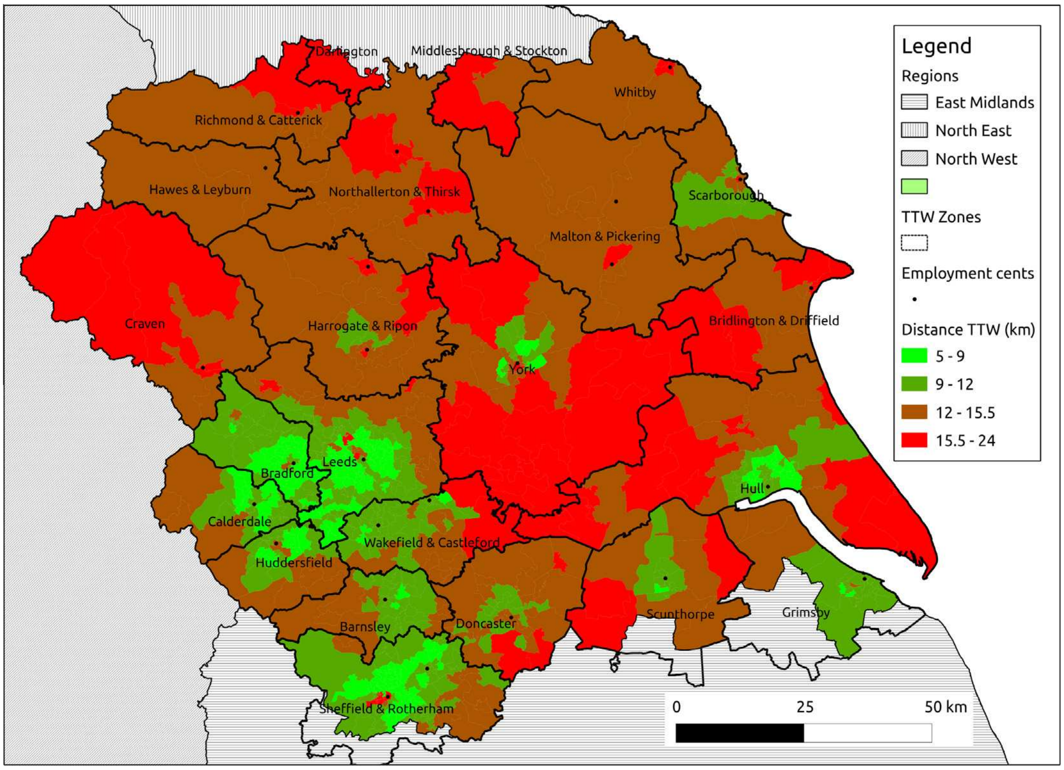

Fig. 10. Average distance travelled to work in Yorkshire and the Humber by MSOA zone. Black lines represent TTW zones.

Table 8
Contingency table of average values for continuous variables related to commuting, cross-tabulated by income bands, based on the spatial microsimulation model for South Yorkshire (n = 531,282).

Income group	Proportion	Age	Dis (km)	N.cars	Income (£/yr)	N.child
v.poor	10%	38	5.8	1.2	5519	0.9
poor	18%	39	8.1	1.2	10,158	1.0
below.av	22%	39	8.3	1.4	13,974	0.8
above.av	18%	39	8.9	1.6	17,902	0.6
affluent	32%	40	16.5	1.9	29,448	0.5

- The dominance of cars is even greater when measuring travel to work in terms of distance travelled: car commuters travel on average further than all other types of commuters bar those who commute by train.
- There are also substantial differences in the age profiles of different commuting modes: walking, which is often associate with older members of society, appears to be more prevalent amongst the young. Bicycle commuters, who are sometimes stereotyped as young (Daley and Rissel, 2011), are not much younger than than the average. Car drivers and home workers tend to be slightly older.
- Car ownership, which is seldom factored-into transport policy assessments, (Kay et al., 2011) varies with the mode of travel to work. Those who catch the bus or walk are least likely to own a car, while a those who drive to work or work from home own on average almost 2 cars per household.

4.2. Geographic variability

The results show a strong relationship between location and distance travelled. The role of location, and distance to employment centres more specifically as a cause of distant commutes was explored using travel to work (TTW) zones, defined by the Office for National Statistics at the wider regional level of Yorkshire and the Humber (Fig. 10).[18] Fig. 10 shows that MSOA areas located in and around the conurbations surrounding Bradford, Sheffield and Hull tend to have low average commuter distances, while rural locations such as the North York Moors are associated with long average commutes. This result differs from that of suburban USA (where urban sprawl accounts for high commuting costs even within major conurbations), but it is hardly new or surprising (Marshall, 2008; Sexton et al., 2012). An unexpected result is the tendency of city centres to be associated with high average commuter distances. This can be seen in red patches surrounded by a sea of green in the centres of Bradford, Leeds, Scarborough and Sheffield. (One hypothesis to explain this is as follows: some city centres attract wealthy individuals, who tend to commute further, often by train.)

[18] The wider regional level of analysis of Yorkshire and the Humber (see Fig. 1) was used in this case because TTW zones are large: only 3 are found in South Yorkshire (Fig. 3), so a larger area is useful to see the overall pattern. Travel to work zones are defined as "zones with a self- containment of at least 75% (which is to say that less than 25% of those who work in an area live outside it, and less than 25% of the employed residents of that area commute to workplaces outside the same area)" (Coombes and Openshaw, 1982).

Table 9
Contingency table of average values for continuous variables related to commuting, cross-tabulated by income bands, based on the spatial microsimulation model for the Ward of Stocksbridge ($n = 6338$) and MSOA Sheffield028, which corresponds to Crookes ($n = 2470$).

Income group	Proportion	Age	Dis (km)	N.cars	Income (£/yr)	N.child
Stocksbridge (13 km from centre)						
v.poor	10%	39	9.5	1.2	5886	1.0
poor	21%	38	12.3	1.0	10,571	0.9
below.av	19%	39	12.3	1.5	14,560	0.7
above.av	20%	39	12.9	1.8	18,513	0.5
affluent	30%	40	17.1	2.0	29,198	0.5
Crookes (2 km from centre)						
v.poor	10%	32	4.0	1.1	5208	0.8
poor	16%	33	5.7	0.9	9972	0.9
below.av	23%	31	7.4	1.1	14,145	0.5
above.av	14%	34	8.7	1.5	17,914	0.5
affluent	37%	36	25.0	1.8	29,932	0.4

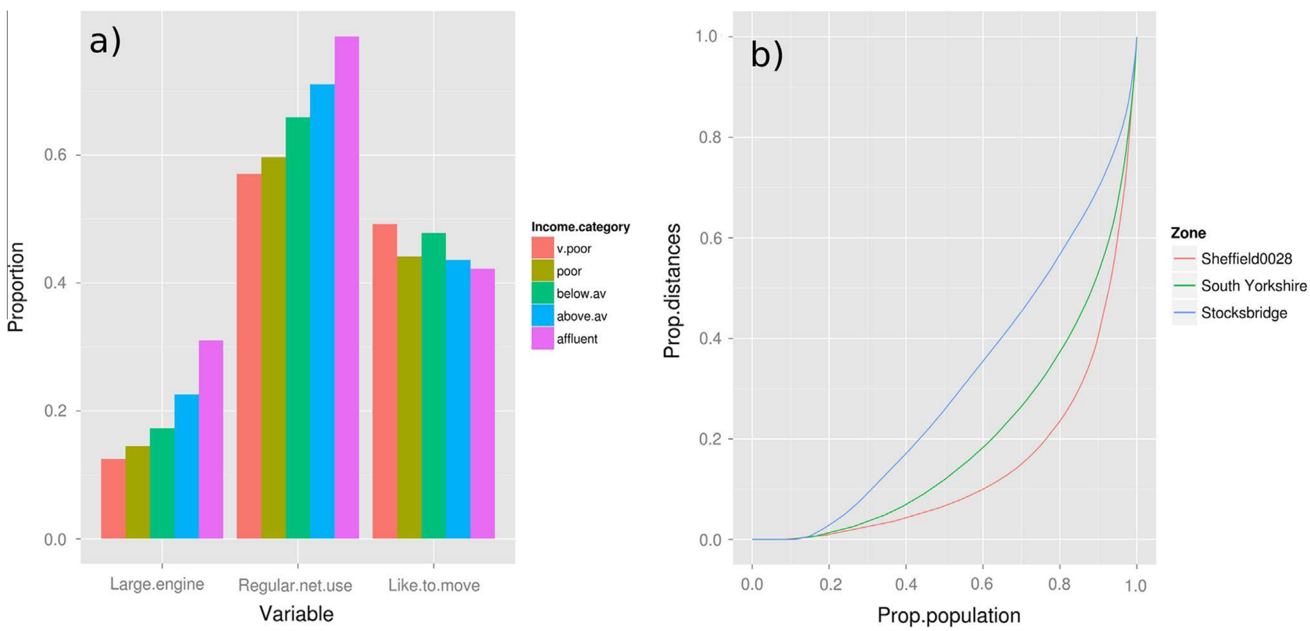

Fig. 11. (a) Variability of vehicles (proportion of primary cars in household whose engine size is 2.0 litres or more), internet use (proportion of commuters who use the internet daily or weekly) and desire to move home depending on equivalised income. These categorical target variables are described in Table 3. (b) Lorenz curves illustrating the individual-level variability in commuter distances for 3 zones. The Gini indices associated with these curves are 0.278, 0.294 and 0.305 for Stocksbridge, South Yorkshire and Sheffield028 respectively.

4.3. Individual-level results

Spatial microsimulation allows one to 'drill down' to the individual level, target specific groups and model who (in addition to where) is most likely to benefit from specific interventions. Table 8, for example, shows simulated differences in commuting patterns between high and low income citizens in South Yorkshire as a whole.[19] Because the individual microdata are also geocoded, we were able to conduct the same analyses for specific zones. Table 9 illustrates how the results of spatial microsimulation allow inter- and intra-zone analysis to be combined. Table 9 indicates that the Sheffield028 (an MSOA zone) is more unequal in terms of income and distance travelled to work than Stocksbridge (a statistical Ward) (see Fig. 6 to see their respective locations). These results, which can be compared with the regional data presented in Table 8, or re-calculated for smaller zones, are thus (to the extent that administrative boundaries allow) 'frame independent' (Horner and Murray, 2002).

[19] The categories "very poor" to "affluent" used here are defined in (Ballas et al., 2005b). Statistical bins are defined as proportions of the the median income, with breaks at 50%, 75%, 100% and 125% of the median (Ballas et al., 2005b, p. 91).

To further explore differences in intra-zone inequality, commuter work travel distances were plotted as Lorenz curves (Fig. 11b). These provide further insight into commuter patterns in each of the zones described in Table 9, and illustrate that a small proportion of the population living in Crookes accounts for a large part of the average trip distance. Stocksbridge, by contrast, has a more even distribution of commuter patterns.

Regarding the categorical target variables described in Table 3, the results imply that wealthy commuters in South Yorkshire drive larger cars, use the internet more frequently, and may be less likely to want to move than those with low incomes (Fig. 11a).

5. Discussion and conclusions

This paper has presented a spatial microsimulation approach to model commuter patterns. Whole individuals from a detailed national survey were allocated to geographic zones at various levels; this provided further insight into intra-zone variability of commuting than is available from the use of aggregated census data alone. In addition, the careful selection of target variables not included in

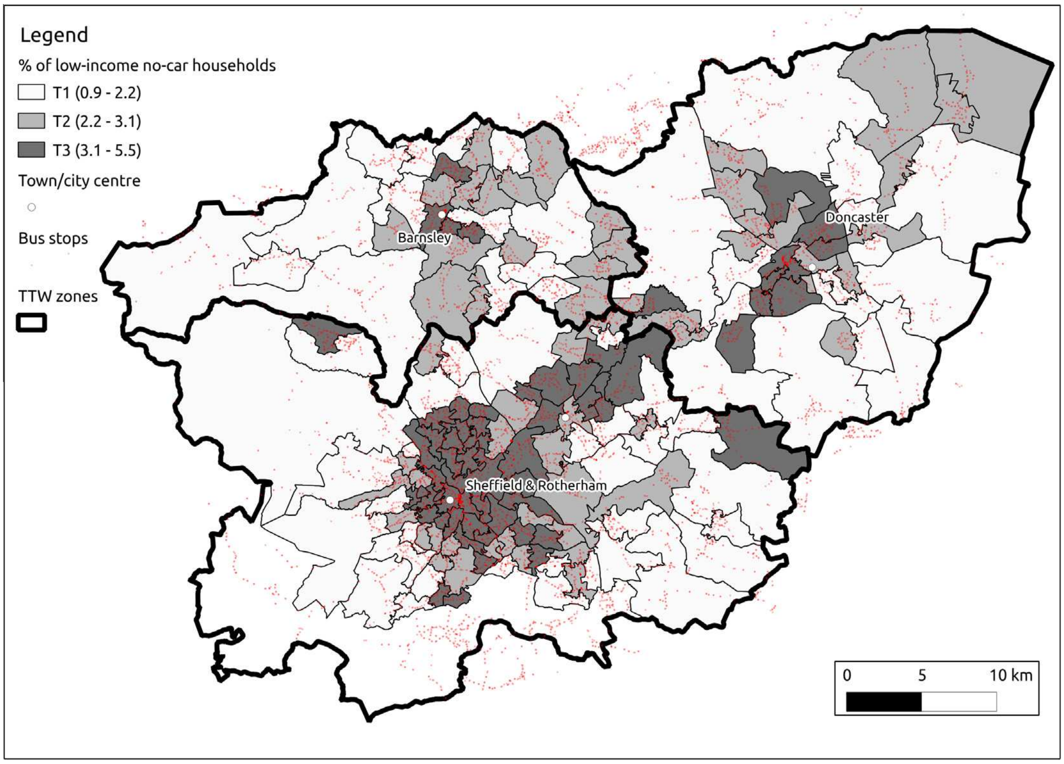

Fig. 12. Proportion of population who is earning less than 50% of South Yorkshire's median income *and* lives in a car free household within the 173 MSOA boundaries of the metropolitan county, according to the spatial microsimulation model. Translucent red dots represent bus stops (data from http://data.gov.uk/dataset/nptdr).

the census provided insight into the relationships between commuting behaviour and a variety of 'target variables' such as income, internet use, desire to move home, type of car and number of children.

From the perspective of the data-constrained policy makers mentioned in the introduction, these results are attractive: they provide a level of detail that is inaccessible for analyses based on geographically aggregated census data alone. The ability to explore the commuter behaviour of subsets of individuals based on age, distance travelled and class (constraint variables) or other variables including size of car or income (target variables) will be useful in various applications: being able to simulate the *characteristics* of commuters who are most likely to benefit from certain interventions and identifying *where* these people live and work clearly has huge potential for transport planning and policy. To illustrate the point, the distribution of low-income households reliant on buses can be simulated and mapped at the county level to help inform the location of new bus routes (Fig. 12). For example, if this type of analysis had been properly conducted and validated during the planning stages of the recently implemented rapid bus routes in Albuquerque mentioned in Tribby and Zandbergen (2012), the system could have been designed such that low income residents benefited from faster access to the city centre. In fact, relatively wealthy households (who probably have more transport options already) benefited most from the scheme (Tribby and Zandbergen, 2012). This illustrates the importance of considering not only aggregate-level impacts, but also taking into account the local and micro-level distributional effects of intervention. The spatial microsimulation approach to modelling commuter patterns outlined in this paper provides a foundation for investigating such effects. In addition, we have demonstrated how spatial microsimulation methods can enrich transport models with policy relevant socio-economic variables at individual and small-area levels. In broader terms, we hope this paper promotes a closer collaboration between the fields of transport modelling, spatial microsimulation and spatial microsimulation.

Another possibility opened-up by the inclusion of the individual level data (but not explored in this paper) is more detailed energy analysis: previous studies have tended to focus on energy use in transport at national or city scales of (Lovelace et al., 2011; Woodcock et al., 2007), omitting important information about its social distribution (see Preston et al., 2013 for a UK example).

Despite these enticing possibilities, is important to remember that the results are *simulated*. Consequently, we must distinguish between linking variables (or constraint variables)—these are constrained by known census aggregates and are therefore trustworthy—and target variables which are more tentative estimates based on correlations between target and linking variables at the national level. As noted in Section 3.1, target variable estimates rely on an often unstated assumption: that the relationships between variables at the national level (e.g. between distance travelled to work and income) tend to remain at local levels. This assumption cannot be expected to hold everywhere, so results arising from target variables are expected to underplay the true level of spatial variability. Where possible, target variable results should be corroborated against independent datasets (Edwards and Clarke, 2009).

Many transport interventions have wide-ranging impacts on commuters. These depend on geographical *and* individual-level factors, and the importance of the latter especially is often overlooked in transport policy (e.g. Tribby and Zandbergen, 2012). The micro-level methods presented in this paper therefore have great potential, to enable researchers and transport planners to better model and predict the impacts arising from various interventions. With the current focus on energy and sustainability in transport (Chapman, 2007), there is a risk that distributional impacts continue to receive little or no attention. Spatial microsimulation has the potential to address this issue, by helping decision makers to design sustainable transport measures that are both effective and fair.

References

Akkerman, A., 2000. The diurnal cycle of regional commuter systems: North Wales, 1991. Geographical Analysis 32 (July). <http://onlinelibrary.wiley.com/doi/10.1111/j.1538- 4632.2000.tb00427.x/abstract>.

Anderson, B., 2013. Estimating small-area income deprivation: an iterative proportional fitting approach. In: Tanton, R., Edwards, K. (Eds.), Spatial Microsimulation: A Reference Guide for Users, Understanding Population Trends and Processes, vol. 6. Springer, Netherlands, pp. 49–67, Chapter 4. http://dx.doi.org/10.1007/978-94-007-4623-7_4.

Ballas, D., Clarke, G., Dorling, D., Eyre, H., Thomas, B., Rossiter, D., 2005a. SimBritain: a spatial microsimulation approach to population dynamics. Population, Space and Place 11 (1), 13–34, http://doi.wiley.com/10.1002/psp.351.

Ballas, D., Clarke, G., Dorling, D., Rossiter, D., 2007. Using SimBritain to model the geographical impact of national government policies. Geographical Analysis 39 (2007), 44–77, http://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.2006.00695.x/full.

Ballas, D., Clarke, G., Hynes, S., Lennon, J., Morrissey, K., Donoghue, C.O., 2013. A review of microsimulation for policy analysis. In: Tanton, R., Edwards, K. (Eds.), Spatial Microsimulation for Rural Policy Analysis: Advances in Spatial Science. Springer, Berlin, Heidelberg, pp. 35–54, Chapter 3. <http://www.springerlink.com/index/10.1007/978-3- 642-30026-4>.

Ballas, D., Clarke, G.P., Dewhurst, J., 2006. Modelling the socio-economic impacts of major job loss or gain at the local level: a spatial microsimulation framework. Spatial Economic Analysis 1 (1), 127–146, http://www.tandfonline.com/doi/abs/10. 1080/17421770600697729.

Ballas, D., Dorling, D., Thomas, B., Rossiter, D., January 2005b. Geography matters: simulating the local impacts of national social policies. No. 3. Joseph Roundtree Foundation. <www.jrf.org.uk/sites/files/jrf/1859352669.pdf>.

Ballou, R.H., Rahardja, H., Sakai, N., 2002. Selected country circuity factors for road travel distance estimation. Transportation Research Part A: Policy and Practice 36 (9), 843–848, <http://www.sciencedirect.com/science/article/pii/S0965856401000441>.

Beckman, R., Baggerly, K., McKay, M., 1996. Creating synthetic baseline populations. Transportation Research Part A: Policy and Practice 30 (6), 415–429, <http://www.sciencedirect.com/science/article/pii/0965856496000043>.

Bhat, C., Guo, J., Srinivasan, S., Sivakumar, A., 2004. Comprehensive econometric microsimulator for daily activity-travel patterns. Transportation Research Record: Journal of the Transportation Research Board 1894, 57–66, http://dx.doi.org/10.3141/1894-07.

Birkin, M., Clarke, M., 1988. SYNTHESIS – a synthetic spatial information system for urban and regional analysis: methods and examples. Environment and Planning A 20 (12), 1645–1671, <http://www.envplan.com/abstract.cgi?id=a201645>.

Birkin, M., Clarke, M., 1989. The generation of individual and household incomes at the small area level using synthesis. Regional Studies 23 (6), 37–41, http://www.tandfonline.com/doi/abs/10. 1080/00343408912331345702.

Birkin, M., Clarke, M., 2011. Spatial microsimulation models: a review and a glimpse into the future. In: Stillwell, J., Clarke, M., Stillwell, J. (Eds.), Population Dynamics and Projection Methods, Understanding Population Trends and Processes, vol. 4. Springer, Netherlands, pp. 193–208, http://dx.doi.org/10.1007/978-90-481-8930-4_9.

Brockmann, D., 2012. Complex systems: spotlight on mobility. Nature 484 (7392), 40–41, http://dx.doi.org/10.1038/484040a.

Buehler, R., 2012. Determinants of bicycle commuting in the Washington, DC region: the role of bicycle parking, cyclist showers, and free car parking at work. Transportation Research Part D: Transport and ... 17 (7), 525–531, http://dx.doi.org/10.1016/j.trd.2012.06.003 http://www.sciencedirect.com/science/article/pii/S1361920912000594.

Buliung, R.N., Kanaroglou, P.S., 2002. Commute minimization in the Greater Toronto Area: applying a modified excess commute. Journal of Transport Geography 10 (3), 177–186, <http://linkinghub.elsevier.com/retrieve/pii/S0966692302000108>.

Buliung, R.N., Kanaroglou, P.S., 2006. A GIS toolkit for exploring geographies of household activity/travel behavior. Journal of Transport Geography 14 (1), 35–51, <http://linkinghub.elsevier.com/retrieve/pii/S0966692304000730>.

Campbell, M., Ballas, D., 2013. A spatial microsimulation approach to economic policy analysis in Scotland. Regional Science Policy & Practice.

Chandola, T., Jenkinson, C., 2000. The new uk national statistics socio-economic classification (ns-sec); investigating social class differences in self-reported health status. Journal of Public Health 22 (2), 182–190.

Chapman, L., 2007. Transport and climate change: a review. Journal of Transport Geography 15 (5), 354–367, <http://linkinghub.elsevier.com/retrieve/pii/S0966692306001207>.

Chin, S.-F., Harding, A., 2006. Regional dimensions: creating synthetic small-area microdata and spatial microsimulation models. National Centre for Social and Economic Modelling, Canberra.

Clarke, G., Harding, A., 2013. Conclusions and future research directions. In: Tanton, R., Edwards, K. (Eds.), Spatial Microsimulation: A Reference Guide for Users. Springer, Berlin, Heidelberg, pp. 259–273, Chapter 16. <http://www.springerlink.com/index/10.1007/978-94- 007-4623-7>.

Clarke, M., Holm, E., 1987. Microsimulation methods in spatial analysis and planning. Geografiska Annaler, Series B: Human Geography 69 (2), 145–164, <http://www.jstor.org/stable/10.2307/490448>.

Coombes, M., Openshaw, S., 1982. The use and definition of travel-to-work areas in Great Britain: some comments. Regional Studies, 37–41, http://www.tandfonline.com/doi/abs/10. 1080/09595238200185161 (August 2011).

Dale, A., Teague, A., 2002. Microdata from the census: samples of anonymised records. In: Williamson, P., Rees, P., Martin, D. (Eds.), The Census Data System. Wiley, Chapter 14.

Daley, M., Rissel, C., 2011. Perspectives and images of cycling as a barrier or facilitator of cycling. Transport Policy 18 (1), 211–216, <http://linkinghub.elsevier.com/retrieve/pii/S0967070X10000995>.

Deming, W.E., Stephan, F.F., 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. The Annals of Mathematical Statistics 11 (4), 427–444, <http://www.jstor.org/stable/2235722>.

DePalma, E., Arnott, R., 2012. Morning commute in a single-entry traffic corridor with no late arrivals. Transportation Research Part B: Methodological 46 (1), 1–29, <http://linkinghub.elsevier.com/retrieve/pii/S0191261511001263>.

DfT, 2011. National Travel Survey 2010. Tech. rep., Department for Transport. <http://www.dft.gov. uk/statistics/releases/national-travel-survey-2010/>.

Edwards, K.L., Clarke, G.P., 2009. The design and validation of a spatial microsimulation model of obesogenic environments for children in Leeds, UK: SimObesity. Social Science & Medicine 69 (7), 34–1127, <http://www.ncbi.nlm.nih.gov/pubmed/19692160>.

Edwards, K.L., Clarke, G.P., Thomas, J., Forman, D., 2010. Internal and external validation of spatial microsimulation models: small area estimates of adult obesity. Applied Spatial Analysis and Policy 4 (4), 281–300, <http://www.springerlink.com/index/10.1007/s12061- 010-9056-2>.

Edwards, K.L., Tanton, R., 2013. Validation of spatial microsimulation models. In: Spatial Microsimulation: A Reference Guide for Users. Springer, pp. 249–258, Chapter 15.

Ehrgott, M., Wang, J.Y., Raith, A., van Houtte, C., 2012. A bi-objective cyclist route choice model. Transportation Research Part A: Policy and Practice 46 (4), 652–663, <http://linkinghub.elsevier.com/retrieve/pii/S0965856411001844>.

Ferguson, M., Maoh, H., Ryan, J., Kanaroglou, P., Rashidi, T.H., 2012. Transferability and enhancement of a microsimulation model for estimating urban commercial vehicle movements. Journal of Transport Geography 24, 358–369, <http://linkinghub.elsevier.com/retrieve/pii/S0966692312001251>.

Fienberg, S., 1970. An iterative procedure for estimation in contingency tables. The Annals of Mathematical Statistics 41 (3), 907–917, <http://www.jstor.org/stable/10.2307/2239244>.

Forrest, D., Glen, J., Ward, R., 1996. The impact of a light rail system on the structure of house prices: a hedonic longitudinal study. Journal of Transport Economics and Policy 30 (1), 15–27, <http://www.jstor.org/stable/10.2307/20053094>.

Gao, W., Balmer, M., Miller, E.J., 2010. Comparison of MATSim and EMME/2 on greater Toronto and Hamilton area network, Canada. Transportation Research Record: Journal of the Transportation Research Board 2197, 118–128.

Gilbert, N., 2007. Agent-based Models (Quantitative Applications in the Social Sciences). SAGE Publications.

Gilbert, N., Troitzsch, K.G., 2005. Simulation for the Social Scientist. Open University Press.

Guo, J., Bhat, C., 2007. Population synthesis for microsimulating travel behavior. Transportation Research Record: Journal of the Transportation Research Board 2014, 92–101, http://dx.doi.org/10.3141/2014-12.

Harland, K., Heppenstall, A., Smith, D., Birkin, M., 2012. Creating realistic synthetic populations at varying spatial scales: a comparative critique of population synthesis techniques. Journal of Artificial Societies and Social Simulation 15 (1), 1, <http://jasss.soc.surrey.ac.uk/15/1/1.html>.

Helminen, V., Ristimäki, M., 2007. Relationships between commuting distance, frequency and telework in Finland. Journal of Transport Geography 15 (5), 331–342, <http://linkinghub.elsevier.com/retrieve/pii/S0966692306001384>.

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 (4), 281–290, <http://linkinghub.elsevier.com/retrieve/pii/S0198971512000336>.

Horner, M., Murray, A., 2002. Excess commuting and the modifiable areal unit problem. Urban Studies 39 (1), 131, <http://usj.sagepub.com/content/39/1/131.short>.

Horowitz, J.L., 1986. Example: modelling choices of residential location and mode of travel to work. In: Hanson, S. (Ed.), The Geography of Urban Transportation. Guilford Press, New York, p. 19, Chapter 9.

Ibeas, A., Cordera, R., DellOlio, L., Coppola, P., Dominguez, A., 2012. Modelling transport and real-estate values interactions in urban systems. Journal of Transport Geography 24, 370–382, <http://linkinghub.elsevier.com/retrieve/pii/S096669231200124X>.

Kavroudakis, D., Ballas, D., Birkin, M., 2012. Using spatial microsimulation to model social and spatial inequalities in educational attainment. Applied Spatial Analysis and Policy 6 (1), 1–23, <http://link.springer.com/10.1007/s12061-012-9075- 2>.

Kay, D., Reynolds, J., Rodrigues, S., Lee, A., Anderson, B., Gibbs, R., Monkhouse, C., Gill, T., 2011. Fairness in a car dependent society. Tech. rep., Sustainable Development Commission, London. <http://www.sortclearinghouse.info/research/740/>.

Please cite this article in press as: Lovelace, R., et al. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. J. Transp. Geogr. (2013), http://dx.doi.org/10.1016/j.jtrangeo.2013.07.008

Knowles, R., 1996. Transport impacts of Greater Manchester's Metrolink light rail system. Journal of Transport Geography 4 (1), 1–14, <http://www.sciencedirect.com/science/article/pii/0966692395000348>.

Lee, A., 2009. Generating Synthetic Microdata from Published Marginal Tables and Confidentialised Files. Tech. rep., Statistics New Zealand, Wellington. <http://www.statisphere.govt.nz/further-resources- and-info/official-statistics-research/series//media/statisphere/Files/official-statistics-research- series/volume-5-2009/generating-synthetic-microdata-from-published- tables-and-confidentialised-files.pdf>.

Levinson, D., El-Geneidy, A., 2009. The minimum circuity Frontier and the journey to work. Regional Science and Urban Economics 39 (6), 732–738, <http://linkinghub.elsevier.com/retrieve/pii/S0166046209000660 http://www.sciencedirect.com/science/article/pii/S0166046209000660>.

Li, T., Corcoran, J., Burke, M., 2012. Disaggregate gis modelling to track spatial change: exploring a decade of commuting in South East Queensland, Australia. Journal of Transport Geography.

Lovelace, R., Ballas, D., 2013. Truncate, replicate, sample: a method for creating integer weights for spatial microsimulation. Computers, Environment and Urban Systems 41, 1–11, <http://linkinghub.elsevier.com/retrieve/pii/S0198971513000240>.

Lovelace, R., Beck, S.B.M., Watson, M., Wild, A., 2011. Assessing the energy implications of replacing car trips with bicycle trips in Sheffield, UK. Energy Policy 39 (4), 2075–2087, <http://linkinghub.elsevier.com/retrieve/pii/S0301421511000620>.

Marshall, J.D., 2008. Energy-efficient urban form. Environmental Science & Technology 42 (9), 7–3133, <http://www.ncbi.nlm.nih.gov/pubmed/18522085>.

Mellander, C., Florida, R., Stolarick, K., 2011. Here to stay the effects of community satisfaction on the decision to stay. Spatial Economic Analysis 6 (1), 5–24, http://www.tandfonline.com/doi/abs/10. 1080/17421772.2010.540031.

Mitton, L., Sutherland, H., Weeks, M., 2000. Microsimulation Modelling for Policy Analysis: Challenges and Innovations, vol. 65. Cambridge University Press.

Morrissey, K., Clarke, G., Ballas, D., Hynes, S., O'Donoghue, C., 2008. Examining access to GP services in rural Ireland using microsimulation analysis. Area 40 (3), 354–364, http://doi.wiley.com/10.1111/j.1475-4762.2008. 00844.x.

Morrissey, K., O'Donoghue, C., 2013. Validation issues and the spatial pattern of household income. In: O'Donoghue, C., Ballas, D., Clarke, G., Hynes, S., Morrissey, K. (Eds.), Spatial Microsimulation for Rural Policy Analysis, Advances in Spatial Science. Springer, Berlin, Heidelberg, pp. 87–102, Chapter 5. http://dx.doi.org/10.1007/978-3-642-30026-4_5.

Mosteller, F., 1968. Association and estimation in contingency tables. Journal of the American Statistical Association 63 (321), 1–28, <http://www.jstor.org/stable/10.2307/2283825>.

Openshaw, S., 1983. The Modifiable Areal Unit Problem. Geo Books Norwich, UK, <http://www.getcited.org/pub/102412488>.

Pibyl, O., Goulias, K., 2005. Simulation of daily activity patterns incorporating interactions within households: algorithm overview and performance. Transportation Research Record: Journal of the Transportation Research Board 1926, 135–141, http://dx.doi.org/10.3141/1926-16.

Preston, I., White, V., Thumim, J., Bridgeman, T., Brand, C., 2013. Distribution of carbon emissions in the UK: implications for domestic energy policy. Tech. Rep. March 2013, Josheph Roundtree Foundation.

Pritchard, D.R., Miller, E.J., 2012. Advances in population synthesis: fitting many attributes per agent and fitting to household and person margins simultaneously. Transportation 39 (3), 685–704, <http://www.springerlink.com/index/10.1007/s11116- 011-9367-4>.

Ravulaparthy, S., Goulias, K., 2011. Forecasting with Dynamic Microsimulation: Design, Implementation, and Demonstration: Final Report on Review, Model Guidelines, and a Pilot Test for a Santa Barbara County Application. Tech. Rep. May, University of California Transportation Center (UCTC), California. <http://scholar.google.com/scholar?hl=en& btnG= Search& q=intitle:Forecasting+with+Dynamic+Microsimulation:+Design, +Implementation,+and+Demonstration#0>.

Redmond, G., Sutherland, H., Wilson, M., 1998. The Arithmetic of Tax and Social Security Reform: A User's Guide to Microsimulation Methods and Analysis, vol. 64. Cambridge University Press.

Rees, P., Martin, D., 2002. The debate about census geography. In: Rees, P., Williamson, P., Martin, D. (Eds.), The Census Data System. Wiley, London, Chapter 2. <http://eprints.soton.ac.uk/15368/>.

Rees, P., Martin, D., Williamson, P., 2002. Census data resources in the United Kingdom. In: Rees, P., Martin, D., Williamson, P. (Eds.), The Census Data System. Wiley, London, Chapter 1. <http://eprints.soton.ac.uk/15359/1/0470846887.pdf>.

Rietveld, P., 2004. Determinants of bicycle use: do municipal policies matter? Transportation Research Part A: Policy and Practice 38 (7), 531–550, <http://linkinghub.elsevier.com/retrieve/pii/S0965856404000382>.

Senior, M.L., 2009. Impacts on travel behaviour of Greater Manchesters light rail investment (Metrolink Phase 1): evidence from household surveys and Census data. Journal of Transport Geography 17 (3), 187–197, <http://linkinghub.elsevier.com/retrieve/pii/S0966692308001300>.

Sexton, S., Wu, J.J., Zilberman, D., 2012. How High Gas Prices Triggered the Housing Crisis: Theory and Empirical Evidence. Tech. rep., University of California, Berkeley. <http://www.chikyu.ac. jp/archive/topics/2011/111219_JunJieWu.pdf>.

Simini, F., González, M.C., Maritan, A., Barabási, A.-L., 2012. A universal model for mobility and migration patterns. Nature, 8–12, <http://www.ncbi.nlm.nih.gov/pubmed/22367540>.

Smith, D.M., Clarke, G.P., Harland, K., 2009. Improving the synthetic data generation process in spatial microsimulation models. Environment and Planning A 41 (5), 1251–1268, <http://www.envplan.com/abstract.cgi?id=a4147>.

Titheridge, H., Hall, P., 2006. Changing travel to work patterns in South East England. Journal of Transport Geography 14 (1), 60–75, <http://linkinghub.elsevier.com/retrieve/pii/S0966692305000451>.

Tribby, C.P., Zandbergen, P.a., 2012. High-resolution spatio-temporal modeling of public transit accessibility. Applied Geography 34, 345–355, <http://linkinghub.elsevier.com/retrieve/pii/S014362281100244X>.

Vega, A., 2012. Using place rank to measure sustainable accessibility. Journal of Transport Geography 24, 411–418, <http://linkinghub.elsevier.com/retrieve/pii/S0966692312001202>.

Vega, A., Reynolds-Feighan, A., 2009. A methodological framework for the study of residential location and travel-to-work mode choice under central and suburban employment destination patterns. Transportation Research Part A: Policy and Practice 43 (4), 401–419, <http://linkinghub.elsevier.com/retrieve/pii/S0965856408002103>.

Verhetsel, A., Vanelslander, T., 2010. What location policy can bring to sustainable commuting: an empirical study in Brussels and Flanders, Belgium. Journal of Transport Geography 18 (6), 691–701, <http://linkinghub.elsevier.com/retrieve/pii/S0966692309001768>.

Voas, D., Williamson, P., 2001. Evaluating goodness-of-fit measures for synthetic microdata. Geographical and Environmental Modelling 5 (2), 177–200, http://dx.doi.org/10.1080/13615930120086078.

Williamson, P., Birkin, M., Rees, P.H., 1998. The estimation of population microdata by using data from small area statistics and samples of anonymised records. Environment and Planning A 30 (5), 785–816, <http://www.envplan.com/abstract.cgi?id=a300785>.

Williamson, P., Mitchell, G., McDonald, A.T., 2002. Domestic water demand forecasting: a static microsimulation approach. Water and Environment Journal 16 (4), 243–248, http://dx.doi.org/10.1111/j.1747-6593.2002.tb00410.x.

Wong, D.W.S., 1992. The reliability of using the iterative proportional fitting procedure. The Professional Geographer 44 (3), 340–348, http://dx.doi.org/10.1111/j.0033-0124.1992.00340.x.

Wong, D.W.S., 2009. The modifiable areal unit problem (MAUP). In: Fotheringham, S., Rogerson, P.A. (Eds.), The SAGE Handbook of Spatial Analysis. Sage Publications, New York, Chapter 7.

Woodcock, J., Banister, D., Edwards, P., Prentice, A.M., Roberts, I., 2007. Energy and health 3 – energy and transport. Lancet 370 (9592), 1078–1088, <http://www.ncbi.nlm.nih.gov/pubmed/17868817>.

Wu, B., Birkin, M.M., Rees, P., 2008. A spatial microsimulation model with student agents. Computers, Environment and Urban Systems 32 (6), 440–453, <http://linkinghub.elsevier.com/retrieve/pii/S019897150800080X>.