The Walkable Accessibility Score (WAS): A spatially-granular open-source measure of walkability for the continental US from 1997-2019

Kevin Credit*[1], Irene Farah†[2], Emily Talen[3], Luc Anselin[3], and Hassan Ghomrawi[4]

[1]National Centre for Geocomputation, Maynooth University
[2]Department of City & Regional Planning, University of California-Berkeley
[3]Center for Spatial Data Science, University of Chicago
[4]Feinberg School of Medicine, Northwestern University

GISRUK 2024

Summary

This paper describes a method, with specific parameters, for developing an open-source measure of walkability that can be applied to any set of spatial demand units. Our method compares favorably with existing proprietary measures of walkability. Exploratory analysis of the resulting measure, available for the entire continental US for every year from 1997-2019, reveals high concentrations of walkability in the dense centres of the largest urban areas.

KEYWORDS: walkability, spatial analysis, accessibility, access score, POI data

1. Introduction

Measuring walkability and its relationship to health, equity, the environment, and economic development has long been of interest to spatial planners, urban designers, and researchers (Jacobs, 1961). To quantify these relationships, researchers have developed a variety of 'accessibility' metrics to measure the aggregate availability of important services and amenities in their surroundings (Ewing and Cervero, 2010; Talen and Koschinsky, 2013). While there is a robust literature on these spatial access measures in the academic community, for practitioners (or members of the public) employing these methods are often beyond the scope of their technical capabilities. Thus, to this point, the most well-known walkability measures, such as Walk Score®, have been developed and provided by commercial entities. For proprietary reasons, the configuration of these metrics is not made openly available to the public - but this makes it difficult to understand exactly what these metrics are capturing, and how they can be improved.

The purpose of this paper is to outline a flexible, open source methodology for computing a Walkable Accessibility Score (WAS) - similar to the commercially-available Walk Score® - that allows researchers to empirically operationalise walkability concepts. We then apply this method to create a comprehensive measure of walkability at the block group scale for the continental United States for every year from 1997-2019 and examine interesting spatial patterns and trends in the data. Ultimately, the goal of this project is to provide this data as an open resource for academics, planners, and members of the public to use to study relationships and better understand their own neighbourhoods. The initial code to

* kevin.credit@mu.ie
† irenef@berkeley.edu

operationalise this method and a link to download the full set of aggregated WAS values at the block group level for the continental US from 1997-2019 are available at the following GitHub repository: https://github.com/kcredit/Walkable-Accessibility-Score/tree/main.

2. Methodology

Spatial accessibility metrics can generally be divided into two broad types: the floating catchment area (FCA) method and its various extensions (Luo, 2004; Wang and Luo, 2005; Saxon and Snow, 2020), and the 'access score' (Isard, 1960). In both, the data are structured as *demand units* (i.e., where the population lives, often census areal units) and *supply sites* (i.e., the facilities where services are provided, often points of interest of some type), and the goal is to aggregate the accessibility *from* each demand unit *to* nearby supply sites. This means that the resulting aggregate measure of accessibility is ultimately a feature of the demand units rather than the supply sites.

Generally, the FCA family of methods considers the ratio of supply to demand and is thus most useful in contexts where limited or differential *supply* is a concern, e.g., physician availability. However, in the walkability application this makes less sense; proximity to many nearby retail businesses improves the walkability of every nearby demand unit no matter the size of the total demand in the area. For this reason, we conceptualise walkability as a function purely of the amount of nearby supply sites using the 'access score' method (Isard, 1960). As shown in Equation 1, the access score (a) for a given demand unit i (Census block groups, in this case) is simply the weighted sum of the distances (d) to a pre-specified number (k) of nearby supply sites j (discounted by some distance decay function g), multiplied by an amenity weight (w), i.e., some measure of the relative importance of each supply site:

$$a_i=\sum_{j=k} w_j * g(d_{ij}) \quad (1)$$

To create a walkable accessibility score (WAS) in this paper, the relevant walkable amenity types to consider (Cerin et al. 2007; Lee and Moudon, 2006; Lovasi et al. 2008) and the form of the distance decay function – logit (Lerner et al. 2008; Vale and Pereira, 2017) – were taken from the literature. In general, proximity to businesses that serve basic needs (such as grocery stores), retail shops, schools, parks, and food services all tend to incentivize walking. **Table 1** shows the specific types of amenities used in the paper, along with their North American Industrial Classification System (NAICS) codes and data source. Lacking any strong evidence from the literature, all amenity types were weighted equally. Distances between demand units and supply sites were calculated using Euclidean distance.

Table 1 Amenity types used in the calculation of the walkable accessibility score (WAS)

Category	*Amenities*	*NAICS*	*Source*
Basic Needs	Grocery and liquor	4451, 4453	InfoUSA
	Drug stores	4461	InfoUSA
General Amenities	Shopping (e.g., furniture, electronics, clothing)	4421, 4431, 4481, 4482, 4483, 4511, 4531, 4532, 4539, 4523	InfoUSA
	Banks	5221	InfoUSA
	Books	451211	InfoUSA
	Schools	NA	GreatSchools USA 2011

	Parks	NA	ArcGIS Online - USA Parks 2021
Food Services	Accommodation and food services	72	InfoUSA
	Bakeries	311811	InfoUSA

3. Results

For the final logit specification, the specific parameters to use cannot be directly gleaned from the literature. This is also true for the choice of *k*-nearest amenities to sum over. Thus, to find the best performing combination of these parameters, we ran a series of systematic comparisons to the 2011 Street Smart Walk Score® for 6 regions: Miami, Chicago, Atlanta, Seattle, and Boston. It is important to note here that Walk Score® data is provided at a multi-scalar geography, e.g., in dense urban areas the demand units are relatively small grid cells, while in less dense suburban areas they are (for example) Census tract centroids. Given the flexibility of our approach, any set of demand units can be used, so while this project's final dataset is produced systematically at the block group level (for ease of interpretability and matching with other datasets), for this comparison we used the same multi-scalar geography as Walk Score®.

As shown in **Table 2**, we calculated the WAS for 12 combinations of *k* values (5, 10, 15, and 150) and distance decay function thresholds (1600m, 2400m, and 3200m) using our chosen data sources, computing the Spearman Rank correlation with the original Street Smart Walk Score® data each time. In the end, summing over the nearest 150 amenities with a 1600m decay threshold provided a correlation of 0.903, which was surprisingly large.

Table 2 Comparison of various distance threshold and *k* parameters for judging similarity to proprietary 2011 Walk Score® values. Best-performing correlation marked in bold.

Maximum Thresholds (> 1% weight)		*Spearman Rank Correlation w/2011 Walk Score®*	
Distance (m)	# Amenities (*k*)	2011 InfoUSA + Parks and Schools	Computation time (µs)
1600	5	0.795	43.6
	10	0.835	21.9
	15	0.860	22.2
	150	**0.903**	60.8
2400	5	0.795	23.6
	10	0.836	21.7
	15	0.860	22.4
	150	0.873	61.7
3200	5	0.795	22.7
	10	0.836	20.7
	15	0.860	22.2
	150	0.869	62.0

As shown in **Figure 1**, the maximum threshold (> 1% weight) of the final specification of the logit decay function based on these tests is effectively 1600m, with a midpoint near 800m.

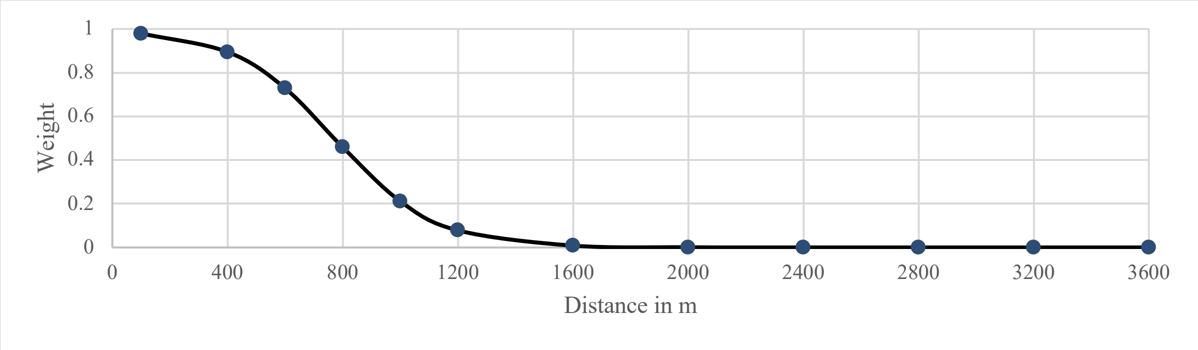

Figure 1 Chosen distance decay function specification.

With the parameters for computing a high-performing open-source walkable accessibility score (WAS) identified, we then computed the WAS at the block group scale for all years (1997-2019) and geographic extents (the continental US) for which we have relevant point of interest (POI) data available on amenities. Initial exploratory analysis of the data indicates that high values of WAS are generally quite concentrated in the dense centres of the country's largest cities; 19 of the top 25 (and 20 of the next 25) average WAS values from 1997-2019 by block group are located in Manhattan, NY; the other cities appearing in top 25 are Seattle, WA (near Pike's Place Market) (1 block group), San Francisco, CA (2), downtown Los Angeles, CA (2), and Santa Fe, NM (1). Chicago, Philadelphia, Miami, and Boston make their first appearances in the block groups ranked 25-50 (along with another in San Francisco). When aggregated to Core Based Statistical Areas (CBSA), as shown in **Figure 2**, these regions also stand out, with New York, Los Angeles, San Francisco, Chicago, and Boston comprising the top 5 in terms of average WAS over this time period.

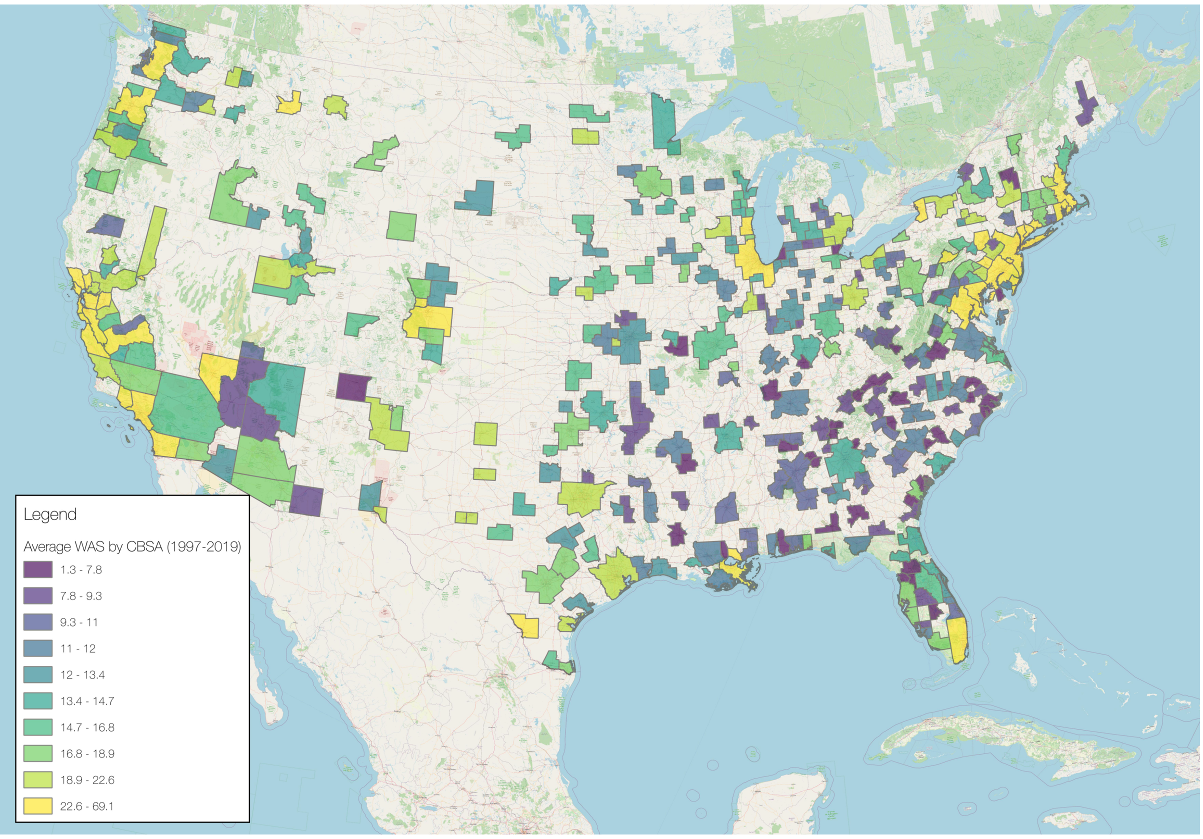

Figure 2 Average Walkable Accessibility Score (WAS), 1997-2019, by Core Based Statistical Area.

4. Conclusions and Future Directions

Beyond the ability to provide the aggregated WAS dataset as an open resource for researchers and the public, these data also offer several interesting avenues for future research that we are just beginning to explore. First, we are interested in analyzing any potential systematic deviations between WAS and Walk Score® and further testing the remaining parameters (e.g., amenity weights, street network distance) and/or others – such as features of street network design – to see if we can obtain an even better fit. Beyond that, we are also interested in conducting external validation of the WAS with more dynamic measures of actual walking behaviour, e.g., footfall data. Finally, these data – as both dependent and independent variables - offer a number of interesting opportunities to examine important urban relationships, including the relationship between socio-economic factors and walkability, the causal impact of infrastructure investments on walkability, and spatio-temporal changes in walkability over time.

5. Acknowledgements

This work was funded under the National Institutes of Health (NIH) project number 5R01AR078342-03.

References

Cerin E, Leslie E, duToit L, Owen N and Frank L (2007) Destinations that matter: Associations with walking for transport. *Health and Place* 13(3): 713–724.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American planning association*, *76*(3), 265-294.

Isard, W. 1960. *Methods of Regional Analysis: Introduction to Regional Science*. MIT Press, Cambridge, MA.

Jacobs J (1961) *The Death and Life of Great American Cities*. New York: Random House.

Lee C and Moudon A (2006) The 3ds + r: Quantifying land use and urban for correlates of walking. *Transportation Research Part D* 11: 04–215.

Lerner, MR, Kocher, JG, and MC Mathieu (2008). US Patent No. US 8892455 B2. Retrieved from https://patents.google.com/patent/US8892455B2/en.

Lovasi G, Moudon A, Pearson A, Hurvitz P, Larson E, Siscovick D, Berke E, Lumley T and Psaty B (2008) Using built environment characteristics to predict walking for exercise. *International Journal of Health Geographics* 7(10): 713–724.

Luo, W. 2004. Using a gis-based floating catchment method to assess areas with shortage of physicians. *Health & Place*, 10(1):1 – 11.

Pivo G and Fisher JD (2011) The walkability premium in commercial real estate investments. *Real Estate Economics* 39(2): 185–219.

Saxon, J. and Snow, D. 2020. A rational agent model for the spatial accessibility of primary health care. *Annals of the American Association of Geographers*, 110(1): 205-222.

Talen, E and Koschinsky, J (2013) The walkable neighborhood: A literature review. *International Journal of Sustainable Land Use and Urban Planning* 1(1): 42–63.

Vale, D. S., & Pereira, M. (2017). The influence of the impedance function on gravity-based pedestrian accessibility measures: A comparative analysis. *Environment and Planning B: Urban Analytics and City Science*, 44(4), 740-763.

Wang, F. and Luo, W. 2005. Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & Place*, 11(2):131 – 146.

Biographies

Kevin Credit is an Assistant Professor at the National Centre for Geocomputation at Maynooth University. His recent work uses quantitative approaches and large open-source datasets to look at topics such as the development of spatially-explicit causal machine learning models, the impact of transit construction on carbon emissions, and the relationship between the built environment and commuting by mode.

Irene Farah is a PhD Candidate in City and Regional Planning at University of California-Berkeley. Her research revolves mostly around themes of inequality, focusing on topics of urban informality, governance, health, and spatial analytics.

Emily Talen is Professor of Urbanism at the University of Chicago. Her research is devoted to urban design and the relationship between the built environment and social equity. Her books include: New Urbanism and American Planning, Design for Diversity, Urban Design Reclaimed, City Rules, and Retrofitting Sprawl: Addressing 70 Years of Failed Urban Form. She is the recipient of a Guggenheim Fellowship (2014-15), and is a Fellow of the American Institute of Certified Planners.

Luc Anselin is the Stein-Freiler Distinguished Service Professor of Sociology and the College at the University of Chicago. He is the developer of the SpaceStat and GeoDa software packages for spatial data analysis. He was elected Fellow of the Regional Science Association International in 2004 and was awarded their Walter Isard Prize in 2005 and William Alonso Memorial Prize in 2006. He was elected to the National Academy of Sciences in 2008 and the American Academy of Arts and Sciences in 2011.

Hassan Ghomrawi is an Associate Professor of Surgery (Pediatric General Surgery), Medicine (Rheumatogoloy), and Pediatrics at Northwestern University. He is a surgical outcomes researcher and health economist with expertise in the use of algorithms and wearable technology to improve appropriateness of surgical care, decision making and cost effectiveness of surgical intervention and medical devices.