Journal of Urban Mobility 5 (2024) 100078

Contents lists available at ScienceDirect

Journal of Urban Mobility

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

A MATSim model methodology to generate cycling-focused transport scenarios in England

David Alvarez Castro [a,*], Alistair Ford [a], Philip James [a], Roberto Palacín [a], Dominik Ziemke [b]

[a] Newcastle University, United Kingdom
[b] Technische Universität Dresden, Germany

ARTICLE INFO

Keywords:
Agent-based modelling
Simulation
MATSim
Sustainable transport
Cycling

Climate change is considered the most pressing environmental challenge of our time, being transport one of the major contributors. Consequently, transport models are required to test different urban mobility policies that can shift travel to more sustainable transport modes (e.g., active modes). This paper focuses on the development of a validated agent-based model (MATSim) applying a novel open-source methodology to generate the main input datasets, easily transferrable to any region in England. Required input datasets (synthetic population and network) are described with a high level of detail, identifying the datasets and tools used to develop them, with special interest in the simulation of cycling routes. A new attribute (quietness) ranking roads for cycling depending on their built-environment characteristics was incorporated into the MATSim bicycle extension. The results obtained in this paper show the baseline transport model of the Tyne and Wear region (England), where discrepancies up to 3.5% in transport mode shares and minimal differences in vehicle counts in urban areas were obtained, and a realistic representation of the routes chosen by the agents using bicycles is obtained. This provides the basis for the development of similar MATSim implementation in other UK regions.

1. Introduction

Given that transport is a major contributor to greenhouse gas emissions, many cities around the globe are introducing sustainable transport legislation to meet the 2015 Paris Agreement of limiting global warming to 2 °C and aiming for 1.5 °C. To achieve this, policies are sought that can change mobility patterns to reduce emissions rapidly. This could involve a portfolio of measures where a combination of changes to the built environment, human behaviours and financial incentives or penalties are considered to enable a shift of travel to more sustainable modes (e.g., active travel).

Before their implementation, urban mobility policies need to be tested to understand their effectiveness and estimate their success or failure. Traditionally, this procedure has consisted of the development of a model, which has been defined by Bandini et al. (2009) as 'an abstract and simplified representation of a given reality, either already existing or just planned to study and explain observed phenomena or to foresee future phenomena'. Models are powerful tools to assess change, as they provide an abstraction of a system that can increase the pace of change by demonstrating feasible and possible options, allowing experimentation with policy alternatives and conversations with stakeholders, and providing an endogenous perspective (Ghaffarzadegan *et al.*, 2011; Ford et al., 2018).

Over time, transport modelling has developed from statistically-based numerical modelling (Tyrinopoulos & Antoniou, 2013) to more activity-oriented and complex modelling techniques such as Agent-Based Models (ABMs) (Krajzewicz et al., 2012; Zia et al., 2013; Maggi *et al.*, 2016; Martinez et al., 2017; Cardinot et al., 2019). ABMs are computer simulations of simulated autonomous agents (individuals) in a simulated space and time (El-sayed et al., 2012). These models allow a detailed representation of the interactions of multiple agents in a realistic synthetic environment where the intent is to recreate the appearance of a complex phenomenon (Martinez *et al.*, 2017).

Some of the advantages of transport ABMs are the possibility to consider the individual interactions between the agents and the environment, providing a new perspective in transport modelling that could not be obtained from previous models (e.g., four-step models). Firstly, the interactions of the agents in space and time allow them to adapt and learn from what others do. Examples were highlighted by Bazzan et al. (2014), where they describe that agents' interactions allow their adaptation and learning capacity to simulate realistic and optimised behaviours. Secondly, agents' interactions with the environment give an insight into how the built environment could affect agents' daily routines. Characteristics, such as the road gradient and type, the existence

* Corresponding author.
E-mail address: d.alvarez2@newcastle.ac.uk (D. Alvarez Castro).

https://doi.org/10.1016/j.urbmob.2024.100078
Received 24 November 2023; Received in revised form 27 March 2024; Accepted 3 May 2024
Available online 7 June 2024

of cycle paths, land-use type and pavement conditions, could be considered when agents decide how (i.e., transport mode and route) and where to go (e.g., destination). The choices made by agents based on the built-environment characteristics are encapsulated within the ‘spatial cognition’ concept, which describes the effect of environmental factors on mobility (Manley et al., 2018; Gr et al., 2019, Manley et al., 2021). Additionally, Heppenstall et al. (2016) suggest that one of the most appealing aspects of ABMs is their ability to represent human behaviour and, through simulation, understand how these behaviours play out over space and time.

The use of ABMs in the transportation sector has increased in the last decade. Maggi et al. (2016) highlight that ABMs present important advantages for analysing urban transport and its sustainability, as well as testing urban mobility decarbonisation policies. From a policy perspective, Ghorbani et al. (2014) underline the importance of ABMs in the study of policy problems, considering them as one of the most instrumental tools for policy analysis. Furthermore, such models offer tools to move forward urban mobility conversations between policy-makers, transport users, and urban populations (Ford et al., 2018). Bastarianto et al. (2023) reviewed the use of transport ABMs from 2006 to 2022 and identified an exponential increase in the number of publications since 2015. The reasons given to explain this increase are two: the significant improvement in computing performance and the use of fully open-source tools.

The use of transport ABMs combined with public policies to reduce GHG emissions is an incipient topic in the literature, where policies to reduce the use of more carbon-intensive modes (i.e., push policies) and to incentivise the use of alternative and more sustainable modes (i.e., pull policies) have been considered. Examples of the former were analysed by Zheng et al. (2012), where a dynamic cordon pricing scheme in Zurich (Switzerland) was simulated. Results show a reduction in travel times and that congestion within the area was alleviated. Another scenario developed by Zheng et al. (2014) analysed the impact of a time-dependent pricing scheme in Sioux Falls (USA), considering the level of congestion in time and the user’s adaptation to the toll cost. Results show effective congestions reductions in the area of study and a modal shift to public transport modes.

Examples of the pull policies are found in Park et al. (2018), who simulated active modes in New York City (USA) to support investment decisions and evaluate the impact of infrastructure changes for walking and cycling. Their results show that the improvement in sidewalk and cycle path conditions could positively increase the number of people using them. Hitge et al. (2023) developed a model to estimate a potential cycling demand in Cape Town (South Africa). Their model showed that 32 % of agents would benefit from cycling, although the percentage was reduced by 8 % when socio-demographic characteristics (e.g., age, gender, household income, household composition and dwelling type) were considered. The combination of both approaches (i.e., pull and push) can also be found. Schlenther et al. (2022) investigate scenarios to reduce the number of motorised vehicles on the road in Hamburg (Germany) combining policies in favour of more sustainable modes and penalising the use of private cars. Results show that better results are achieved when push policies (e.g., car use penalty) are applied than when improving the attractiveness of other modes (e.g., public transport), and these studies also demonstrate the effectiveness of the use of ABMs in testing potential policy interventions.

Additionally, the models can be classified into two main groups depending on how the efficiency of policies are analysed. The first computes the variations of emissions directly (e.g., differences in NOx, C02 emissions), where policies are applied and compared against an initial base scenario. The second provides the agents with the use of new transport (e.g., car sharing, micro-mobility) or sustainable modes (e.g., walking, cycling) to reduce the use of private motor cars and therefore, reduce emissions. For the first case, Minh Duc et al. (2020) proposes a model that simulates the traffic and air quality in Hoan Kiem district (Vietnam) that can be used as a decision support tool for local authorities when implementing new policies. Kilani *et al.* (2022) developed a passenger transport model for the North of France to analyse the impacts of policies focusing on the limitations of emissions and congestion. Gurram *et al.* (2019) developed an exposure-modelling framework that integrates agent-based activity and travel simulation with air pollution modelling for Tampa (Florida), estimating the mean daily population exposure concentration of NOx of different subgroups of the population living in the area of study.

For the latter case, Müller et al. (2022) presents a model applied in Vienna considering different transport modes including car sharing for different groups in society based on socio-demographic attributes. Leblond (2020) presented a new agent-based simulation software to simulate e-scooters in the city of Rennes that can be used as a decision support tool for designing mobility services. Ziemke et al. (2019b) developed a MATSim contribution to model bicycle traffic from a more realistic perspective, considering characteristics such as presence of cycling infrastructure, road surfaces and gradients.

Several ABM platforms dedicated to transport mobility have been developed, such as MATSim (Horni *et al.*, 2016), SUMO (Krajzewicz et al., 2012), TRANSIMS (*Smith et al.*, 1995) and SimMobility (Adnan *et al.*, 2016), which have been applied to different cities and regions including Berlin (Ziemke et al. 2019a), Zurich (Rieser-Schüssler, 2016), Singapore (Erath *et al.*, 2016), Munich (Kickhofer, 2016) and London (Serras et al., 2016). Currently, the use of mesoscopic ABMs (e.g., MATSim) to simulate mixed interactions of micro-mobility modes (e.g., walking, bicycles, e-bikes, e-scooters) is in discussion, as spatial interactions that occur at the micro-level (e.g., traffic safety, intersections, road lane changes) between themselves and other vehicles (e.g., cars, buses) could be missed. Tzouras et al. (2023) highlight this aspect focusing on the simulation of e-scooters, where a dilemma between modelling their behaviour and interactions at a link level and predicting long-term travel behaviour using microscopic models (TRANSIMS, SimMobility) is identified. The former could simplify the interactions, while the latter currently does not have the required capacities to model bicycle or pedestrian traffic or is not capable to simulate large-scale networks. Their conclusion considers the development of a hybrid model that could combine the analysis in network and link levels.

From a MATSim perspective, different upgrades have been incorporated to simulate more realistic mobility patterns for those modes. This is the case of filtering behaviour (Agarwal et al., 2016), which allows cyclists to overtake faster vehicles when stopped in congestion, and the aforementioned bicycle extension (Ziemke et al., 2019b) that allows to simulate this mode in a more detailed way and considering infrastructure and built-environmental characteristics when choosing the route. Additionally, another step forward in the use of MATSim with micro-mobility modes is the research made by Becker et al. (2020), who developed a model combining three types of shared mobility (car-sharing, bike-sharing and ride-hailing) for a city scale transport system.

Transport ABMs require the development of two main input datasets: synthetic population (or *demand*) and network (or *supply*). The first is a digital representation of the individuals living in the study area (e.g., socio-demographic attributes and activity plans). The second is a digital geospatial representation of the road and public transport network used by the individuals. While the latter is easily generated using open access data (e.g., Open Street Map) and open-source tools in the majority of the cases (e.g., PT2MATSim for MATSim models and netGenerate for SUMO), the former depends on the availability of socio-demographic data and detailed travel surveys from the study area, besides the possibility to use and/or adapt existing tools to generate it.

In the British context, despite the existence of official datasets to generate a synthetic population, there is no methodology to integrate these data and include the necessary number of socio-demographic attributes. Although SPENSER (Lomax et al., 2022) allows the definition of a synthetic population for any region in the UK, the number and diversity of individual attributes is scarce (i.e., age, sex, household location and ethnic group). This is an important drawback for synthetic populations applied into ABMs, as limited differences between agents can be identified, and thus the diversity of travel behaviours will be limited. This is relevant as research has identified different mobility patterns depending on the socio-demographic characteristics of the individuals (Mwale *et al.*, 2022). Research studies have found that single individuals spend more time in leisure time (Lee et al., 2004) and their physical activity is greater (Puciato *et al.*, 2021) than married people, while the presence of children in the family makes changes in the use of time, work situation and composition and size of social networks (Zwerts et al., 2007), encourages the use of the car (McCarthy et al., 2017) and affects more women's travel patterns than men (McGuckin et al., 2005; Ng *et al.*, 2018).

Economic spending power is another factor that influences the behaviour of individuals, which is derived from their economic activity and occupation type. Close (2020) conducted a survey in 2019 of adults living in the Tyne and Wear region (UK) to identify their willingness to use shared and emerging mobility services. Outcomes show that only specific groups in the society are interested in those modes, principally younger residents aged under 40 and those with household incomes of over £60,000. Additionally, those with higher levels of education think more actively about environmental concerns and use more diverse transport modes than other groups in the society, especially the youngest. In 2019, the UK Government released a report about inequalities in mobility and access in the transport system, showing that lower income households travel less overall in the UK, making nearly 20 % fewer trips and travel 40 % less distances than the average household (Lucas et al., 2019).

Lastly, the access to different transport modes also defines and conditions human travel behaviour. The possibility to access a car brings the possibility to go anywhere and whenever, while its lack conditions their movements and possibilities if other transport modes are not an alternative, which could generate barriers to employment, education and healthcare, besides producing social isolation (Lucas et al., 2019). Socio-demographic attributes also affect the use of cars, as highlighted by *Tiikkaja et al.* (2021), where it is stated that women have less access to the household car than men. Linked to the use of a car is the possession of a driving licence, which depends on socio-demographic attributes such as age and sex, among others (Department for Transport, 2023; National Travel Survey, 2023b). Based on the UK National Travel Survey (2023 a), the ownership of a bicycle differs on age, with young individuals (aged 5 to 10) being more likely to have access to bicycles (83 %) than any other individuals in different range of ages, although a peak between those aged 40 to 49 is observed (50 %).

These challenges present the need for a comprehensive and diverse set of data inputs for both demand (e.g. synthetic population) and supply (e.g. active travel networks) in the transport system. To facilitate this, this paper describes the development of a validated MATSim model of the Tyne and Wear region (England) that is representative of a normal transport working day. The novelty proposed in this paper is the development and application of a new, and open-source methodology to generate a detailed and heterogeneous synthetic population with 11 socio-demographic attributes that can be applied to any region in England. Additionally, a new network attribute (i.e., quietness) was developed to simulate more realistic cycling routes, taking into account several built-environment characteristics at once. Therefore, the MATSim bicycle extension developed by Ziemke et al. (2019b) was updated to take into account this new attribute. It is expected that the updated extension described in this paper could contribute to simulate more realistic cyclists' movements, in line with previously cited research.

The paper is organised as follows: Section 2 explains the proposed methodology to develop and validate the model. Section 3 explains how input datasets were generated applying the novel open-source methodology in the Tyne and Wear region (England). Section 4 describes the model calibration and validation stages to achieve a simulation model of the area of study representative of a regular working day. Section 5 discusses the findings and key aspects obtained from the validated MATSim model. Section 6 identifies and describes potential future works to improve the model. Sections 7 highlights the conclusions achieved.

2. Methodology

The methodology proposed to validate a MATSim model consists of the development of a synthetic population (blue box) and road network (green box). Once these two main components are generated, they can be imported into MATSim (grey box), where simulations (yellow box) are calibrated and validated (orange box), showing a realistic geospatial and temporal baseline representation of the transport mobility during a regular day in the study area. Fig. 1 summarises the methodology employed.

The foundation of a mobility ABM is a synthetic population (demand) of agents that represents the transport demand in the model region. Therefore, a synthetic population of the North East of England (the main region where people interacting with the Tyne and Wear area live) was created. A synthetic population constitutes a simplified digital twin of the real population, with individual socio-demographic characteristics (e.g., age, sex, economic activity, etc.) of all of the citizens in the geographic area of the study. Each individual in the synthetic population is also assigned an activity plan that represents the activities performed (e.g., star-end time and location, purpose of the trip and transport mode) on a normal working weekday, depending on their own socio-demographic characteristics. The development of a synthetic population is a key input to most agent-based simulations (Borysov et al., 2019), so its accuracy is crucial for a realistic representation of the population and their urban mobility interactions.

The second objective is the development of the network (supply) used by the individuals to move between activities by the diverse range of transport modes. It consists of a digital geospatial representation of the road network in the area of study, where the type of road, allowed modes and flow capacities are considered along with other attributes. The public transport network is incorporated within the network, considering the stops, routes and timetables of buses, trains and light rail, and the elevation and road characteristics for cycling.

Once demand and supply datasets are developed, they can be imported into an ABM platform to simulate different scenarios where agents interact in space and time. For this project, MATSim (Horni et al., 2016) was selected as the simulation platform. MATSim is an open-source framework that allows the implementation of large-scale agent-based transport simulations (MATSIM, 2023), which has been widely used by academics and industry for more than 15 years. This iterative model uses demand and supply datasets to simulate the interactions of agents in space and time. During each iteration, each agent receives a score (scoring phase – see Section 4) depending on their mobility performance (e.g., arriving late at destination reduces the score). Before starting a new iteration, a percentage of the population implements some changes in their activity plans (e.g., starting times, transport modes and routes chosen) (re-planning phase). After several iterations, the simulation achieves a relaxed state and simulation results can be analysed. The initial baseline scenario needs to be calibrated until results reflect a business-as-usual case in the area of study, validating the results against other sources of data (e.g., origin-destination matrices, sensor and statistical data).

3. Input datasets

The following sections describe the development and application of a novel methodology to generate a very detailed synthetic population and a cycling-focused network in the Tyne and Wear region.

3.1. Synthetic population (Demand)

As described in Section 1, the lack of a specific and detailed synthetic

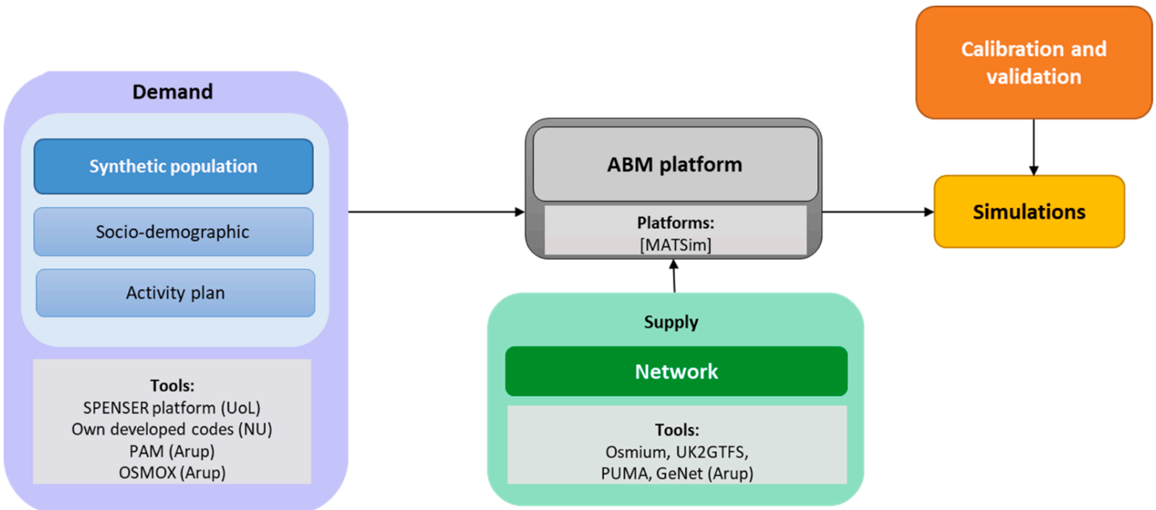

Fig. 1. Flow diagram of the methodology. A synthetic population with daily routines (blue box) alongside the network (green) are used as inputs into the ABM platform (grey). Simulations (yellow box) are calibrated and validated (orange box).

population methodology for an English context makes the generation of a detailed representation of the individuals from a specific region challenging. Therefore, a novel methodology combining SPENSER (Lomax et al., 2022) with a new Python tool (Alvarez Castro, 2022) were considered in this research and used in MATSim for the first time.

First, a synthetic population (households and individuals) was generated using SPENSER (Lomax et al., 2022), a tool developed by the University of Leeds that integrates UK 2011 census data and projects it into the future, based on survival probability and local authorities constraints (e.g., the development of new housing, new jobs, or transport infrastructure). The synthetic population generated represents the North East region of England for the year 2019 (the latest regular year before the Covid pandemic) and contains information about households (circa 1.2 million) and individuals (circa 2.6 million). For each household, there is information about the type of household, tenure, number of people, car access and census Output Area (OA) (the lowest level of geographical area for census statistics, made up of between 40 and 250 households (Office for National Statistics, 2023)) where it is located, among other attributes. Unfortunately, the number of attributes for the individuals is scarce, being limited to the age, sex, ethnic group and household ID to which they belong, making the outcome too simple and homogeneous to be used in transport ABMs.

In order to generate a more heterogeneous synthetic population, the inclusion of new attributes that allow the agents to be more diverse was required. Therefore, using the code-based method described by Alvarez Castro (2022), eight extra attributes where added based on open access UK 2011 census data projected to 2019 using Office of National Statistics (ONS) and National Travel Survey (NTS) datasets. These are marital status [married, couple or single], children dependency [true, false], driving license [true, false], car access [true, false], bicycle access [true, false], economic activity [employed, unemployed, inactive], occupation [nine categories for employed and unemployed; five for inactive] and annual gross income [numeric value]. The underlying methods are open access and can be applied to any region of England. The results obtained were internally validated against statistical 2019 ONS data. For all variables, the difference between synthetic and observed values grouped by sex were in the range 1–5 %.

Activity plans were assigned to the synthetic individuals using NTS travel diaries (Department for Transport, 2022a), the primary source of data on personal travel patterns by residents of England within Great Britain. The survey collects information on how, why, when and where people travel as well as factors affecting travel and individual socio-demographic characteristics, from 2002 to 2022 at Local Authority level (ibid). Individuals from the synthetic population were matched with NTS individuals based on similar socio-demographic characteristics (e.g., age range, marital status, dependents, economic activity, income, driving licence, car access, bicycle access), assuming that individuals with similar characteristics have similar mobility patterns.

Once the activity plans were assigned to each of the agents in the synthetic population, the purpose (16 types (home, work, education ((0–15), (16–18), (18+)), shop, supermarket, leisure, leisure sport, medical, eating out, other, accompanying education, accompanying work, accompanying leisure and accompanying other)); the start-end time (hh:mm); the distance travelled (kilometres); and the transport mode used (seven types (car, car passenger, bus, railway, light rail, walking and cycling)) were obtained for each of the activities. The only information missing was the location of the origin and destination for each trip, as the spatial granularity of the NTS dataset is at Local Authority level. Within the use of ABMs, there is the need-to-know specific locations (x, y coordinates), where the agents go to their activities. To identify the activity locations, OpenStreetMap (OSM) buildings (Geofabrik, 2023) were classified in eleven categories (work, education (0–15), (16–18), (18+), shop, supermarket, leisure, leisure sport, medical, eating out and other), using OSMOX (Arup, 2023) (Fig. 2).

Then, locations to activity plans were assigned depending on the activity purpose:

- Workplace: these activities were assigned using UK census origin-destination matrices (Office for National Statistics, 2011) and the distances travelled from the activity plan. Origin and destination matrices count the number of people moving between Middle layer Super Output Areas (MSOA) (level of geographical area comprising between 2000 and 6000 households (Office for National Statistics, 2023)). Once the destination of the workplace was identified at MSOA level, a building within the area dedicated to work purposes was assigned based on the building's capacity. This capacity value was calculated based on the volume of the building (floor area and number of floors derived from OSM) and the employment building density (UK Homes & Community Agency, 2015)), a UK guideline that defines an average number of square meters that an employee requires depending on the workplace type (e.g., retail, office, food store).
- Schools for children under 16 years old were assigned based on the distances travel from home (information from the activity plan assigned before) and the number of pupils each school has (Spooner

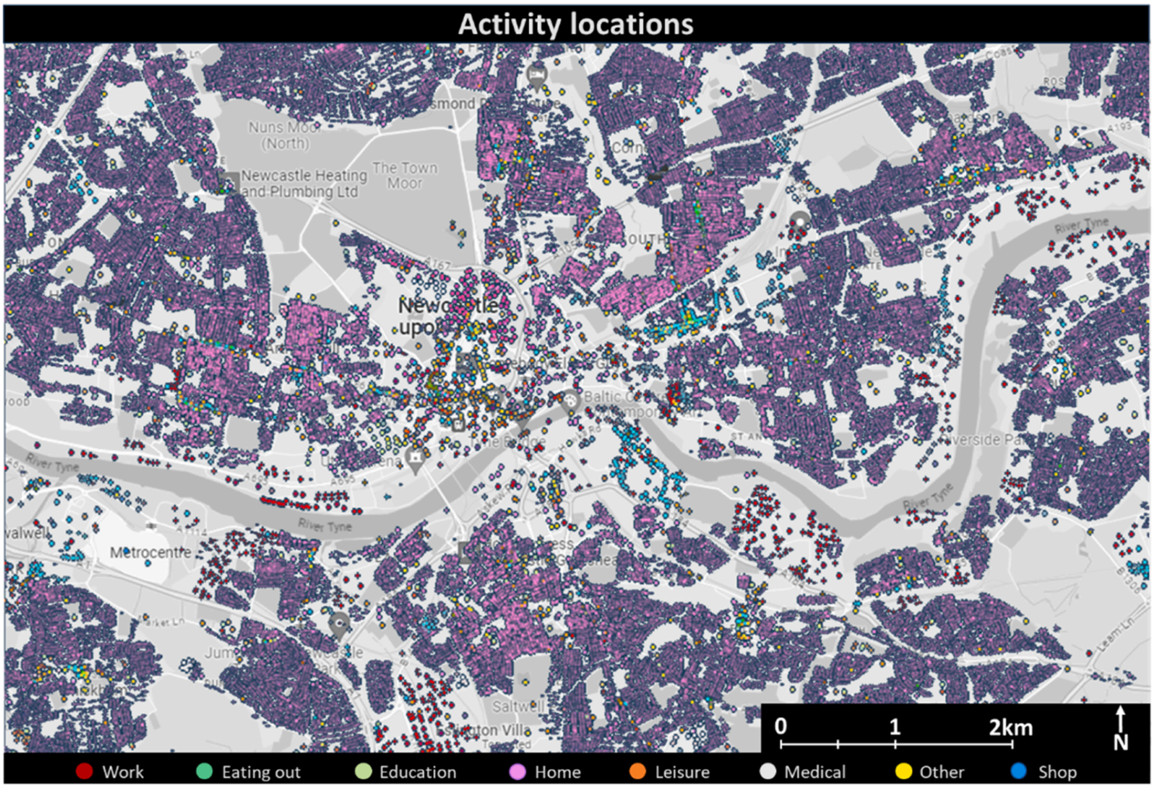

Fig. 2. Map showing the activity locations of the agents, based on the type of activity classified using OSMOX tool. Data obtained from Open Street Maps (OSM). OSM background map.(openstreetmap.org/copyright).

et al., 2021). This information was obtained from an open access dataset from the UK Government (2023).

- Accompanying activities: when an activity is shared with other member from the same household (i.e., when activity purpose is accompanying someone to work, education, or leisure), the location assigned was the same as the one visited by the other household member. In the case that another member of the household with the possibility to share the activity with is not found, the 'accompanying' activity is changed to 'leisure' or 'other' activity, randomly.
- Remaining activities (education for those older than 15 years, shop, supermarket, leisure, leisure sport, medical, eating out, other): these activities were located based on the distances provided in the activity plan, applying Spatial Interaction Modelling (SIM) techniques. Characteristics such as the competitiveness (e.g., number of all amenity types), attractiveness (e.g., number of amenities of the trip purpose in the area) and accessibility (inverse of the distance to the area and inverse of the distance to home) are considered. A measure combining these factors is calculated first for the surrounding OA or MSOA zones (OA when distance is below 10 km and MSOA in the rest of the cases) that can be reached within the distance from the activity plan. Based on a probability value, an OA or MSOA zone is selected. Then, a similar measure is calculated for the buildings located within the selected zone that are designed for the purpose of the activity, and one of them is selected.

Internal validation was performed comparing obtained results against NTS from 2019 and North East region in 2019 (Fig. 3). All graphs show similar patterns, with small discrepancies: differences are always below 5 % in trips duration (top) and 10 % in trip purposes and transport mode (middle and bottom). These differences may be due the small amount of household surveys done in 2019 (around 14.000) and North East specific surveys (around 900) when compared against the 1.2 million households from the synthetic population. Fig. 3 shows the results, where the percentage of trips by duration (top), by purpose (middle) and transport mode (bottom) are compared. Blue bars represent results obtained from the synthetic population, purple bars represent survey data of individuals from 2019, while pink bars represent those from the North East of England in 2019 only.

A population file (population.xml) containing all individuals of the synthetic population, their socio-demographic and their activity plans was generated in MATSim format, using PAM (Arup, 2020).

3.2. Network (Supply)

A road network dataset was created using OSM data from the North East of England (Geofabrik, 2023). Public transport information was obtained using General Transit Feed Specification (GTFS) data from two open access sources Department for Transport (DfT) (2023) and Rail Delivery Group (2023), which were merged using UK2GTFS tool (University of Leeds, 2022).

Both datasets, roads and GTFS, were merged and simplified using PUMA (Arup, 2022) and GeNet (Kozlowska et al., 2023), tools developed by Arup to create road networks for MATSim. The latter was also used to add elevation data to nodes and gradients to links by using a 10-metre resolution Digital Elevation Model (DEM) from DEFRA (Department for Environment Food & Rural Affairs, 2023) and viewfinderpanorama (de Ferranti et al., 2023). Additionally, roads were classified based on their characteristics for cycling, using Cyclestreets (2022) data. This attribute, named 'quietness', ranks roads as a score depending on their characteristics (e.g. road type, length, width, quality, surface, the existence of segregated cycle paths barriers, kerbs, crossing and junctions), ranging them from 0.0 (very poor) to 1.0 (excellent). This information was transferred into the network generated previously, by matching links with the same OSM identifier. To the best of the authors' knowledge, the use of this attribute is a novel implementation within MATSim models, which could allow simulating more realistic

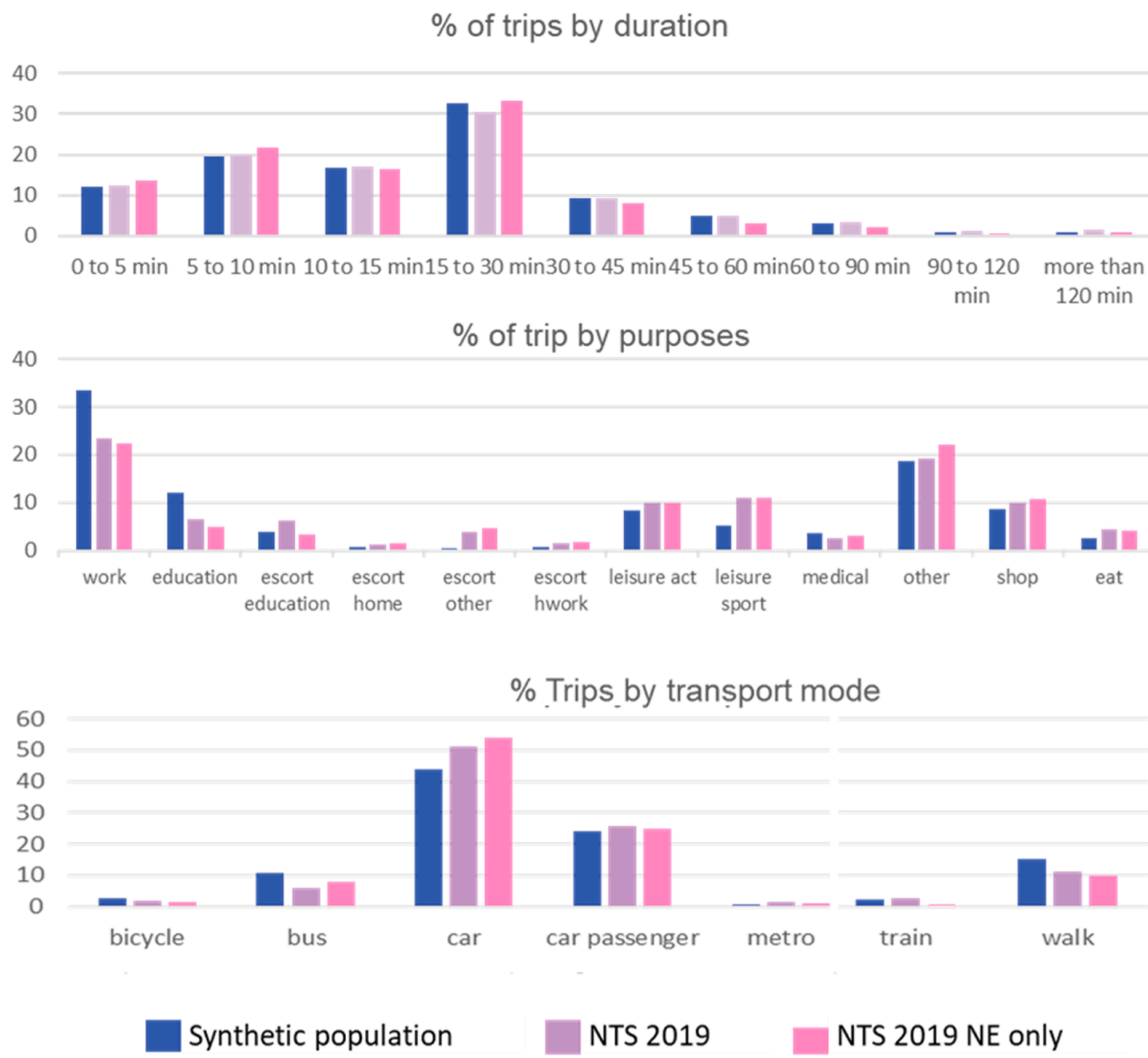

Fig. 3. Internal validation results. Percentage of trips by duration (top), by purpose (middle) and transport mode (bottom).

cycling routes, as more and diverse characteristics of the built-environment are considered when cyclists choose their routes.

These attributes (i.e., gradient and quietness) are fundamental factors to consider when cycling, as physical efforts are required when riding, in opposition to other vehicle types (e.g., cars, public transports). Research has found that characteristics such as slopes (Menghini et al., 2010; Hood et al., 2011; Li et al., 2012), pavement surface conditions and smoothness (Landis et al., 1997; Hölzel et al., 2012; Milakis et al., 2014), and the existence of continuous cycle paths (Sener et al., 2009; Li et al., 2012) are important factors that influence the use of bicycles. Therefore, their consideration in transport simulations are essential to simulate realistic scenarios where cycling is a predominant objective to be achieved.

The obtained result is a set of files containing a detailed network with cycling rating and gradient values (network.xml), public transport schedules (schedules.xml) and public transport vehicles information (vehicles.xml).

4. Model calibration and validation

Once the demand and supply datasets were generated, a MATSim scenario with a 20 % sample population (circa 200,000 agents) using Qsim controller was set up.

Seven different transport modes were defined (car driver, car passenger, bus, railway, light rail, bicycle and walk), where the use of car as a driver was restricted to only those agents with car access, based on their sociodemographic attributes. Public transport modes were simulated based on their schedules (i.e., deterministic), allowing walking access and egress to all transit stops, and also cycling to railway stations.

The replanning phase was defined with three different strategies, allowing 10 % of the agents to change the route, another 10 % to change the transport mode and another 10 % to modify the starting time of activities. The remaining 70 % of the agents choose an activity plan from previous iterations.

The scoring function used in the model to compute the satisfaction of each agent's plan when interacting in space and time with other agents and the environment was calculated using the Charypar-Nagel Utility Function, which is computed as (Charypar et al., 2005; Nagel et al., 2016) (Eq. (1)):

$$S_{plan} = \sum_{q=0}^{N-1} S_{act,q} + \sum_{q=0}^{N-1} S_{trav,\ mode(q)} \tag{1}$$

Eq. (1) Definition of the scoring function applied in the model

Where $S_{act,q}$ is the utility (satisfaction) that the agent obtains when performing activity q (normally positive), while $S_{trav,\ mode(q)}$ is the (dis) utility that the agent obtains when travelling between activities (normally negative). N is the total number of activities performed by the agent.

Additionally, MATSim's bicycle extension (Ziemke et al., 2019 b) was enabled and updated in order to consider the 'quietness' value included in the road network. The reasons of using this attribute instead of the default attributes of comfort and type of road included in the bicycle extension are mainly two. Firstly, the former is a better-defined value, based on good local knowledge. Secondly, the quietness attribute was obtained for almost the whole area of study, while the OSM comfort attribute was not exhaustively available in the input data for the study region.

A new marginal utility of 'quietness' (βquietness (a)) was included in the code, similarly as those generated for the comfort and infrastructure attributes in the original code. The effectively used marginal utility of 'quietness' for a link a is computed as follows (Eq. (2)):

$$\beta_{quietness\ (a)} = \beta^{maxquietness\ (a)} * (1 - quietness(a)) \tag{2}$$

Eq. (2) Definition of the marginal utility of quietness

Where $\beta^{maxquietness\ (a)}$ is always 1.0 and $quietness(a)$ is the quietness value of each link. This functionality has included in MATSim since version 15.0.

Table 1 summarises the main parameters and values used to set up the model for calibration and validation.

The model was validated against official statistics identifying transport modes when commuting to work (Department for Transport, 2022b) and official datasets that count the number of vehicles passing through specific roads per hour (Gateshead Council, 2023). Results show minimal differences (up to 3.5 %) in the percentage of commuting trips made by each transport mode (living in the area of study (orange bars) and working in the area of study (green)) (Fig. 4).

The number of vehicle counts per hour in different zones of the study area show some discrepancies depending on the type of road. Motorways were found to have less vehicles during off peak times, as freight and trips from other regions passing through the area of study were not considered. In the case of urban areas, very similar numbers of vehicles per hour were obtained when compared with real data. Fig. 5 shows the results achieved from several locations, distinguishing motorways (1, 2) and urban areas (3, 4). In all cases, red bars represent real data while blue simulated results. Fig. 6 shows the full day results from each station, where it can be observed a close representation of the number of vehicles simulated (y-axis) and in the real world (x-axis), with some discrepancies principally located in motorways, as discussed before.

Table 1
Definition of main parameters and values used to calibrate and validate the model.

Parameter	Value, constraint and assumptions
Population sample	20 % of those agents interacting with the Tyne and Wear region within the North-East of England
Number of iterations	1500
Controller	Qsim
Transport modes	car, car passenger, bike, walk, pt (bus, rail, subway, ferry)
Cars	Only those agents with access to car in their socio-demographic attributes (car_Access = True) were allowed to use the car in the simulation (considerCarAvailability (true)). Cars were allowed to overtake bicycles (linkDynamics = PassingQ)
Public transport	Simulated as deterministic (modules SBBPT and SwissRailRaptor). Maximum vehicle capacity was not considered (useCapacityConstraints = false), Economic cost was not considered Access and egress to public stops allowed on foot and by bicycle.
Bicycle	Updated bicycle extension enabled where road gradient and quietness attributes are considered when choosing a route. Marginal utility of gradient (−0.02) and quietness (−0.035) values.
Walking	Simulated as teleported mode.
Strategies	Reroute (0.1) TimeAllocatorMutator (0.1) SubtourModeChoice (0.1) ChangeExpBeta (0.7)
Strategy criteria	80 % of all iteration
Transport ASC	car: −0.37 car passenger: −1.7 bike: −1.1 walk: 0.0 bus: −7.2 rail: −0.001 subway: −0.001 ferry: −0.001

Also, an analysis of the cycling routes was performed, in order to identify if cyclists behave as expected based on the parameters used in the updated MATSim bicycle extension (i.e., gradient and quietness attributes). Fig. 7 shows the results of a cyclist when the bicycle contribution is disabled (top) and enabled (bottom). Cycle paths are represented as green lines, while the followed route is in black. Additionally, a route profile generated by Google Maps (Google, 2024) is attached to each route (top right), showing elevation variations in each case.

The differences between the two variants are clear, as in the first case only the travel time is considered and the shortest route is chosen, while in the second case, characteristics of the environment (i.e., gradient) and built-environment (i.e., quietness) are taken into account besides the travel time, being clearly visible the use of cycle paths (green lines). Terrain profiles (on the top right corner of each route) show a steeper route in the first case, while the second achieves a smoother route, especially when crossing the river. When the bicycle extension is disabled, the cyclist decided to choose the shortest trip, which made the cyclist to cross the river using a low bridge. Contrary, when the bicycle extension is enabled, the cyclist decided to use a higher bridge to avoid descending to the river and ascending again in the other side. In aggregated terms, the first case requires to ascend and descend 37 and 32 metres, respectively, while in the second only 28 and 27 metres.

5. Discussion

The application of the new open-source methodology to generate a synthetic population in the Tyne and Wear region achieved a heterogeneous result with 11 socio-demographic attributes (e.g., household location, age, sex, marital status, children dependency, driving license, car access, bicycle access, economic activity, occupation and annual gross income), and an activity plan (purpose of trip, transport mode, start-end time and geospatial locations (projected point locations)) based on their characteristics. The socio-demographic attributes obtained differences between 1 and 5 % when validated with official UK datasets, while the activity plans shown differences up to 5 % in trip durations and 10 % in trip purposes and transport modes.

This implemented methodology allows the generation of a very detailed and heterogeneous digital representation of the individuals in the study area, as activity plans assigned to each agent were based on a great variety of socio-demographic attributes, implying a great diversity of activity plans during the simulation stage. As discussed previously, scientific research have found that age, gender, income, work status and family size are some of the factors that influencing travel demand (Mwale et al., 2022). As such, this approach provides new tools for Agent-Based transport modellers in the UK context.

The development of the network has also been updated with the inclusion of a new parameter (i.e., quietness), which ranks roads for cycling based on their built-environment characteristics. This attribute, calculated by Cyclestreets (2022), can be considered more accurate than the standard attributes used in the MATSim bicycle extension (i.e., road type and comfort), as the former is a better-defined value based on local knowledge, while the latter are not exhaustively available for the whole study region. Consequently, the MATSim bicycle extension was updated to take into account this new attribute, similarly as former attributes were coded, being available for any researcher from version 15.0. The validation shows that this updated extension could contribute to simulate more realistic journeys by bicycle users.

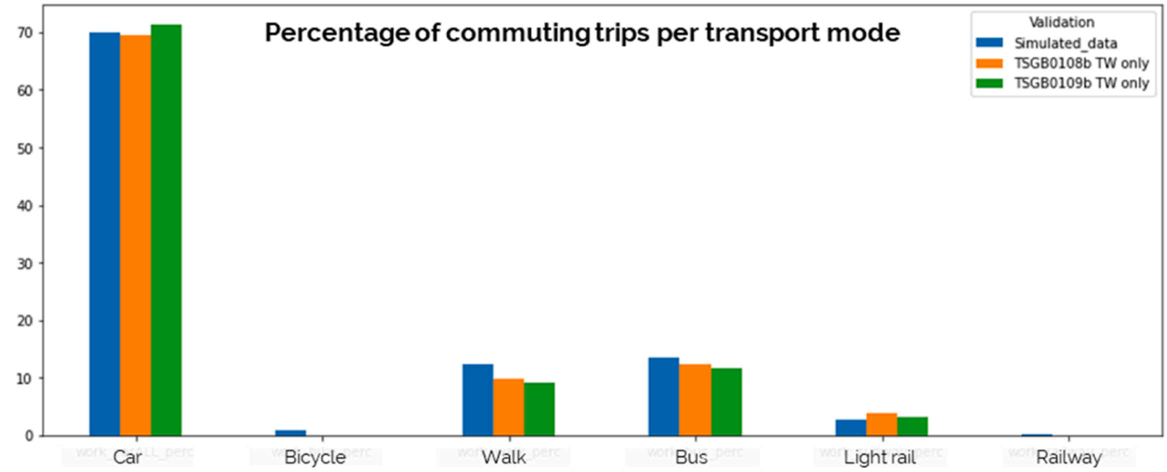

Fig. 4. Validation results of the percentage of commuting trips per transport mode

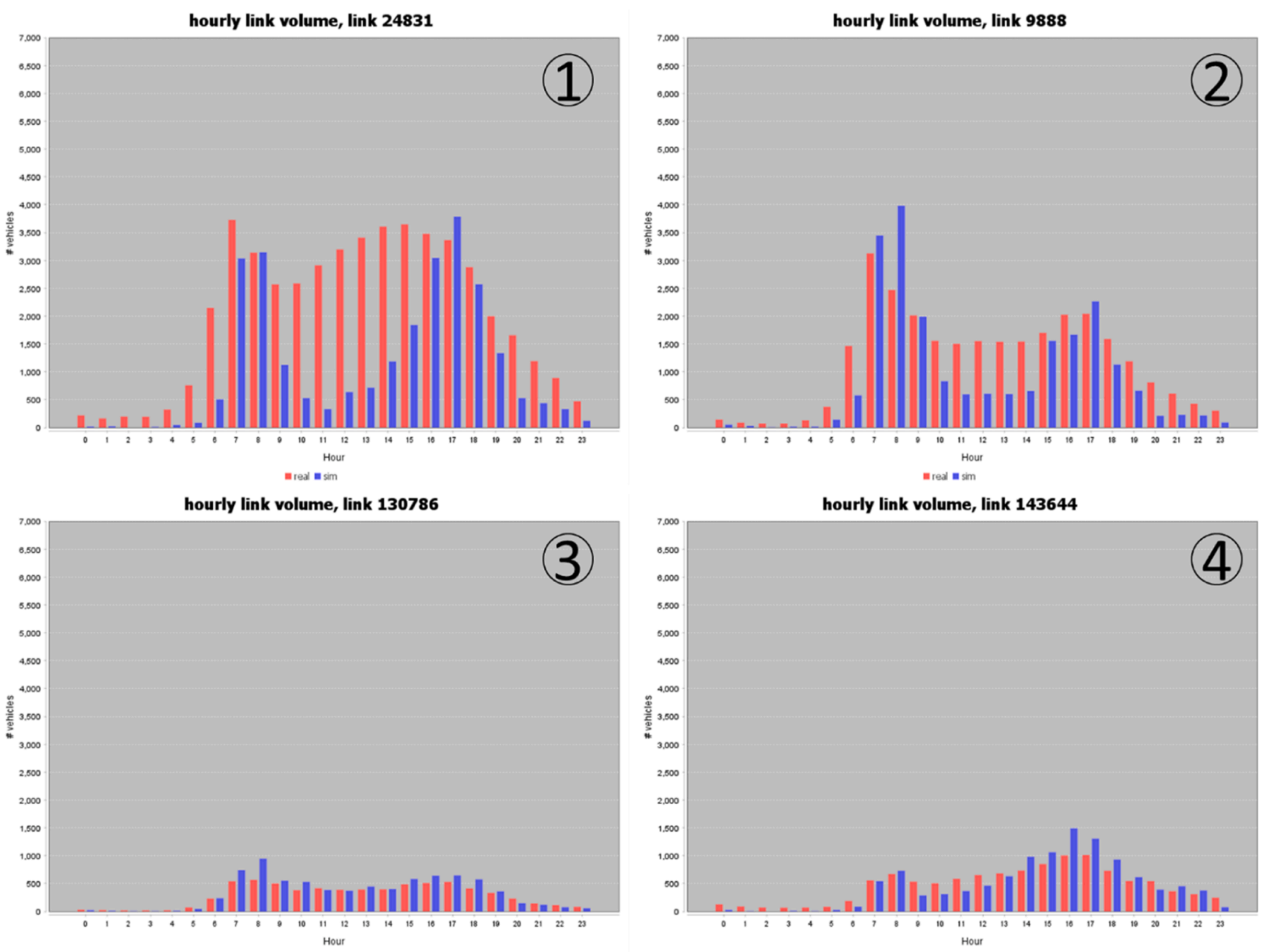

Fig. 5. Vehicle counts at different periods of the day in some motorway areas (1 and 2) and urban zones (3 and 4).

Results obtained from the validated MATSim model represents a reasonable facsimile of a normal working day in the area of study. The model was validated against official datasets of the area of study until differences in the percentage of commuting trips by transport mode were below 3.5 %-points and vehicle counts in urban areas reflect a real pattern per hour, although differences were observed between motorways and urban zones. In the case of motorways, fewer vehicles during off-peak periods were observed due principally to the lack of freight and other journeys passing through the study area starting and/or ending outside of it. In the case of urban zones, accurate and precise results were

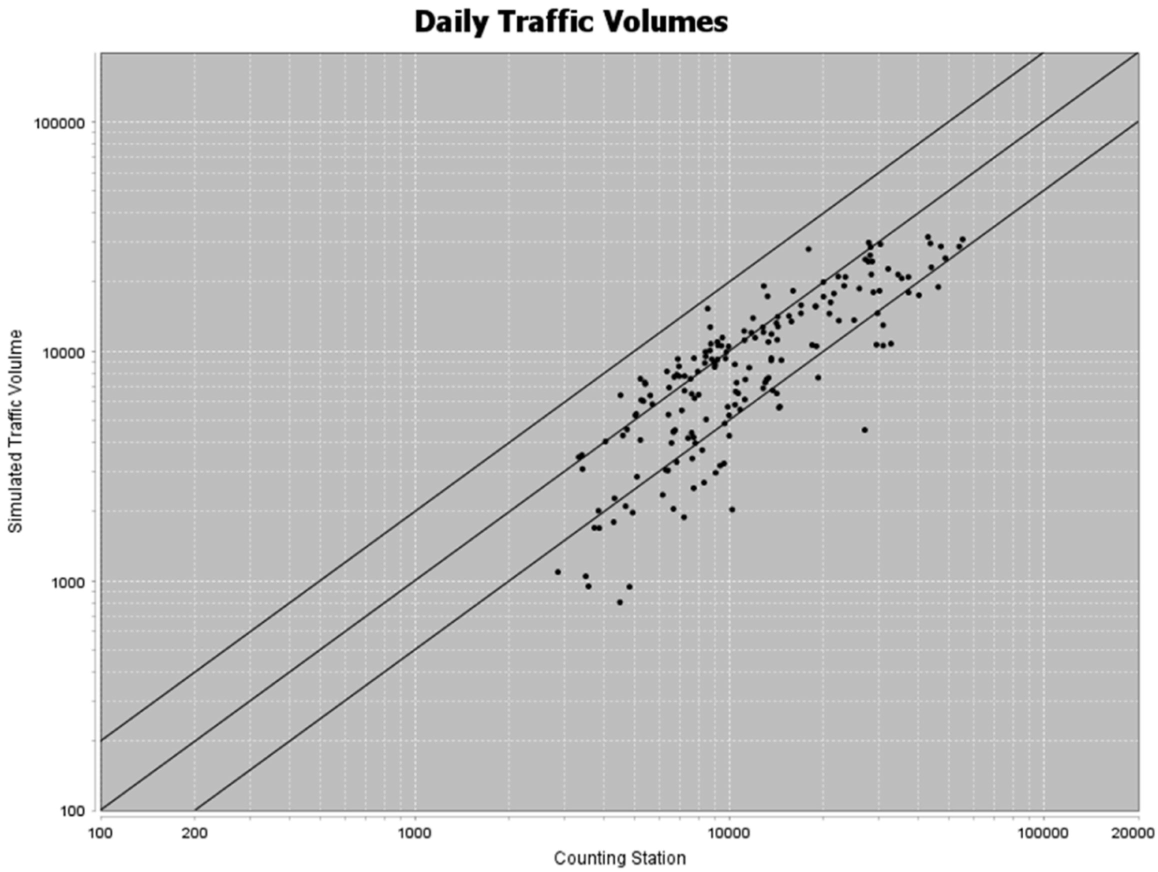

Fig. 6. Daily traffic volumes comparisons between real counting stations (x-axis) and simulated results (y-axis).

obtained during the complete simulated day.

Despite the lack of freight and other trips passing through the study area could be considered as a limitation of the generated model, this could be assumed in case the validated model is used predominately to test policies to reduce the use of cars only in urban areas. Those policies could be focussed on shifting car users to more sustainable modes, such as active modes. In that case, the main trips to shift would come from short trip distances, principally from urban areas, where the global traffic flow in space and time is analogous to the normal mobility in a regular day.

Additionally, the validated model showed the behaviour of cyclists when choosing their routes based on the network attributes (i.e., gradient and quietness). It was achieved a realistic simulation of the routes followed by cycling agents, choosing flat routes and the use of cycle paths, when available, assuming a normal cyclists' behaviour when compared with several real cyclist cases in the study area.

The whole methodology used in this study can be easily transferable to any other region in England by adapting the input data to the desired area. All datasets used are open access except the NTS travel diaries (Department for Transport, 2022a), which requires a special license from the UK Data Service. All tools used are open access except PUMA, but a similar outcome can be achieved using the open access PT2MATSim tool (PT2MATSim, 2020) within MATSim.

6. Future work

The steps described above represent the baseline scenario of the regular transport mobility of the study region using a new and transferrable methodology to MATSim scenarios in England. It shows, in space and time, the normal behaviour of the agents when performing their daily routines, a valuable source of information that can be used to analyse the current status of daily transport mobility.

The model could be improved by incorporating more socio-economic attributes to the synthetic population. Characteristics such as the health status (e.g., very good, good, bad, very bad) could identify and represent different mobility realities within the society, and therefore, simulate more realistic and integrated scenarios, where vulnerable and/or disabled individuals are taken into account.

On the supply side, additional network characteristics alongside quietness, such as street lighting, urban greenspace, or noise levels, could be included to further enrich the representation of road infrastructure for active travel and thus route choices made by cycle users. Data on such factors is limited, so additional data collection would be necessary.

Additionally, the model can be used to test the efficiency of different urban mobility policies to identify an effective combinations of infrastructure interventions and human behaviour changes to reduce private and polluting vehicles in urban areas in favour of active modes. Several push or pull policies could be defined and implemented by modifying the input datasets of the model. Some examples could be the implementation of Low Traffic Neighbourhoods (LTNs) or road user charging schemes. The heterogeneous synthetic population of agents, and the inclusion of the new quietness attribute, allow a more realistic simulation of such policies within the supply and demand datasets input to MATSim.

7. Conclusion

This paper has shown the application of a new methodology to generate the main MATSim input datasets and to develop a baseline scenario for any region of England with a special interest in cycling, applying open-source tools and open access datasets, when possible.

A very detailed synthetic population with 11 socio-demographic attributes and activity plan (based on their socio-demographic

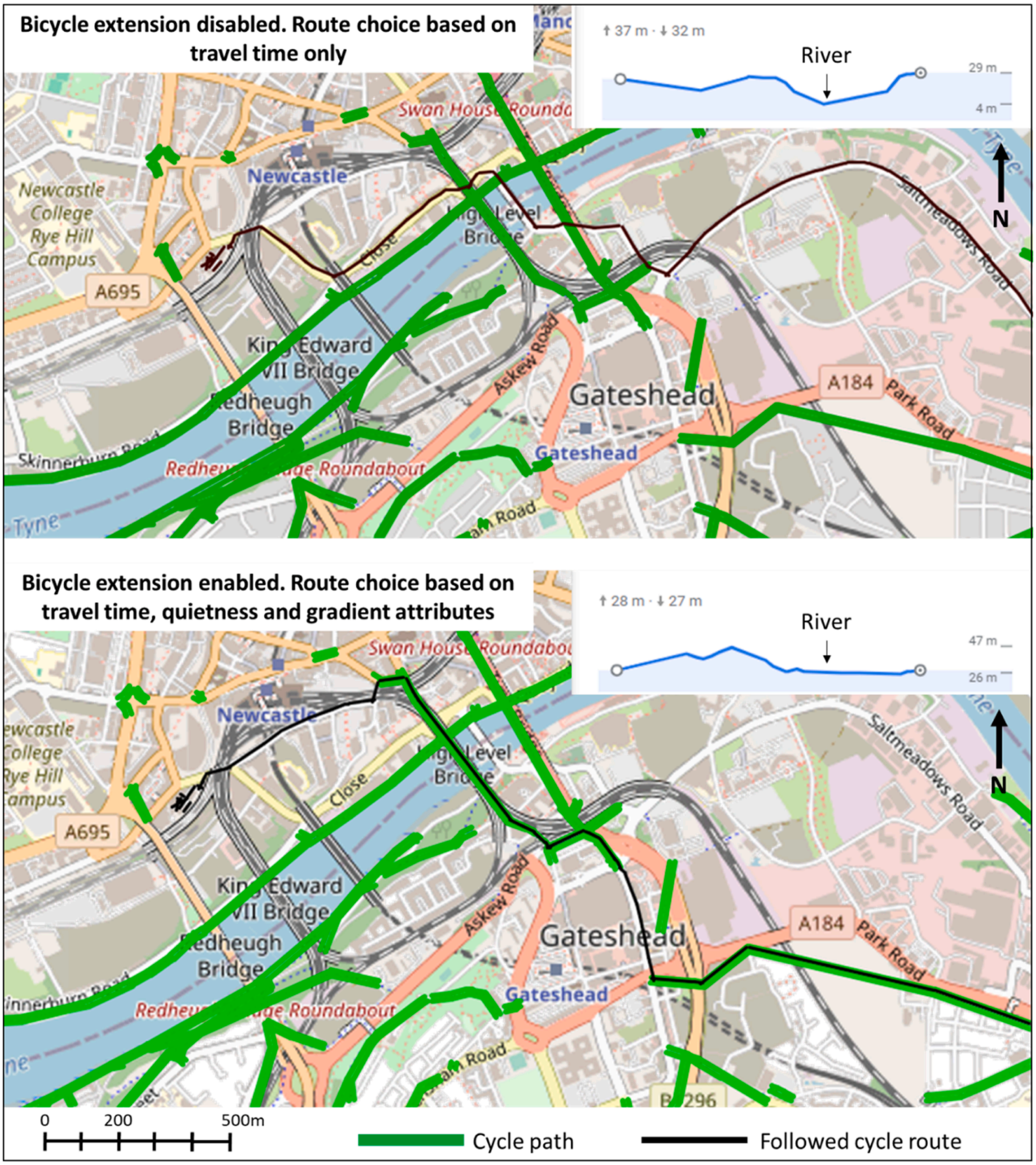

Fig. 7. Comparison of the cycling route followed by an agent when the updated bicycle contribution was disabled (top) and enabled (bottom). OSM background map. (openstreetmap.org/copyright).

characteristics) for each individual in the study area was generated. The level of heterogeneity achieved allows simulating a great variety of different mobility patterns during the simulation stage, which implies a more realistic representation of the normal mobility in the study area.

Besides, a cycling-focused network was generated, as a new attribute ranking the roads for cycling depending on their built-environment characteristics was included. This attribute brings the possibility to simulate more realistic cycling routes, as more attributes of the roads are taken into account. The updated bicycle extension considers it and can be calibrated using the gradient and quietness attributes instead of the former road type and comfort.

The validated MATSim model generated is representative of the normal mobility in the study area, as a consequence of the input datasets generated using the novel methodology proposed in this paper. Minimum differences in the share modes were obtained as well as vehicle counts in urban areas.

Such a tool could provide a basis for decision-making on active travel policies in urban areas of England.

Data access statement

Data access statement The tools, documentation and data collection support required to generate a detailed synthetic travel demand for any region in England are available at:

- SPENSER: https://hub.docker.com/r/nismod/spenser
- synthetic_population_dev: https://github.com/DACNC/synthetic_population_dev
- activity_plans_dev: https://github.com/DACNC/activity_plans_dev
- PAM: https://github.com/arup-group/pam

Due to ethical concerns, the required dataset to generate the activity plans (National Travel Survey, 2002-2022: Special Licence Access) cannot be made openly available. Further information about the data and condition of access are available at: http://doi.org/10.5255/UKDA-SN-7553-11 The tools and documentation to generate the transport network are available at:

- OSMOX: https://github.com/arup-group/osmox
- GeNET: https://doi.org/10.5281/zenodo.8274051

Due to privacy reasons, PUMA tool is not open access. Alternatively, open access PT2MATSim tool is available at: https://github.com/matsim-org/pt2matsim OpenStreetMap (OSM) dataset to generate the network is available at: https://download.geofabrik.de/ GTFS datasets to obtain information about public transport services in England are available at

- Rail Delivery Group: https://data.atoc.org/ (information about trains only)
- Traveline: https://www.travelinedata.org.uk/travelineopen-data/traveline-national-dataset/

(other public transport modes) Open access cycleability rating ('quietness') dataset is available at: https://bikedata.cyclestreets.net/cycleability/Open access Digital Elevation Model (DEM) datasets are available at:

- Department for Environment Food and Rural Affairs (DEFRA): https://environment.data.gov.uk/survey
- Viewfinderpanorama: http://viewfinderpanoramas.org/dem3.html

MATSim tool is available at: https://github.com/matsim-org

CRediT authorship contribution statement

David Alvarez Castro: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alistair Ford:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Philip James:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Roberto Palacín:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Dominik Ziemke:** Writing – review & editing, Validation, Supervision, Software, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results

Acknowledgement

The authors would like to thank the Geospatial Systems Centre for Doctoral Training at Newcastle University (UK), funded by EPSRC, for their support.

Funding

This research was funded by EPSRC Centre for Doctoral Training in Geospatial Systems, grant number EP/S023577/1.

References

Adnan, M., Pereira, F.C. and Azevedo, C.L. (2016) 'SimMobility : A multi-scale integrated agent-based simulation platform', (July).

Agarwal, A., & Lämmel, G. (2016). Modeling seepage behavior of smaller vehicles in mixed traffic conditions using an agent based simulation. *Transportation in Developing Economies, 2*(2). https://doi.org/10.1007/s40890-016-0014-9

Alvarez Castro, D. (2022) 'synthetic_population_dev'. Available at: https://github.com/DACNC/synthetic_population_dev.

Arup (2020) *PAM*. Available at: https://github.com/arup-group/pam (Accessed: 21 June 2021).

Arup (2022) 'PUMA'. Available at: https://github.com/arup-group/puma.

Arup (2023) 'Osmox'. Available at: https://github.com/arup-group/osmox.

Bandini, S., Manzoni, S., & Vizzari, G. (2009). Agent based modeling and simulation: An informatics perspective. *Jasss, 12*(4).

Bastarianto, F. F., et al. (2023). Agent-based models in urban transportation: Review, challenges, and opportunities. *European Transport Research Review, 15*(1). https://doi.org/10.1186/s12544-023-00590-5

Bazzan, A. L. C., & Klügl, F. (2014). A review on agent-based technology for traffic and transportation. *Knowledge Engineering Review, 29*(3), 375–403. https://doi.org/10.1017/S0269888913000118

Becker, H., et al. (2020). Assessing the welfare impacts of shared mobility and Mobility as a Service (MaaS). *Transportation Research Part A: Policy and Practice, 131* (September 2019), 228–243. https://doi.org/10.1016/j.tra.2019.09.027

Borysov, S. S., Rich, J., & Pereira, F. C. (2019). How to generate micro-agents? A deep generative modeling approach to population synthesis. *Transportation Research Part C: Emerging Technologies, 106*(July), 73–97. https://doi.org/10.1016/j.trc.2019.07.006

Cardinot, M., et al. (2019). Evoplex: A platform for agent-based modeling on networks. *SoftwareX, 9*, 199–204. https://doi.org/10.1016/j.softx.2019.02.009

Charypar, D., & Nagel, K. (2005). Generating complete all-day activity plans with genetic algorithms. *Transportation, 32*(4), 369–397. https://doi.org/10.1007/s11116-004-8287-y

Close, Jundi (2020). *Demand Modelling and Assessment through a Network Demonstrator (DeMAND) Project Contents Notice*. Available at: https://cp.catapult.org.uk/wp-content/uploads/2021/12/DeMAND-Project_attitudes-towards-shared-mobility.pdf.

Cyclestreets (2022) *Bikedata data to support getting more people cycling. Data created by CycleStreets, using data from OpenStreetMap © OpenStreetMap*. Available at: https://bikedata.cyclestreets.net/(Accessed: 10 November 2023).

Department for Environment Food and Rural Affairs (2023) *Defra survey data download*. Available at: https://environment.data.gov.uk/survey (Accessed: 3 November 2023).

Department for Transport (2022a) 'National Travel Survey, 2002-2021: Special Licence Access', *UK Data Service*, 11th Editi. doi: http://doi.org/10.5255/UKDA-SN-7553-11.

Department for Transport (2022b) *National Travel Survey*. Available at: https://www.gov.uk/government/collections/national-travel-survey-statistics#:~:text=National Travel Survey (NTS) is,of England within Great Britain. (Accessed: 5 July 2023).

Department for Transport (2023) *Driving licence holding and vehicle availability. Table NTS0201. Full car driving licence holders by age and sex, aged 17 and over: England, 1975 onwards*. Available at: https://www.gov.uk/government/statistical-data-sets/nts02-driving-licence-holders (Accessed: 19 October 2023).

Department for Transport (DfT) (2023) *Bus open data implementation guide*. Available at: https://www.gov.uk/government/publications/bus-open-data-implementation-guide/bus-open-data-implementation-guide (Accessed: 3 November 2023).

El-sayed, A.M. et al. (2012) 'Social network analysis and agent-based modeling in social epidemiology', pp. 1–9. doi: 10.1186/1742-5573-9-1.

Erath, A. and Chakirov, A. (2016) 'The Multi-Agent transport simulation MATSim. Singapore', (2012), pp. 379–382.

de Ferranti, J. and Hormann, C. (2023) *Digital elevation data*. Available at: http://viewfinderpanoramas.org/dem3.html (Accessed: 28 November 2023).

Ford, A., et al. (2018). Land-use transport models for climate change mitigation and adaptation planning. *Journal of Transport and Land Use, 11*(1), 83–101. https://doi.org/10.5198/jtlu.2018.1209

Gateshead Council (2023) *Traffic and accident data unit*. Available at: https://www.gateshead.gov.uk/article/4464/Traffic-and-Accident-Data-Unit (Accessed: 4 November 2023).

Geofabrik (2023) *OpenStreetMap Data Extracts*. Available at: https://download.geofabrik.de/(Accessed: 27 November 2023).

Ghaffarzadegan, N., Lyneis, J., & Richardson, G. P. (2011). How small system dynamics models can help the public policy process. *System Dynamics Review, 27*(1), 22–44. https://doi.org/10.1002/sdr.442

Ghorbani, A., et al. (2014). Enhancing ABM into an inevitable tool for policy analysis. *Journal of Policy and Complex Systems, 1*(1), 61–76. Available at: https://research.tudelft.nl/en/publications/enhancing-abm-into-an-inevitable-tool-for-policy-analysis.

Google (2024) *Google maps*. Available at: https://www.google.com/maps/@53.9590656,-1.0551296,13z?entry=ttu (Accessed: 9 January 2024).

Gr, J. et al. (2019) 'A cognitive model for routing in agent-based modelling A cognitive model for routing in agent-based modelling', 250005(July). doi: https://doi.org/10.1063/1.5114245.

Gurram, S., Stuart, A. L., & Pinjari, A. R. (2019). Agent-based modeling to estimate exposures to urban air pollution from transportation: Exposure disparities and impacts of high-resolution data. *Computers, Environment and Urban Systems, 75* (January), 22–34. https://doi.org/10.1016/j.compenvurbsys.2019.01.002

Heppenstall, A., Malleson, N., & Crooks, A. (2016). Space, the final frontier": How good are agent-based models at simulating individuals and space in cities? *Systems, 4*(1). https://doi.org/10.3390/systems4010009

Hitge, G., & Joubert, J. W. (2023). The survivability of cycling in a co-evolutionary agent-based model. *Transportation*, (0123456789)https://doi.org/10.1007/s11116-023-10422-z

Hölzel, C., Höchtl, F., & Senner, V. (2012). Cycling comfort on different road surfaces. *Procedia Engineering, 34*, 479–484. https://doi.org/10.1016/j.proeng.2012.04.082

Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters, 3*(1), 63–75. https://doi.org/10.3328/TL.2011.03.01.63-75

Horni, A., Nagel, K., & Axhausen, K. W. (2016). The multi-agent transport simulation MATSim. *The Multi-Agent Transport Simulation MATSim*. https://doi.org/10.5334/baw

Kickhofer, B (2016). Munich (eds.). In A Horni, K Nagel, & KW. Axhausen (Eds.), *The multi-agent transport simulation MATSim* (pp. 383–384). Ubiquity Press. https://doi.org/10.5334/baw.58. Available at:.

Kilani, M., Diop, N., & De Wolf, D (2022). A multimodal transport model to evaluate transport policies in the North of France. *Sustainability (Switzerland), 14*(3), 1–16. https://doi.org/10.3390/su14031535

Kozlowska, K. et al. (2023) 'arup-group/genet: v4.0.0 (v4.0.0.0)'. https://doi.org/10.5281/zenodo.8274051.

Krajzewicz, D., et al. (2012). Recent Development and Applications of SUMO - Simulation of Urban MObility. *International Journal On Advances in Systems and Measurements, 5*(3), 128–138. Available at: http://elib.dlr.de/80483/.

Landis, B. W., Vattikuti, V. R., & Brannick, M. T. (1997). Real-time human perceptions: toward a bicycle level of service. *Transportation Research Record, 1578*(1), 119–126. https://doi.org/10.3141/1578-15

Leblond, V. (2020) 'A new agent-based software for designing and optimizing emerging mobility services : Application to city of Rennes', (December).

Lee, Y. G., & Bhargava, V. (2004). Leisure time: Do married and single individuals spend it differently? *Family and Consumer Sciences Research Journal, 32*(3), 254–274. https://doi.org/10.1177/1077727X03261631

Li, Z., et al. (2012). Physical environments influencing bicyclists' perception of comfort on separated and on-street bicycle facilities. *Transportation Research Part D: Transport and Environment, 17*(3), 256–261. https://doi.org/10.1016/j.trd.2011.12.001

Lomax, N., et al. (2022). An open-source model for projecting small area demographic and land-use change. *Geographical Analysis, 54*(3), 599–622. https://doi.org/10.1111/gean.12320

Lucas, K. et al. (2019) 'Inequalities in mobility and access in the UK transport system future of mobility: Evidence review foresight, government office for science 2 inequalities in mobility and access in the UK transport system', p. 80. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/784685/future_of_mobility_access.pdf.

Maggi, E., & Vallino, E. (2016). Understanding urban mobility and the impact of public policies : The role of the agent-based models Research in Transportation Economics Understanding urban mobility and the impact of public policies : The role of the agent-based models. *Research in Transportation Economics, 55*(June), 50–59. https://doi.org/10.1016/j.retrec.2016.04.010

Manley, E., et al. (2021). A spatial model of cognitive distance in cities A spatial model of cognitive distance in cities. *International Journal of Geographical Information Science, 00*(00), 1–23. https://doi.org/10.1080/13658816.2021.1887488

Manley, E., & Cheng, T. (2018). Exploring the role of spatial cognition in predicting urban traffic flow through agent-based modelling. *Transportation Research Part A, 109*(November 2017), 14–23. https://doi.org/10.1016/j.tra.2018.01.020

Martinez, L. M., & Viegas, J. M. (2017). Assessing the impacts of deploying a shared self-driving urban mobility system: An agent-based model applied to the city of Lisbon. *International Journal of Transportation Science and Technology, 6*(1), 13–27. https://doi.org/10.1016/j.ijtst.2017.05.005

'MATSIM' (2023). Available at: https://www.matsim.org/about-matsim.

McCarthy, L., et al. (2017). Factors influencing travel mode choice among families with young children (aged 0–4): A review of the literature. *Transport Reviews, 37*(6), 767–781. https://doi.org/10.1080/01441647.2017.1354942

McGuckin, N. and Nakamoto, Y. (2005) 'Differences in trip chaining by men and women', *Research on Women's Issues in Transportation - Volume 2: Technical Papers*, pp. 49–56.

Menghini, G., et al. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice, 44*(9), 754–765. https://doi.org/10.1016/j.tra.2010.07.008

Milakis, D., & Athanasopoulos, K. (2014). What about people in cycle network planning? Applying participative multicriteria GIS analysis in the case of the Athens metropolitan cycle network. *Journal of Transport Geography, 35*, 120–129. https://doi.org/10.1016/j.jtrangeo.2014.01.009

Minh Duc, P., et al. (2020). HoanKiemAir: Simulating impacts of urban management practices on traffic and air pollution using a tangible agent-based model. In *Proceedings - 2020 RIVF International Conference on Computing and Communication Technologies, RIVF 2020*. https://doi.org/10.1109/RIVF48685.2020.9140787

Müller, J., et al. (2022). Integration of Different Mobility Behaviors and Intermodal Trips in MATSim. *MDPI Sustainability, 14*(1). https://doi.org/10.3390/su14010428

Mwale, M., Luke, R., & Pisa, N. (2022). Factors that affect travel behaviour in developing cities: A methodological review. *Transportation Research Interdisciplinary Perspectives, 16*(March), Article 100683. https://doi.org/10.1016/j.trip.2022.100683

Nagel, K., et al. (2016). A Closer Look at Scoring'. *The Multi-Agent Transport Simulation MATSim*, 23–34. https://doi.org/10.5334/baw.3

National Travel Survey (2023a) *Cycling, motorcycling, school travel, concessionary travel and road safety. Table NTS0608 Pedal cycle ownership by age, aged 5 and over: England, 2002 onwards*. Available at: https://www.gov.uk/government/statistical-data-sets/nts06-age-gender-and-modal-breakdown (Accessed: 19 October 2023).

National Travel Survey (2023b) *Region and Rural-Urban Classification. Table NTS9901 Full car driving licence holders by sex, region and rural-urban classification of residence, aged 17 and over: England, 2002 onwards*. Available at: https://www.gov.uk/government/statistical-data-sets/nts99-travel-by-region-and-area-type-of-residence (Accessed: 19 October 2023).

Ng, W.-S., & Acker, A. (2018). Understanding urban travel behaviour by gender for efficient and equitable transport policies. *International Transport Forum*, 1–19 (2018–01)Available at: https://www.econstor.eu/handle/10419/194064.

Office for National Statistics (2011) *Location of usual residence and place of work by method of travel to work (MSOA level)*. Available at: https://www.nomisweb.co.uk/census/2011/wu03EW (Accessed: 13 February 2024).

Office for National Statistics (2023) *Census 2021 geographies*. Available at: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies#:~:text=1. Output Areas (OAs),and household changes since 2001 (Accessed: 10 October 2023).

Park, B. H., et al. (2018). A high resolution agent-based model to support walk-bicycle infrastructure investment decisions: A case study with New York City. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018-Novem* (pp. 3779–3784). https://doi.org/10.1109/ITSC.2018.8569232

PT2MATSim (2020) 'PT2MATSim'. Available at: https://github.com/matsim-org/pt2matsim.

Puciato, D., & Rozpara, M. (2021). Physical activity and socio-economic status of single and married urban adults: A cross-sectional study. *PeerJ, 9*, 1–19. https://doi.org/10.7717/peerj.12466

Rail Delivery Group (2023) *Rail Industry Data*. Available at: https://data.atoc.org/(Accessed: 3 November 2023).

Rieser-Schüssler, N., et al. (2016). Zürich. In A Horni, et al. (Eds.), *Multi-Agent transport simulation MATSim* (pp. 56–59). London: Ubiquity Press. https://doi.org/10.5334/baw.56.

Schlenther, T., et al. (2022). Simulation-based investigation of transport scenarios for Hamburg. *Procedia Computer Science, 201*(C), 587–593. https://doi.org/10.1016/j.procs.2022.03.076

Sener, I. N., Eluru, N., & Bhat, C. R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation, 36*(5), 511–539. https://doi.org/10.1007/s11116-009-9201-4

Serras, J. et al. (2016) 'London. In: Horni, A, Nagel, K and Axhausen, KW. (eds.) The multi-agent transport simulation MATSim', Ubiquity Press, pp. 447–450. doi: https://doi.org/10.5334/baw.73.

Smith, L., Beckman, R. and Baggerly, K. (1995) 'TRANSIMS: Transportation analysis and simulation system', *Los Alamos National Lab, NM (United States)*.

Spooner, F., et al. (2021). A dynamic microsimulation model for epidemics. *Social Science and Medicine, 291*(October), Article 114461. https://doi.org/10.1016/j.socscimed.2021.114461

Tiikkaja, H., & Liimatainen, H. (2021). Car access and travel behaviour among men and women in car deficient households with children. *Transportation Research Interdisciplinary Perspectives, 10*(April), Article 100367. https://doi.org/10.1016/j.trip.2021.100367

Tyrinopoulos, Y., & Antoniou, C. (2013). Factors affecting modal choice in urban mobility. *European Transport Research Review, 5*(1), 27–39. https://doi.org/10.1007/s12544-012-0088-3

Tzouras, P. G., et al. (2023). Agent-based models for simulating e-scooter sharing services: A review and a qualitative assessment. *International Journal of Transportation Science and Technology, 12*(1), 71–85. https://doi.org/10.1016/j.ijtst.2022.02.001

UK Government (2023) *Get information about Schools*. Available at: https://get-information-schools.service.gov.uk/Downloads?Skip=&SearchType=DateFilter&FilterDate.Day=11&FilterDate.Month=09&FilterDate.Year=2020 (Accessed: 14 February 2024).

UK Homes and Community Agency (2015) 'Employment density guide: 3rd Edition', (November), p. 40. Available at: https://www.kirklees.gov.uk/beta/planning-policy/pdf/examination/national-evidence/NE48_employment_density_guide_3rd_edition.pdf.

University of Leeds (2022) 'UK2GTFS'. Available at: https://itsleeds.github.io/UK2GTFS/.

Zheng, N., et al. (2012). A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice, 46*(8), 1291–1303. https://doi.org/10.1016/j.tra.2012.05.006

Zheng, N., Rerat, G., & Geroliminis, N. (2014). A time-dependent area-based pricing scheme for multimodal urban networks with user adaptation: An agent-based approach. In *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014* (pp. 471–476). https://doi.org/10.1109/ITSC.2014.6957734

Zia, K., et al. (2013). An agent-based parallel geo-simulation of urban mobility during city-scale evacuation. *Simulation, 89*(10), 1184–1214. https://doi.org/10.1177/0037549713485468

Ziemke, D., Kaddoura, I., & Nagel, K. (2019a). The MATSim open Berlin scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. *Procedia Computer Science, 151*(2018), 870–877. https://doi.org/10.1016/j.procs.2019.04.120

Ziemke, D., Metzler, S., & Nagel, K. (2019b). Bicycle traffic and its interaction with motorized traffic in an agent-based transport simulation framework. *Future Generation Computer Systems, 97*(May 2018), 30–40. https://doi.org/10.1016/j.future.2018.11.005

Zwerts Davy Janssens Geert Wets, E. (2007) 'How the Presence of Children Affects Parents' Travel Behavior', (February 2010), pp. 21–25.