**Conference or Workshop Item:**
Prédhumeau, M and Manley, E (Accepted: 2023) Building an Environment for Spatial Modelling of Canadian Cities Using Open Data: A Replicable Workflow Applied to Winnipeg. In: The 18th International Conference on Computational Urban Planning and Urban Management (CUPUM), 20-22 Jun 2023, Montreal, Canada. (Unpublished)

This is an author produced version of a conference paper originally presented at The 18th International Conference on Computational Urban Planning and Urban Management (CUPUM) in Montreal, Canada, 20-22 June 2023.

# Building an Environment for Spatial Modelling of Canadian Cities Using Open Data: A Replicable Workflow Applied to Winnipeg

Manon Prédhumeau[1] and Ed Manley[1]

University of Leeds, School of Geography, Leeds, LS2 9JT, UK,
m.predhumeau@leeds.ac.uk

**Abstract.** Microscopic approaches, such as agent-based modelling, are increasingly used by urban planners to model cities. The development of accurate microscopic models is facilitated by the increasing availability of real-world open data, essential for agent-based models to be widely used and replicated. Whereas open data is gaining popularity in Canada, only few agent-based models have been built for urban planning using open data. This paper presents a workflow for building an environment for agent-based modelling of Canadian cities based exclusively on open data and tools. Combining data from Census, OpenStreetMap, and local portals, we propose a synthetic population and an environment model that form a basis for agent-based modelling of any Canadian city. The approach is implemented and validated on Winnipeg. Results indicate that the model offers a valuable basis for urban analysis. The proposed workflow has the potential to support human-centred urban planning and to facilitate replicable modelling in Canada.

**Keywords:** synthetic population, microsimulation, public data, human-centred, urban planning

## 1 Introduction

Cities are complex systems which, beyond being defined by their land use or transport systems, are essentially made up of human beings. In this perspective, microscopic approaches like agent-based modelling are increasingly used to represent cities and support evidence-based urban planning [1]. An agent-based model (ABM) consists in heterogeneous individual agents who act and interact autonomously in a shared environment. ABMs in urban planning consider geographically explicit individual entities, to study interactions between the city inhabitants and the aggregate phenomena that emerge from these interactions.

A limitation of ABMs is that they require accurate and high granularity data in order to provide a reliable and useful model of the population and city. The growing availability of open data, through national and local portals or crowd-sourcing [2], helps to overcome this limitation. Open data encourages the development of robust ABMs, that are grounded in real-world locations and can be widely used and replicated. ABMs using open data have been successfully designed to support urban planning with simulating infectious disease outbreaks [3], with optimising police patrols for alcohol-related violent crime [4] or with transport planning in Paris [5], Berlin [6], or the state of California [7].

The Open Data Inventory, which measures the coverage and openness of countries statistical offerings ranked Canada 15th out of 187 in 2020 [8]. However, the data coverage and openness is pretty heterogeneous in Canada: while statistics are very good regarding population and economic, data coverage and openness is less good on facilities and built environment.

Despite the availability of open data in Canada, only few ABMs have been built for urban planning using open data. Perez et al. [9] developed a geospatial ABM of the island of Montreal to simulate the decision-making process of location of new household for incoming immigrant populations. The model was built using data from the Census, from the National Household Survey as well as land use and transportation network data from local portals. The model has been extended to the City of Toronto, Vancouver, and the City of Calgary by Anderson et al. [10]. However, these ABMs only considers the city's immigrant population, and the limited agents attributes (immigration category, two age ranges, number of children, income, education level and language) makes the models unsuitable for other use cases. Malik and Abdalla [11] proposed a simple model of settlement pattern of students in Waterloo region. The model consists of road and public transport networks and grocery stores, as well as agents representing students but without any socio-demographic attributes.

A detailed ABM has been proposed by Miller and Roorda [12] with the TASHA (Toronto Area Scheduling Model for Household Agents) model, designed to study individual activity schedules and travel patterns for the Greater Toronto area. TASHA includes a synthetic population sampled from the Transportation Tomorrow Survey data. However, this survey is conducted only in the Greater Golden Horseshoe Area and the approach is therefore not applicable at the national level. A synthetic population has been developed by Fatmi and Habib [13] for Halifax to simulate individuals' decisions along their life-course. The authors used data from Census and local open databases to synthesise a population with socio-demographic attributes, vehicle ownership, residential location, transportation mode transition and life-stage transition. However, the activities point locations and public transport services were retrieved from proprietary data. Finally, a synthetic population has been proposed for the Atlantic region for the year 2006 by Hafezi and Habib [14, 15]. Individuals and households were synthesised using the Canadian Census, with the attributes: gender, age, ethnicity, immigrant status, citizenship, household size, tenure, dwelling type, and household income. However, geolocation is missing and there is no integration with an environment model.

This paper proposes a workflow combining various open data sources to build a basis for spatial agent-based modelling of any Canadian city. The workflow uses data from Census, population projections, OpenStreetMap, Microsoft Building Footprints and local portals, to generate a synthetic population of individuals and households, and a synthetic environment including the road network, public transport services, buildings and facilities. The approach is implemented and validated on Winnipeg City and its potential for urban analysis and planning is discussed. The objective is to show that by combining several sources of open data available nationally in Canada into a spatial micro model, urban planners may benefit from a valid and unified modelling base.

The paper is organised as follows. Section 2 presents the open data sources and their integration into a model through a unified workflow. Section 3 describes the approach implementation and validation for the city of Winnipeg. Section 4 discusses possible applications for urban planning.

## 2 Proposed Workflow

### 2.1 Open Data Sources

The proposed workflow integrates several open data sources available for all Canada, that include:

1. OpenStreetMap for the road network and facilities locations,
2. Microsoft Building Footprints for a complete representation of buildings,
3. Census Profile and Public Use Microdata Files,
4. Population projections,
5. National surveys and provincial open data portals,
6. Local open data portals for public transport services and building addresses.

**OpenStreetMap** (OSM) is a volunteered geographic information mapping project that aims to develop a free and accessible map of the world [16]. OSM is widely used as a source for road networks and points of interest (POIs). Zhang and Malczewski [17] found that OSM is a reliable source for road networks in Canada (comparable to Digital Mapping Technologies Inc., a proprietary data source), although urban areas received more contributions than rural areas. The authors found that 77% of OSM Canadian roads were within a 5 m distance error, and 90% were within 30 m distance error. In addition to road network, OSM contains polygons and POIs identifying buildings and commercial businesses as well as churches, schools, post offices, pubs, and amenities like parks, tennis courts, food stalls, ATMs, etc. Unfortunately, the buildings coverage is less complete that the road network [18]. Moreover, no study has been conducted to assess the points of interest coverage in Canada.

**Microsoft Building Footprints** have been published in 2019 by Microsoft, in collaboration with Statistics Canada [19]. Deep Neural Networks have been used to detect building footprints from Bing imagery. The dataset consists in 11,842,186 building footprints across all Canadian Provinces and Territories, open for download in GeoJSON format, and usable for research and analysis.

**Census Profile and Public Use Microdata Files** are published every 5 years by Statistics Canada. Public Use Microdata Files (PUMF) give access to non-aggregated data on individuals and households in the Canadian population. The Individual PUMF [20] provides access to anonymised records from the Census questionnaire for a 2.7% sample of the Canadian population. Each record presents 123 individual variables and is geolocalised at the province level to preserve confidentiality. The Hierarchical PUMF [21] contains non-aggregated data for a 1% sample of the Canadian households. The file contains individuals' records related into households, with 95 variables, and can be used to study individuals in relation to their households.

**Population Projections** for provinces and territories [22] are published every 5 years by Statistics Canada. It consists in population projections by age and sex at the provinces and territories level, based on assumptions on the population growth. The projections give a perspective of the future Canadian population demography.

**National Surveys and Provincial Data Portals** provide reliable open data sources regarding Canada population. National surveys are regularly conducted by Statistics Canada on education, health, housing, transportation, etc. Another open data source for national statistics regarding transportation is the Canadian Motor Vehicle Traffic Collision Statistics reported by Transport Canada every year. These national statistics can be refined with provincial statistics if available. Apart from Nunavut, all provinces and territories provide access to some local statistics through Open Government portals [23]. Moreover, some provincial public insurances like Manitoba Public Insurance or Insurance Corporation of British Columbia report statistics regarding transportation.

**Local Open Data Portals** are available for most of the urban areas in Canada [23]. These local portals provide access to very heterogeneous datasets and generally include buildings addresses, location of public facilities, parks and parking areas. When public transport is available, government local portals or public transport agencies usually provide geolocation of stops and service schedules via the standardised General Transit Feed Specification (GTFS) format. A list of open GTFS data from across the world is also available through the Mobility Database Catalogs [24].

### 2.2 Workflow

The proposed workflow (Figure 1) consists in generating two models which constitute the basis for an ABM of a given urban area: a synthetic environment and a synthetic population.

**Environment Modelling** The synthetic environment for the considered urban area is modelled in several steps. First, OSM data is downloaded for the considered province using *Geofabrik*, which provides OSM data extracts for parts of the world, including Canadian provinces [25]. The OSM relation identifier for the desired urban area is found using *OSM main website* [16], and *osmium* command line tool is used to extract OSM data within the desired urban area boundaries. This OSM extract is then used as a basis to model the road network, the public transport services, and the buildings and facilities.

The road network, and public transport network and service are modelled using *PT2MATSim* [26]. PT2MATSim provides an OSM converter to convert

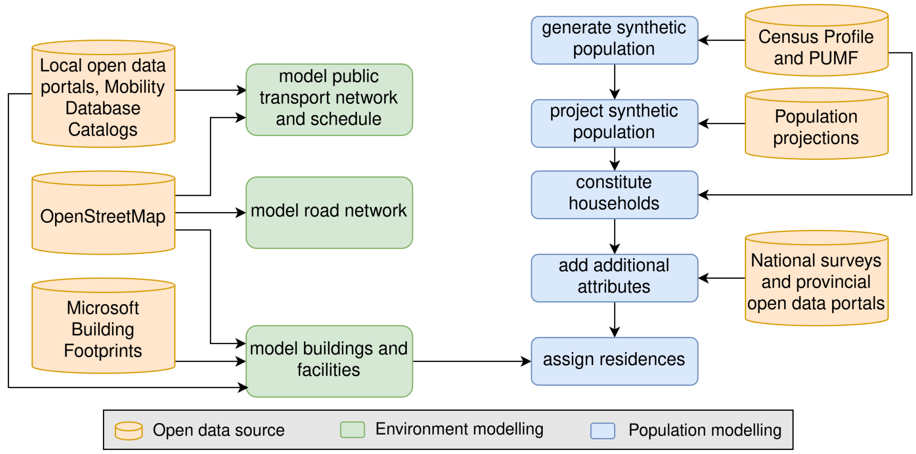



Fig. 1: Proposed workflow to build an ABM basis for any Canadian city, combining open data sources.

OSM roads into a multimodal network composed of links and nodes representing the road network, buses lanes and tram lanes. The public transport GTFS feed is retrieved from the considered local open data portal (or from the Mobility Database Catalogs). The standardised GTFS file contains the public transport schedule, the name and direction of the routes, the stopping points and the timetable of the public transport service at the stopping points. PT2MATSim provides a GTFS converter to extract an unmapped transit schedule for a given date from GTFS data. Finally, the transit schedule and stop sequences are mapped to the multimodal network with PT2MATSim so that the public transport services are associated to routes on the network.

While OSM has good coverage considering roads, data is much more heterogeneous for buildings and facilities. We therefore propose to merge three data sources: building polygons from Microsoft Building Footprints [19], addresses from local portals [23] and facilities from OSM points of interest [16].

As the geolocated building footprints have been obtained by automatic image detection, some polygons do not correspond to buildings, but to garden sheds or crop fields. To extract only the building polygons, the polygons with an area between 30 m$^2$ and 350,000 m$^2$ are filtered. A spatial join is then applied between the filtered polygons and the addresses obtained from the local portal in order to keep only the buildings with addresses.

The OSM building tags and points of interest are then retrieved and a spatial join is applied with the building polygons in order to add OSM tags to the nearest building. Then, for each building, the building type is deduced from the associated OSM tags, among industrial, civic amenity, sport facility, sustenance, education, transport, financial, healthcare, entertainment, shop, office or other use. If the building is explicitly referenced as residential or has no tag associated, it is considered as residential. A building may be related to several types, e.g. a shop on the ground floor and residences on the upper floors.

Finally, residential buildings are converted into residences by dividing the building area and levels by the average living area. If the number of levels is specified, then the average living area used is the average Canadian condominium size, i.e. 80 $m^2$. If no levels are specified, the average house area is used to derive the number of residences in the building. Assuming that some houses have levels, an approximated area of 125 $m^2$ is used for a house residence. The living area values are derived from the Canadian Housing Statistics Program [27] and can be refined if local values are available.

**Population Modelling** Using 2016 Census data [28], a synthetic population is generated at the DA level for the province of the considered urban area. A hybrid approach called Quasirandom Integer Sampling of Iterative Proportional Fitting is applied [29]. This method uses Individual PUMF data [20] as a seed and assigns each individual in the sample a weight such that the weighted population fits distributions defined by the Census Profile [28] for attributes aggregated at DA level. To obtain an integer population from the fractional weights, individuals are sampled according to a discrete distribution without replacement derived from the aggregated data. Each individual in the synthetic population has a set of socio-demographic attributes: sex, age range, education level, labour force status, income, household size, primary maintainer of the household or not, and is located in a DA.

The 2016 synthetic population is then projected to represent the actual 2023 population. Population projections by age and sex for the considered province are used [22]. For each age group and sex, the province synthetic population is resampled (i.e. individuals are duplicated or deleted) to cover the population count difference between 2016 and 2023.

The 2023 synthetic individuals are then assigned together to constitute households. One household is created for each individual identified as primary household maintainer. Then, each household is completed with non-maintainer individuals within the same DA to match the household size. The assignment is done by randomly drawing individuals according to the distribution of non-maintainer individuals' age group and sex by primary maintainer's age group and sex inferred from the Hierarchical PUMF [21].

The 2023 synthetic population dataset used is public and freely available for all Canada, at the DA level at: https://doi.org/10.5281/zenodo.7085408 [30]. Local synthetic populations can be extracted from the dataset for specific urban areas by filtering the considered DAs. Each synthetic individual has the attributes listed in Table 1.

Individuals' attributes are then complemented with additional characteristics using national and provincial statistics. Health-related attributes can be added for instance by linking the synthetic population to the Canadian Community Health Survey conducted annually. Survey outputs like the "Perceived health, by age group and sex" table are published at the province level [31] and can

Table 1: Individual attributes in the synthetic population

| Variable | Definition | Categories |
|---|---|---|
| index | Individual identifier | Integer unique for the province |
| HID | Household identifier | Integer unique for the province |
| sex | Sex | 0: female / 1: male |
| prihm | Is primary household maintainer | 0: no / 1: yes |
| agegrp | Five-year age group | 0: 0 to 4 years<br>1: 5 to 9 years<br>...<br>16: 80 to 84 years<br>17: 85 years and over |
| age | Age in full years | Integer $\in$ [0;120] |
| area | Dissemination area code | 8-digit code |
| hdgree | Highest qualification | 0: no certificate, diploma or degree<br>1: secondary school or equivalent<br>2: postsecondary degree |
| lfact | Labour force status | 0: employed<br>1: unemployed<br>2: not in labour force |
| hhsize | Number of individuals in the household | 0: 1 person<br>1: 2 persons<br>...<br>4: 5 persons or more |
| totinc | Total income before taxes and deductions | 0: $<$ 20,000 \$<br>1: 20,000 \$ to 59,999 \$<br>2: 60,000 \$ to 99,999 \$<br>3: $\geq$ 100,000 \$ |
| hhtype | Type of relation between household members | 0: Couples without children<br>1: Couples with children<br>2: One-parent-family<br>3: One-person<br>4: Other kind of household |

be used to probabilistically assign a health status attribute to synthetic individuals. Another example is the driving license ownership attribute, that can be linked to synthetic individuals using the Canadian Motor Vehicle Traffic Collision Statistics [32]. In addition to road safety statistics, this program annually reports the national number of licensed drivers by age group and sex. According to the proportion of licensed drivers by age group and sex, synthetic individuals can be probabilistically assigned a new attribute indicating if their own a driving license or not. National surveys have the advantage of providing statistics that are available for all Canada. However, if similar statistics are available at a finer geographical level, they may allow for a more accurate attribution of attributes values.

Finally, households are randomly assigned a residential building within their DA. For each DA, the building polygons that fall within the DA boundaries are extracted. If enough residences exist for all households in the DA, then one residence is randomly assigned to each household. If there are not enough residences for all households, then all residences are assigned, and the remaining households are placed in random already-assigned residences. Finally, in case there is no residential building identified in the DA, households are assigned to non-residential buildings.

## 3 Implementation and Validation on Winnipeg

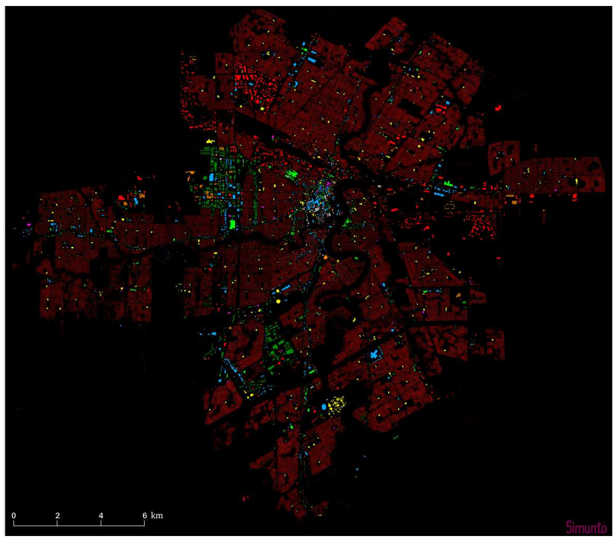
As of 2021, Winnipeg was the sixth-largest city in Canada with a population of 749 607, and the largest city of the province of Manitoba. The proposed workflow has been implemented in python to generate an ABM basis for Winnipeg. The model has then been validated by cross-checking with various open data sources.

### 3.1 Implementation

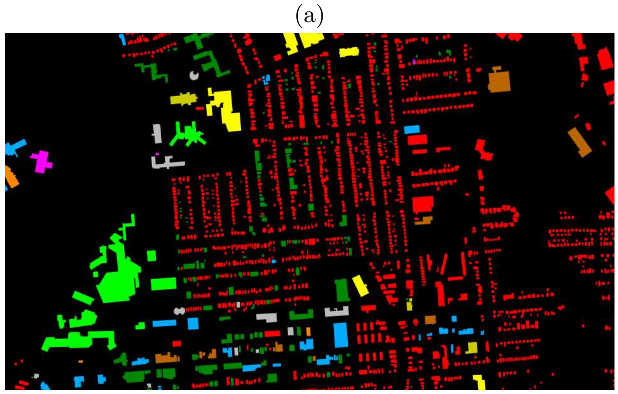
Winnipeg data has been extracted from Manitoba OSM data [25] using the city's OSM identifier, i.e. 1790696. The public transport in Winnipeg is provided by Winnipeg Transit, with 640 buses serving over 5 000 bus stops. Combining OSM data and GTFS data from Winnipeg Transit's Open Data Web Service [33], a model of the road network and public transport service has been generated for each day of the week. The public transport model consists of 5 345 stops served by 84 bus routes. The road network and public transport routes and stops on Tuesday appear in Figure 2.


Fig. 2: Model of Winnipeg road network (in grey) and public transport Tuesday service (in orange). Visualisation with Simunto.

Winnipeg building footprints were extracted from the 632 982 Manitoba building footprints in the Microsoft Building Footprints dataset [19]. After being filtered by their footprint area, the polygons have been joined with the City of Winnipeg Property Addresses and Coordinates provided though Winnipeg open data portal [34], in order to filter out buildings without addresses. Finally, OSM tags have been spatially joined to the building polygons, and buildings have been assigned one or more functional types, as shown in Figure 3.



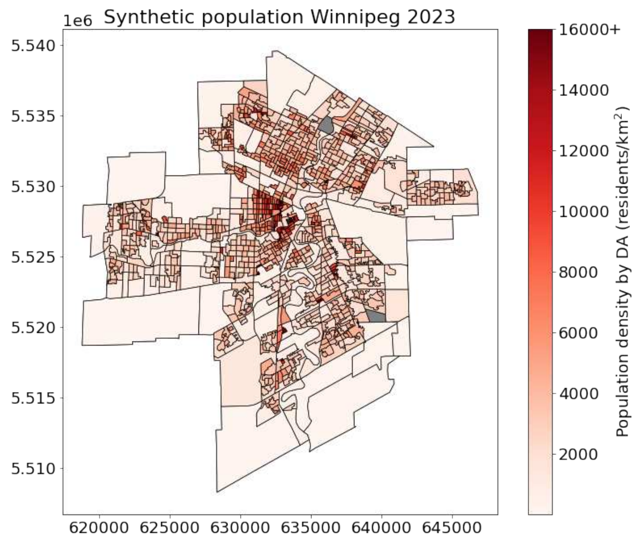

(a)



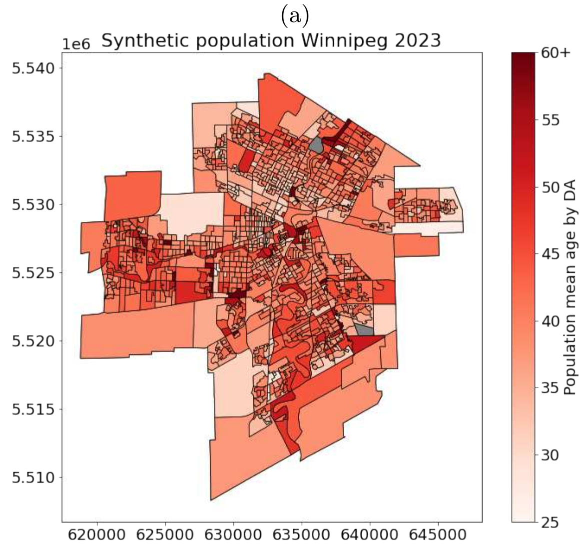
(b)

Fig. 3: Model of Winnipeg buildings and facilities for (a) all Winnipeg and (b) Central St. Boniface district. Residential buildings in red, shops and sustenance in blue, education and civic amenities in yellow, sport and entertainment facilities in pink, healthcare buildings in light green, industrial and transport in orange, other land use in dark green and mixed land use in grey. Visualised with Simunto.

A 2023 Winnipeg synthetic population has been extracted from the 2023 Canadian synthetic population [30] by filtering the DAs related to Winnipeg CSD (CSD code 4611040). The synthetic population is composed of 774 883 individuals in 316 649 households. The synthetic population density by DA and mean age by DA are shown in Figure 4.




(a)




(b)

Fig. 4: Winnipeg 2023 synthetic population (a) density by DA and (b) mean age by DA. Grey DAs are not populated.

### 3.2 Validation

The road network model has been visually compared to data from Winnipeg open data catalog referencing road network [35]. The road network from OSM, in light grey, is superposed over the one from the open data portal, in red, on Figure 5 to highlight their differences. Almost all roads are present in both datasets. The missing road sections in the OSM model, shown in red, were under construction at the time the data were extracted from OSM.

The public transport model includes the 84 fixed bus lines referenced by Winnipeg Transit [36]. The model contains 5 345 bus stops, including the 5 160 stops from the GTFS data with some duplicated stops when serving routes from both directions.


Fig. 5: Comparison between the road network model from OSM (grey) and the one from Winnipeg open data portal (red). Visualised with Simunto.

253 982 building footprints were extracted from Microsoft Building Footprints data for Winnipeg, of which 205 768 buildings had an address and a realistic size. As a comparison, the OSM export only refers to 85 755 buildings for Winnipeg, which confirms the lack of completeness of OSM regarding buildings. The number of residences estimated in the model is close to the total private dwellings and private dwellings occupied by usual residents reported by the 2021 Census for Winnipeg (Table 2). The assignment of households to buildings within a DA could be validated if data on building populations were available, which is not the case. The number of facilities by type in the model is shown in Table 2. A direct comparison with an external source is not possible

due to the lack of available data on facilities in Winnipeg. For shops, sustenance and entertainment facilities, counts from the Longitudinal Employment Analysis Program for Winnipeg Census Metropolitan Area (CMA) in September 2022 are reported in order to provide a comparative order of magnitude.

Table 2: Facilities counts by type in the model and comparative counts from Statistics Canada.

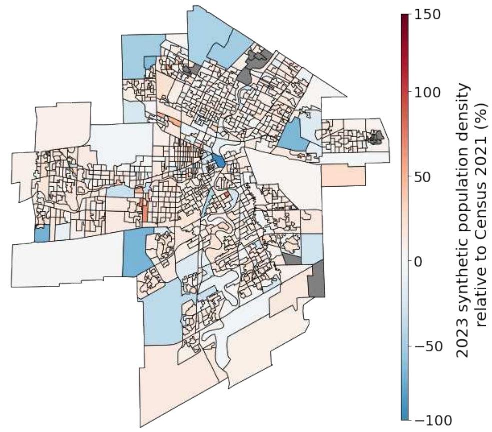
| Facility type | Model | Comparative count *(source)* |
|---|---|---|
| Residence | 303 908 | 315 465 private dwellings (300 431 occupied) *(Census 2021)* |
| Non-residential | 16 822 | 18 883 *(StatCan[1])* |
| Education | 1 282 | - |
| Healthcare | 246 | - |
| Shop | 4 310 | 1,718 retail trade businesses *(StatCan[1])* |
| Sustenance | 831 | 1,238 accommodation and food services *(StatCan[1])* |
| Entertainment | 149 | 320 art, entertainment and recreation busin. *(StatCan[1])* |
| Sport | 919 | - |
| Civic amenity | 852 | - |
| Industrial | 891 | - |
| Financial | 131 | - |
| Transportation | 248 | - |
| Office | 139 | - |
| Other amenity | 75 | - |
| Other land use | 6 451 | - |

[1] Statistics Canada: Longitudinal Employment Analysis Program - Experimental estimates for business openings and closures for Canada, provinces and territories, census metropolitan areas, seasonally adjusted - Winnipeg CMA - September 2022

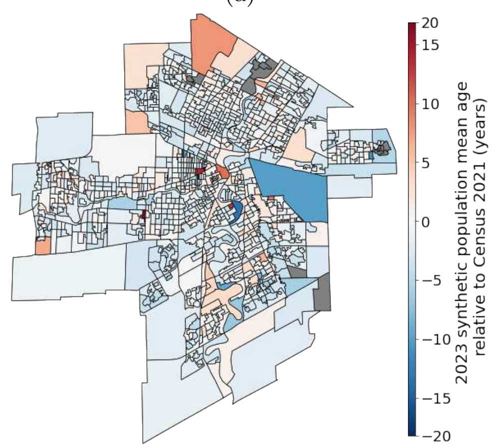
The 2023 synthetic population density and mean age by DA are compared to those reported by 2021 Census in Figure 6. For 90% of the DAs, the absolute difference in density is less than 20%. The mean age by DA ranges from 20 to 71 years in the model, and from 25 to 71 years in 2021 Census. Half of DAs have a mean age differing from less than 2 years compared to 2021 Census, and in 90% of DAs the difference is less than 4 years. Grey DAs are not populated or not directly comparable due to boundaries update since last Census.

## 4 Discussion and Conclusions

Results presented for Winnipeg show that the proposed workflow can produce an accurate model of the population and urban environment. By using a realistic population model, this workflow has the potential to support human-centred urban analysis and planning. The microscopic approach supports the analysis of urban characteristics at various aggregation levels, not only spatially, but also through socio-demographic categories.



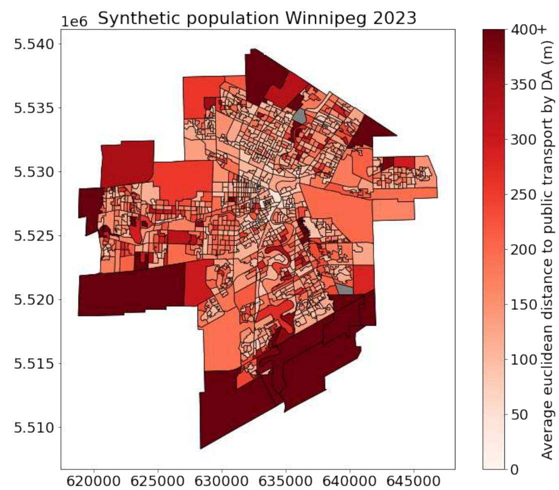

(a)



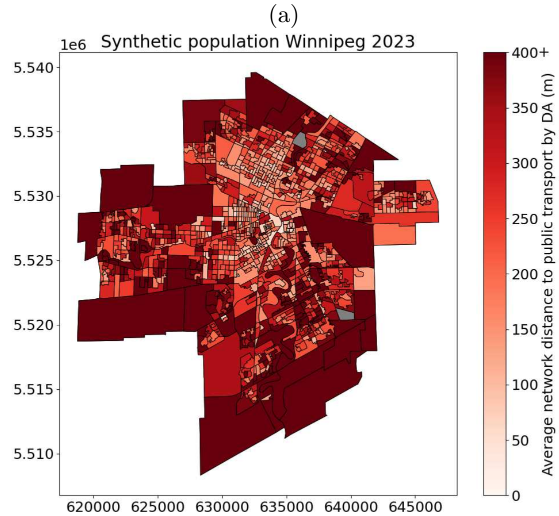

(b)

Fig. 6: Winnipeg 2023 synthetic population (a) density by DA and (b) mean age by DA, relative to the 2021 Census.

Moreover, to model the city at a precise spatial level, it is important to model individuals at the building level instead of assuming that everyone lives at the DA centroid or is evenly distributed in the DA. Indeed, the built environment is sometimes very heterogeneous within a DA and studying spatial inequalities in access to transport or facilities requires geolocating individuals at a high spatial granularity. As an example, the maps presented in Figure 7 report measures of accessibility to public transport in Winnipeg based on the proposed model. The first map shows that from most of Winnipeg's less densely populated DAs the nearest public transport stop is on average over 400 m away as the crow flies. More unexpectedly, this can also be observed for some denser DAs. Overall, the mean distance to the nearest public transport stop is over 400 m away as the crow flies in 2% of Winnipeg DAs. The second map highlights the importance of considering the network distance instead of the as-the crow flies distance. When considering the shortest path network distance to reach the nearest public transport stop, the mean walking distance is over 400 m in 14% of DAs. The accessibility measured is consistent with observations reported by the International Institute for Sustainable Development in 2018 [37].




(a)




(b)

Fig. 7: Average distance (in meters) to public transport by DA from 2023 Winnipeg model: (a) euclidean distance and (b) network distance. Grey DAs are not populated.

In addition, the accessibility to public transport is not the same for all socio-demographic categories. For seniors (aged 65 or older), who walk slower or whose physical mobility is decreasing, the walk to the bus stop can be a barrier to using public transport [38]. The proposed model can be used to measure seniors' accessibility to public transport. Figure 8, for instance, shows 19 DAs where seniors represent more than 25% of the population and where the walking distance to the nearest bus stop is over 400 metres for 50% of the residents.

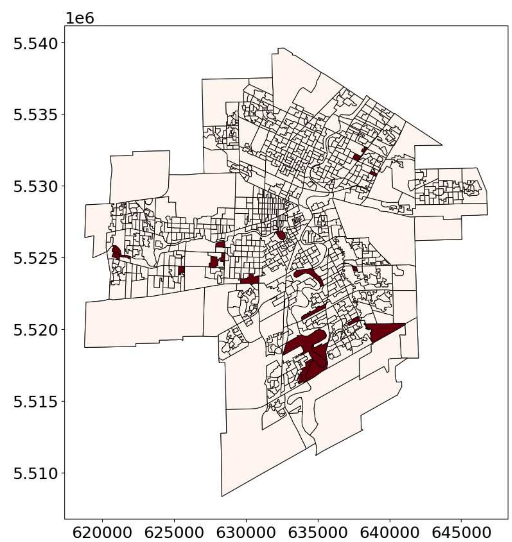



Fig. 8: The DAs in dark red have more than 25% of elderly people and for 50% of the inhabitants the walking distance to the nearest bus stop is over 400 metres.

The proposed model provides a basis for an ABM to support urban planning in Canada. Individuals can be turned into agents by complementing them with individual behaviours such as short- or long-term mobility, evacuation in case of crisis, or consumption of goods or services. Accurate individual behaviours can be added to the model according to the individuals' characteristics (place of residence, age, sex, etc.). The model can be used as a basis to study longer-term changes, i.e. it can be combined with a resettlement model to explore long-term relocation, or with a microsimulation framework to simulate population ageing.

Depending on their case study, users can assign workplaces (and schools) to the synthetic individuals according to their age, residential location, labour force status and additional attributes related to the specific case study. In combination with an activity-based model describing the daily schedules of agents [12], the model can be used to simulate urban mobility with open-source platforms such as MATSim. The ABM could then be used to explore what-if scenarios and gain insights into the effects of public policies before they are implemented. The creation of a new bus route, the construction of new facilities or the application of a low-carbon emission transport policy are examples where the model has potential for urban planning.

As the workflow is based exclusively on open data available in all Canada, the modelling can be replicated for other Canadian cities. The ABMs developed from the model can therefore be easily transferred to other regions and compared.

The model presents several limitations. It is based on 2016 Census data, which was the most recent data available at the time the model was built. A more accurate model could be developed by applying the same approach to 2021 Census data when it becomes available. The model validation was performed with locally available open data, but further evaluation could be performed with more local statistics available.

We intend to extend the Winnipeg model with individuals' daily activity patterns to produce an activity-based model. The model will then be used to explore future transportation scenarios with MATSim, such as the integration of autonomous on-demand vehicles into public transport systems.

## 5 Acknowledgements

This research has been conducted as part of the RAIM project (Responsible Automation for Inclusive Mobility), funded by the ESRC-Canada AI initiative (ES/T012587/1). This work was undertaken on ARC4, part of the High Performance Computing facilities at the University of Leeds, UK.

## References

1. Crooks, A., Heppenstall, A., Malleson, N. & Manley, E. *Agent-Based Modeling and the City: A Gallery of Applications*, 885–910 (Springer, Singapore, 2021).
2. Crooks, A. *et al.* Crowdsourcing urban form and function. *International Journal of Geographical Information Science* **29**, 720–741 (2015).
3. Hunter, E., Mac Namee, B. & Kelleher, J. An open-data-driven agent-based model to simulate infectious disease outbreaks. *PLOS ONE* **13**, e0208775 (2018).
4. Redfern, J. *et al.* An open-data, agent-based model of alcohol related crime. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–6 (2017).
5. Hörl, S. & Balac, M. Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies* **130**, 103291 (2021).
6. Ziemke, D., Kaddoura, I. & Nagel, K. The matsim open berlin scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. In *10th International Conference on Ambient Systems, Networks and Technologies and 2nd International Conference on Emerging Data and Industry 4.0 / Workshops*, vol. 151, 870–877 (2019).
7. Balac, M. & Hörl, S. Synthetic population for the state of California based on open-data: examples of San Francisco Bay area and San Diego County. In *100th Annual Meeting of the Transportation Research Board (TRB)* (2021).

8. Open Data Watch. Open Data Inventory - Canada Country Profile. Available at https://odin.opendatawatch.com/Report/countryProfileUpdated/CAN?year=2020 (2020). Accessed: 2022-12.
9. Perez, L., Dragicevic, S. & Gaudreau, J. A geospatial agent-based model of the spatial urban dynamics of immigrant population: A study of the island of Montreal, Canada. *PLOS ONE* **14**, 1–23 (2019).
10. Anderson, T., Leung, A., Dragicevic, S. & Perez, L. Modeling the geospatial dynamics of residential segregation in three canadian cities: An agent-based approach. *Transactions in GIS* **25**, 948–967 (2021). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12712. https://onlinelibrary.wiley.com/doi/pdf/10.1111/tgis.12712.
11. Malik, A. & Abdalla, R. Agent-based modelling for urban sprawl in the region of Waterloo, Ontario, Canada. *Modeling Earth Systems and Environment* **3** (2017).
12. Miller, E. J. & Roorda, M. J. Prototype Model of Household Activity-Travel Scheduling. *Transportation Research Record* **1831**, 114 – 121 (2003).
13. Fatmi, M. R. & Habib, M. A. Baseline Synthesis and Microsimulation of Life-stage Transitions within an Agent-based Integrated Urban Model. In *8th International Conference on Ambient Systems, Networks and Technologies and 7th International Conference on Sustainable Energy Information Technology* (2017).
14. Hafezi, M. H. & Habib, M. A. Synthesizing Population for Microsimulation-based Integrated Transport Models Using Atlantic Canada Micro-data. In *The 1st International Workshop on Information Fusion for Smart Mobility Solutions (IFSMS'14)*, 410–415 (2014).
15. Hafezi, M. H. & Habib, M. A. Development and Evaluation of and Algorithm to Produce the Population in Regional Level and Dissemination Area Level. In *Canadian Transportation Research Forum 50th Annual Conference*, 15 (2015).
16. OpenStreetMap contributors. OpenStreetMap. Available at https://www.openstreetmap.org (2017). Accessed: 2022-05.
17. Zhang, H. & Malczewski, J. Accuracy evaluation of the Canadian OpenStreetMap road networks. *International Journal of Geospatial and Environmental Research* **5**, Article 1 (2018).
18. Zhou, Q., Zhang, Y., Chang, K. & Brovelli, M. A. Assessing OSM building completeness for almost 13,000 cities globally. *International Journal of Digital Earth* **15**, 2400–2421 (2022).
19. Microsoft and Statistics Canada. Canadian Building Footprints. Available at https://github.com/Microsoft/CanadianBuildingFootprints (2019). Accessed: 2022-12.
20. Statistics Canada. Individuals File, 2016 Census of Population – Catalogue no. 98M0001X. Available at https://www150.statcan.gc.ca/n1/en/catalogue/98M0001X (2016). Accessed: 2022-08.
21. Statistics Canada. Hierarchical File, 2016 Census of Population – Catalogue no. 98M0002X. Available at https://www150.statcan.gc.ca/n1/en/catalogue/98M0002X (2016). Accessed: 2022-08.
22. Statistics Canada. Projected population, by projection scenario, age and sex, as of July 1 (x 1,000) – Table 17-10-0057-01. Available at https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710005701 (2018). Accessed: 2022-08.
23. Government of Canada. Open Government Across Canada. Available at https://open.canada.ca/en/maps/open-data-canada (2022). Accessed: 2022-12.
24. MobilityData. Mobility database catalogs. Available at https://database.mobilitydata.org/ (2019). Accessed: 2022-07.
25. Geofabrik. Download OpenStreetMap data for Canada. Available at http://download.geofabrik.de/north-america/canada.html (2022). Accessed: 2022-05.

26. Poletti, F. PT2MATSim v. 22.11. Available at https://github.com/matsim-org/pt2matsim (2021). Accessed: 2022-10.
27. Statistics Canada. Living area and assessment value per square foot of residential properties by property type and period of construction, provinces of Nova Scotia, Ontario and British Columbia - Table 46-10-0028-01. Available at https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=4610002801 (2019). Accessed: 2022-12.
28. Statistics Canada. 2016 Census Profile for Canada, provinces, territories, CDs, CSDs and DAs - Catalogue no. 98-401-X2016044. Available at https://www150.statcan.gc.ca/n1/en/catalogue/98-316-X2016001 (2016). Accessed: 2022-08.
29. Smith, A., Lovelace, R. & Birkin, M. Population Synthesis with Quasirandom Integer Sampling. *Journal of Artificial Societies and Social Simulation* **20**, 14 (2017).
30. Prédhumeau, M. & Manley, E. Synthetic population for Canada at the DA level for 2016, 2021, 2023 and 2030. (2022). URL https://doi.org/10.5281/zenodo.7085408.
31. Statistics Canada. Health characteristics, annual estimates – Table 13-10-0096-01. Available at https://doi.org/10.25318/1310009601-eng (2022). Accessed: 2022-08.
32. Transport Canada. Number of Licensed Drivers by Gender and by Age - Canadian Motor Vehicle Traffic Collision Statistics. Available at https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2020 (2020). Accessed: 2022-08.
33. Winnipeg Transit. Winnipeg Transit's Open Data Web Service. Available at https://api.winnipegtransit.com/home/api/v3 (2022). Accessed: 2022-07.
34. Winnipeg open data portal. City of Winnipeg Currently Active Official Property Addresses and Associated Coordinates. Available at https://data.winnipeg.ca/City-Planning/Addresses/cam2-ii3u (2022). Accessed: 2022-05.
35. Winnipeg open data portal. Single lane road network of the City of Winnipeg. Available at https://data.winnipeg.ca/City-Planning/Road-Network/2eba-wm4h (2022). Accessed: 2022-05.
36. Winnipeg Transit. Winnipeg Transit's Webpage - Regular service information. Available at https://info.winnipegtransit.com/en/service/regular-service/ (2022). Accessed: 2022-12.
37. Wiebe, K. Measuring Winnipeggers' Convenient Access to Public Transit. Tech. Rep., International Institute for Sustainable Development (IISD) (2018). Available at https://policycommons.net/artifacts/614388/measuring-winnipeggers-convenient-access-to-public-transit/1594668/.
38. Hess, D. B. Walking to the bus: perceived versus actual walking distance to bus stops for older adults. *Transportation* **39**, 247–266 (2012).