AGILE: GIScience Series, 7, 42, 2026. https://doi.org/10.5194/agile-giss-7-42-2026
Proceedings of the 29th AGILE Conference on Geographic Information Science, 16–19 June 2026.
Eds.: Evelyn Uuemaa, Alexander Kmoch, Holger Virro, and Simon Scheider
This contribution underwent peer review based on a full paper submission.

Turning Geospatial Data into Planning Decisions: Evaluating a Participatory Accessibility Model for Urban Drinking Water

Oluwatimilehin Adenike Shonowo[1], Xingyi Du[1], Diego Pajarito Grajales[1], Sebastian Hafner[1], Joao Porto de Albuquerque[1], Qunshan Zhao[1], Peter Elias[2], Kehinde Baruwa[2], Oluwatoyin Odulana[2], Sabitu Saad[3], Taiwo G. Amole[4], Solomon Abioye Olusegun[5], Jose Luis Fernandez-Marquez[6]

[1] Urban Big Data Centre, University of Glasgow, Glasgow, United Kingdom
[2] Department of Geography, University of Lagos, Lagos, Nigeria
[3] Science Department, School of Continuing Education, Bayero University Kano, Kano, Nigeria
[4] Aminu Kano Teaching Hospital, Bayero University Kano, Kano, Nigeria
[5] DonateWater Youths Climate Action Foundation, Lagos, Nigeria
[6] University of Geneva, Geneva, Switzerland

Correspondence: Oluwatimilehin Adenike Shonowo (o.shonowo.1@research.gla.ac.uk)

Abstract. Spatial accessibility models are widely used in infrastructure planning, yet quantitative evaluation against lived access conditions remains rare. This study implements a quality-weighted E2SFCA model of drinking water access in Kano and Lagos, Nigeria, and evaluates grid-level predictions against participatory validation data. Results show limited overall agreement (macro F1: 0.20–0.23) and context-dependent class bias. Severe deprivation is over-detected in Kano but under-identified in Lagos. Findings highlight the necessity of participatory validation when translating geospatial accessibility modelling into defensible planning interpretation.

Submission Type. model; analysis; case study

BoK Concepts. [AM] Analytical Methods; [GC] Geocomputation; [GS] GI and Society

Keywords. Spatial accessibility modelling; E2SFCA; Participatory GIS; Urban drinking water access; African cities

1 Introduction

Urban water planning increasingly relies on geospatial data to identify underserved populations and prioritise infrastructure investment. However, transforming spatial data into planning decisions requires more than mapping locations. Accessibility models estimate potential interaction between population demand and service supply, yet whether these estimates reflect lived access conditions remains an empirical question. In rapidly urbanising cities of Sub-Saharan Africa, where infrastructure reliability and informal water markets shape access, proximity does not guarantee adequacy (Penchansky and Thomas, 1981; WHO and UNICEF, 2023).

The Enhanced Two-Step Floating Catchment Area (E2SFCA) method provides a structured framework for estimating spatial accessibility by integrating supply–demand ratios, network-based travel time, and distance decay (Luo and Qi, 2009). Widely applied in health service research, E2SFCA generates continuous accessibility surfaces under defined behavioural and infrastructural assumptions. Yet such models remain structural abstractions. They capture physical reachability but not service intermittency, affordability constraints, queueing dynamics, or adaptive coping strategies common in urban water systems (WHO, 2020).

Recent work emphasises integrating citizen-generated data into spatial analysis to enhance contextual validity and accountability (Fritz et al., 2019; Mazumdar et al., 2017). However, systematic quantitative evaluation of accessibility models against participatory validation data remains limited in urban water planning contexts.

This study evaluates a participatory accessibility modelling framework for drinking water in Kano and Lagos, Nigeria. An E2SFCA model incorporating service-level weights derived from waterpoints quality categories was implemented using pedestrian network travel times within a 60-minute threshold. Grid-level accessibility scores were classified into Low, Medium, and High deprivation categories. Model outputs were then compared against participatory validation data aggregated at grid level. Predictive agreement was assessed using class-specific recall and F1 metrics.

By explicitly evaluating model performance rather than assuming validity, this paper strengthens the link between geospatial accessibility modelling and planning interpretation. The results demonstrate how participatory validation reveals context-specific classification biases, informing more cautious and evidence-based use of accessibility outputs in urban water decision-making.

2 Methodology

2.1 Conceptual Framework

Drinking water accessibility was conceptualised as a spatial interaction between population demand and service-qualified infrastructure supply. In line with classical accessibility theory, accessibility reflects the potential for interaction between population locations and service opportunities (Hansen, 1959).

Operationally, this spatial interaction was implemented using a floating catchment approach that integrates supply, demand and distance decay within a defined threshold.

Unlike purely proximity-based accessibility models, the present framework incorporates service availability characteristics into the supply term. Waterpoints were differentiated by functionality, drinkability, and infrastructure type, and operationalised using quality-weighted supply coefficients. Accessibility therefore reflects not only spatial reachability but also relative service adequacy.

In line with multidimensional access theory (Penchansky and Thomas, 1981), the model captures only selected dimensions of access. Affordability, queueing time, seasonal reliability, and volumetric capacity were not explicitly modelled due to data constraints.

2.2 Data Preparation and Service-Level Construction

2.2.1 Waterpoint Data Cleaning

Waterpoint records were obtained from the Donate Water Dataset: Youth-Driven Citizen Science on Water Access in Nigeria-Combined Dataset (Waves 1–3, 2023–2024), a nationwide citizen science dataset collected via the Donate Water mobile application across October 2023 to May 2024 (DOI 10.5281/zenodo.19181608).

2.2.2 Drinkability Imputation

Where drinkability status was recorded as “Don’t know”, a rule-based imputation was applied using observed water condition attributes. Water described as “clear” was classified as drinkable, while visibly contaminated states (e.g., muddy, coloured, dirty, or no water available) were classified as non-drinkable. This deterministic proxy-based classification substitutes observable water state for laboratory-confirmed potability and should be interpreted as structured imputation rather than empirical verification.

2.2.3 Improved and Unimproved Classification

Water sources were classified as improved or unimproved using context-adapted criteria informed by JMP-aligned typologies and local participatory interpretation to ensure consistency across both cities.

2.2.4 Service-Level Categorisation and Weighting

Each waterpoint was assigned to one of three service categories:

1. **Optimal Water**: Fully functional, improved, and drinkable sources
2. **Moderate Water**: Functional sources not meeting all optimal criteria
3. **Limited Water**: Non-drinkable sources

Service categories were operationalised using heuristic supply weights:

1. Optimal = 1.0
2. Moderate = 0.5
3. Limited = 0.2

These weights represent relative service adequacy rather than discharge capacity. In the absence of facility yield data, uniform baseline capacity was assumed across facilities, with quality differentiation captured through ordinal weighting.

2.2.5 Population Demand Surface

Population demand was derived from the WorldPop 100 m gridded population dataset (WorldPop and CIESIN, 2018), representing the estimated spatial distribution of resident population within each study city.

2.3 Accessibility Modelling using E2SFCA

Accessibility was estimated using the E2SFCA method (Luo and Qi, 2009), integrating supply availability, population demand, and Gaussian distance decay within a defined time-based network catchment threshold.

2.3.1 Travel-Time Threshold Contextualisation

A 60-minute catchment threshold was adopted to reflect empirically observed upper-bound water collection times in the sampled wards of the case study areas.

In Kano, household survey evidence from Rimin Kebe (Ungogo LGA) indicates that prolonged water collection journeys are common. Bello et al. (2020) report that 34% of respondents trekked one hour or more to access water sources, with an additional 30% trekking between 30 and 60 minutes. This supports a 60-minute threshold as a realistic upper limit under infrastructural scarcity.

In Lagos, extended access times have similarly been documented. Fagbohun et al. (2022) report that households in Papa Ashafa ward (Agege LGA) recorded average times of up to 60 minutes to access available water facilities, particularly in areas with limited public supply. The documented access duration substantiates the plausibility of hour-level collection times in underserved neighbourhoods.

While the World Health Organization (WHO, 2020) recommends a 30-minute benchmark for acceptable basic water access, the extended threshold adopted in this study reflects empirically observed coping realities. A stricter threshold would truncate supply–demand interaction in peripheral and underserved areas, potentially underrepresenting effective accessibility.

2.3.2 Walking Profile Configuration

Accessibility was modelled using the OpenRouteService (ORS) “foot-walking” routing profile to reflect prevailing pedestrian water collection practices in urban Sub-Saharan Africa. Network distances were computed along pedestrian-permitted infrastructure, including residential streets and walkable pathways, while excluding motorways, highways, and restricted-access roads.

Travel time was derived from network distance assuming a constant pedestrian walking speed of 5 km/hour (≈1.39 m/s). This parameter lies within the empirically reported range of average adult walking speeds (approximately 1.3–1.4 m/s) documented in cross-national observational studies (Levine and Norenzayan, 1999).

While water collection may involve additional physical burden (e.g., carrying containers or traversing unpaved terrain), applying a uniform walking speed ensured methodological consistency across both case study areas and avoided introducing unverified local speed assumptions.

2.3.3 Distance Decay Specification

A Gaussian distance decay function was applied:

$$W(d) = e^{-\frac{d^2}{\beta}} \quad (1)$$

where d represents walking travel time and β is derived from the maximum threshold duration. This formulation ensures progressively declining influence of facilities with increasing distance.

2.3.4 Accessibility Computation

Step 1: Supply-to-demand ratio

$$R_i = \frac{S_i}{\sum_{\{k \in d_{\{ik\}} \leq d_0\}} W\left(d_{\{ik\}}\right) P_k} \quad (2)$$

Step 2: Grid-level accessibility

$$A_j = \sum\left\{i \in d_{\{ij\}\leq} d_0\right\} \; W\left(d_{\{ij\}}\right) R_i \quad (3)$$

Where:
S_i denotes service-weighted supply at facility i;
P_k denotes population demand at location k;
d_ik represents network-based walking time between facility i and demand location k;
d_ij represents network-based walking time between facility i and grid cell j;
d_0 is the maximum catchment threshold (60 minutes);
R_i is the facility-level supply-to-demand ratio;
A_j is accessibility at grid cell j;
$W(d)$ is the Gaussian distance-decay weight.

2.4 Standardisation and Deprivation Classification

Accessibility scores were standardised using Min–Max normalisation within each city, producing an index bounded between 0 and 1.

Initial deprivation thresholds were derived using equal quantile classification (tertiles), ensuring approximately balanced class distribution. These thresholds were subsequently reviewed in consultation with local stakeholders to assess alignment with observed service realities. Adjustments were limited and preserved the overall distributional structure of the original quantile classification. Refinements were applied independently within each city.

During distributional inspection in Kano, a small cluster of grid cells exhibited extremely low accessibility values (< 0.000001) associated with sparsely populated peripheral areas. These were treated as artefactual edge effects and removed prior to final classification. Distributional inspection of Lagos did not reveal a comparable cluster of extreme near-zero values; therefore, no exclusion threshold was applied.

Final outputs classify grid cells into three deprivation levels: Low, Medium, and High.

2.5 Participatory Validation

Participatory validation was conducted through structured community workshops in Dorayi Karama (Kano) and Ajegunle Ikorodu (Lagos). Participants were introduced to the drinking water accessibility modelling framework and the interpretation of deprivation classes (low, moderate, and high) using visual examples and descriptions.

Model outputs were displayed through the Integrated Deprived Area Mapping System (IDEAMAPS) ecosystem platform as colour-coded grid-level maps, and participants were trained to assess whether the mapped classifications reflected the lived realities of water access in their communities. Validation involved both interactive map-based agreement/disagreement assessment and guided discussion on whether the model outputs accurately captured local access conditions, including reliability, travel burden, and service adequacy.

Responses were aggregated at the grid level using a majority-vote approach to derive participatory validation classes used for model evaluation.

2.6 Model Scope and Limitations

The model estimates physical accessibility and structural reliability under network-constrained walking conditions rather than full-service adequacy.

Although affordability was recorded in the dataset, it was captured only as a binary descriptor (“free” vs. “not free”) without price magnitude or household income information. The absence of socioeconomic context prevented construction of a meaningful affordability index; therefore, affordability was excluded from quantitative modelling.

Queueing conditions were recorded descriptively without consistent measurement of waiting duration. Because waiting time could not be quantified reliably, it was not incorporated into the accessibility model. In high demand areas, omission of waiting time may systematically overestimate effective accessibility, as actual service acquisition requires both travel and queuing time.

Additional assumptions include:

1. Uniform facility capacity
2. Heuristic service weighting
3. Deterministic proxy-based drinkability classification
4. Static network conditions without seasonal variability

These limitations may influence accessibility estimates depending on local conditions.

A formal sensitivity analysis of threshold and service-weight assumptions was beyond the scope of this short paper and is planned for the full paper.

2.7 Data and Software Availability

All analyses were implemented in Python using GeoPandas, Pandas, NumPy, and associated geospatial libraries within Jupyter Notebooks. Travel-time matrices were computed using the OpenRouteService API and OpenStreetMap network data.

The modelling workflow, preprocessing scripts, and reproducible analysis pipeline are publicly available here.

3 Results

3.1 Kano: Model Output and Participatory Validation

The original E2SFCA model classified 100 m grid cells in Kano into Low, Medium, and High deprivation categories using tertile-based thresholds following Min–Max standardisation. Spatially, the model produced a strong concentration of High deprivation across peripheral and sparsely serviced zones, with Medium and Low deprivation distributed more locally around clusters of functional and drinkable improved waterpoints.

Participatory validation was conducted across 1,474 grid-level observations as seen in Fig.1. Comparison between model output and validation data yielded an overall accuracy of **26%**. The macro-averaged F1 score was **0.20**, indicating limited overall agreement across classes.

Class-level performance reveals marked asymmetry:

1. **Low deprivation:** recall = 0.00, F1 = 0.00
2. **Medium deprivation:** recall = 0.16, F1 = 0.23
3. **High deprivation:** recall = 0.99, F1 = 0.37

In Kano, the model demonstrates strong sensitivity to severe deprivation but limited discrimination across lower deprivation levels, with a tendency towards high-deprivation classification. The confusion matrix suggests this dominant prediction pattern inflates recall for severe deprivation while suppressing performance in the Low and Medium categories, potentially leading to over-prioritisation of interventions in comparatively less deprived areas.

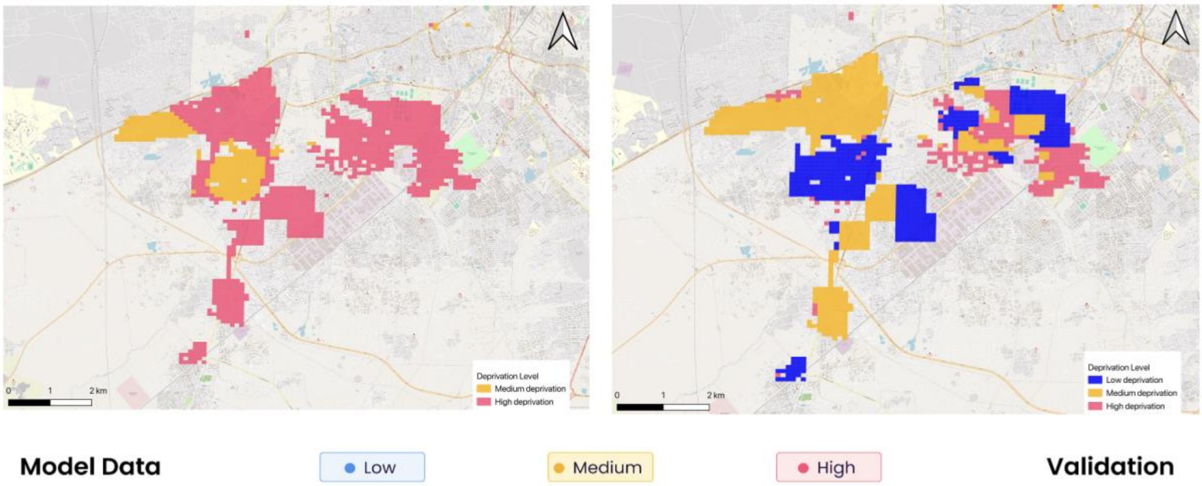

Figure 1. Kano: (a) Model Output Deprivation Classification; (b) Participatory Validation Classification

3.2 Lagos: Model Output and Participatory Validation

In Lagos, 543 grid cells were classified into three deprivation levels. Validation was aggregated across 2,094 participatory assessments corresponding to 423 grid cells as shown in Fig. 2. The overall model accuracy was **29.31%**, slightly higher than in Kano. The macro-averaged F1 score was **0.23**, indicating marginally stronger quantitative performance relative to Kano.

Class-level metrics demonstrate a contrasting performance structure:

1. **Low deprivation:** recall = 0.50, F1 = 0.27
2. **Medium deprivation:** recall = 0.46, F1 = 0.40
3. **High deprivation:** recall = 0.01, F1 = 0.02

Unlike Kano, Lagos did not exhibit dominance of a single predicted class. Medium deprivation achieved the highest F1 score (0.40), reflecting better balance between precision and recall. However, performance in identifying High deprivation was extremely weak (recall = 0.01), indicating substantial under-detection of validated severe deprivation zones.

The model performs better in intermediate classes but substantially underrepresents severe deprivation. This under identification of High deprivation contrasts sharply with the over identification observed in Kano. For planning, under detection of High deprivation implies a risk of under-prioritising the most constrained areas when model outputs are used for targeting.

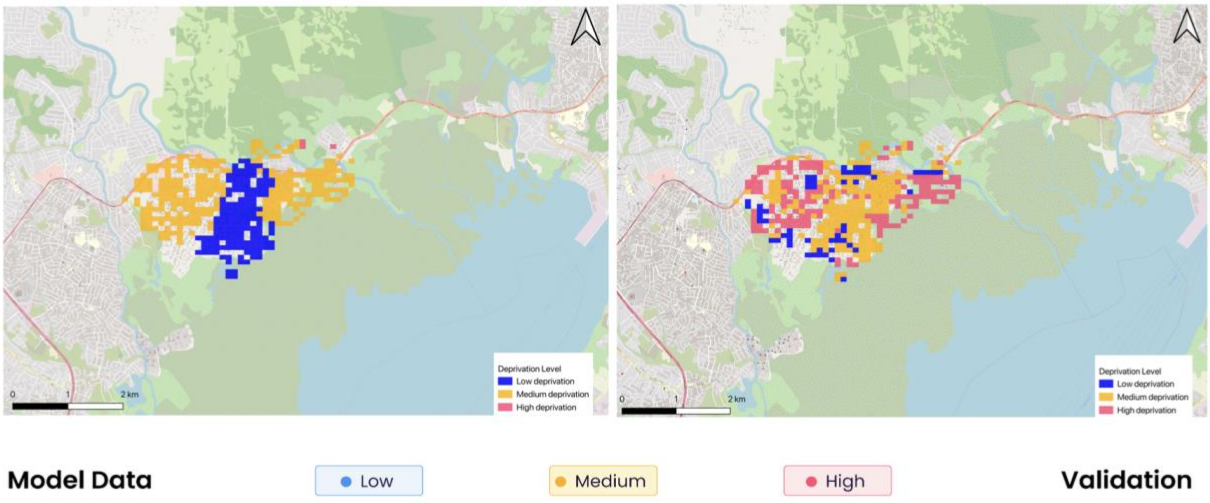

Figure 2. Lagos: (a) Model Output Deprivation Classification; (b) Participatory Validation Classification

3.3 Cross-City Quantitative Comparison

Performance varies substantially across contexts, with opposing classification tendencies between Kano and Lagos, highlighting context sensitivity of structural accessibility models.

Table 1. Model Performance Comparison between Kano and Lagos.

Metric	Kano	Lagos
Accuracy	26%	29.31%
Macro F1	0.20	0.23
Low Recall	0.00	0.50
Medium Recall	0.16	0.46
High Recall	0.99	0.01

4 Conclusion

This study operationalised an E2SFCA model to estimate drinking water access deprivation across two Sub-Saharan African urban contexts, Kano and Lagos, Nigeria, and evaluated model outputs against participatory validation data.

Validation results reveal limited predictive alignment in both cities as seen in Tab. 1, with macro F1 scores of 0.20 (Kano) and 0.23 (Lagos). However, performance patterns differed substantially between contexts. In Kano, the model demonstrated high sensitivity to severe deprivation but systematically over-classified grid cells into the High category, failing to identify Low deprivation. In contrast, Lagos exhibited more balanced intermediate classification performance but under-detected High deprivation zones. These contrasting structures indicate that model behaviour is context-dependent and class-sensitive.

The findings suggest network-constrained physical accessibility, even when service quality is weighted, does not fully capture lived conditions. Participatory validation highlights classification blind spots that are not uniformly distributed across deprivation levels. Rather than invalidating spatial accessibility modelling, these results demonstrate the importance of integrating community validation to expose bias in deprivation classification.

Building on the baseline model and participatory validation presented in this short paper, subsequent sensitivity analyses using validation-informed threshold recalibration and participatory-informed service weights strengthen the modelling framework, and the findings will be presented in the full paper.

Declaration of Generative AI in writing

The authors declare that Generative AI tools were used solely for language editing and grammar improvement, not for scientific content, research data, or substantive conclusions.

References

Bello, N. I., Bello, A., Abubakar, A. S., and Abdullahi, I. K.: Water consumption determinants in Rimin Kebe area, Ungogo Local Government, Kano State, Nigeria, *International Journal of Research and Innovation in Social Science*, IV, 652–655, 2020.

Fagbohun, P. O., Shotunde, T. O., Oladiboye, E. O., and Ajaegbo, G. A.: Water consumption capacity and the level of accessibility among the residents of Agege-Lagos, *International Journal of Innovative Science and Research Technology*, 7, 1781–1789, 2022.

Fritz, S., See, L., Carlson, T., Haklay, M., Oliver, J. L., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L. A., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., and West, S.: Citizen science and the United Nations Sustainable Development Goals, *Nature Sustainability*, 2, 922–930, 2019, https://doi.org/10.1038/s41893-019-0390-3

Hansen, W. G.: How accessibility shapes land use, *Journal of the American Institute of Planners*, 25, 73–76, 1959, https://doi.org/10.1080/01944365908978307

Levine, R.V., Norenzayan, A.: The pace of life in 31 countries. *Journal of Cross-Cultural Psychology* 30(2), 178–205(1999), https://doi.org/10.1177/0022022199030002003

Luo, W. and Qi, Y.: An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians, *Health & Place*, 15, 1100–1107, 2009, https://doi.org/10.1016/j.healthplace.2009.06.002

Mazumdar, S., Wrigley, S., & Ciravegna, F. (2017). Citizen Science and Crowdsourcing for Earth Observations: An Analysis of Stakeholder Opinions on the Present and Future. *Remote Sensing*, *9*(1), 87. https://doi.org/10.3390/rs9010087

Olusegun, S. A., David, R., Fernandez-Marquez, J. L., Grey, F., Treytl, V., Bärnthaler, M., Fraisl, D., Santolini, M., & Kokshagina, O. (2026). DonateWater Dataset: Youth-Driven Citizen Science on Water Access in Nigeria - Combined Dataset (Waves 1–3, 2023–2024) (2.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.19184253

Penchansky, R. and Thomas, J. W.: The concept of access: definition and relationship to consumer satisfaction, *Medical Care*, 19, 127–140, 1981.

World Health Organization (WHO): *Domestic water quantity, service level and health*, WHO Press, Geneva, 2020.

World Health Organization (WHO) and United Nations Children's Fund (UNICEF): *Progress on household drinking water, sanitation and hygiene 2000–2022: Special focus on gender*, WHO/UNICEF Joint Monitoring Programme, Geneva, 2023.

WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00660