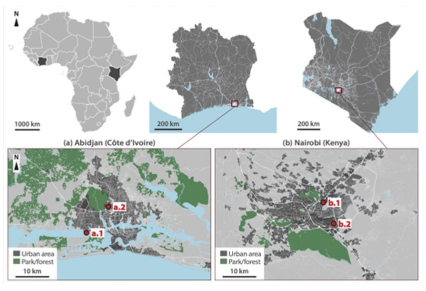
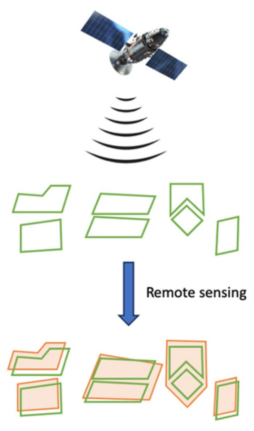
Algorithmic performance of remote sensing in informal settlements in Africa

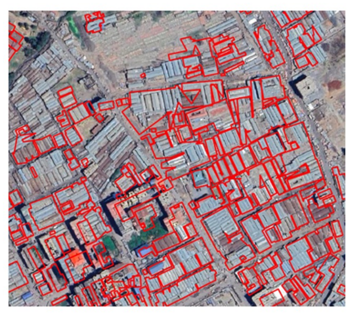
Jordane Provin, Science and environmental engineering.
Supervisor: Vitor Pessoa. Laboratory manager: Jérôme Chenal.

INTRODUCTION:

The rapid demographic growth in sub-Saharan Africa has led to a surge in population, but the supply of affordable housing falls short of demand, resulting in the emergence of informal settlements within urban areas. These settlements, characterized by residents lacking legal claims to the land they occupy per the United Nations definition [1], pose a unique challenge for existing geospatial algorithms. Most algorithms are trained on data from well-planned, regular neighborhoods in Western countries, making them less accurate when applied to the irregular structures of informal settlements common in Africa and Asia [2]. While studies have highlighted the imprecision of these algorithms [3, 4, 5], the lack of verified field data in such areas limits thorough evaluation. Field studies in cities like Abidjan and Nairobi aim to test the algorithms' ability to represent housing characteristics and morphology accurately in informal zones. Additionally, the research underscores the importance of considering local-scale imprecision in algorithm applications, particularly for planning and modernization efforts in informal settlements, which can have significant social impacts.

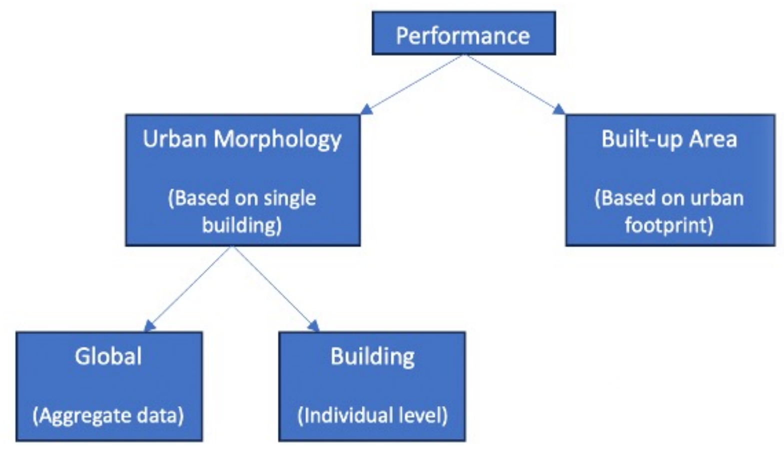






METHOD:

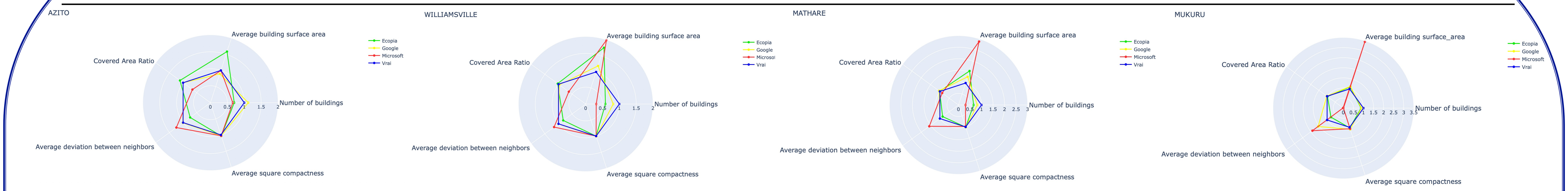
Regarding the data and methods employed, we used cartographic information describing the footprints of buildings in each site. These data were extracted from very high-resolution satellite images (approximately 50 cm per pixel) from various providers, such as Ecopia (in 2022), Microsoft (2014-2020, the neighborhood already existed during this period), and Google (in May 2023, the latest version). To ensure data accuracy, we compared the results with manually corrected building footprint maps following field verifications, using the Geographic Information System (GIS) software QGIS.



Settlement morphology (entropy & density)		Global detection performance
Global	Building	
Number of buildings	Number of neighbours within 100 m	Percentage of false positive (false detection)
Average building surface area	Building surface	Percentage of false negative (building parts not detected)
Settlement's Covered Area Ratio (CAR)	Covered Area Ratio (CAR) for each building	False positive / output / building / False negative
Average deviation between neighbours	Mean & Maximal deviation	
	Building orientation	

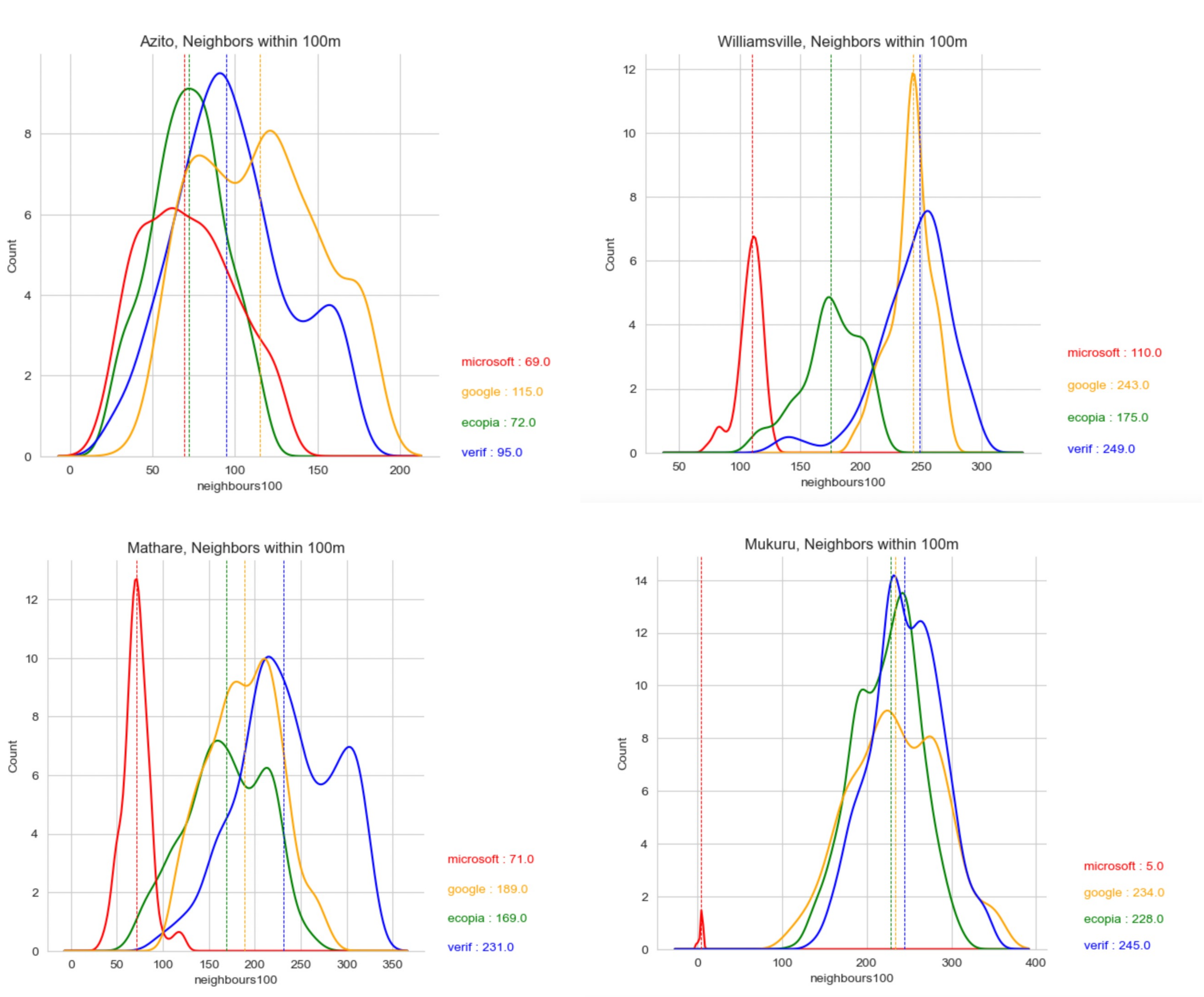
RESULTS & ANALYSIS:

Global analysis:



We observe that the number of detected buildings tends to remain lower than reality, confirming the simplification of shapes by the algorithms. However, the average building area often approaches or exceeds reality, indicating a strong tendency to group multiple distinct buildings into one. Overall, Microsoft's algorithm is the furthest from reality, in addition to the general tendency of algorithms to "structure" informal settlements.

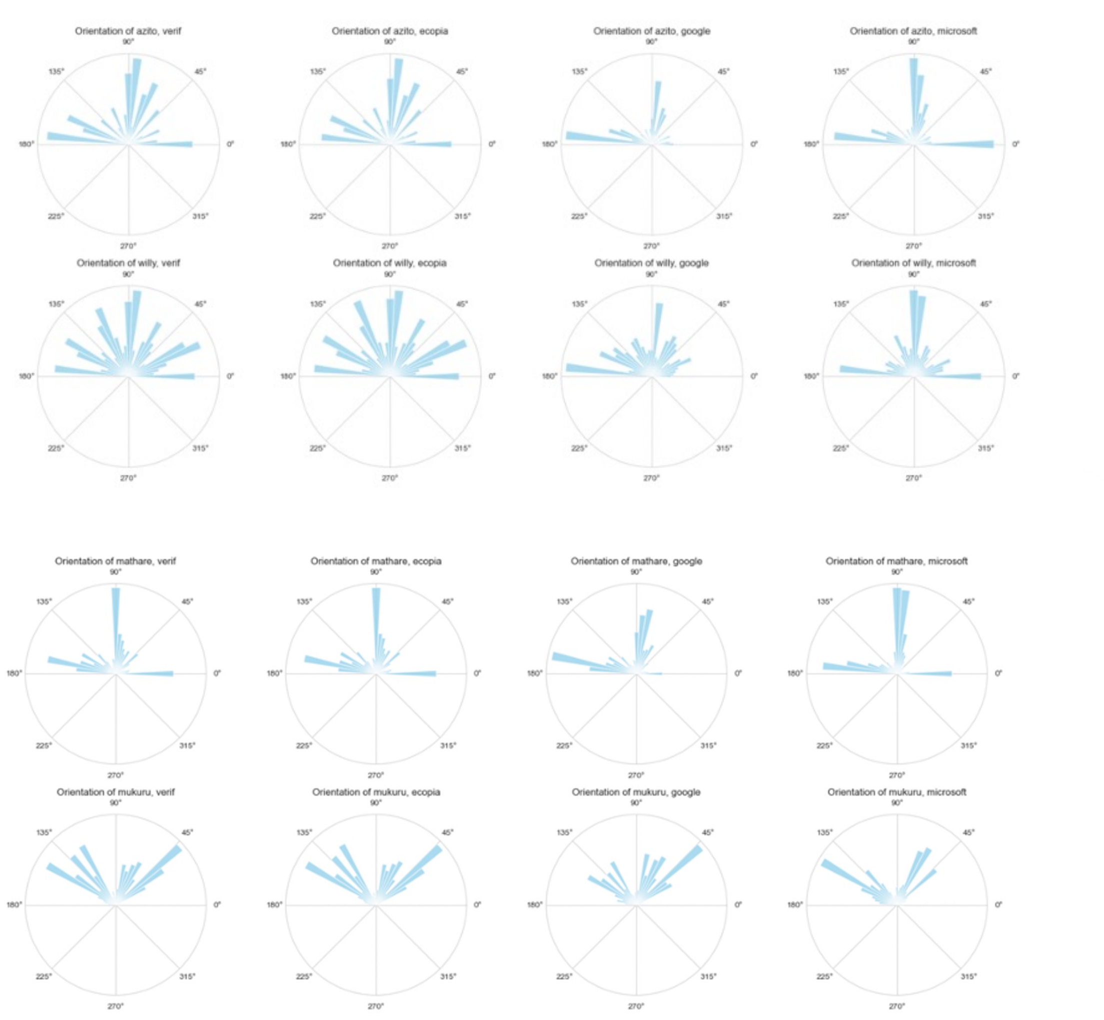
Local analysis:



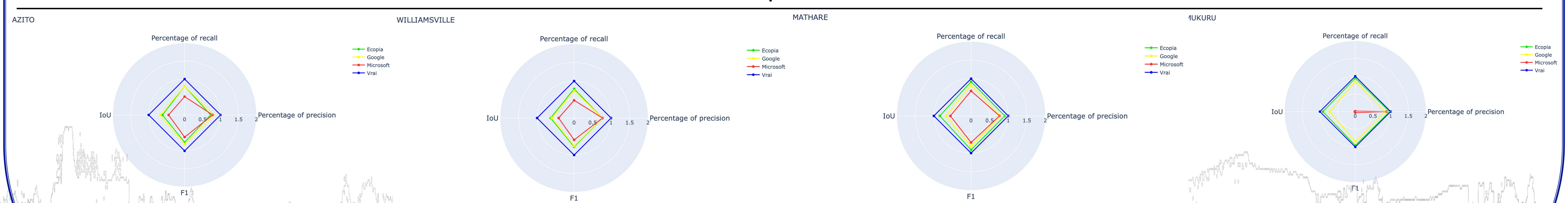
Number of neighbors: The study aimed to assess algorithmic accuracy in estimating the number of neighbors in informal settlements, hypothesizing that algorithms would underestimate this due to spatial simplification. Analysis of the results confirmed this hypothesis, revealing a general underestimation of the average number of neighbors within a 100-meter radius for all algorithms, with Google displaying the closest approximation to reality across sites.

Building Area: The examination of building area at a local level also confirms our hypothesis. We anticipated that algorithms would identify a higher number of larger buildings, while the actual scenario would include more smaller buildings, due to the spatial simplification tendency. This is supported by statistical analysis, indicating that the algorithms' calculated average building area is significantly larger than true values. In fact, the median area values indicate an average ranging from 30% to 250% of the real building areas.

Orientation: Algorithms, being trained in remote sensing from satellite images, tend to create a consistent orientation of buildings, thus approaching planned neighborhoods. This is reflected in the graphs, indeed, they reveal a greater variety of orientations in reality, while algorithms simplify this diversity. Microsoft's algorithm deviates the most from reality in this regard, while Google's algorithm is the closest.



Detection performance:



General performances in accuracy and detection of algorithms are summarized in these polar graphs. Overall, Ecopia is the algorithm that comes closest to reality in terms of precision and correct detected buildings. Google follows, and Microsoft is the farthest from reality. Moreover, in terms of morphological accuracy, Ecopia is the algorithm closest to reality, with an accuracy of 80 to 95%, while Google has an average accuracy of around 75%, and Microsoft only 60%, which is relatively low. Therefore, although Google offers a better overall morphological analysis, its inaccuracy is more pronounced at the building level (local).

CONCLUSION:

In conclusion, the use of geospatial data must be done with caution. If the data is used to obtain information about the number of buildings, households, and occupied space in neighborhoods, algorithms like Google and Ecopia come close to reality. However, when it comes to urban planning, such as creating public roads or demolishing informal housing, at the local level, this inaccuracy becomes significant and has significant economic and social impacts.

Finally, the hypothesis based on the idea that remote sensing algorithms trained on structured and planned cities in developed countries tend to simplify the morphology of buildings in informal neighborhoods. In addition, the layout to detecting buildings where there are none and vice versa, was confirmed by the results obtained as we have seen. There is therefore a real lack of precision in the maps representing the buildings in these areas. Any use of these data must be made with caution as regards the lacks precision at the local scale, which can have important social impact and consequences for the planning and renovation of these areas.

REFERENCES:

1- Glossary of the habitat 3, United Nations, 2017.
2- Taubenböck H, Graff NJ, Wurm M. The dynamics of poor urban areas – analyzing morphologic transformations across the globe using Earth observation data, 2020.
3- Lilford, Because space matters: conceptual framework to help distinguish slum from non-slum areas, 2019.
4- Kholi D, , Sliuzas R, Stein A. Uncertainty for image interpretations of urban slums, 2016.
5- Rizwan Ahmed Ansari, Krishna Mohan B. Textural segmentation of remotely sensed images using multiresolution analysis for slum area identification, 2019.
6- Photo, Mukuru kwa Njenga, Nairobi. Geoffrey Mboya.