Urban Data Science for Sustainable Transport Policies in Emerging Economies (Short Paper)

Nick Malleson[1]
School of Geography, University of Leeds, UK
http://www.nickmalleson.co.uk/
N.S.Malleson@leeds.ac.uk

Hang Nguyen Thi Thuy
VNU Vietnam Japan University, Hanoi, Vietnam
nguyen.t.thuyhang@gmail.com

Thanh Bui Quang
Faculty of Geography, VNU University of Science, Hanoi, Vietnam
qthanh.bui@gmail.com

Minh Kieu
Faculty of Engineering, University of Auckland, New Zealand
minh.kieu@auckland.ac.nz

Phe Hoang Huu
R&D Consultants, Hanoi City, Vietnam
hoanghuuphe@gmail.com

Alexis Comber
School of Geography, University of Leeds, UK
A.Comber@leeds.ac.uk

Abstract

In the city of Hanoi, Vietnam, as with other rapidly-developing cities, transport infrastructure is failing to keep pace with the burgeoning population. This has lead to high levels of congestion, air pollution, and a broad inequity in the accessibility of large parts of the city to residents. The emerging discipline of Urban Data Science has a valuable role in providing policy makers with robust evidence on which to base policy, but the discipline faces problems with the application of techniques that are based on assumptions that do not hold when applied to emerging economies.

This paper presents the preliminary outputs of a new programme of urban data science work that is being developed specifically for Hanoi. It leverages a spatial microsimulation approach to up-sample a bespoke travel survey and create a synthetic representation of the transport preferences of all residents in the city. These new data are used to assess the impacts that changes in the broader socio-economic context, such as increasing prosperity amongst residents, could have on rates of car ownership and hence on the problems of congestion and pollution. The results begin to highlight parts of the city where the impacts of improved economic conditions coupled with changes to wider transport policies might lead to greater use of personal cars in the future.

2012 ACM Subject Classification Computing methodologies → Modeling and simulation

Keywords and phrases Urban data science, synthetic population generation, simulated annealing, transport, modelling

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

Funding This work has received funding from the British Academy under the Urban Infrastructures of Well-Being programme [grant number UWB190190]

[1] Corresponding author

© Nick Malleson, Hang Nguyen Thi Thuy, Thanh Bui Quang, Minh Kieu, Phe Hoang Huu, Alexis Comber;
licensed under Creative Commons License CC-BY
42nd Conference on Very Important Topics (CVIT 2016).
Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:6
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

The structure of cities and transport systems are closely related and road networks play a key role in meeting the transport needs of urban areas. However, Hanoi, Vietnam, like many major cities in emerging economies, suffers serious traffic congestion and air pollution due to rapid urbanization, increases in private transport, and the informal infrastructures formed during the emergence of urban sprawl. The field of Urban Data Science (UDS) consists of "quantitative workflows for gathering, processing, and analyzing data in a spatiotemporal context that applies statistics and computer science to urban questions" [2], and could be extremely valuable as a means of better understanding the dynamics of cities such as Hanoi. Hence the application of UDS to questions about sustainable transport infrastructure could help to generate robust, effective policy to reduce the burdens of traffic congestion and pollution.

However, there are some fundamental difficulties with the conceptualisation of UDS that present challenges for the application of core UDS techniques that have emerged in the Global North to cities in emerging economies [1], such as Hanoi. For example, data that are commonly used in the application of UDS to urban mobility include those that are created from the use of smart cards and intelligent transportation systems [2], i.e. "analytics-powered, intelligent traffic management" [7]. But this does not translate to systems where the commonly-used means of transport either do not record information about journeys digitally (e.g. via cash-based ticketing systems) or are fundamentally organised in much more ad hoc manner that lack any formal means of recording journeys or even publishing timetables (such as the matatu system in Nairobi [14]). Similarly, commonly-used transport modelling methods, including both aggregate traffic assignment [11] and mircosimulation models [8], will struggle to account for the behaviour of vehicles such as motorbikes that do not follow the behaviours that would be expected from car drivers but are extremely common in many emerging economies.

This paper reports on the work as part of a wider project that aims to use UDS techniques to provide policy makers at the highest level of government with new data and computer models to support evidence-based policy to create a more efficient, equitable, and sustainable transport system that meets Hanoi's expanding population needs. In the context of limited data availability with regards to the dynamics of the transport infrastructure – particularly when compared to cities of a similar size that are characteristic of those in the Global North – we fall back on the creation of a *synthetic* population that is designed to represent all individuals in the city of Hanoi. The population is created through the use of simulated annealing as a means of up-sampling a new survey of 1,500 households (conducted specifically for the project) by combining it with data from the 2019 Vietnam population census. Initial results begin to provide an insight into the preferences of residents for different types of motor vehicle use, highlighting areas that are at particular risk of becoming more car dependent as households become more affluent, or as the nature of the transport infrastructure changes. More broadly, the project aims to explore the relevance of commonly-used UDS techniques in the context of a rapidly developing city in an emerging economy.

2 Research Context

In Hanoi, motorbikes are the preferred transportation mode: over 90% of the vehicles driven in Hanoi are motorbikes and there are on average 2.5 motorbikes per person [13]. Since the introduction of the Doi-Moi policy [5] in the 1980s, the number of motorbikes has increased 10-fold and there are now more than 4 million motorbikes in Hanoi alone [4, 5].

Simultaneously, public transport infrastructure has developed slowly. As public transport does not meet the city's requirements, increases in personal traffic are inevitable, resulting in acute welfare problems, especially air quality. Pollution is chronic, with PM2.5 and ozone concentration regularly exceeding safe levels. In response, the City has developed fast buses, a skytrain system, tightened the standards for new vehicles and imposed petrol quality controls. Nevertheless, the Real-time Air Quality Index, measured by the U.S. embassy, recently found pollution in Hanoi at levels sufficient for people with heart and respiratory problems to stay indoors. Some officials proposed a radical plan to ban motorbikes in large parts of the city, but this was met with strong public opposition. Surveys linked to models can answer questions about how, where, and when motorbikes should be banned (if at all), about the impacts on local communities, whether public transport can cope, and whether there are better alternatives. Importantly, they can also provide information about the factors that are encouraging or prohibiting peoples' vehicle ownership preferences; this paper pays particular attention to the factors that might lead to greater car ownership amongst residents, especially if motorbikes are no longer an option for travel.

3 Data & Methods

The aim of this work is to up-scale a new survey of transport behaviours and preferences conducted by the project team in Hanoi, Vietnam. We do this through linkage to the most recent Vietnamese population census using simulated annealing to create a synthetic population of all individuals in Hanoi that contains core census variables as well as variables in the new survey. The following sections outline the methods used; drawbacks and caveats are discussed in Section 5.

3.1 Transport Survey

A new survey is currently being conducted in Hanoi that asks people for basic demographic information as well as details about their travel behaviour (e.g. common journeys) and preferences (e.g. aspirations for ownership of different types of vehicle). The COVID-19 pandemic has interrupted the survey on multiple occasions, but at the time of writing 1,500 households, out of a target of 10,000, have responded. The key variables that are relevant for this paper include, among others:

Demographics Sex, age group, occupation.

Vehicle ownership Types of vehicles owned by the household.

Travel behaviour Details about regular journeys made: start/end locations, frequency, mode of transport, reasons for choice of mode.

Vehicle ownership aspirations Whether the household would like to own additional vehicles and what factors prevent them from ownership.

3.2 Vietnamese Census of Population

Vietnam's most recent population and housing census was conducted in 2019 [3]. It found that the population of Vietnam had grown to 96M people. Hanoi, the case study area, is the second largest city after Ho Chi Minh City with a population of 8M people; increasing by 1.5M between 2010 and 2019[2]. At the time of writing, the project has access to counts

[2] https://vietnam.opendevelopmentmekong.net/topics/vietnams-population-and-census/

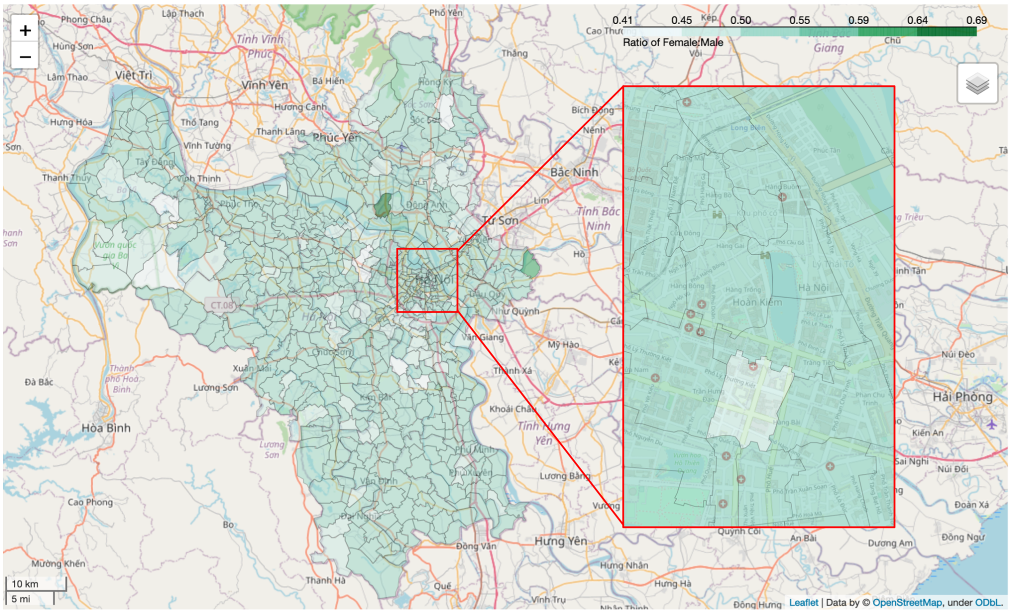

Figure 1 The ratio of females to males in Hanoi from the 2019 Vietnam census.

of people per district level broken down by sex and age group (as separate variables, not cross-tabulations). As an example to demonstrate the level of geography available in the census, Figure 1 illustrates the ratio of males to females in each district in the study area.

3.3 Synthetic Population Generation

The survey (Section 3.1) aims to include responses from 10,000 households which will be one of the largest household travel surveys conducted in Hanoi. However, the city is so populous that, naturally, the geography of the respondents is very sparse (there will be very few respondents per district). Therefore to up-scale the survey in order to make inferences about transport behaviour across a much wider spatial area, we use population synthesis to combine the survey results with the population census to create a synthetic population. The new population aims to be representative, both in terms of demographics and transport behaviours, of the true underlying population.

Synthetic population generation (also sometimes referred to in geography as 'microsimulation') was originally inspired by the work of Orcutt [12]. It aims to construct a data set of individual units (people in this case) over a large area by cloning individuals from a survey (the travel survey) such that the aggregates match some known aggregate data (the population census). The resulting synthetic population contains attributes from both the aggregate and survey data [6]. The assignment of individuals to areas is conducted using an iterative optimisation algorithm simplified from simulated annealing [9] as implemented in the Flexible Modelling Framework software[3] [6].

[3] https://github.com/MassAtLeeds/FMF

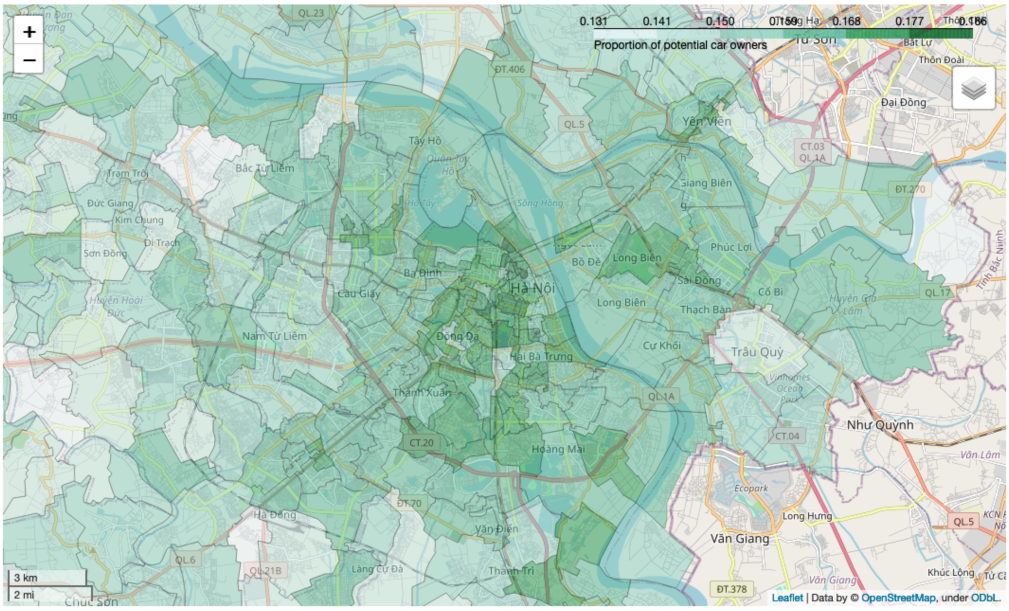

Figure 2 The proportion of people in the synthetic population who would like to own a car but are prohibitted from purchasing one due to the cost.

4 Results

The survey is extremely rich, so there are a wide variety of variables that are attached to the synthetic population and could be analysed. Here we examine one factor; the propensity for individuals, who do not currently own a car, to purchase one. Individuals are extracted from the population who meet three criteria: (i) they do not currently own a car; (ii) they would like to own a car in the future; (iii) cost is the main factor that is prohibiting them from owning one. Figure 2 then maps the proportion of synthetic individuals who meet the criteria above. Although these are preliminary results and it is too soon to draw firm conclusions, it is interesting that towards the centre of the city there are larger proportions of synthetic individuals who would like to purchase a car if they could afford to. This issue is important for policy because, as the economy in Vietnam expands and more people become able to afford a motor car, transport policies will need to encourage alternative means of transport to prevent an unsustainable rise in car use.

5 Discussion & Conclusions

The field of Urban Data Science (UDS) has shown promise as a means of better understanding the dynamics of cities in order to make them better places to live. However, assumptions about data characteristics and availability do not necessarily translate well to the urban context in developing economies. In Hanoi, for example, there are very limited digital data that describe the use of the transport network. Therefore this paper leverages a synthetic population generation framework to up-scale a new transport survey, allowing inference about transport behaviours over a much wider spatial area than would be possible otherwise. Preliminary results suggest that the distribution of residents who have the *propensity* to own a motor car (i.e. they would own one if they could) varies considerably across the city. This has the potential to inform transport policy, providing robust data to support sustainable transport policy.

This is preliminary work and there are many caveats that need to be resolved. To begin with, the survey needs to be distributed to a wider population in Hanoi. Secondly, there is a discrepancy between *households* and *individuals*. The survey collects information about households, but currently the synthetic population generation algorithm creates synthetic individuals, not households. Future work aims to take an additional step that will allow the synthetic individuals to be grouped into households, following [10]. Thirdly, the currently available census data contain only counts of people by age group. Hence age group is the only constraint used in the creation of the synthetic population, which means that the work assumes that all people in a particular age group will have similar behaviours and preferences with respect to transport use. This is obviously a very weak assumption. To make the analysis more robust, future work will make use of census data that contain a much richer set of cross-tabulated variables, as well as additional variables that are present in both the survey and the census that can be held back from the synthetic population generation process and used as a means of validation.

References

1 Michael Batty. Defining Urban Science. In Wenzhong Shi, Michael F. Goodchild, Michael Batty, Mei-Po Kwan, and Anshu Zhang, editors, *Urban Informatics*, pages 15–28. Springer Singapore, Singapore, 2021. doi:10.1007/978-981-15-8983-6_3.

2 Geoff Boeing, Michael Batty, Shan Jiang, and Lisa Schweitzer. Urban Analytics: History, Trajectory, and Critique. *SSRN Electronic Journal*, 2021. doi:10.2139/ssrn.3846508.

3 General Statistics Office. *Completed Results of the 2019 Viet Nam Population and Housing Census*. Statistical Publishing House, Vietnam, 2020.

4 Arve Hansen. Hanoi on wheels: emerging automobility in the land of the motorbike. *Mobilities*, pages 1–18, 2016. doi:10.1080/17450101.2016.1156425.

5 Arve Hansen. Transport in transition: Doi moi and the consumption of cars and motorbikes in Hanoi. *Journal of Consumer Culture*, 17(2):378–396, 2017. doi:10.1177/1469540515602301.

6 Kirk Harland. *Microsimulation Model User Guide*. Number 06/13 in National Centre for Research Methods Working Paper. University of Leeds, Leeds, UK, 2013.

7 Jens Kandt and Michael Batty. Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities*, 109:102992, 2021. doi:10.1016/j.cities.2020.102992.

8 Le-Minh Kieu, Dong Ngoduy, Nicolas Malleson, and Edward Chung. A stochastic schedule-following simulation model of bus routes. *Transportmetrica B: Transport Dynamics*, 7(1):1588–1610, 2019. doi:10.1080/21680566.2019.1670118.

9 S Kirkpatrick, C D Gelatt, and M P Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. doi:10.1126/science.220.4598.671.

10 Nik Lomax and Andrew P. Smith. An introduction to microsimulation for demography. *Australian Population Studies*, 1(1):73–85, 2017.

11 A. D. May. *TRAFFIC FLOW FUNDAMENTALS*. 1990.

12 Guy H. Orcutt. A New Type of Socio-Economic System. *The Review of Economics and Statistics*, 39(2):116, 1957. doi:10.2307/1928528.

13 Tan Hong Van, Jan-Dirk Schmoecker, and Satoshi Fujii. Upgrading from motorbikes to cars: Simulation of current and future traffic conditions in Ho Chi Minh City. *Proceedings of the Eastern Asia Society for Transportation Studies*, 2009:335–335, 2009. doi:10.11175/eastpro.2009.0.335.0.

14 Sarah Williams, Adam White, Peter Waiganjo, Daniel Orwa, and Jacqueline Klopp. The digital matatu project: Using cell phones to create an open source data for Nairobi's semi-formal bus system. *Journal of Transport Geography*, 49:39–51, 2015. doi:10.1016/j.jtrangeo.2015.10.005.