SPATIALLY-CONSTRAINED CLUSTERING OF GEOSPATIAL FEATURES FOR HEAT VULNERABILITY ASSESSMENT OF FAVELAS IN RIO DE JANEIRO

Baptiste Clemence, Thomas Hallopeau, Laurent Demagistri & Joris Guerin
UMR Espace-Dev
French National Research Institute for Sustainable Development
Montpellier, 500 rue Jean-François Breton - 34393, France
{baptiste.clemence,thomas.hallopeau,laurent.demagistri,joris.guerin}@ird.fr

ABSTRACT

Informal settlements face disproportionate exposure to climate-related health hazards. However, existing methodologies lack systematic approaches to link diverse settlement characteristics with environmental health outcomes. We develop a data-driven framework to assess heat vulnerability in Rio de Janeiro's favelas by combining spatially-constrained clustering with land surface temperature (LST) analysis. Using remote sensing and geospatial features, we identify two distinct favela typologies: recent, well-connected settlements on flat terrain (Cluster 0) and historical, poorly-connected communities on vegetated slopes (Cluster 1). Analysis of 16 extreme heat events reveals systematic temperature differences of 2-3°C between clusters, with flat-terrain favelas experiencing significantly higher heat exposure. Our findings demonstrate that settlement morphology critically influences heat vulnerability, providing a replicable framework for targeted urban planning and public health interventions in informal settlements globally.

1 INTRODUCTION

In 2022, the global urban population exceeded 56% and is projected to reach 68% by 2050 (United Nations, 2024). Informal settlements, particularly slums, house over one billion people (United Nations, 2025). They develop outside regulatory planning frameworks, often in response to rapid demographic growth, land pressure, and socio-spatial exclusion. They are characterized by high building density, precarious housing, limited access to public services, poor construction quality, increased exposure to natural hazards, and a lack of green spaces (Kuffer et al., 2016). For all those reasons, they present major public health challenges (Corburn and Sverdlik, 2019).

Informal settlements, called favelas in Brazil (Abreu, 1987), exhibit significant heterogeneity in their physical characteristics, leading to different responses to health stressors such as pandemics or extreme heat events (Mendonça and Steinmetz, 2017; Kaiser et al., 2025). Understanding which settlement characteristics influence health risk exposure is essential for targeted interventions and urban planning. However, existing methodologies lack a systematic, multivariate approach to link the diverse informal settlements geospatial features with health-related environmental outcomes.

This research addresses this gap by developing a data-driven methodology to study heat vulnerability of informal settlements in Rio de Janeiro, Brazil. This is a valuable case study given the city's substantial informal settlement population and recurring public health challenges during heat waves (de Faria Peres et al., 2018). Our approach establishes a typology of Rio's favelas through spatially-constrained clustering of remote sensing and geospatial features, including textural descriptors (De Mattos et al., 2021; Graesser et al., 2012), spectral indices, and OSM road network data (Hallopeau et al., 2025; Owen, 2013). We then analyze how the derived settlement typology relates to land surface temperature (LST) variability during extreme heat events, providing a framework for identifying heat-vulnerable favela types.

Table 1: Description of the variables used for the favela typology

Variable	Description	Source
NDVI	Normalized Difference Vegetation Index.	S2 (16/08/2022)
Entropy	Measure of disorder or textural diversity	S2 (16/08/2022)
Slope	Average slope, in degrees.	DEM GLO-30 ESA 2019
Convexity	Representing terrain curvature	DEM GLO-30 ESA 2019
Nodes	Number of nodes in the road network.	OSM 2025
Road length	Total length of road segments, in meters.	OSM 2025
Mean connections	Average connectivity of road nodes	OSM 2025
Min connections	Min connectivity value of nodes in the cell.	OSM 2025
Max connections	Max connectivity value of nodes in the cell.	OSM 2025

2 Typology of Favelas

2.1 Data

We use the reference data of favela boundaries from the Pereira Passos Institute (IPP, year 2022). They cover the entire metropolitan area of Rio de Janeiro. A spatial grid with a resolution of 150 meters is generated across the entire area covered by our study, following the method proposed by Hallopeau et al. (2025), inspired by Owen (2013). The grid can be visualized in Figure 1.

For each grid cell, we compute several descriptors from remote sensing and geospatial data, described in Table 1. In a previous study, Hallopeau et al. (2025) demonstrated the effectiveness of using these features for a *favela vs. residential* cell classification task, which supports that they are relevant descriptors to establish our typology. Following Hallopeau et al. (2025) methodology, we retain only grid cells with at least 70% favela coverage to ensure the analysis focuses on areas where favelas are dominant.

As the resolution of the favela boundaries is sometimes very precise in the IPP polygons, some very close favelas are referenced as different settlements in the IPP database. We solve this issue by merging polygons less than 20 metres apart, which result in the creation of favela complexes with a single identifier.

2.2 Method

To establish a typology of the favelas, we aim to apply a clustering algorithm to separate the cells into unified groups of settlements presenting similar characteristics. In order to guarantee that cells belonging to the same favela are allocated to the same cluster, we must apply a constrained clustering algorithm, allowing us to incorporate must-link constraints to certain pairs of cells. To do this, we use the COP-KMeans algorithm, a variant of the classic K-Means unsupervised classification algorithm that incorporates constraints into the clustering process (Wagstaff et al., 2001). This method is particularly suitable when prior knowledge is available and the goal is to guide the automatic grouping process, while allowing part of the data to structure itself based on similarity. We use a Must-Link constraint based on the favela unique identifiers to group cells belonging to the same favela or complexes of favelas in order to try and preserve the continuity of the urban fabric.

Before training the model, NDVI, slope, and convexity are spatially aggregated to a 150-meter grid using the mean. Shannon entropy and OSM variables are computed at the same 150 m resolution. All variables are then normalized. We selected the number of clusters (k=2) following the elbow technique and silhouette coefficient analysis (Hastie et al., 2009; Rousseeuw, 1987) (Figure A.1).

2.3 Results

Figure 1 shows the spatial distribution of the two identified clusters. Cluster 1 is mainly concentrated in the city center (eastern part of the city), while Cluster 0 extends towards the west. This spatial division reflects contrasting urban dynamics linked to the history and topography of Rio de Janeiro (Abreu, 1987).

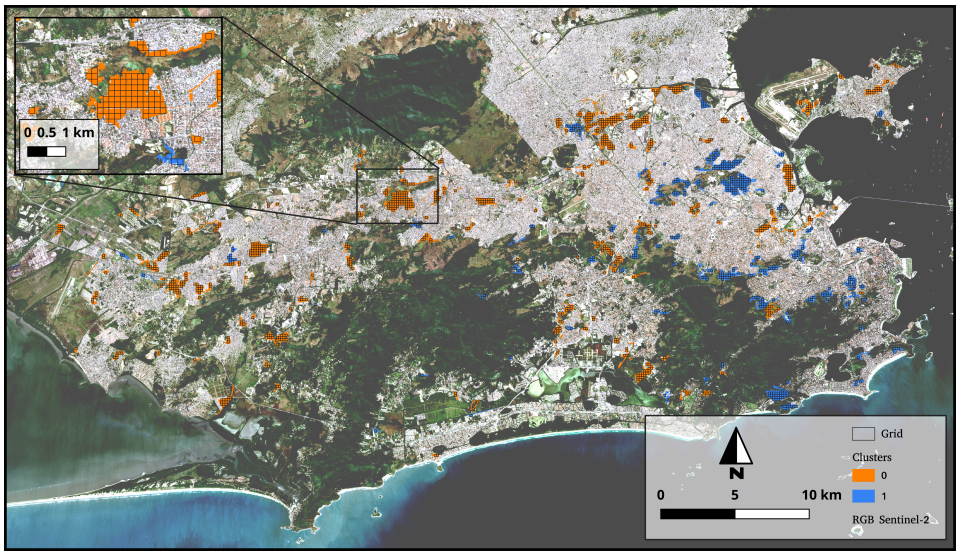

Figure 1: Spatial distribution of favela clusters identified by COP-KMeans

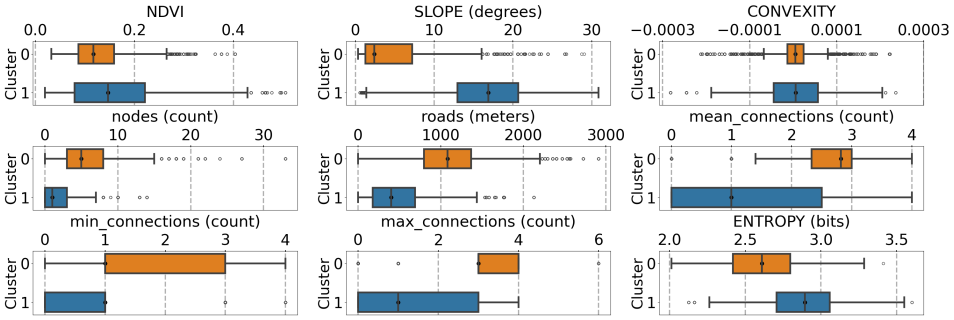

Figure 2: Distribution of explanatory variables by cluster before normalization

Figure 2 highlights the differences in explanatory variable values between the two clusters. The key differentiating variables are: (1) The slope, with a median slope of 17° in Cluster 1 and only 2° in Cluster 0. (2) The road network, which is much less dense in Cluster 1, with a median road length of 400 meters per cell, compared to 1100 meters in Cluster 0. The number of road nodes is also lower in Cluster 1, with a median of 1, compared to 5 in Cluster 0, indicating a less interconnected network. Other variables present similar distributions and are not differentiating factors.

3 Heat Vulnerability Analysis

3.1 Data and Methods

We extract LST data directly from Landsat 8 and 9, Collection 2, Level-2 products, specifically using the thermal band `ST_B10`. From an initial dataset of 40 images acquired during warm seasons between 2014 and 2025, only the 16 dates where the median LST of the entire image was ≥ 40°C under clear-sky conditions were retained. This selection criterion ensures that the analysis focuses on the hottest days, which are likely to correspond to heat-wave events. We then compare the distribution of LST between each clusters, for each of the 16 dates.

3.2 Results

The comparison of clusters' responses during extreme heat events is presented in Figure 3. The box plots show a difference in LST between clusters, regardless of the selected date. LST values of favelas are 2 to 3°C higher in Cluster 0 compared to Cluster 1. While the temperature distributions overlap, 75% of temperature pixels in Cluster 1 remain cooler than the median of Cluster 0.

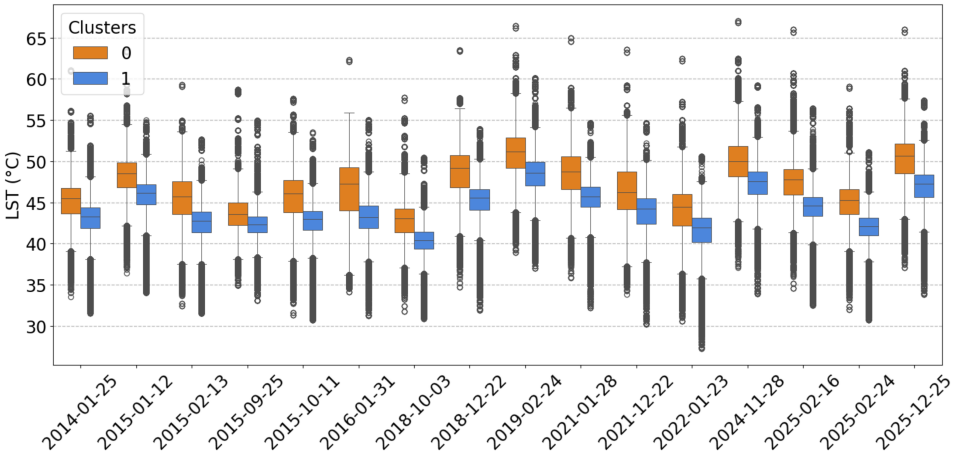

Figure 3: Comparison of LST distribution between clusters during extreme heat events.

4 Discussion

This study reveals two distinct profiles of favelas :

- Cluster 0 includes relatively recent favelas located on flat terrain. These areas are more organized, with a denser road network. Their urban structure, characterized by a high presence of artificial surfaces (buildings, roads) and bare soil, promotes increased absorption of solar radiation. This results in higher LSTs compared to the other cluster, as urban materials store and re-emit more heat than vegetated surfaces (Li et al., 2020).
- Cluster 1 consists of the oldest favelas, historically established on vegetated slopes. These areas exhibit a slightly higher NDVI, indicating a greater presence of vegetation. LSTs are systematically cooler in this cluster, across all selected dates. This finding aligns with Alves and Lopes (2017), who emphasized that vegetated slopes benefit from the cooling effect of evapotranspiration and shading, thereby reducing heat absorption.

5 Limitations and Future Work

While our binary typology effectively captures the primary distinction between historical hillside settlements and recent flat-terrain favelas, we acknowledge several limitations that open avenues for future research.

Typology refinement. The choice of k=2 clusters, though supported by silhouette and elbow analyses, may oversimplify the heterogeneity within Rio's informal settlements. Future work could incorporate complementary geospatial features such as building height, building density per cell, or roof material characteristics to refine the typology.

Spatial and temporal resolution. The 150-meter resolution may not capture fine-scale intra-favela variations. On the other hand, our daytime LST focus might sometimes reflect material heat storage rather than true air temperature. Higher-resolution data and nighttime Landsat acquisitions would better characterize urban heat island intensity across typologies.

Integration with health outcomes. Linking our favela typology with epidemiological data, such as mortality during the COVID-19 pandemic (Gracie et al., 2021) or excess deaths rates during heat waves (Monteiro dos Santos et al., 2024), would provide direct evidence of the health implications of different settlement types. Such integration would strengthen the policy relevance of our approach and enable targeted health protection strategies.

Despite these limitations, our methodology provides a replicable and scalable framework for linking settlement typology to urban environment health outcomes, with potential applications to other urban contexts and health-related vulnerabilities in informal settlements worldwide.

REFERENCES

United Nations, "The sustainable development goals report," 2024. [Online]. Available: https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024.pdf

United Nations, "World urbanization prospects: The 2025 revision," 2025. [Online]. Available: https://www.un.org/development/desa/pd/content/world-urbanization-prospects-2025-summary-results

M. Kuffer, K. Pfeffer, and R. Sliuzas, "Slums from space: 15 years of slum mapping using remote sensing," *Remote Sensing*, vol. 8, no. 6, p. 455, 2016. [Online]. Available: http://dx.doi.org/10.3390/rs8060455

J. Corburn and A. Sverdlik, "Informal settlements and human health," in *Integrating Human Health Into Urban and Transport Planning: A Framework*, M. Nieuwenhuijsen and H. Khreis, Eds. Cham: Springer International Publishing, 2019, pp. 155–171. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-74983-9_9

M. d. A. Abreu, *Evolução urbana no Rio de Janeiro*. Rio de Janeiro: IPLANRIO/Zahar, 1987.

F. Mendonça and G. Steinmetz, "Le climat urbain et la santé: les changements climatiques et la dengue dans les villes brésiliennes," *Publications de l'AIC*, vol. 30, pp. 1–6, 2017. [Online]. Available: https://www.academia.edu/34378816/LE_CLIMAT_URBAIN_ET_LA_SANT%C3%89_LES_CHANGEMENTS_CLIMATIQUES_ET_LA_DENGUE_DANS_LES_VILLES_BR%C3%89SILIENNES

Z. Kaiser, A. Sakil, R. Baikady, A. Deb, and M. Hossain, "Building resilience in urban slums: Exploring urban poverty and policy responses amid crises," *Discover Global Society*, vol. 3, no. 1, p. 8, 2025. [Online]. Available: http://dx.doi.org/10.1007/s44282-025-00142-3

L. de Faria Peres, A. J. de Lucena, O. C. Rotunno Filho, and J. R. de Almeida França, "The urban heat island in rio de janeiro, brazil, in the last 30 years using remote sensing data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 64, pp. 104–116, 2018. [Online]. Available: https://doi.org/10.1016/j.jag.2017.08.012

A. C. De Mattos, G. McArdle, and M. Bertolotto, "Mapping slums with medium resolution satellite imagery: A comparative analysis of multi-spectral data and grey-level co-occurrence matrix techniques," *arXiv*, 2021. [Online]. Available: https://arxiv.org/pdf/2106.11395

J. Graesser, A. Cheriyadat, R. R. Vatsavai, V. Chandola, J. Long, and E. Bright, "Image based characterization of formal and informal neighborhoods in an urban landscape," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1164–1176, 2012. [Online]. Available: https://ieeexplore.ieee.org/document/6236225

T. Hallopeau, Y. Fouzai, L. Demagistri, J. Guérin, V. P. de Matos, R. Gracie, and N. Dessay, "Addressing data imbalance in urban informal settlement mapping from earth observation using ensemble learning: A case study in rio de janeiro," *Science of Remote Sensing*, p. 100273, 2025. [Online]. Available: https://hal.science/hal-05337270v1/file/1-s2.0-S2666017225000793-main.pdf

W. D. Owen, K.K., "An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics," *Applied Geography*, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.apgeog.2012.11.016

K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained k-means clustering with background knowledge," vol. 1, pp. 577–584, June 2001. [Online]. Available: https://web.cse.msu.edu/~cse802/notes/ConstrainedKmeans.pdf

T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY: Springer, 2009.

P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0377042787901257

Y. Li, S. Schubert, J. P. Kropp, and D. Rybski, "On the influence of density and morphology on the urban heat island intensity," *Nature Communications*, vol. 11, p. 2647, 2020. [Online]. Available: https://www.nature.com/articles/s41467-020-16461-9

E. D. L. Alves and A. Lopes, "The urban heat island effect and the role of vegetation to address the negative impacts of local climate changes in a small brazilian city," *Atmosphere*, vol. 8, no. 2, p. 18, 2017. [Online]. Available: https://www.mdpi.com/2073-4433/8/2/18

R. Gracie, A. S. d. A. Silva, C. Bigler, G. Douglass-Jaimes, E. M. Campos, and T. Williamson, "Painel Unificador Covid-19 nas Favelas: metodologia para dar visibilidade a territórios periféricos," in *Covid-19 no Brasil: cenários epidemiológicos e vigilância em saúde - Série Informação para ação na Covid-19*, E. Fiocruz, Ed. Scielo Books, 2021, pp. 251–270. [Online]. Available: https://doi.org/10.7476/9786557081211.0016

D. Monteiro dos Santos, R. Libonati, B. N. Garcia, J. L. Geirinhas, B. B. Salvi, E. Lima e Silva, J. A. Rodrigues, L. F. Peres, A. Russo, R. Gracie, H. Gurgel, and R. M. Trigo, "Twenty-first-century demographic and social inequalities of heat-related deaths in brazilian urban areas," *PLOS ONE*, vol. 19, no. 1, pp. 1–30, 01 2024. [Online]. Available: https://doi.org/10.1371/journal.pone.0295766

A Optimal number of clusters

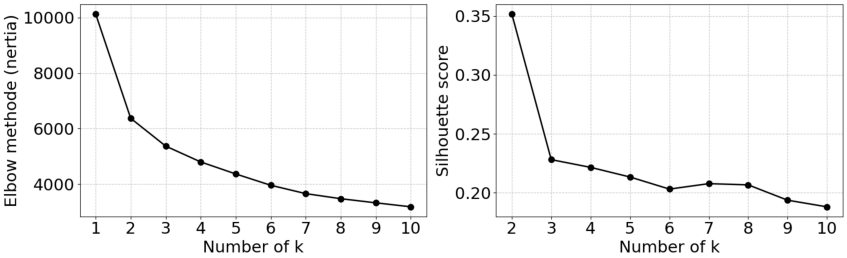

Figure A.1: Elbow method and Silhouette score

The selection of the number of clusters is performed with elbow method and silhouette analysis Figure A.1. The decrease in inertia in the elbow method suggests 2 to 3 clusters. The absence of a local maximum on the silhouette curve seems to indicate uniformly distributed data, data with little variance or the need to explore other metrics. We choose a 2-cluster partition in this study as a first approach, which already allows us to identify a typology.