International Labour Organization

Methods for spatial sampling of urban neighbourhoods by socio-economic status in Indian cities

Methods for spatial sampling of urban neighbourhoods by socio-economic status in Indian cities

Pratyush Tripathy, Krishnachandran Balakrishnan, Teja Malladi, Amruth Kiran and Gautam Bhan

Indian Institute for Human Settlements

Contents

Note on Methods: Approaches to Neighbourhood Identification and Categorisation

This brief note lays out an approach to neighbourhood selection and houselisting. While it is presented as a generalized step-by-step framework that can be used in any city where socio-economic classification of neighbourhoods is needed, it is tailored for cities where household level public data is not easily accessible or where such information excludes sections of the city considered informal/unplanned/peri-urban or irregular. Given this, the approach reflects an emphasis on the need to develop methods for data scarce urban contexts, or those whose urbanization is marked by tensions with formal logics of planning, property and labour. These cities are likely to be in the global south, though aspects of this framework maybe equally useful in the North.

The note is structured as follows. The next section uses the current study in Bangalore and Chennai to remind us of our sampling requirements. The following section lays out an approach to identifying what needs to be done to attain a city-wise profile of neighbourhoods to meet such a requirement. The final section then illustrates an application of the approach to Bangalore, showing neighbourhood identification and categorization.

Categorisation: Socio-economic Status, Location, and Planning Typologies

Our proposed research requires three kinds of categorization:

- Neighbourhoods of high, medium and low socio-economic status across formal and informal neighbourhoods.
- Neighbourhoods in sub-regional areas of large city regions
- Neighbourhoods of different planning forms ranging from dense organic settlements to gated apartment complexes.

Socio-economic Categorisation

The distribution of paid and unpaid domestic work is likely to vary significantly in households of different socio-economic status with income-richer households hypothetically engaging in more paid domestic work, and for different kinds of domestic work, that income-poor households. This implies that any study on domestic work must sample households across different economic strata. The methodological challenge is how to do so.

The first step is to create a distribution of households by measures of socio-economic status. This can be, at its simplest, income, but many countries use index based, weighted multi-dimensional measures that could also be used as a income, asset and wealth measure of socio-economic status. If such data is available for analysis, then this is the simplest scenario. Distributions can create typologies of High/Medium/Low categories of households either proportionally (one third/one third/one third), or by other principles (standard deviations from official poverty lines or minimum incomes; standard deviations from median income etc). South African cities, for example, have such data.

What is such household level information is not available? Then we must move up the scale. If socio-economic classifications of neighbourhoods but not households are available, then those can be used as proxies. Those neighbourhoods classifications – lets say they go from A for rich neighbourhoods to G for poor ones) are often made by governments on the basis of property tax, or infrastructural assessments, that act as proxies for socio-economic status.

Here, an assumption is being made that households within a neighbourhood classified as Category A are more like each other than those in Category B. The validity of this assumption will weaken the narrower the differences between neighbourhoods (Category A and B are likely to be similar) and strengthen the wider the differences (Category A and G, for example, are unlikely to be similar). If studies use neighbourhood classifications, then the typologies of High/Medium/Low must be blunter, comparing A against E and G, rather than B and C. New Delhi, for example, ranks all its neighbourhoods from A to G precisely like this, allowing a typology of socio-economic neighbourhoods to be created.

What if neither household nor neighbourhood classifications exist? In this case, neighbourhoods must be assessed by criteria observable at scale. Later in this note, we propose a geo-spatial method that combines physical structure and street typology to classify neighbourhoods that has been developed by IIHS.

Formal-informal neighbourhoods

In many cities, not all neighbourhoods are reflected or recognized on official maps, plans and in census or other data. Typically, these are thought of by the state as informal, irregular, unauthorized settlements. The fact that they are a significant part of urrbanisation in cities of the global south is indicated by how many local language names settlements like these have in different cities: favelas in Rio, *colonias* populares in Mexico City, *musseques* in Luanda, *amchi wastis* in Pune, *ashwa'iyyat in* Cairo, shacks or *'mjondolos* in Durban, *sukumbhashi bastis* in Dhaka, *katchi abadis* in Karachi, *kampung liars or hak miliks* in Kuala Lumpur, and the *sahakhums* in Phnom Penh.

If we rely only on official data, we must ensure that it is indeed universal in its coverage of the actually existing city and not only legal

and planned neighbourhoods. In Delhi, 45% of citizens are estimated to live in informal or unauthorized layouts, to take just one example, that are also informal in different ways.

There is no exact way to precisely included what is, by definition, outside precise definition or measurement. Yet we must attempt to correct exclusion biases. One way is to see the ways in which the state accounts for informality. Many city governments do have lists of "slums," or unauthorized layouts, or colonies. These can be used as an addition to database of neighbourhoods generated in the step above, with deliberate additional sampling. The difficulty will be in assigning a socio-economic assessment to these neighbourhoods. Not all informal neighbourhoods are poor, though when understood as the state as "slums," a legitimate assessment of them being in the "low" category in the SES distribution can certainly be made. Local knowledge will be key to assign informal neighbourhoods to H/M/L tiers in the SES distribution.

Sub-Regional Analysis

For many city-regions, but especially rapidly expanding regions in the global south, there are multiple cities within cities. Often, there can be clear demarcations of different sub- city variants (Victoria Island versus the Mainland in Lagos; Electronic City and Whitefield versus Central Bangalore; Park Station and Soweto in Johannesburg; walled city of Old Delhi and New Delhi in Delhi, or the Kasbah and the planned city in Marakech, for example). These can be different phases of urban development across decades (including through colonial transition), suburbanization, or even municipal expansion and peri- urban growth.

In some cities, these different phases of urban development have remarkably different urban forms. If so, it is important to separate these sub-regions. Household SES variation and formal-informal categorization will exist within each of the sub-regions but patterns of domestic work may differ. This categorization need not feel relevant in all cities, so is not required but it certainly should be examined by all researchers. We believe it is critical in both Bangalore and Chennai, and will illustrate the two sub-regions [labelled as Core and Periphery in our case] we identify in Bangalore later in this note.

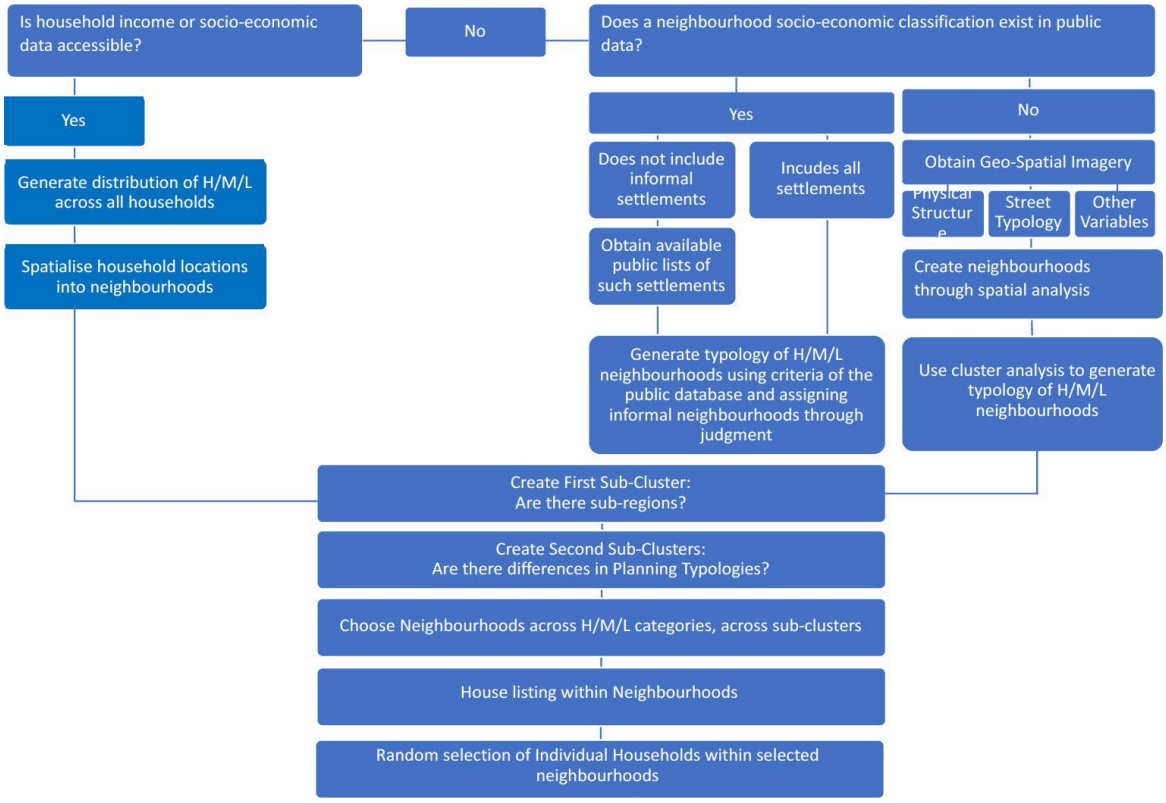
Planning Typologies

Every city has distinct planning typologies that combine different planning regulatory categories (urban villages, neighbourhoods, heritage areas, industrial townships, etc) with different built forms (organic layouts, planned layout grids, plotted development, apartment buildings, row housing, group housing, etc). We hypothesize that the nature of domestic work will be different at least in some of these – for example, in a multi-story apartment building than in plotted row houses.

The categorization of planning typologies must be done by researchers familiar with the different types of urban form and planning categories that exist in a city. Again, this categorization need not feel relevant in all cities, so is not required but it certainly should be examined by all researchers. We believe it is critical in both Bangalore and Chennai, and will illustrate with choices of planning typologies we chose in Bangalore later in this note.

Process of Creating Sampling Clusters and Sub-Clusters

How are these four categorisations to be operationalised in a research design? The flow chart below sets up an approximation that must be contextualized to each city but offers a way to work through the steps.



Illustrating the Process: Bangalore

In this section, we show the maximum extent of this process for a city where neither household nor neighbourhood SES data is easily available. A reminder that this is the full iteration of the flowchart above – it can, and hopefully will, be much simpler in places where household or neighbourhood classifications exist. This particular method of using satellite imagery and street networks for neighbourhood delineation has been developed at the Indian Institute for Human Settlements. The methods of using open street maps data to extract street hierarchy are available in the citation in the footnote.[1]

First, we outline the sub-regional analysis. We separate Bangalore into core and periphery (Part A). Then we separate neighbourhood into type 1, 2, and 3 that is our approximation of H/M/L (Part B). Then we add informal and unauthorized layouts (Part C). We present them together in final sampling framework work. In each part, we show the data needed and the analysis. We can undertake this analysis for other cities in the project, if such capacity does not exist in local research organisations.

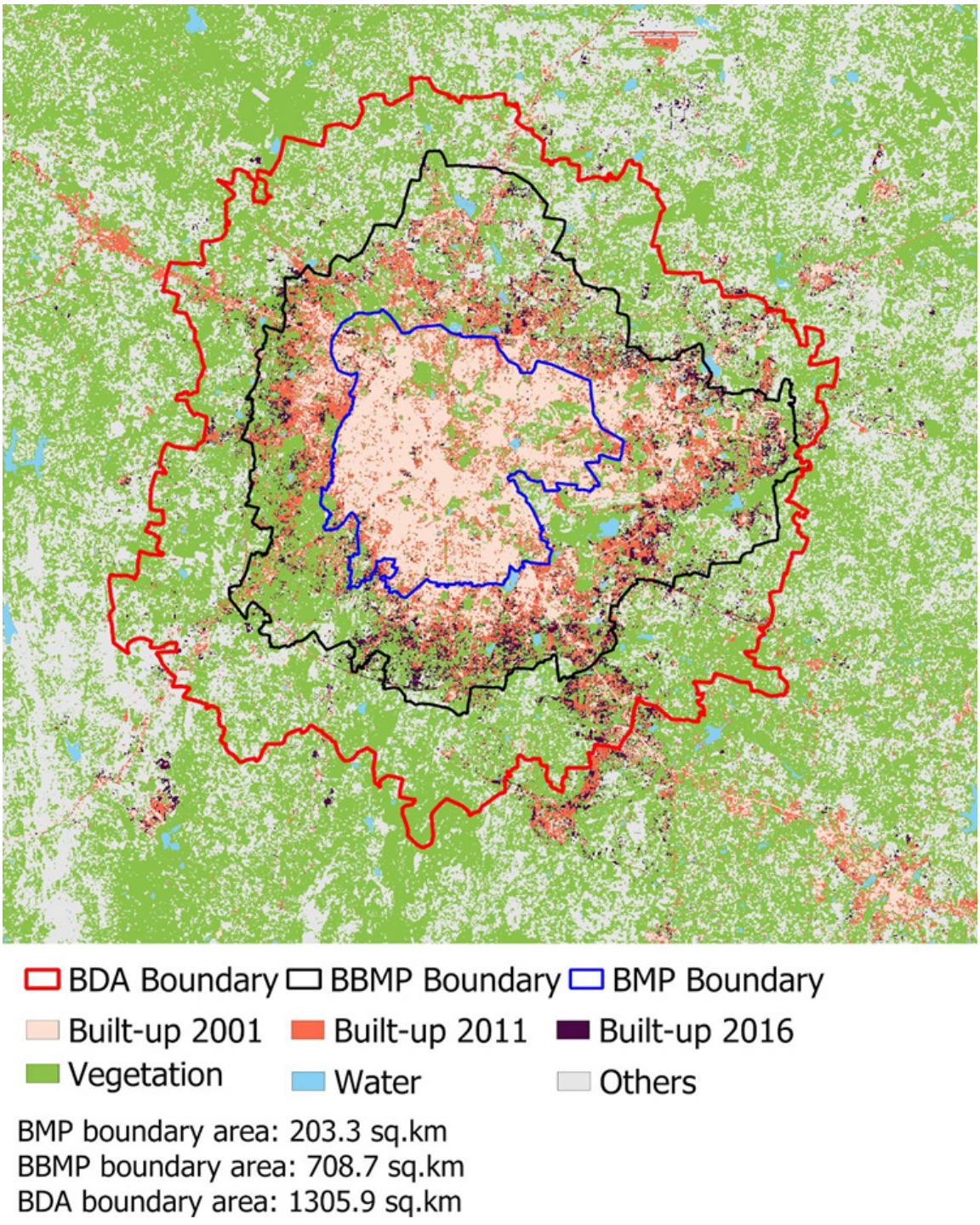
A. Core and Periphery

Changes in city's administrative boundaries, temporal land cover change or built-up growth can indicate city and region's growth over time and can help distinguish between old and new developments. For the purposes of this research, we are proposing the old Bangalore Mahanagar Palike (BMP) boundary as the core city region and the areas that are outside the BMP boundary and within the new Bruhat Bangalore Mahanagar Palike (BBMP) boundary as periphery. The last demarcation exercise in Bangalore was conducted in 2007, increasing the number of wards from 100 to 198 wards.

In the case where the cities administrative boundaries have not changed or these boundaries do not help in distinguishing between, land cover change, especially the built-up growth can be used as an indicator to identify the new developments in the city. A land cover map highlighting the built-up growth over the last two decades in Bangalore, layered with the administrative boundaries is shown in Figure 1. Here built-up represents the impervious surface; vegetation includes grassland, cropland, and forest area. The water class shows man made tanks and natural water bodies etc., and other class shows all the barren land and others.

The availability of high spatial and temporal resolution imagery will vary across different cities. We have used Landsat imagery, an open for all dataset, which is available for over the last three decades at 30m spatial resolution is used here for mapping land cover change. This could be replicated in other places. However, based on the availability, it is recommended to use high

1 Tripathy, P., Rao, P., Balakrishnan, K., & Malladi, T. (2021). An open-source tool to extract natural continuity and hierarchy of urban street networks. Environment and Planning B: Urban Analytics and City Science, 48(8), 2188-2205. http://dx.doi.org/10.1177/2399808320967680



▶Figure 1. Land cover map of Bangalore for the year 2016 prepared from the Landsat satellite data (30m cell size), overlaid with BMP, BBMP and BDA boundaries.

resolution imagery to map the changes at better resolution.

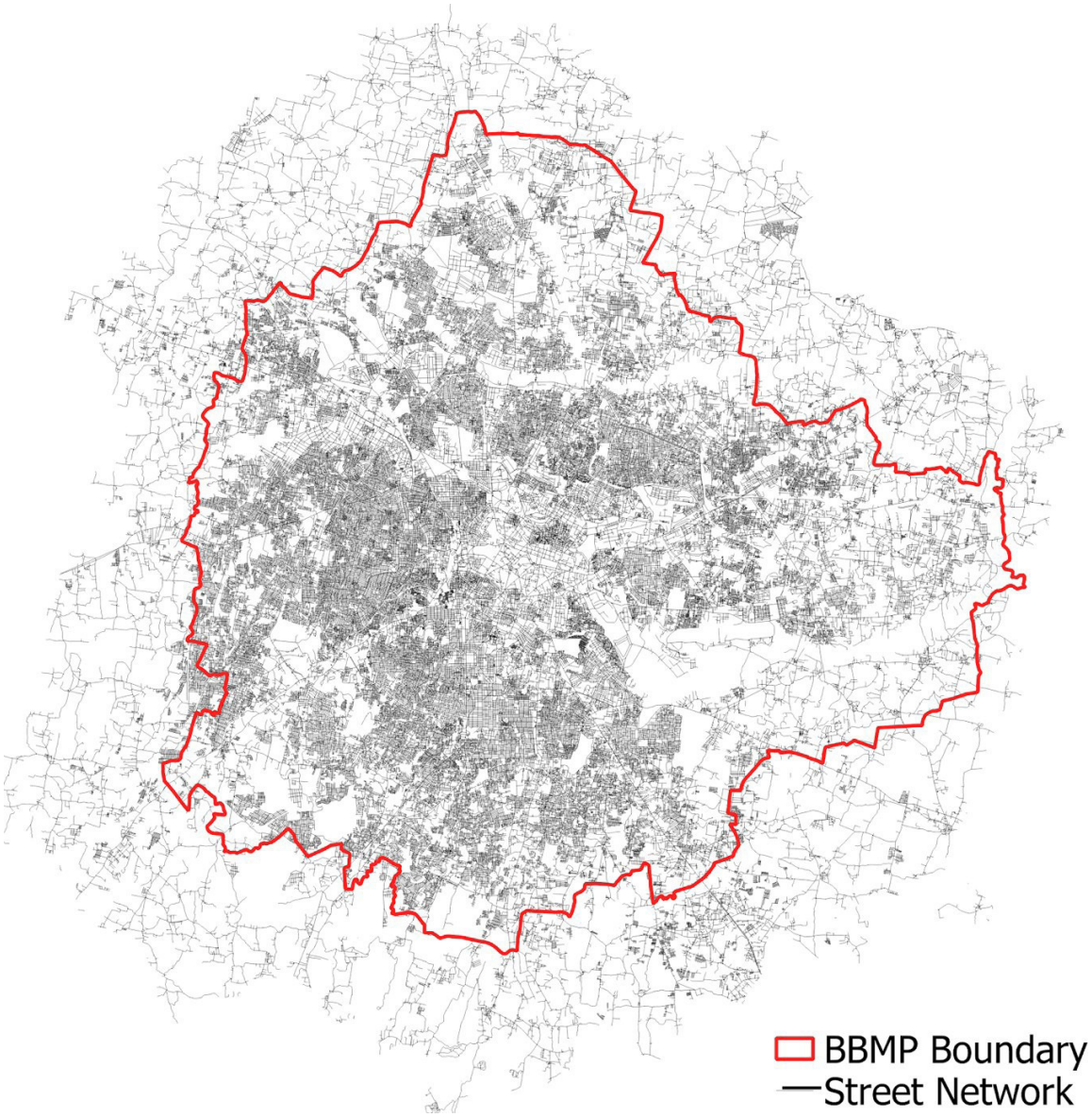
B. Neighbourhood identification and delineation

In the following sections, we demonstrate how physical characteristics can be used as proxy for determining socio-economic, wealth, planned and unplanned neighbourhoods at a finer resolution than administrative wards. As a method we propose to use Street typology, Built- up and Vegetation density as indicators to identify different neighbourhoods types.

We believe that this method is easily replicable as all these datasets are easily and readily across most cities in the world.

Street Typology

Street networks form the skeleton of the urban form. Street Network for most cities can be downloaded for free from the OpenStreetMap project. Figure 2 shows the street network of Bangalore.



▶Figure 2. Bangalore street network overlaid with BBMP boundary.

Neighbourhood Identification

To identify different neighbourhood types, we used the following seven layers using the K-Means clustering algorithm:

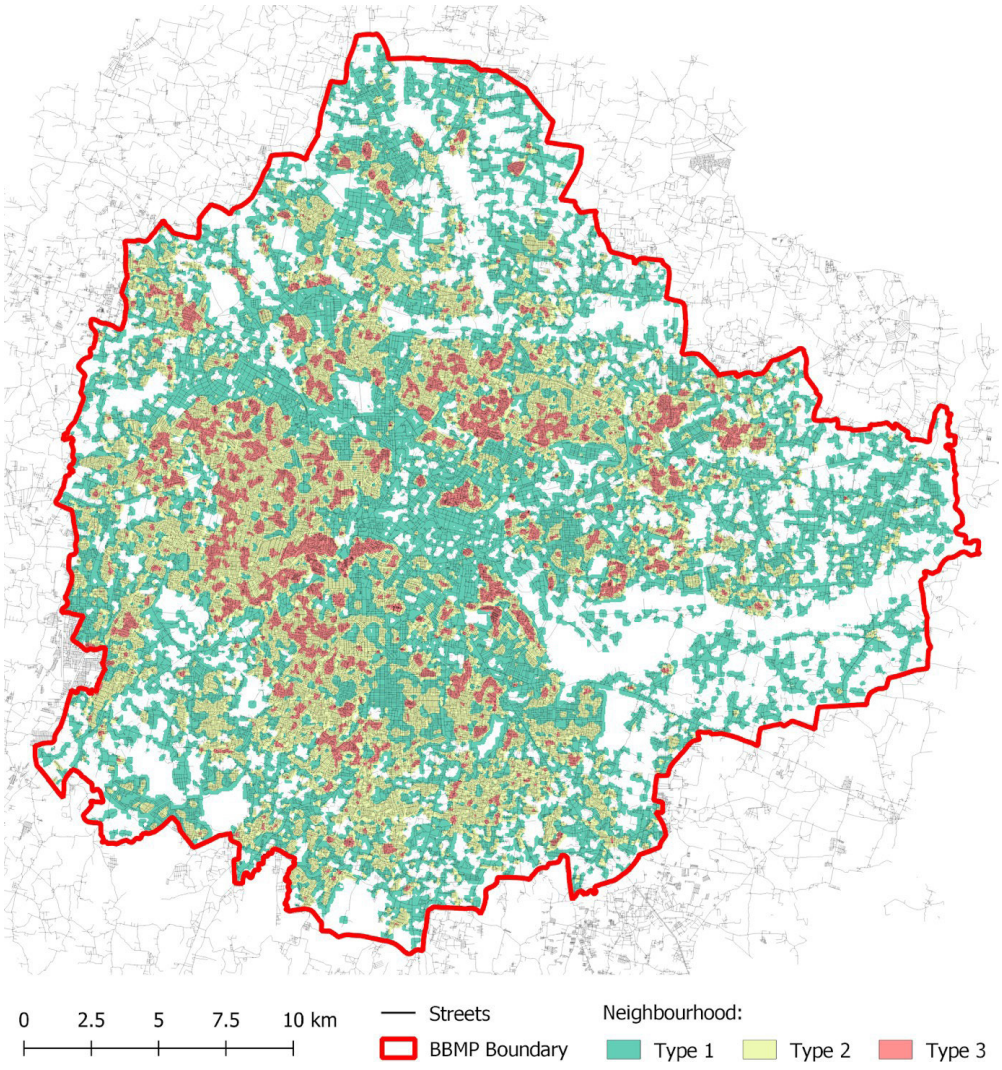
- Edge density
- Edge straightness
- Edge slope
- Node density
- Node degree
- Smoothened built-up density
- Smoothened vegetation density

Details of how to measure 1-7 are in Annexure 1 of this document. For interest in specific details of each step, please read that Annexure first before proceeding.

The K-Means clustering algorithm finds groups within the data based on the data distribution in multi-dimensional space (number of dimensions is the number of input features).

The number of output classes or groups is an input parameter for the K-Means algorithm. If the number of classes is less, the output consists of large continuous patches and it fails to capture the intra-urban heterogeneity and is opposite when the number of classes is large, the output results in too many classes. The number of classes for K-Means clustering is very subjective, and we have used generated output with seven classes for this research.

Based on visual interpretation and with researchers familiarity with the Bangalore neighbourhoods, we merged seven classes into three representing different types of neighbourhoods as shown in Figure 3.

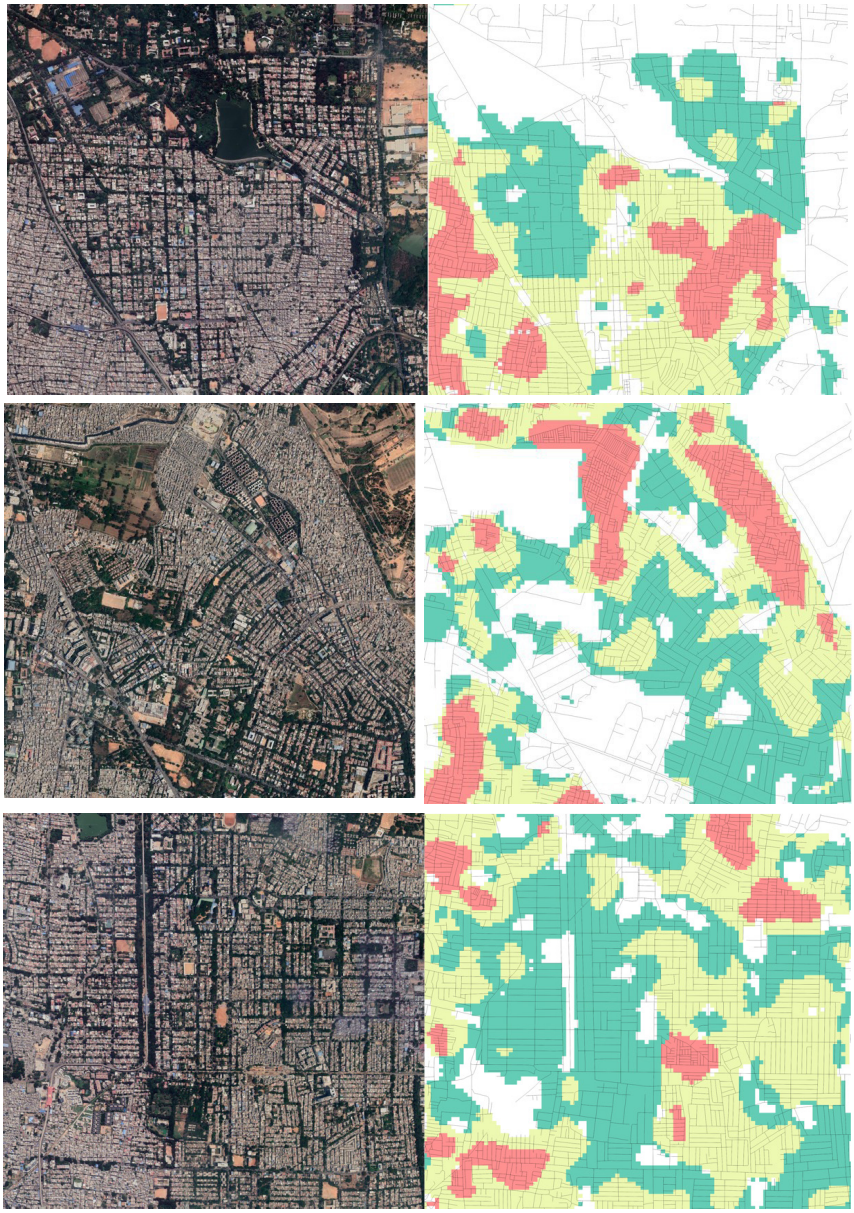


▶**Figure 3. Neighbourhood distribution map of Bangalore.**

The neighbourhoods types generated are:

1. *Neighbourhood type 1* - This type consists of the neighbourhoods that have large plot size, less street density, planned / organised streets, low built density, high vegetation density. Our supposition is that these can be considered as neighbourhoods with High Socioeconomic Status.
2. *Neighbourhood type 2* - This type consists of the neighbourhoods that have mixed street typology, moderate built-up and vegetation cover that lie between type 1 and type 3 fall in this class. Our supposition is that these can be considered as neighbourhoods with Medium Socioeconomic Status.
3. *Neighbourhood type 3* - This type consists of the neighbourhoods that are dense with small plot size, high-street density, unplanned and unorganised street layout, minimal vegetation and high built up density. Our supposition is that these can be considered as neighbourhoods with Low Socioeconomic Status.

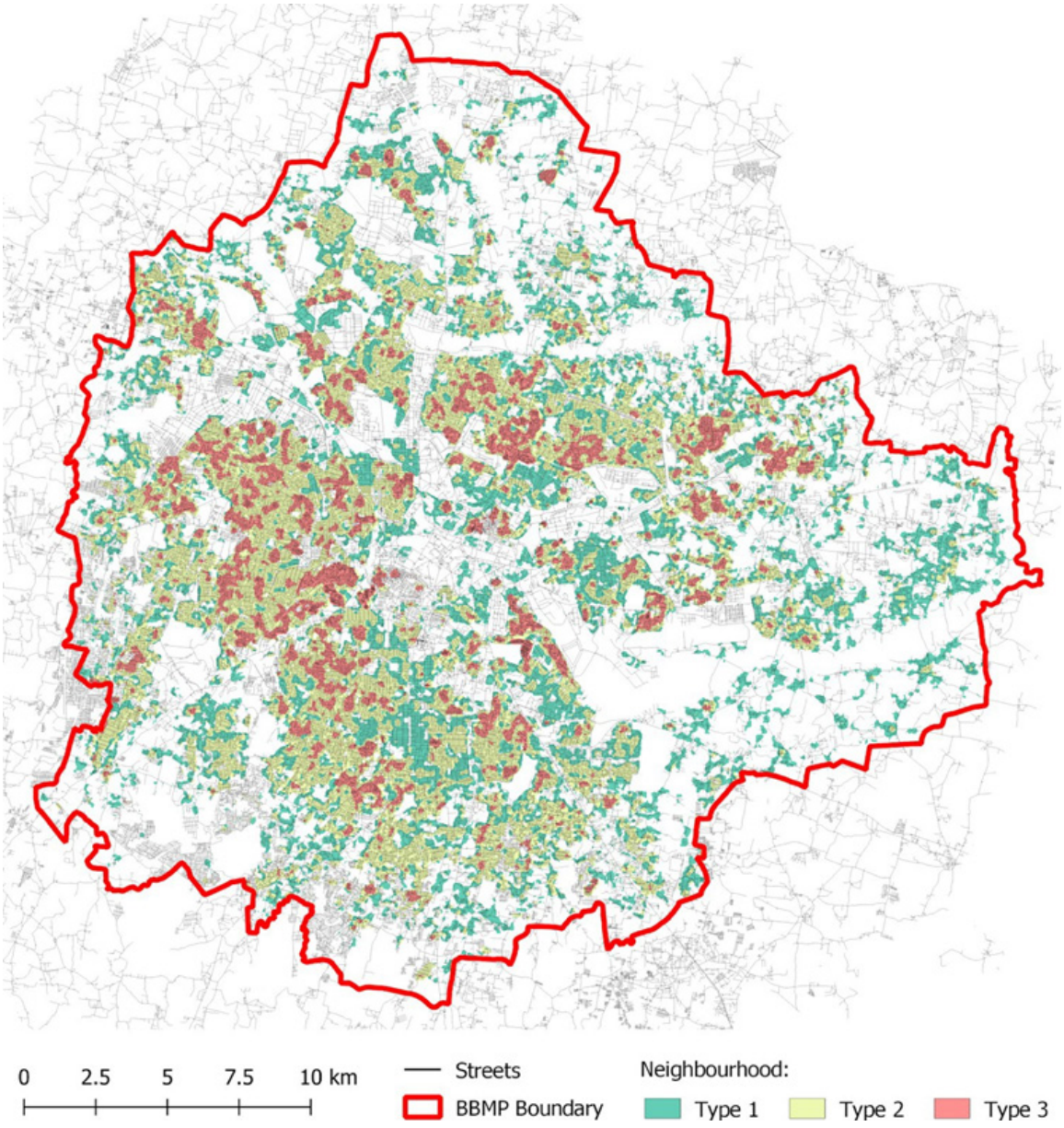
An example of three types of neighbhourhoods and how they differentiated in an area of the satellite map is shown in Figure 4.



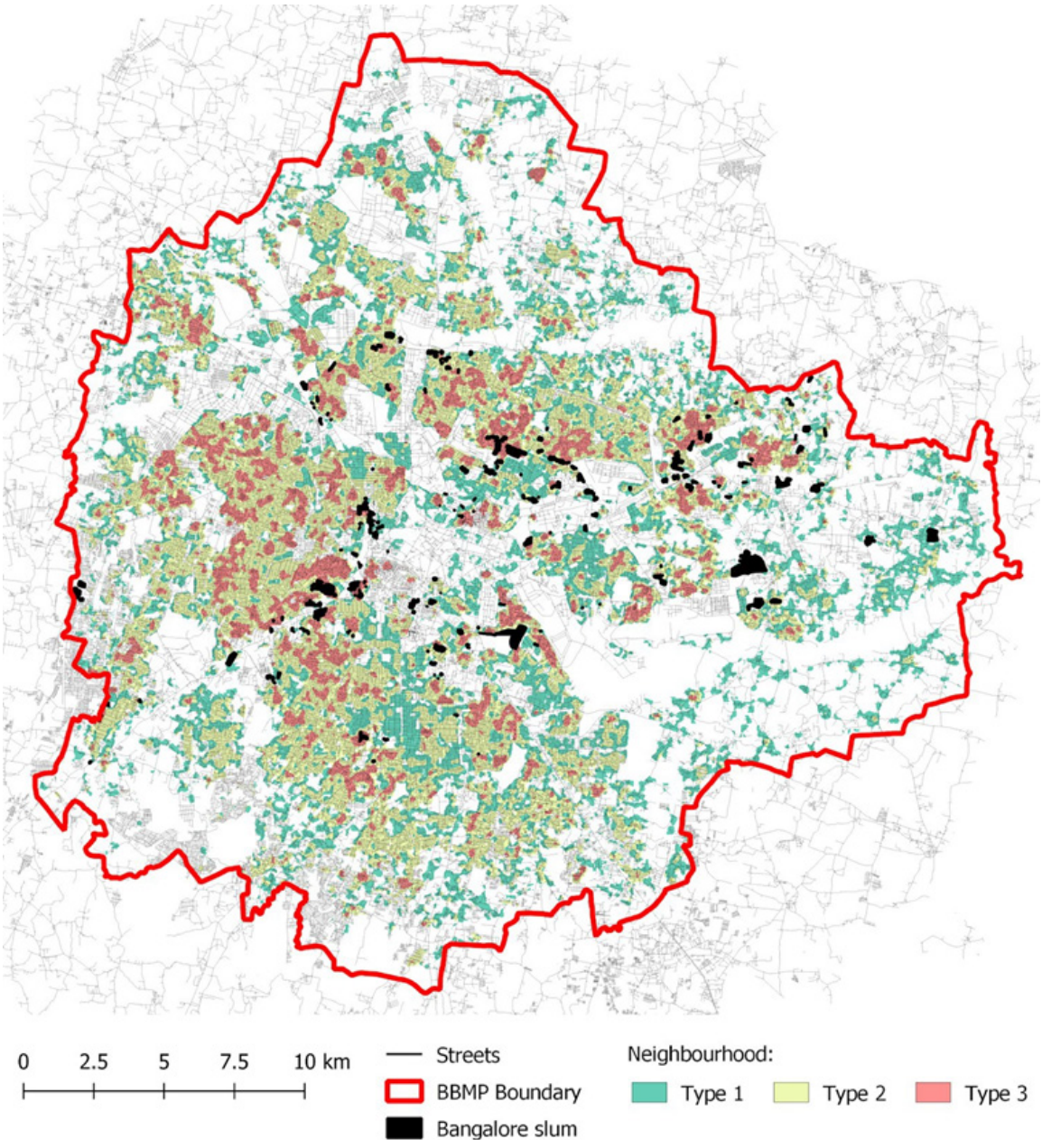
▶Figure 4: Generated neighbourhoods map with corresponding Google Earth image

C. Planning Legalities

We used the residential land use class from the 2015 Existing Land Use Master Plan data to identify recognised residential neighborhoods from the city as shown in Figure 5. We have used the notified informal settlements which represent the unplanned and informal neighbourhoods in the city as shown in Figure 6 to map both the settlement types.



▶Figure 5. Neighbourhood distribution map of Bangalore for residential areas only (existing land use 2015).



▶Figure 6. Neighbourhood distribution map of Bangalore for residential areas only (existing land use 2015), overlaid with informal settlements

Figure 6 allows us thus to have all our neighbourhoods (formal and informal), classified into high, medium and low. These then act as a basis for different sampling strategies across them, when read along with the core-periphery layout in Figure 1. On this, we would add differentiation on planning typologies as needed.

Annexure

Assessing Street typology and Physical Structure

Street networks have been discussed widely using graph theory. The primal graph theory represents the intersection of the streets as nodes and the streets as edges (Figure 1).

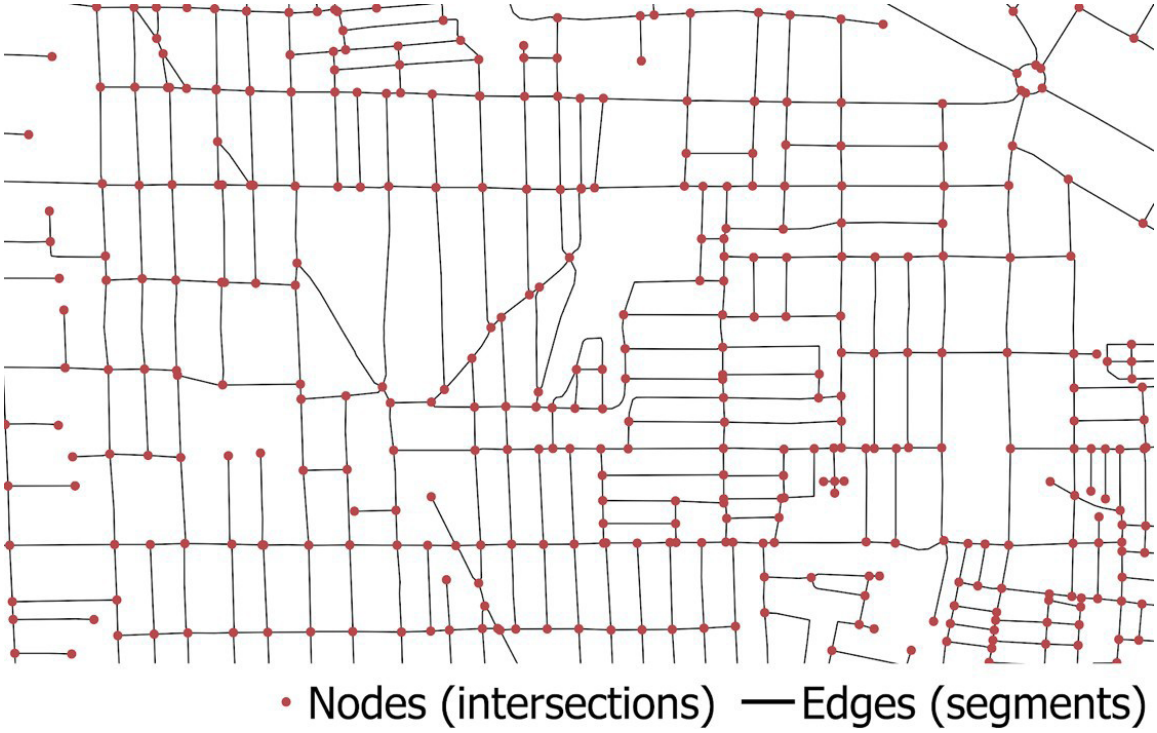
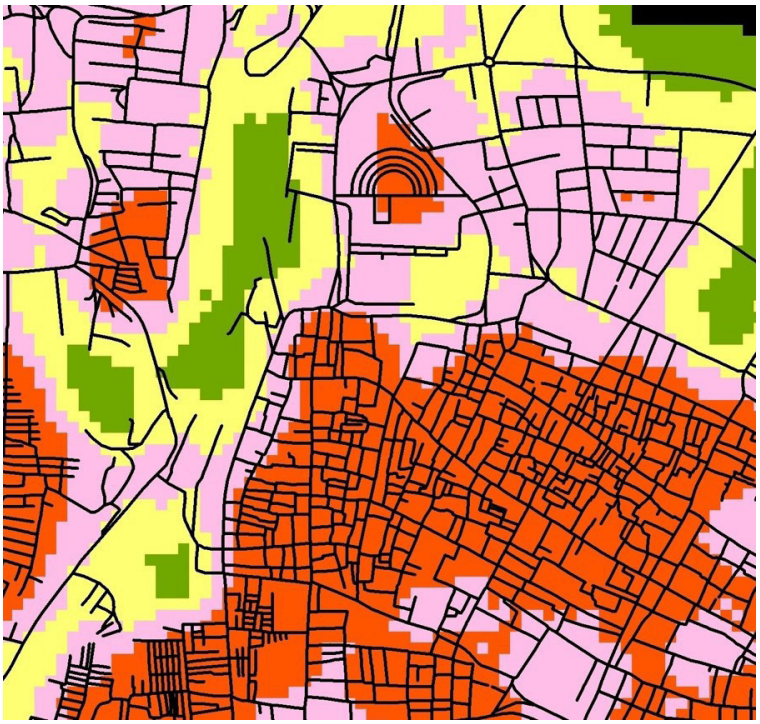
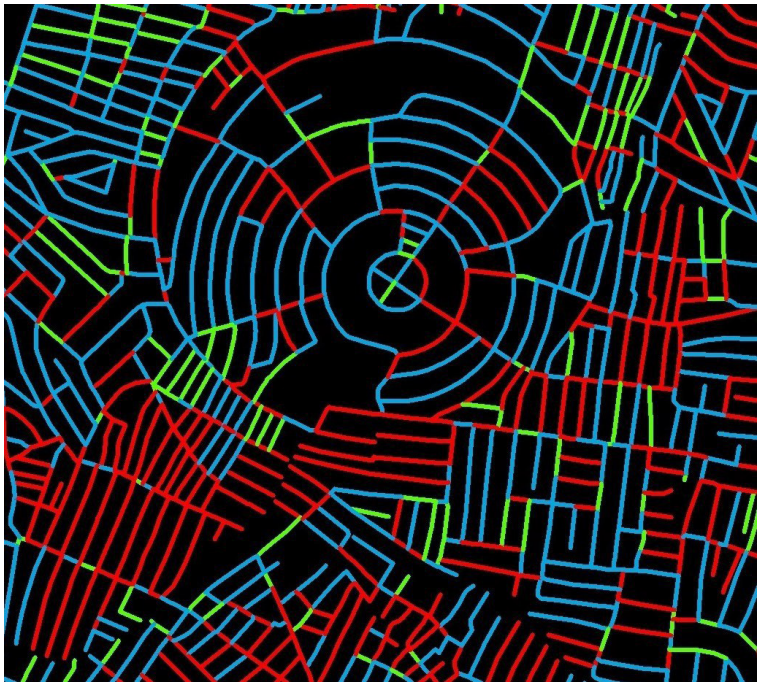
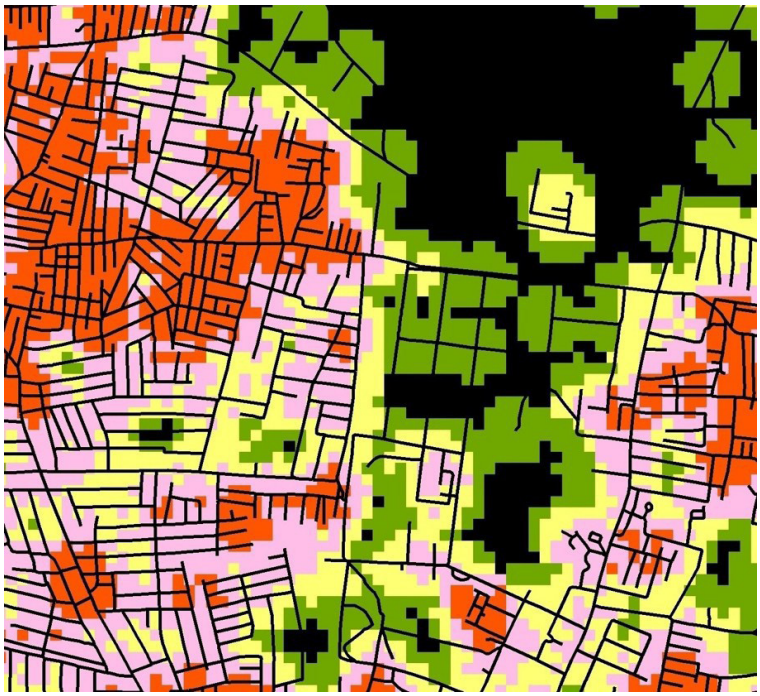
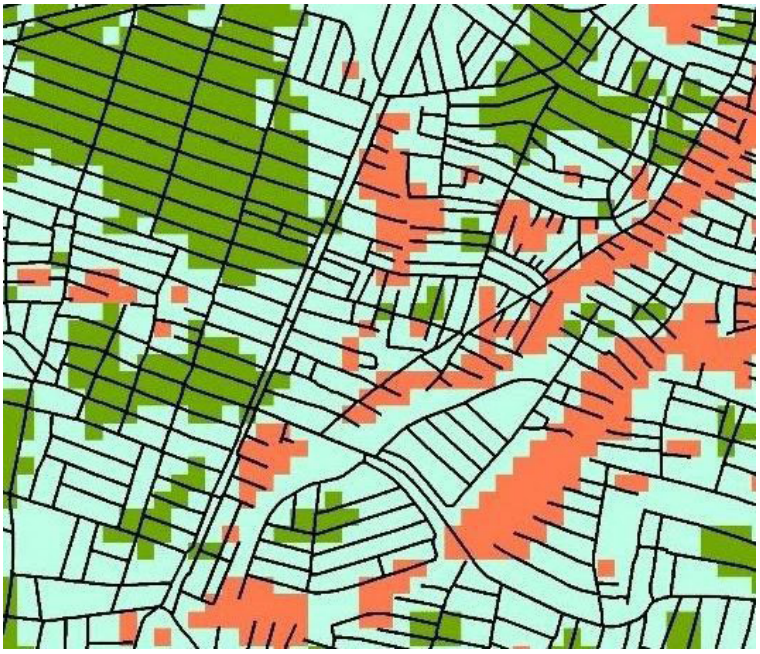


Figure 1. Neighbourhood level streets represented as a combination of nodes and edges.

In the present approach, we consider three properties of the edges, and two properties of the nodes as below.

Description	Graphical representation
1. Edge density This shows the density of the street segments per unit area for a given radius. We used a 90m radius to compute the density. The figure to the right shows four intervals of edge density for better visualisation. The Chickpete area which has high streets density (shown in orange colour) as compared to the area towards the north (green in colour), which is quite spacious.	
2. Edgestraightness This shows the orthogonality of street segments, which we used to quantify the curvature of the segments. It is calculated by computing the ratio between the actual segment length and the shortest distance between its two ends. The straightest segment will have a straightness value one, and the lower values depict the amount of curvature. The image to the right has been classified for better visualisation. We can see that the straight streets are coloured in green, followed by curved streets in blue then crooked streets in purple, and extremely curved in red colour (Majestic Bus Stop).	

Description	Graphical representation
3. Edge slope This represents the slope of the street segments. We extracted the minimum and maximum elevation value in the street segments from the digital elevation model (DEM), which shows the height of any point in map from the mean sea level (MSL) in meters. We used the difference in the minimum and maximum height, and the length of segments to compute the slope using trigonometry.	
4. Node density This shows the density of nodes per unit area within a given radius. We used a 90m radius to compute the density. While this is very similar to the edge density, it still adds value when there are fewer streets but lot of street intersections. This fails when there are long parallel streets without any intersection, which is backed by the edge density. The map to the right has been classified for better visualisation.	

Description	Graphical representation
5. Node degree This shows the number of adjacent segments at each intersection. Eg. planned neighbourhoods with a grid-like layout will have node degree value four, T-points will have value three, dead ends will be represented by value one. The map to the right shows the node degree overlaid with the street network layer. Planned neighbourhoods can be seen in green colour and dead-end streets can be seen in pink colour.	

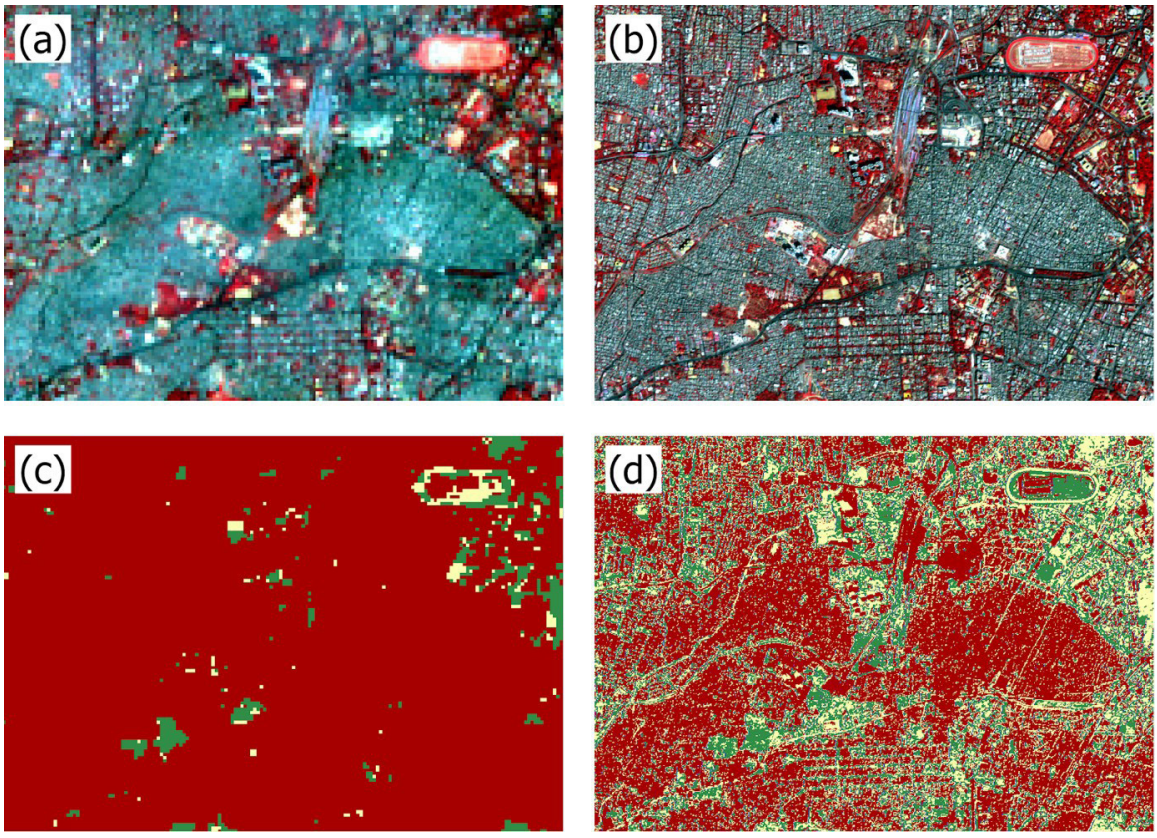
As a next step, we are in the process of developing an open tool that other researchers can use for developing similar outputs for their area of interest.

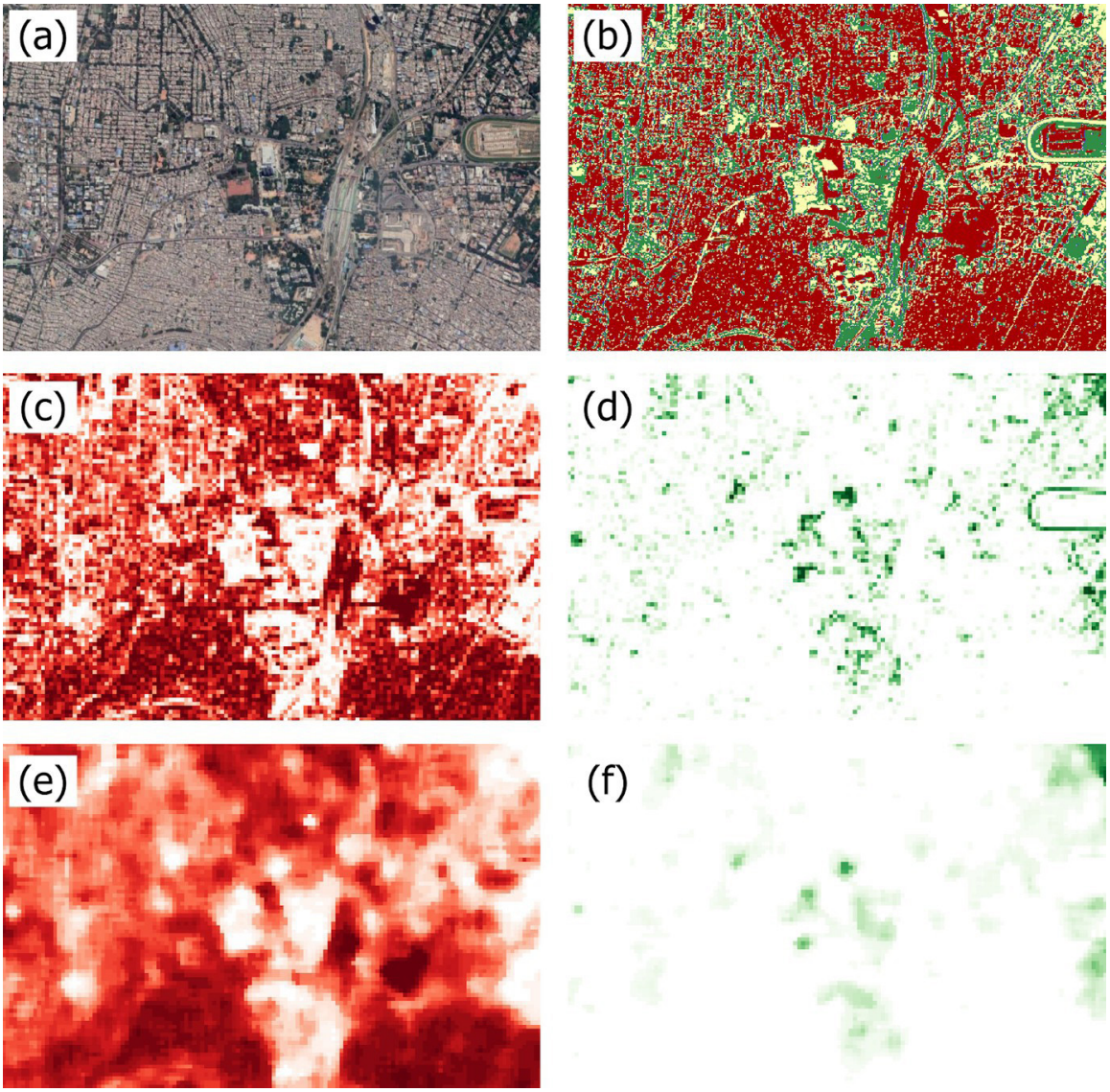
Finer Resolution Land Cover for calculating Built-up and Vegetation Density

As mentioned in the earlier sections, higher resolution land-cover imagery could help in mapping land cover at finer resolution. In this research we use LISS-IV multispectral imagery of 5m resolution for land cover mapping. LISS IV resolution is six times finer than the free to use Landsat 30m resolution imagery. Figure 2 shows a comparison of land cover data prepared from the Landsat and LISS-IV imagery.

As seen in Figure 2(a) and 2(b), the LISS-IV data enables better interpretation of surface features than the coarser Landsat data. The difference in the land cover data derived from LISS-IV is significantly better than the Landsat derived data. After the land cover data creation, we used a 6X6 hopping kernel method to calculate built-up and vegetation density at 30 m resolution. The density maps represent the number of built-up or vegetation cell for every kernel and the values are ranging from 1 to 36. A 5X5 kernel was then used on the density map to smoothen the output density rasters.

▶Figure 2. Comparison of (a) Landsat raw data, (b) LISS-IV raw data, (c) Landsat derived land cover, and (d) LISS-IV derived land cover for the same area.





▶Figure 3. Map showing (a) Google satellite basemap, (b) land cover derived from LISS-IV 5m data, (c) density of built-up computed at 30m, (d) density of vegetation at 30m, (e) smoothed built-up density map using 5X5 kernel (f) smoothened vegetation

In case of unavailability of high resolution imagery or lack of resources, researchers could still proceed with Landsat imagery or use free Sentinel imagery provided by European Space Agency (ESA) available at 10m resolution as alternati

iihs®
INDIAN INSTITUTE FOR
HUMAN SETTLEMENTS

Contact details

ILO DWT for South Asia and Country Office for India
Core 4B, 3rd Floor, India Habitat Centre
Lodhi Road, New Delhi – 110003, India

T: +91 11 4750 9200
E: delhi@ilo.org

© International Labour Organization 2022