

Meeting User Requirements for Mapping and Characterizing Deprived Urban Areas in Support of Pro-Poor Policies

Monika Kuffer¹, Sabine Vanhuyse², Stefanos Georganos² and Jon Wang¹

¹University of Twente, ITC, The Netherlands

²Université libre de Bruxelles, Belgium

Abstract

Spatial data on Low-and-Middle-Income-Country (LMIC) cities, and deprived areas within cities, are often not readily available in support of local and global information needs. To address this information gap, we propose the systematic semi-automated SLUMAP framework that provides policy-relevant information on deprived urban areas in Sub-Saharan Africa (SSA), based on free open-source software (FOSS). First, we assess user needs for spatial information on deprivation (ranging from local communities to global research and policy support). Second, we show how free or low-cost image datasets can be used for mapping the location of deprived areas at the city scale and providing an overall assessment of their spatial patterns. This is implemented as a grid-based approach using machine learning and assessing the contribution of a large number of spectral and spatial features derived from open or low-cost imagery. Third, we show how higher (spatial and spectral) resolution data can provide a detailed characterization of such areas, with a GEOBIA/machine-learning workflow and deep learning techniques. We illustrate the experiments and results on the city of Nairobi (Kenya) and discuss transferability to SSA cities.

Keywords: slum, earth observation, sustainability, spatial inequalities, machine learning

1 Introduction

Urbanization rates are rising in most Low-and-Middle-Income Countries (LMICs) (UN, 2019). Most of this increase is happening in areas commonly known as slums, informal settlements and areas of inadequate housing, hereafter “deprived urban areas”. In particular, African cities are rapidly growing, while there is an insufficient provision of low-income serviced housing. The urban Sustainable Development Goal 11 (SDG 11) has the “proportion of urban population living in slums, informal settlements or inadequate housing” as its first indicator to measure progress towards sustainability. Unfortunately, data for this indicator is commonly not readily available for supporting local or global monitoring. Existing datasets supporting the SDG 11.1.1 indicator are country-level estimates without a reference to individual cities. Thus, existing data failed to provide insights into the spatial patterns of deprived urban areas and their dynamics within cities. Earth Observation (EO) data has, in principle, the capability to map deprived urban areas (e.g., Wang, Kuffer, Roy, & Pfeffer, 2019), as data archives are

growing and computational power is increasing. However, EO methods fall short in providing scalable, transferable, and low-cost solutions that respond to user needs (Kuffer et al., 2020). Therefore, to date, EO has not been used as an effective tool to provide relevant information to various users on urban development issues, specifically for monitoring deprived urban areas and accounting for the local SDG 11.1.1 indicator. To make appropriate use of the growing amount of EO data and advancements in methods, it is essential first to understand user needs. However, there is a general communication gap between the EO experts and potential users of EO data, hereafter “users”. We observe that the EO community is developing methods that are mostly based on very high resolution (VHR) commercial EO data, often for a small subsection of a city. At the same time, urban development questions typically require an understanding of patterns at the city or regional scale. In addition, resource constraints and understanding of advanced methods is hindering the knowledge transfer from research to users. The paper aims to provide an overview of spatial information needs in deprived urban areas and develop solutions for meeting these needs. Examples from Nairobi (Fig. 1) illustrate recent developments in machine learning and FOSS solutions for developing a systematic semi-automated SLUMAP framework that provides policy-relevant information on deprived urban areas.



Figure 1:
A deprived area in Nairobi
(Photo: Ángela Abascal Imízcoz).

2 Methodology

2.1 Mapping user needs and requirements

The first step towards shortening the gap between existing inconsistent/unavailable datasets and essential geospatial resources in deprived urban areas is to develop an adequate understanding of user needs and requirements at (inter)national and local levels. This was done through the assessment of data requirements by way of an online survey, fortified by additional discussions and workshops that covered diverse users:

- An online survey of users utilizing deprived urban area-related spatial data (N = 112). The survey included different professions and sectors working with ‘slum’ related data, including civil society, government, international and research organizations.

The backgrounds of respondents were GIS, EO and data science (40%), urban planning (28%), social (20%), environmental (9%) and health (4%) professions.

- Interactions with community-based organizations in Nairobi and Lagos, in form of online meetings due to COVID restrictions.
- Workshops (at the World Urban Forum and hybrid local/online workshops in Lagos and Accra), expert discussions with local and national authorities (Kenya and Lagos).

The questions and discussion points included spatial, temporal, contextual information requirements of users as well as requirements about data access, aggregation, uncertainties and ethics/privacy. Two major aspects discussed in the user interactions that are key for routine and accurate production of maps of deprived urban areas at continental scale are highlighted (Fig. 2), namely (i) the need for a low-cost mapping system and (ii) the local data requirements (characterization) for the city and community-level data on deprived urban areas. To show how data can be produced that respond to the user needs, we use the case of Nairobi to explore the potential of several HR and VHR sensors (i.e., Sentinel-1/2, SPOT6/7, WorldView-3 and Google Earth (GE) images) for mapping and characterizing deprived urban areas. The city-scale mapping using Sentinel-1/2 data is responding to the user needs for a low-cost mapping system. This allows for developing a standardized and scalable mapping system and drastically increases scalability and repeatability (routine mapping). The local characterization explores the potential of VHR images to respond to the user needs on urban morphology, environmental/ health aspects (e.g., garbage piles) and automatizing building mapping (in support of local planning needs).

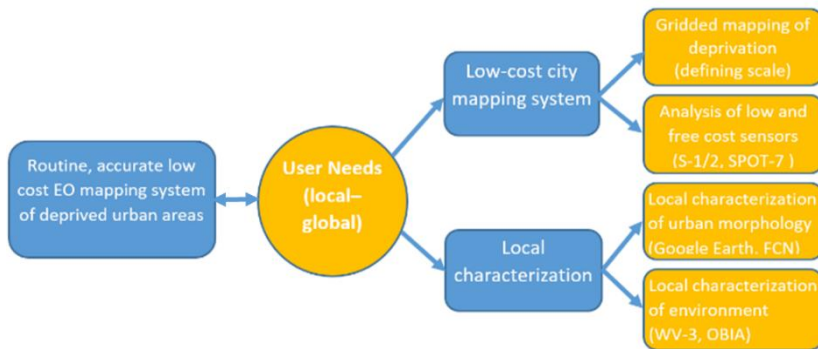


Figure 2: A semi-automated SLUMAP framework.

2.2 Using open vs. low-cost imagery at the city scale

To assess the potential of free-cost Sentinel-1/2 for mapping the morphological deprivation probability at the city scale, we develop a machine learning workflow using FOSS software GRASS GIS in a Jupyter Notebook and R. A grid-based approach is implemented. Gridded mapping has proved successful for mapping slums with VHR GE images (Duque et al., 2017). Besides, it tends to have a high transferability potential, as reflected by the increasing number of available global gridded layers (e.g., WorldPop, GHSL, GUF, etc.), and it responds to privacy concerns (e.g., ‘blurs’ the boundaries). We apply our workflow to an area of interest

covering the city of Nairobi (652 km²). We use Sentinel-1/2 and SPOT7, which is considered low-cost commercial imagery (Fig. 3) and compare the results. Ancillary open global datasets (i.e., SRTM, OSM and a preliminary version of the World Settlement Footprint 2019 – unpublished to date - which is an improvement of the World Settlement Footprint 2015 (Marconcini et al., 2020)) are also included in the experiments. First, a wide set of over 2000 spectral, spatial and ancillary features are extracted. For optical imagery, these features are mainly based on vegetation indices, water or moisture indices, built-up indices, image transforms, texture metrics (e.g., GLCM, Structural Feature Set) and a few metrics calculated on an unsupervised classification (such as the Mean Patch Size). For SAR they are mostly based on intensity, coherence, textures and filtered bands. Ancillary features include geomorphometric features, built-up and street density. Statistics are calculated in 50m x 50m grid cells, and feature selection (using the VSURF - Variable Selection Using Random Forest - algorithm (Genuer, Poggi, & Tuleau-Malot, 2015)) is implemented prior to random forest (RF) classification, for parsimony. The classification scheme includes 8 land-use/land-cover classes: (1) High to mid-density built area, (2) Low density built area, (3) Industry/large structures, (4) Paved ground/Bare ground, (5) Vegetation, (6) Water, (7) Deprived urban areas (typical), and (8) Deprived urban areas (atypical). For our focus classes (7 and 8), a detailed class description is provided in the textbox. For training and testing, 3962 manually labelled samples (i.e., grid cells) representing the dominant class are used. Several feature combinations are assessed, and their respective performances are compared based on accuracy metrics (i.e., precision, recall and F1 score).

Textbox:	Definition of deprived classes:
(7)	Very compact arrangement of low-rise buildings, generally forming 'organic' patterns. No structured street layout, except for a few main streets. Little or no vegetation.
(8)	Arrangement of buildings with a density that varies from compact to mid-dense, and a pattern that is more regular than in class 7.

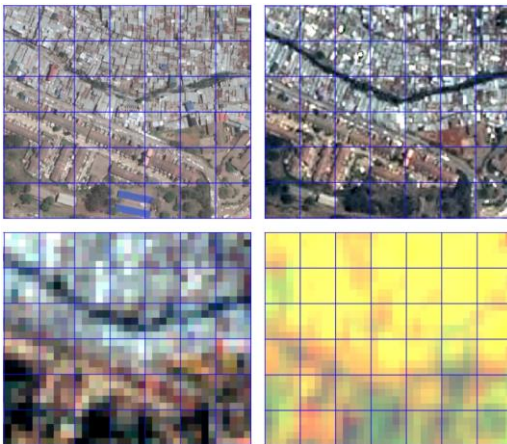


Figure 3:

Interface between deprived and non-deprived urban areas. Top left: GE imagery. Top right: SPOT7 (RGB). Bottom left: S2 (RGB). Bottom right: S1 intensity (VV, VH, VV/VH).

2.3 Local characterization of deprived areas

Next, we investigate the characterization of intra-deprived areas environments (i.e., garbage piles, built-up morphology). First, we make use of VHR superspectral data collected by the WorldView-3 satellite (8 multispectral and 8 SWIR bands) to map the urban environment in

deprived urban areas. State-of-the-art machine learning classifiers and processing methods such as Geographic Object-Based Image Analysis are deployed (Georganos, Grippa, Lennert, et al., 2018; Georganos, Grippa, Vanhuyse, Lennert, Shimoni, & Wolff, 2018). Moreover, we assess the created land cover/ land use (LULC) models for deprived areas. The assessment focuses on maximizing their interpretability and transferability and alleviate the data management and processing burden (Georganos, Grippa, Vanhuyse, Lennert, Shimoni, Kalogirou, et al., 2018). For example, this included defining a suitable grid size that reflects the urban patterns of deprived areas but still allows data aggregation to ease processing. We extract various indicators at a grid level (i.e., 25 meters) derived from the modelled LULC of these regions. For instance, these indicators may be pertinent to open space, building density, or the proportion of garbage piles. The training data on garbage piles were collected in collaboration with local community-based groups in Mathare (a deprived area in Nairobi), a key environmental issue that emerged in interaction with communities. A similar effort is presently ongoing in other communities. Second, we extract building footprints and map the morphological patterns using GE imagery. These morphological patterns allow differentiating deprived urban areas from better-off areas at the city scale. To achieve this, we largely rely on open tools and free data. There are two major steps in this approach: (1) extracting building footprints from GE imagery by using deep learning techniques (modified U-Net architecture) using a global training dataset provided by Wuhan University that containing labelled building footprint (gpcv.whu.edu.cn/data/building_dataset.html), and (2) measuring the morphological configuration of buildings with the open-source tool MOMEPEY (<http://docs.momepy.org/en/stable/>).

3 Results

3.1 User requirements for evidence-based policy-making

The results of the user need assessment (Fig. 4) shows that data on deprived urban areas are not available or accessible for users. In workshops, it was stressed that data, if at all available, are often not usable (e.g., not covering the area of interest) or are too dated. Most data needs relate to routine and up-to-date information about the location of deprived urban areas, building information and more detailed characterisation of their environment.

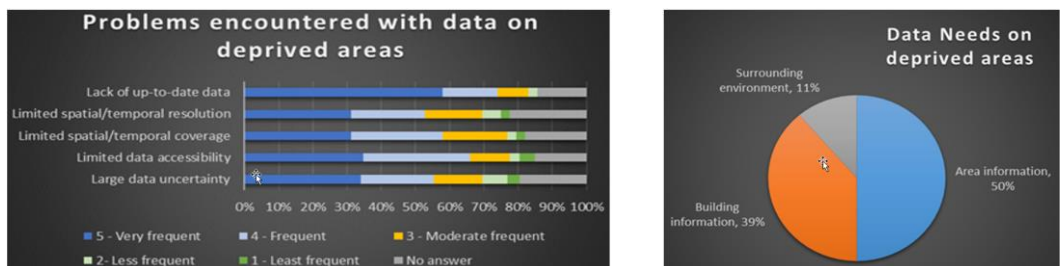


Figure 4: Existing data gaps and data needs on deprived urban areas: assessment of problems encountered by users (left) and user's data needs on deprived areas (right).

3.2 City-scale results

For each of the assessed feature combinations, the original set can be drastically reduced to a small number of important predictors with VSURF. The results of the random forest classifications are validated with an independent test set, focusing on the two deprived urban area classes. The best combination of SPOT7 and ancillary predictors achieves higher accuracy than the best combination of Sentinel1/2 and ancillary predictors (Tab. 1). However, the difference is not as marked as could be expected given the spatial resolution gap. The morphological deprivation probability is computed by summing the class probability of classes 7 and 8 (Fig. 5). Considering that Sentinel images are free datasets with wide temporal availability, they constitute a valuable option for mapping the morphological deprivation probability at the city scale, allowing for frequent updates, as required by users.

Table 1: Accuracy assessment of the best feature combinations involving Sentinel-1 (S1), Sentinel-2 (S2), SPOT7, and ancillary global datasets.

Class	Metric	S2	S2	SPOT7	SPOT7
		S1	S1	Ancillary	Ancillary
Class 7	Precision	0.94	0.96	0.86	0.94
	Recall	0.89	0.89	0.89	0.93
	F1	0.91	0.92	0.87	0.94
Class 8	Precision	0.79	0.84	0.79	0.88
	Recall	0.82	0.89	0.78	0.89
	F1	0.80	0.86	0.79	0.89

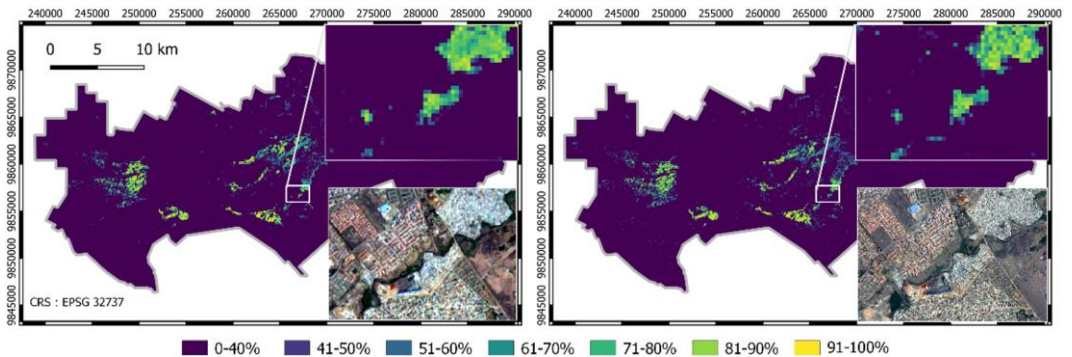


Figure 5: Morphological deprivation probability classes in 50x50m grid cells (Nairobi). Left: S1-S2-Ancillary (with S2 RGB subset). Right: SPOT7-Ancillary (with SPOT7 RGB subset).

3.3 Local characterization based on LULC

Taking as an example Mathare, Nairobi, Fig. 6 illustrates the potential of our modelled LULC for characterizing the local environment in a deprived area. Notably, garbage pile density (Fig. 6.a) is a very important socio-economic and health indicator as it can be associated with disease exposure, water/sanitation and act as a socio-economic proxy for the surrounding neighbourhoods (Engstrom, Hersh, & Newhouse, 2017). The lack of openness (Fig. 6.c) can also be detected. The detection of vehicles (Fig. 6.e) reflects socio-economic activity to a

degree. The RF Out of Bag Overall Accuracy of the map product for Mathare using all valuable WV-3 resources (multispectral + shortwave infrared) surpassed 87%. Finally, these indicators can be extracted in a gridded format (25 m), as illustrated in Fig. 6.g, which maps the spatial distribution of garbage piles across Mathare.

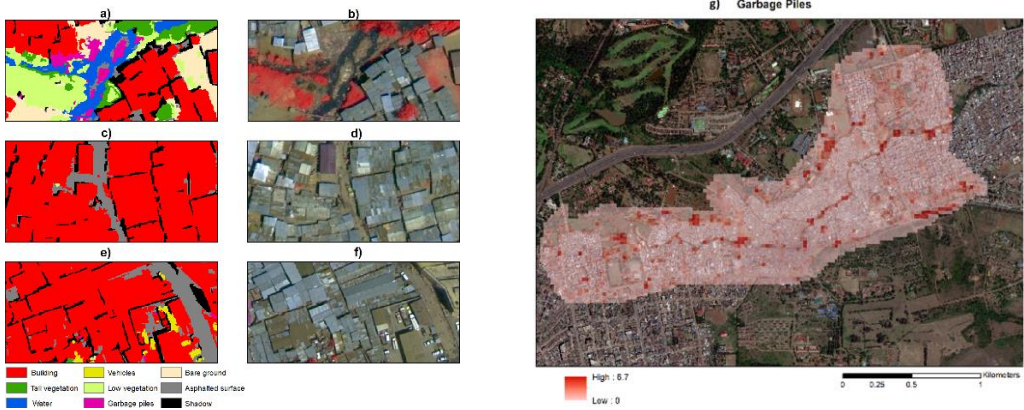


Figure 6: Subsets of mapped LULC in deprived areas in Mathare, Nairobi: a) garbage piles; c) lack of openness; e) detection of vehicles (b,d,f RGB); g) garbage piles density (%) at a 25 meter spatial resolution.

3.4 Local characterization based on building footprints extraction and urban morphology

Fig. 7 shows buildings extracted from different places within the city of Nairobi. Visually, the building configuration exhibits a significant difference especially comparing the building patterns in Fig. 7(b), where building patterns in deprived areas can be quite different from the other places shown in Fig. 7.a, c, d.

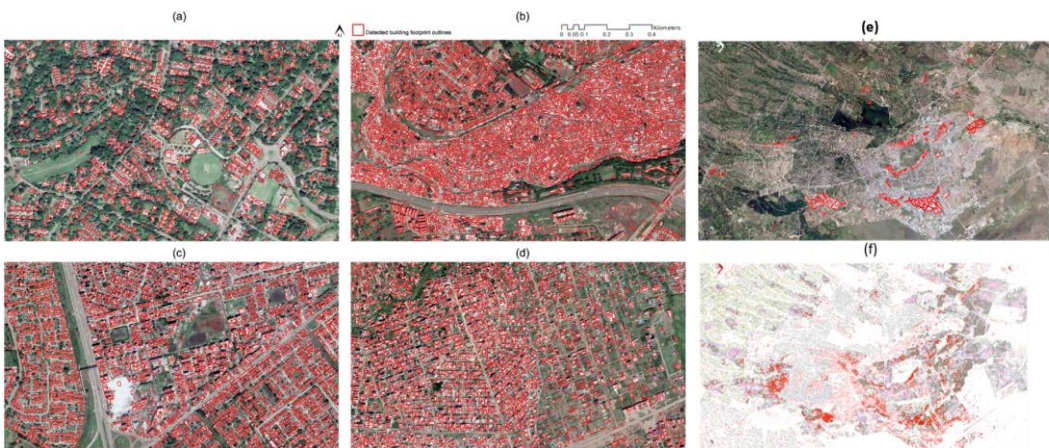


Figure 7: (a-d) Building footprints extracted from different neighbourhoods; (e) manually delineated deprived areas (dated); (f) building clusters based on building morphological metrics (Nairobi).

Once the morphological building patterns are explicitly measured, similar building patterns are classified within the same morphological clusters. The morphological cluster highlighted in red (Fig. 7.f) reflects the distribution of deprivation areas delineated in Fig. 7.e.

4 Conclusions

Deprived areas emerge with the rapid urbanization occurring in LMICs and the insufficient provision of low-cost urban housing. An increasing number of people migrate to cities, with complex drivers such as climate change. Global datasets do not account for these areas, and local data often ignore them. Our results show the capability of the SLUMAP framework that builds upon a FOSS solution to respond to user needs for routine and accurate mapping of deprived urban areas. To protect privacy, exact settlement boundaries are not shown, which could be used against communities (e.g., land tenure conflicts). A fine-scale local characterization makes use of commercial (WV-3) and freely available (GE) VHR data to meet the local needs for detailed environmental characterization, such as garbage piles mapping and morphological characterization of built-up density patterns at local and city scale. The SLUMAP framework is transferable to other SSA cities to provide data allowing for inter-city comparisons.

References

- Duque, J., Patino, J., Betancourt, A., 2017. Exploring the Potential of Machine Learning for Automatic Slum Identification from VHR Imagery. *Remote Sensing* 9, 895. <https://doi.org/10.3390/rs9090895>
- Engstrom, R., Hersh, J. S., & Newhouse, D. L. (2017). *Poverty from space : using high-resolution satellite imagery for estimating economic well-being*. Retrieved from Washington, D.C.: <http://documents.worldbank.org/curated/en/610771513691888412/Poverty-from-space-using-high-resolution-satellite-imagery-for-estimating-economic-well-being>
- Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2015). *V\$URF: An R Package for Variable Selection Using Random Forests* Retrieved from
- Georganos, S., Grippa, T., Lennert, M., Vanhuyse, S., Johnson, B. A., & Wolff, E. (2018). Scale Matters: Spatially Partitioned Unsupervised Segmentation Parameter Optimization for Large and Heterogeneous Satellite Images. *Remote Sensing*, 10(9), 1440.
- Georganos, S., Grippa, T., Vanhuyse, S., Lennert, M., Shimoni, M., Kalogirou, S., & Wolff, E. (2018). Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience & Remote Sensing*, 55(2), 221-242. doi:10.1080/15481603.2017.1408892
- Georganos, S., Grippa, T., Vanhuyse, S., Lennert, M., Shimoni, M., & Wolff, E. (2018). Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting. *IEEE Geoscience and Remote Sensing Letters*, 15(4), 607-611.
- Kuffer, M., Thomson, D. R., Boo, G., Mahabir, R., Grippa, T., Vanhuyse, S., . . . Kabaria, C. (2020). The Role of Earth Observation in an Integrated Deprived Area Mapping “System” for Low-to-Middle Income Countries. *Remote Sens.*, 12(6), 982.
- Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., . . . Strano, E. (2020). Outlining where humans live, the World Settlement Footprint 2015. *Scientific Data*, 7(1), 242. doi:10.1038/s41597-020-00580-5
- UN. (2019). *World Urbanization Prospects. The 2018 Revision*. New York, US: United Nations.
- Wang, J., Kuffer, M., Roy, D., & Pfeffer, K. (2019). Deprivation pockets through the lens of convolutional neural networks. *Remote Sens. Environ.*, 234, 111448.