DEVELOPMENT OF A MULTI-CITY DEPRIVED AREA MAPPING ECOSYSTEM

Ryan Engstrom[1], Dana Thomson[2], Julia Ek[1], and Monika Kuffer[3]

1. Department of Geography, George Washington University, Washington DC USA
2. Department of Social Statistics and Demography, University of Southampton, Southampton UK
3. Department of Urban and Regional Planning and Geo-Information Management, Faculty of Geo-Information Science and Earth Observation, University of Twente, Twente, Netherlands

ABSTRACT

The number of people living in deprived urban areas within low and middle income countries (LMICs) is large and predicted to continue to grow. Mapping these areas over time and space in a consistent manner is important for monitoring the Sustainable Development Goals (SDGs) and degree of deprivation between cities. This work describes the development of a mapping ecosystem to do this. The first steps are to define urban extents and collect the data needed for developing deprived area models. The goal of the mapping ecosystem is to produce maps of ranges in deprived areas that are comparable between cities and can be used by local governments and researchers to map within city variations in deprivation. By developing and describing this ecosystem, we hope to replicate this in other cities and countries so that this type of work can be expanded.

Index Terms— Deprived Areas, Sentinel 2, Contextual features, modeling

1. INTRODUCTION/BACKGROUND

The UN estimates that more than 1 billion people currently live in slum-like conditions worldwide but overwhelmingly in low- and middle-income countries (LMICs), and face inadequate housing, insecure tenure, unplanned housing, pollution, environmental risk, and/or social exclusion [1]. Future predictions from the UN indicate that the number of people living in these areas will continue to increase [2]. Data on deprived neighborhoods are generally difficult to find, and if they are available, mapping is often performed using one-off definitions that vary from city to city. This leads to difficulty in comparing the deprived area maps from one city to another, and to make general statements about changes across LMIC cities. Furthermore, the lack of consistency in these types of measurements makes it difficult to monitor and address areas of inadequate housing (Sustainable Development Goal (SDG)11.1.1)), and impacts the ability of governments, researchers, and development organizations to monitor progress over time and across space. The Integrated Deprived Area Mapping Systems (IDEAMAPS) Network, representing all major approaches to deprived area mapping, came together to define new methods and processes to tackle these issues [3].

IDEAMAPS has one year of seed funding from the United Kingdom Research and Innovation (UKRI) Fund to grow its network of researchers, practitioners, and decision-makers, and to pilot new integrated methods for mapping deprived areas in Accra (Ghana), Lagos (Nigeria), and Nairobi (Kenya). The long-term goals of this project are to create a deprived area mapping ecosystem that facilitates new types of data exchange among key stakeholders, distribute analysis-ready datasets of deprived areas and their characteristics, and new modelling approaches that will ensure repeatability of deprived area mapping methods over multiple cities, in multiple countries. The Network includes researchers and professionals who map deprived areas using currently silo-ed techniques including (1) aggregating survey or census data, (2) field mapping and observations, (3) manually digitizing satellite imagery, and (4) machine-learning modelling with earth observations (EO) [4]. The IDEAMAPS approach leverages strengths of each siloed approach, using data integration capacities and scalability of machine-learning models, the physical and environmental characteristics represented in EO data, the area-level social characteristics represented in census and survey data, and the accuracy and context-relevance of field observations and imagery digitized by local experts.

IDEAMAPS outputs show deprivation on a continuous scale in fine-scale raster format to support multiple use cases by key stakeholders (e.g., city governments, community-based organizations, development agencies). Use cases include identification of the most deprived area to prioritize for planning and upgrading, or monitoring of SDG targets. We eventually want to be able to model types of deprivation across cities to support use cases focused on, for example, poverty alleviation and equity policies, versus tenure regularization, versus infrastructure investments. The project also aims to support use cases that require different, context-specific sub-classifications for deprived areas that reflect separate issues of legality of tenure or land uses, socio-economic poverty, and/or environmental hazards. A "slum", "informal settlement", and "vulnerable area" might refer to specific distinct, but overlapping, areas in a given city. While we are keen to support a range of stakeholders to access spatially detailed, accurate information about area-level deprivation, we also recognize the severe risks to many deprived communities to be identified on maps (i.e., increasing their risk of eviction, fines, and harassment). It is therefore important to balance provision of open spatial data with strong, clear data protection practices. The work presented here is the development of the data and mapping ecosystem part of this project.

2. METHODS

2.1. Study Area

This project is focusing on the cities of Accra (Ghana), Lagos (Nigeria), and Nairobi (Kenya). These cities were chosen to pilot the IDEAMAPS methods and data ecosystem because they reflect a variety of deprived areas types (e.g., from "pocket slums" in Lagos, to high-density sprawling slums in Nairobi), are highly populated and growing rapidly, and are places where we had extensive experience and strong existing collaborations. For each city, the first step was defining the city extent. This was not a trivial task because there are many ways to determine where a city ends both politically and/or based on the built environment. Obtaining political boundaries is typically more straightforward as government bodies, such as the National Statistics Office, define clear local government jurisdictions by subnational administrative units. Defining the morphological or functional city boundary based on the built environment is more difficult as cities tend to grow and connect with other built areas over time.

To create IDEAMAPS urban extents, a combination of data sources from the Database of Global Administrative Areas (GADM), the Global Human Settlements Layer (GHSL) and WorldPop were used. In the first step, we selected the GADM boundary for the city using the highest-level administrative unit(s) which shared the city name. GADM provides sub national spatial data for each country's administrative areas, with generally two to five levels of subdivision. In the second step, we selected the urban center boundary defined in the GHSL Urban Centers database [5]. An urban center consists of contiguous grid cells of 1 sq km with a grid population density of at least 1,500 inhabitants per sq km (based on GHS-POP dataset) or a 50% or more grid built-up share of available land (based on GHS Layer-BUILT dataset) and a minimum total population of 50,000 for the entire urban center. The third step involved converting the WorldPop Urban Buildings raster into a polygon. This data set is a 100m grid representing cells that contain either urban buildings, not urban buildings or no buildings derived from the Maxar and Ecopia buildings footprints layer that is generated from very high spatial resolution imagery for much of Sub-Saharan Africa [6]. The next step included cleaning all of the polygons based on the GHS Urban Centre and then merging the three feature classes to create one boundary shapefile. Finally, we created a 1km buffer around the polygon and any internal "islands" were removed. This resulted in a contiguous extent slightly outside the actual boundary of all city definitions to clip other geospatial data sets used within the study.

2.2. Satellite Image Data

Satellite imagery can be a major source of contextual information about deprived areas. To ensure transferability of IDEAMAPS methods across LMIC city contexts, it was important that EO data be freely available, have global coverage, and have enough spatial detail to provide insight into deprived areas within cities. Data from Sentinel-2 are the highest spatial resolution imagery that are currently available at the temporal and spatial scales that are required for this work at no cost. A key challenge with these data, as with any optical data set, is obtaining cloud free coverage of an entire city. While we can create mosaics using Google Earth Engine and other mosaicing tools, we have found that our mosaic results from these sources have cumulus cloud, haze, and cloud shadows issues for cities such as Lagos. Therefore, we have been working with iMMAP to create mosaics using the geomedian and more advanced image preprocessing techniques [7], [8].

It is important for us to have cloud-free imagery in order to produce fine-scale (e.g., 10m) contextual features. Contextual features can be defined as representing the statistical quantification of different levels of image features, such as low level edge patterns, pixel groups, gaps, which forms high level image textures, and the raw spectral signatures calculated over groups of pixels or neighborhoods. Contextual features, being a proxy of unplanned urbanization, provide information about the spatial patterns that emerge across neighborhoods within cities, and have been used previously to improve estimates of population, poverty, and slums [9]–[12]. Previous work focused primarily on extracting contextual features from very high spatial resolution imagery, however recent work has indicated that contextual features extracted from lower resolution Sentinel-2 data provide similar information and can be used within a variety of models [10], [13]. The features we calculate are: Fourier Transform (FT) which is used to detect high or low frequency of lines, Gabor Filter, a linear Gaussian filter used for edge detection [14]. Histogram of Oriented Gradients (HOG), which captures the orientation and magnitude of the shades of the image

[15], Lacunarity (LAC), which describes the extent of gaps and holes in a texture. Low-lacunarity geometric objects are homogeneous because all gap sizes are the same, whereas high-lacunarity objects are heterogeneous [16], Line Support Regions (LSR), which characterize line attributes [17], Normalized Difference Vegetation Index (NDVI), the most widely used vegetation index that provides information about the health and amount of vegetation, PanTex, which is a built-up presence index derived from the grey-level co-occurrence matrix [18], Structural Feature Sets (SFS), which are statistical measures to extract the structural features of direction lines [19], These contextual features are calculated over areas of 3x3, 5x5, and 7x7 pixels or $30m^2$, $50m^2$, and $70m^2$ respectively and reporting back at the original $10m^2$ spatial resolution.

2.3. Geospatial Data

In addition to the Earth Observation data, we compiled dozens of free, open-source sub-city-level geospatial datasets representing eight of the nine IDEAMAPS Domains of Deprivation: (1) SES and (2) Housing at the household-level, (3) social hazards and assets, (4) physical hazards and assets, (5) unplanned urbanization, and (6) contamination representing within neighborhood area phenomena, and (7) infrastructure, (8) facilities and services, and (9) governance representing phenomena that connect neighborhoods across a city (IDEAMAPS Network, 2019). All datasets were processed into a consistent 100m grid; coarser rasters were resampled, finer-scale rasters were aggregated, and statistics were calculated for vector data by grid cell (e.g., distance to point or line, and presence of polygon). The IDEAMAPS database continues to grow, and as datasets are processed, we release them to our website, https://ideamapsnetwork.org/data/. The data sets thus far include elevation, population, land cover, markets, nighttime lights, armed conflicts, modelled health and poverty indicators from household survey data, and much more for each city. These data are then converted to uniform geographic projection and clipped to the urban extent as described previously.

Our database includes bespoke training data sets of ~300 point locations in ~30 neighborhoods for just Lagos thus far. These data are collected by local slum community mappers and university students who not only verify whether the point is located in a slum/non-slum/mixed area, but also observe the presence of more than 30 characteristics within 20m of that point (e.g., open drains, pay-for-use toilets, road quality). The characteristics in these datasets align with the Domains of Deprivation Framework and variables present among the other spatial data sets in our database. In time, we envisage that the IDEAMAPS ecosystem would gather training data from the public, rather than bespoke surveys, by providing users with useful data summaries in small user-defined areas in exchange for their classification of the area as either deprived, non-deprived, mixed, or non-residential. We are presently treating all training data as sensitive and only redistributing with data use agreements to protect vulnerable populations

2.4. Modeling

There are a number of ways to model deprived areas from EO, other spatial data, and community driven data collection [4]. However, deprived area modeling has typically been done as a one-off project for individual cities. The reason for this that the compiling and processing of all of the geospatial data and collect the training data required to produce a deprived area model is a very large task, regardless of the method used. As a result, both the model inputs and methods used are typically inconsistent from city to city which makes comparison of the outputs and gauging the importance of input variables difficult between cities. This project seeks to overcome these limitations by providing similar training and input data for all of the cities, such that model inputs and results are informed by a theoretical framework. This will allow modelers using different approaches to compare methods, inputs, results, and transferability both within cities and between cities. Furthermore, we anticipate that our classification of EO and other spatial data by conceptual domain, and collection of training data which relate to these domains, will drive methodological innovations around the modelling of dominant type(s) of deprivation that exist at a fine geographic scale.

3. MODEL OUTPUTS

Generally, deprived area model outputs are binary, resulting in a map of deprived versus not deprived areas. While it is easy to perform an accuracy assessment and map binary outputs, these distinct values do not reflect actual ranges in deprivation found within cities and large uncertainties along boundaries are very common. Different stakeholders have different definitions of deprived areas and have a large range of use cases from the neighborhood- to national-scale which is why it is important for deprivation to be presented on a continuous scale, and to support data users to classify their own categories as needed. Not only will identification of key deprivations in a given area support different types of policy and program interventions, it can help local authorities to catalogue which areas in their city meet locally-relevant deprived area classifications (e.g., slum vs. informal settlement) to track local, as well as global, indicators in a meaningful way. The output of deprivation models at 100m grid cells enable the aggregation of results into any meaningful area or administrative boundary. We chose a 100m resolution to strike a balance between provision of spatially detailed information while obfuscating the exact boundaries of slums, informal settlements, and other particularly vulnerable areas.

3.1. Two major goals

We have two major goals for creating the model outputs: first, creating consistent maps for city comparisons, and second, supporting diverse local stakeholders to generate new context-specific information for their work in planning, development, advocacy, and monitoring. While these two goals might seem at odds, we believe they are achievable by mapping deprivation on a scale and facilitating post-analysis classification using different cut-offs to create use-case-specific maps of deprivation for a given city or sub-city region. Furthermore, because we are providing similar input data for all cities, we will be able to determine the ability of models to transfer between cities, create multi-city models, and predict into cities where we have limited model training data.

4. FUTURE WORK

The IDEAMAPS Project focuses on a large societal challenge and will take time develop the datasets, methods, trusting relationships among currently siloed groups, and an ecosystem that facilitates a sustained exchange of information among stakeholders. Some of the first steps of building the IDEAMAPS mapping ecosystem are being developed with our creation of city extents, a transferable mapping framework, data collection protocols, and image processing. We have been working with modelers to develop and test inputs and data formats for increasing the ease of use of these data sets. Outside of the mapping ecosystem we have been working with government officials and researchers within the study area cities to determine what they are interested in using these data for. As part of this we are working to get modelers and local groups together to be able to better understand the inputs, outputs, limitations, and uncertainties associated with mapping deprived areas. By doing this we believe that this can create the ecosystem where geospatial data, modelers, and local officials and researchers can collaborate together.

While the focus on this project has been to develop collaborations with local governments and people with similar interests in Accra, Ghana, Nairobi Kenya and Lagos Nigeria, we are hopeful to spread this to other cities in LMICs. We are open to collaboration with others and are hopeful that the ecosystem we are developing here can be replicated in other cities and countries in the future.

5. REFERENCES

[1] UN-Habitat, *Unpacking the Value of Sustainable Urbanization*. 2020.

[2] UNDESA, *World Urbanization Prospects*, vol. 12. 2018.

[3] D. Thomson *et al.*, "Need for an Integrated Deprived Area ' Slum ,'" *Soc. Sci.*, vol. 9, no. 80, p. 17, 2020.

[4] M. Kuffer *et al.*, "The role of earth observation in an integrated deprived area mapping 'system' for low-to-middle income countries," *Remote Sens.*, vol. 12, no. 6, 2020.

[5] L. Florczyk, et al. "GHS-UCDB R2019A - GHS Urban Centre Database 2015, multitemporal and multidimensional attributes." European Commission, Joint Research Centre (JRC), 2015.

[6] A. J. Dooley, C. A. and Tatem, "Gridded maps of building patterns throughout sub-Saharan Africa, version 1.0. University of Southampton." University of Southampton: Southampton, UK., 2020.

[7] D. Roberts, N. Mueller, and A. McIntyre, "High-Dimensional Pixel Composites from Earth Observation Time Series," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6254–6264, 2017.

[8] D. Frantz, "FORCE-Landsat + Sentinel-2 analysis ready data and beyond," *Remote Sens.*, vol. 11, no. 9, 2019.

[9] R. Engstrom, J. Hersh, and D. Newhouse, "Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being," Washington, DC, 8284, 2017.

[10] J. Hersh, R. Engstrom, and M. Mann, "Open data for algorithms: mapping poverty in Belize using open satellite derived features and machine learning," *Inf. Technol. Dev.*, 2020.

[11] R. Engstrom, et al., "Evaluating the relationship between spatial and spectral features derived from high spatial resolution satellite data and urban poverty in Colombo, Sri Lanka," *2017 Jt. Urban Remote Sens. Event, JURSE 2017*, pp. 8–11, 2017.

[12] R. Engstrom, D. Pavelesku, T. Tanaka, and A. Wambile, "Mapping poverty and slums using multiple methodologies in Accra, Ghana," in *2019 Joint Urban Remote Sensing Event, JURSE 2019*, 2019.

[13] R. Masaki, Takaaki, Newhouse, David, Rudra, Ani, Bedada, Adane, Engstrom, "Small Area Estimation of Non-Monetary Poverty with Geospatial Data," Washington, DC, 2020.

[14] Mehorta, R., Namuduri, K.R., Ranganathan, N, "Gabor filter-based edge detection," *Pattern Recognit.*, vol. 25, no. 12, pp. 1479–1494, 1992.

[15] N. Dalal and W. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR05*, vol. 1, no. 3, pp. 886–893, 2004.

[16] S. W. Myint, V. Mesev, and N. Lam, "Urban Textural Analysis from Remote Sensor Data: Lacunarity Measurements Based on the Differential Box Counting Method," *Geogr. Anal.*, vol. 38, no. 4, pp. 371–390, Oct. 2006.

[17] C. Ünsalan and K. L. Boyer, "A system to detect houses and residential street networks in multispectral satellite images," *Comput. Vis. Image Underst.*, vol. 98, no. 3, pp. 423–461, 2005.

[18] M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A robust built-up area presence index by anisotropic rotation-invariant textural measure," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 1, no. 3, pp. 180–192, 2008.

[19] X. Huang, L. Zhang, and P. Li, "Classification and Extraction of Spatial Features in Urban Areas Using High-Resolution Multispectral Imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, 2007.