Creating a raster agent-based modelling environment from volunteered geographic information.

Hannah Gumble[*1]

[1]The Bartlett Centre for Advanced Spatial Analysis, University College London.

March 30th, 2021

Summary

This work attempts to lay the groundwork for future applied fine-grained modelling that uses geospatial data. A methodology is presented using OpenStreetMap (OSM) data to devise a raster-based agent-based model environment. The resolution of raster-based environment size on model initiation is discussed along with OSM's limitations in this context. In processing, clear data gaps impacted the methodology, especially missing classifications of attributes and lack of information on street furniture. A clear relationship between environment granularity and model initiation times was documented. Further research should seek to understand environment granularity and the impact on simulation times.

KEYWORDS: Agent-based Modelling, Crowdsourced data, OpenStreetMap, Simulation Environments and Raster data.

1. Introduction

The use of geographic information in agent-based modelling (ABM) is increasing (Crooks *et al.* 2019). Few modelling applications focus on fine-grained streetscapes, however the importance of understanding interactions at this level has grown in the wake of Covid-19. The pandemic has prompted the trial of e-scooters in the UK (Bradshaw, 2020), which has created a need to understand individual's interactions and how modes such as e-scooters can co-exist in current streetscapes. This paper examines the viability of processing OpenStreetMap (OSM) data into a fine-grained ABM simulation environment suitable for streetscape-level simulation.

An ABM environment is a critical component: it is the space in which the agents exist, interact and in some instances informs agent behaviours. There are numerous ways in which physical space can be modelled, from real world data to hypothetical approaches such as synthetically generated raster grids or tessellations (Brown *et al.* 2005). Environments can be dynamic, changing with the simulation results/ behaviours or static remaining the same throughout the simulation. There is no single recommended approach to construct an appropriate environment leading to high variation in the representation of space between models.

Mapping has evolved from national mapping agencies undertaking widescale updates to a crowdsourced approach, whereby individuals create/ contribute to maps to suit their needs (Goodchild, 2007). Perhaps, the most well-known open-source mapping application is OSM. Due to localised updates, data quality in some areas may be rich enough to support the development of fine-grained modelling environments; in other areas, the data quality may not be rich enough to support certain simulations. Data contributors' motivations and resulting accuracy have been the subject of numerous publications (Haklay, 2010); although accuracy is explored, understanding the potential effect on ABM simulations and best practises for its use remain limited.

The useability of raster versus vector data varies by scale and context of application. Vector data has richer data quality; however, it can suffer issues with locational accuracy. In contrast, a raster approach

---

[*] Hannah.gumble.17@ucl.ac.uk

relies on generalisation and attribute agglomeration (Heywood *et al.* 2006) resulting in a grid like structure. It must be noted that data used is informed by the purpose, scope and application of the model.

When developing a model, there can be a trade-off between computational demand and model detail. In some instances, high resolution data can be demanding on computational power/ simulation performance, particularly if agents adapt their behaviour at each step based on the environment.

Given this, this research aims to:
- Present a possible workflow to convert OSM data to a raster agent-based environment.
- Evaluate the impact of environment resolution and scale on model initialisation performance.
- Highlight some of the limitations in OSM when looking to create a raster-based ABM environment.

2. Methodology

The hypothetical context of the ABM used to underpin this study is a fine scaled street-based mobility application. A variety of platforms support the incorporation of GIS data, although capabilities vary. The platform used for this study is MESA, a modular python framework still in its infancy compared to NetLogo, Repast and MASON.

Data was downloaded from openstreetmap.org for an area in the City of London (Figure 1). The structure of downloaded data includes a different layer for each attribute classification i.e., highways, buildings and barriers provided in a mixture of point and polyline layers. A visual representation of the data used in this application can be seen in Figure 2.
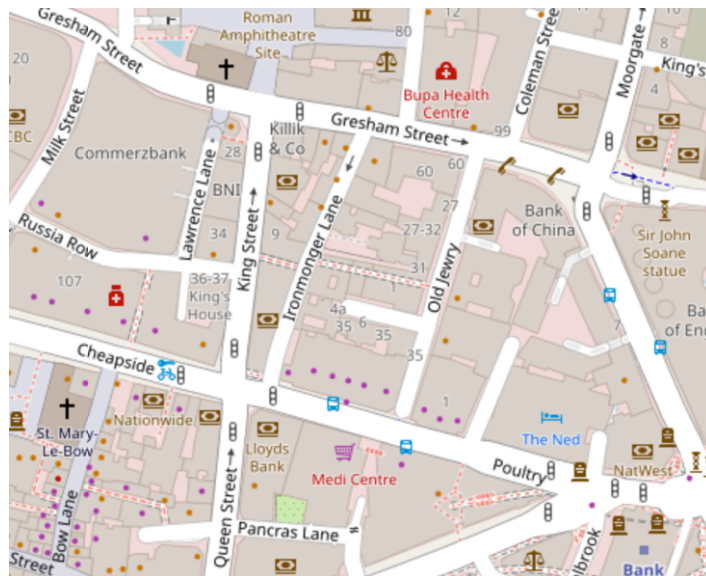

Figure 1: Output from OpenStreetMap, Case study: London.

Figure 2: Data downloaded from OSM visualised in ArcGIS, all data is available as polylines. Figure 1 is a subsection of this area.

The overall workflow is shown in Figure 3. Steps undertaken in the methodology can be split in to four processing stages: preparing OSM data, building processing, highway processing and raster creation.
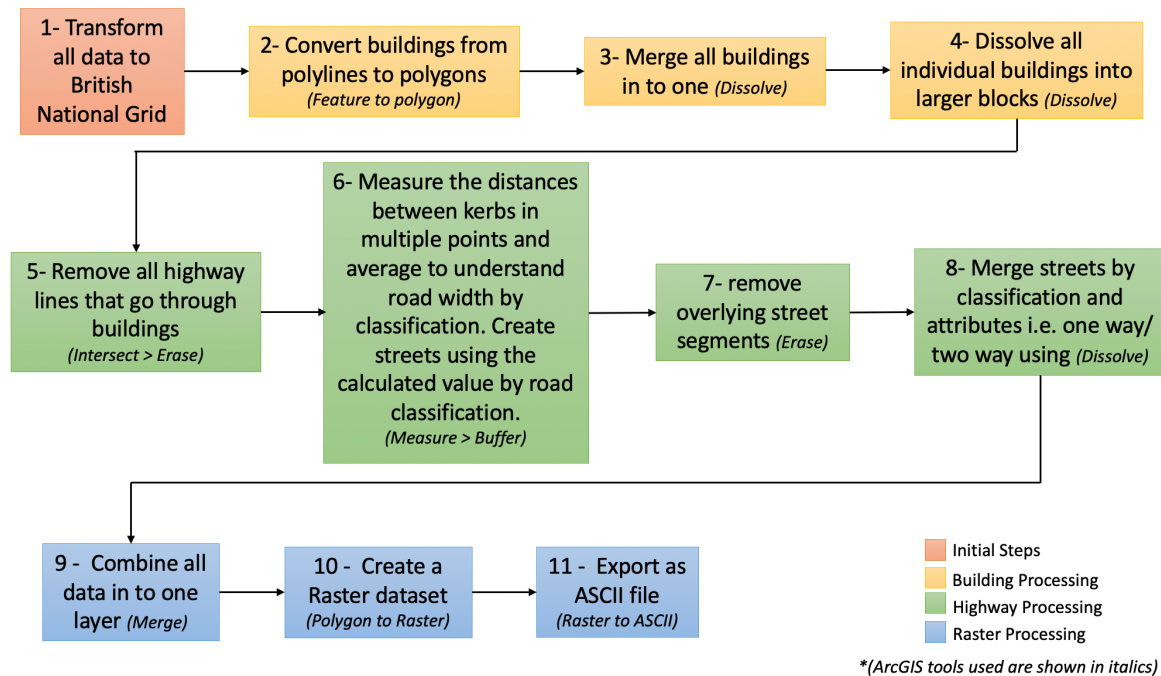

Figure 3: Workflow of the methodology used in this study to process OSM data.

It is important to highlight that average road width was calculated measuring kerb to kerb data; however, this data was not spatially consistent. Furthermore, it does not account for bus laybys. Given this, it was supplemented with satellite imagery analysis. Buffer distances used can be seen in Table 1.

Table 1: Buffered road widths (full) used by road classification.

Road Classification	Buffered road widths (Meters)
Residential	1.7
Tertiary	3.2
Unclassified	2.3
Service	1.7

Processed OSM data and three example output raster grids with increasing cell size are shown in Figure 4 and Figure 5. As expected, it can be seen in Figure 5 that the smaller the cell size the truer to the input vector data.
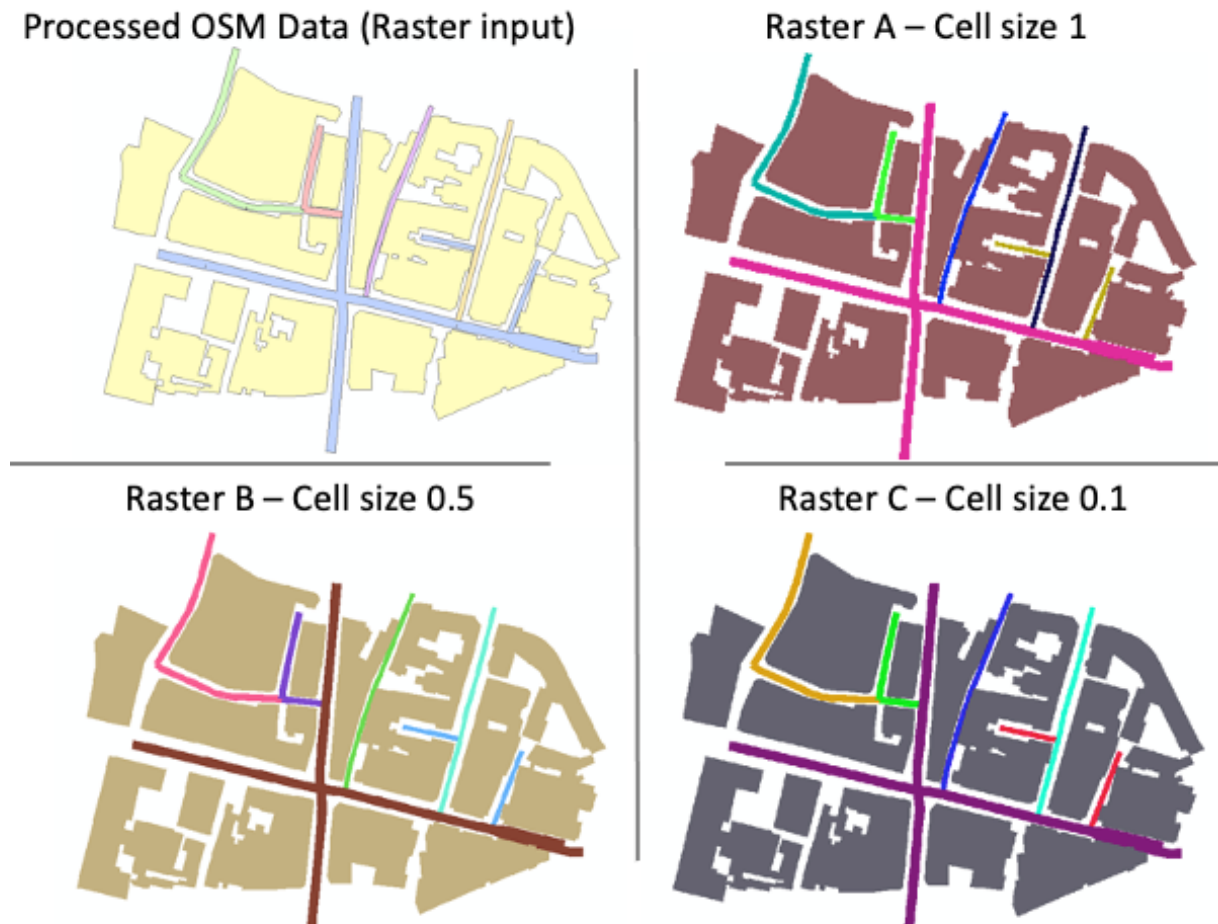
Figure 4: Processed OSM data and output rasters.

Raster A – Cell size 1   Raster B – Cell size 0.5   Raster C – Cell size 0.1
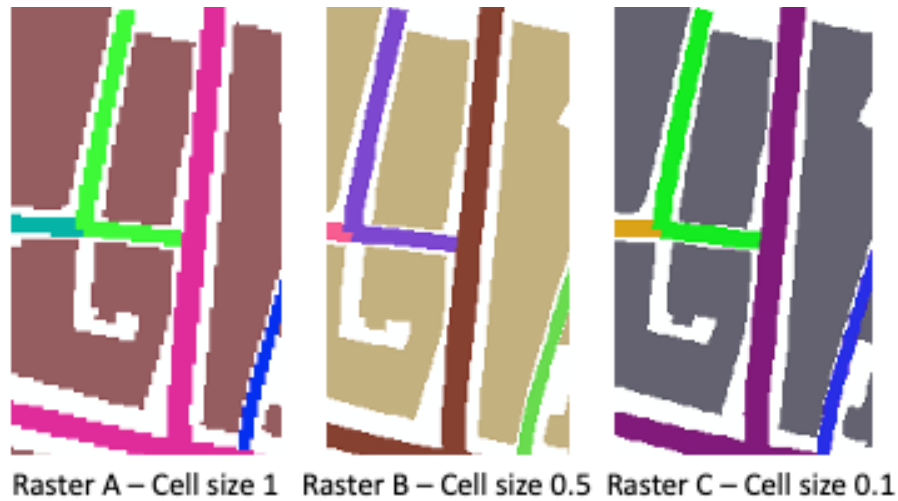
Figure 5: Small area comparison of cell size.

3. Analysis

3.1. Methodology Evaluation

When looking closely, it can be seen that with larger resolutions features such as roads become disjointed, staggering diagonally across cells. Raster environments like this affect agent pathing. For example, the movements of a simulated car travelling across a road would appear non-linear in visualisation due to the staggered nature of the road in the environment shown. In the GIS processing the impact of the environment scale had minimal impact on processing time. However, it must be questioned if this methodology is only suitable for smaller study areas with fewer features.

When implementing the environment into a MESA framework using the visualisation server a key limitation was uncovered. The size of raster environment could not be portrayed above a given threshold. As an outcome of this, the model had to be tested without the visualisation element. In some instances this may be suitable, however for the application discussed here the visual interaction of agents would be a key input for the model building attempting to debug or verify the model's behaviours.

3.2. Impact of raster resolution on model performance.

To understand the raster resolution and simulation size on model performance, comparison has been undertaken assessing raster dimensions and cell size on model initialisation times e.g. how quickly it takes the model to load upon start up. Results in seconds can be seen in Table 2. There is a clear relationship between smaller cell sizes and increasing model load time (Figure 6).

Table 2: Model initialisation time for different rasters produced.

Cell Size	No. of Rows	No. of columns	Model load time (seconds)
1	364	286	1
0.9	440	318	1.4
0.8	445	357	2
0.7	520	408	2.5
0.6	607	476	3.7
0.5	728	572	5.8
0.4	910	715	9.2
0.3	1213	953	16.6
0.2	1820	1429	37.7
0.1	3640	2859	255.5
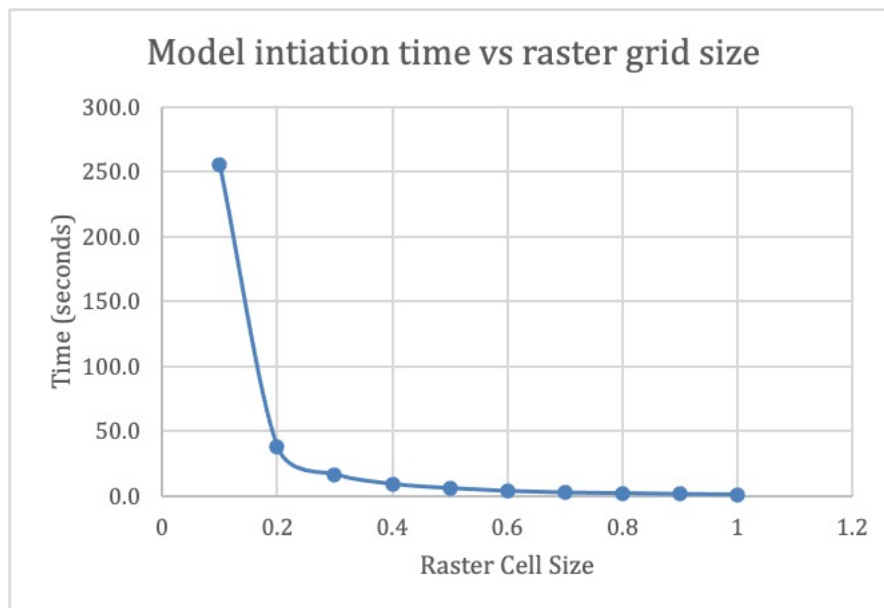
Figure 6: Graph highlighting the relationship between raster cell size and model load times using default settings.

3.3. OpenStreetMap Evaluation

Missing attributes were a hinderance for this application. Having sparce records of road widths, even though they are visualised on the online map, was a key limitation when trying to simulate a street environment. The highway data represented the centre of the road. To turn the streets from a single line into roads that agents can occupy, a buffer was used, with the widths determined by kerbs available for each road type specified. Some records remained unclassified, highlighting the problem of inconsistent contributions.

When examining an environment at street level, accurate distribution of road space in mobility can be critical. However, key streetscape features such as cycle lanes, street-side parking accommodation or bus laybys are not always indicated. Cycle lanes are commonly recorded in the highway attribute table with some records specifying side of the road. However, there is little notation of dimensions which makes spatial quantification difficult.

Waltersdorfer and Beddow (2019) contend that OSM data is not very detailed about the spaces 'in-between' buildings and roads. Observations made in this study support this statement as a large amount of detail is missing, such as street furniture. Although the addition of fine-grained urban features would be a vast undertaking, standalone trees within OSM are currently being mapped (Tag:natural=tree). This detail would enrich the possibility for fine-grained street-level ABM simulation environments to be developed.

4. Conclusion

A workflow to create raster-based ABM environments from OSM data is proposed. A clear pattern was documented between increasing raster granularity and increased model initialisation time. This can be important for applications where the environment informs agent behaviours or environmental features change throughout the simulation. To fully understand the impact of environment granularity, future studies should analyse the effect on simulation times and outputs. Furthermore, there is a need to examine different capabilities between MESA and other ABM frameworks. OSM remains a rich dataset for creating some ABM environments due to its localised updates and accessibility. However, the detail required is dependent on model application. To simulate individuals' interactions there needs

to be a higher level of street detail. Spatial inconsistences of mapped features in OSM, and to some degree crowdsourced data, is a key limitation at this scale. This study focuses on one location; therefore, care should be taken when applying the methodology to a different geographical context. The inclusion of additional data sources and/or greater understanding of location specific limitations could be used to extend and possibly create a more universal methodology.

5. Acknowledgements

6. References and Citations

Bradshaw, T (2020) 'Coronavirus drives UK authorities to embrace electric scooters', *Financial Times,* 3rd September. Available at: <https://www.ft.com/content/becb4445-1062-41c8-9ffa-5732a6e55ae8> (Accessed: 13th February 2021).

Brown, D., Riolo, R., Robinson, D., Morth, M. and Rand, W. (2005), 'Spatial process and data models: Toward integration of agent-based models and GIS', *Journal of Geo-graphical systems,* 7, pp. 25–47.

Crooks, A., Malleson, N., Manley, E and Heppenstall, A (2019) *Agent-Based Modelling and Geographical Information Systems: A Practical Primer (Spatial Analysis and GIS),* London: SAGE Publications.

Goodchild, M. F. (2007), 'Citizens as sensors: The world of volunteered geography', *Geo-Journal*, 69(4), pp.211–221.

Haklay, M (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets, *Environment and Planning B: Planning and Design,* 37, pp682-703.

Heywood, I., Cornelius, S and Carver, S. (2006) *An introduction to geographical information systems.* Harlow: Pearson Education.

Waltersdorfer, D, and Beddow, C. (2019) *From Sideline to Centre: The Potential for OpenStreetMap to seize the Curb.* Available at:< https://blog.mapillary.com/update/2019/02/12/potential-for-openstreetmap-to-seize-the-curb.html> (Accessed: 7th February 2021)

Biography

Hannah Gumble is a doctoral student at The Bartlett Centre for Advanced Spatial Analysis (CASA), University College London. Hannah studied Geography at the University of Plymouth before obtaining her MSc in Spatial Data Science and Visualisation. Her research interests are micromobility, agent-based modelling, spatial analysis and transportation.