A method for determining the Socio Economic Status distribution for the case of Colombo, Sri Lanka

Author
Rosalie de Vries

Supervisors
Dr. Eric Koomen
Maaike van Aalst

Deltares
Nationale GI Minor
Vrije Universiteit Amsterdam
Date: 28-01-2019

Abstract

Determining the Socio Economic Status distribution in urban settlements is a crucial step in the process of formulation of pro-poor policies. However, the use of conventional methods for determining the Socio Economic Status (SES) distribution such as field surveys can be time-consuming, costly and may not be reliable when measured by different people who have different opinions on what is good, moderate and bad. This paper explores the possibility of implementing a practical standardized method for determining the SES differentiation on district and building-level for the case of Colombo, Sri Lanka.

Contents

1. Introduction

For a flood risk management study in Colombo, Sri Lanka, the risk is calculated through multiplication of the damage and the probability. This flood risk method does not make a distinction on socioeconomic status (SES) or wealth, but calculates the absolute economic damage based on damage curves for two types of buildings: shanty buildings, on story buildings or multiple story buildings.

Absolute estimated damages per building type are estimated and aggregated to calculate the flood risk for an area. This is problematic as absolute damages borne by poor people are very low, although their relative well-being is much more affected compared to the well-being of wealthier people that might have much more absolute flood damages. As a result, measures that decrease the absolute flood risks most, will be preferred, while this might not be the measures that decrease the 'relative' flood risks for poor and vulnerable people (Kind, Botzen, & Aerts, 2017)

To properly take the relative flood risk damages into account (instead of absolute aggregates), an estimation of spatial socioeconomic status data is needed. Based on census data, there are some estimates of income at district level available, but to collect more information about the distribution in a specific district, the aim is to classify different SES levels based on different characteristics and indicators derived at building/street level through open data sources.

The idea is to use the available SES proxy data from the general statistics office to identify and map the GN districts (lowest level from which information is available in Colombo) on socioeconomic status. From the GN districts that have the lowest SES level and that are located within the flooding extent, one or two districts will be selected and the characteristics of these districts will be examined through classification methods identified in the literature phase. Different methods and approaches to establish a SES classification method exist and the aim is to combine and find the classification method that is feasible , up scalable and gives a comprehensive understanding of SES distribution at local and building level for the case of Colombo, Sri Lanka. The research question explored in this article is:

"How can socioeconomic status at local and building level be identified from open and local sources for the city of Colombo, Sri Lanka?"

The corresponding sub-questions are:

- Which methods exist to identify socio-economic status distribution from spatially explicit data sources at local and building level?
- Which data requirements do these methods pose?
- To which extent does open data from global sources meet these data requirements and allows the mapping of socioeconomic status at local and building level for the city of Colombo, Sri Lanka?
- To which extent does more specific data from local sources meet these data requirements and allows the mapping of socioeconomic status at local and building level for the city of Colombo, Sri Lanka?
- How can the information from the local sources with the information from the open sources be combined to identify SES distribution at a finer spatial scale?

2. Methods

2.1 Literature review

Socioeconomic status can be defined in different ways. In this article the SES definition of (Baker, 2014) is used. Baker, 2014 states: “Socioeconomic status (SES) is defined as a measure of one’s combined economic and social status and tends to be positively associated with better health.”

There is a broad range of methods available in capturing SES differentiation. Methods vary in terms of level of automation. Methods with a great deal of user involvement used manual image interpretation. As an example, (Motholo, 2014) examined the relationship between socio-economic status and features derived from visual image interpretation.

More automated methods are pixel-based, and field-based classification. Zhang et al., (2017) wrote an article describing an automated methodological framework which classifies urban land use in a study area within Haidian District, Beijing, China in 2016. Zhang et al chose to use a combination of the per field method and the per pixel method. Both methods hold different advantages. When using VHR images, the per pixel method provides abundant and detailed information on the spectral, textural, contextual and spatial configuration of urban land cover. The advantage of the per field method is that multiple sources of social sensing data (social media, phone, digital maps, GPS) make it possible to examine the socioeconomic and demographic characteristics of urban land (Zhang et al., 2017).

The field method is a useful method to determine the socioeconomic status at local level. However, this data is not accurate enough to make a building level classification.

More automated studies used a combination of texture and Object Based Image Analysis (OBIA) features. Textures and object-based features are hand-crafted features which can be extracted if combined with spectral information. These methods reach a higher accuracy than pixel-based methods especially when working with VHR images because the relation between pixels becomes essential (Kuffer, Pfeffer, & Sliuzas, 2016). Williams, Quincey, & Stillwell, (2016) used OBIA to classify roof objects of the informal settlements to estimate the population.

However expert knowledge is needed to allocate each roof type to a certain SES status. Besides, for the case of Colombo, it is difficult to differentiate between different roof types using VHR images, because most roofs contain the same shape and texture. Therefore this method is not used for determining the SES classification.

All the methods mentioned above are based on indicators which define the SES distribution. Therefore, several physical and household/resident characteristics were examined. The physical indicators were classified under one of the three levels following Kohli et al. (2012): environmental, settlement and object level.

<u>Environmental level</u>

Several environmental indicators were identified that potentially explain a differentiation in socioeconomic status. First of all, poorer people tend to live in more hazardous locations, where formal settlement is not/less possible, for example along highways or in flood zones (Davis, 2006). Slums may also be adjacent to planned areas of major ring roads (Kohli et al., 2012). This can be mapped by calculating the distance to the planned areas or major ring roads.

<u>Settlement level</u>

At settlement level, a classification is made between the formal sector and the informal sector. The informal sector grows organically in the areas where the formal sector does not want to be located or where it is officially forbidden to be located. This areas are often along rivers or in wetlands.

Other indicators at settlement level are building density and the amount of greenery. Building density and vegetation are often used as indicators for determining the SES differentiation because richer areas tend to have more open spaces and greenery than poorer areas (Davis, 2006). However both dense formal neighbourhoods and low density slums are sufficiently common worldwide to ensure ambiguity. Therefore, this indicator can only be used in combination with other indicators to produce a reliable SES index.

Object level

Indicators at object level can be divided into building related indicators and network related indicators. An indicator at object level is for example, the building footprint. Smaller building footprints tend to have a lower SES status than buildings with a higher building footprint (Friesen, Taubenböck, Wurm, & Pelz, 2018); (Graesser et al., 2012). Besides the building footprint, is building orientation also an important indicator for defining the SES differentiation (Kuffer et al., 2016). Slums are, for example, mostly irregular orientated while formal areas are orientated in the same direction. (Kuffer et al., 2016). When looking at the properties of the building, the roof type is also very useful for determining the SES differentiation. Certain types of roofs, such as rusty iron sheets, are more common in slums than in formal settlements (Kuffer, Pfeffer, Sliuzas, Baud, & van Maarseveen, 2017).

Indicators related to the road network are for example the road material (Wong, 2013) and the irregularity of the road network (Sliuzas, Kuffer, & Planning, 2008). The geometry of irregular networks can be calculated by the number of nodes in the network. The unpaved paths can be classified using training samples.

Table 1. Physical characteristics to determine the Socio Economic Status distribution

Level	**Indicators**	**Interpretation element**	**Observation**	**Parameter**	**References**
Environment	Location	Pattern, secondary data	Along highways, on flood zones	Distance to features	(Davis, 2006)
	Neighbourhood characteristics	Pattern, secondary data	Surrounding the planned areas	Distance to planned settlement	(Kohli et al., 2012)
Settlement level	Shape	Pattern	Encircling the major ring road	Geometry – buffer	
	Density	Texture	Denser compared to planned areas Low vegetation	Texture – contrast Geometry – area of vegetation	(Davis, 2006)
Object level	Building	Building footprint	Range of values	Geometry – area	(Friesen et al., 2018); (Graesser et al., 2012)
		Material	Roof material	Spectral – layer mean values	(Kuffer et al., 2017)
		Orientation	Irregular arrangement of buildings	Geometry – angle of buildings	((Kuffer et al., 2016)
	Access network	Shape	Irregular	Geometry – number of nodes, length	(Sliuzas et al., 2008)
		Type	Road material, e.g. unpaved paths	Spectral – training samples	(Wong, 2013)

Household and resident characteristics

Not only physical characteristics define the SES distribution, but household and resident characteristics can define the SES distribution as well. An example is level of education of adult household members (Habitat, 2003). But also the type of cooking (Rutstein & Johnson, 2004), source of drinking water (Rutstein & Johnson, 2004) and households by tenure (Habitat, 2003). These are included in table 2.

Table 2. Household and resident characteristics to determine the Socio Economic Status distribution

Level	Indicators	Interpretation element	Observation	Parameter	References
Household level	Household characteristics	Type of cooking fuel	Firewood, kerosene, gas, electricity	Number of households with e.g. kerosene.	(Falkingham & Namazie, 2002)
		Source of drinking water	Protected well within premises, unprotected well, tap within unit	Number of households with e.g. protected well.	(Rutstein & Johnson, 2004) (Habitat, 2003)
		Households by tenure		Number of households which are e.g. rented	(Falkingham & Namazie, 2002)
	Resident characteristics	Level of education of adult household members	Primary, secondary, degree and above, no schooling	Number of residents who are in the possession of e.g. a primary degree.	(Habitat, 2003)

2.2 Conceptual framework

Resident information is often available at district level while physical characteristics are available at building level. While we wanted to combine the residential and the physical characteristics the following method is developed.

First, the SES differentiation per Grama Niladhari division (further abbreviated to GN divisions) is determined by analyzing local SES indicators at GN level (The GN divisions are subdivisions of the 13 districts of Colombo). The correlation between the indicators is calculated and the average value of the three indicators with the highest correlation is used to make a GN level classification.

The second step is to select two GN districts based on the differentiation within the GN division and the similarities with the survey classification (data explained in chapter 2.2). Both GN districts have to contain a substantial flood risk.

The third step is to distribute the people in the different socioeconomic-status levels at GN district level over the specific buildings per district. While information about the percentage of people in low, moderate and high is now obtained per GN district, an assumption need to be made about the average household size per category to calculate the absolute household numbers per category.

The fourth step is to convert the absolute household numbers to percentages. The percentages are then multiplied with the amount of houses to define how many houses are categorized as low, moderate and high.

The fifth step is to define the locations of the buildings in low, moderate and high. This is done by analyzing different indicators such as building density, building size and distance to roads. The indicators which have the highest (visual) correlation with each other and are validated with google street view are merged into one classification index.

The sixth step is to allocate either ‘low’, ‘moderate’ or ‘high’ to each building based on the mean index values per building and the GN classification which contains the amount of buildings in each category. This is done by sorting the mean index values from high to low (where a high value means a low SES status and a low value a high SES status) and classify the highest X values as ‘low’, the lowest X values as ‘high’ and the other values as ‘moderate’, where X is the amount of buildings.

The last step is to combine this map with the already classified shanty buildings of the Survey data. The classification of the shanty buildings overrules the other indicators, and are automatically categorized as ‘low’.

The overall method is shown in figure 1. This method can be used for different districts. However it should be taken into consideration that the building level indicators can be different for each district.

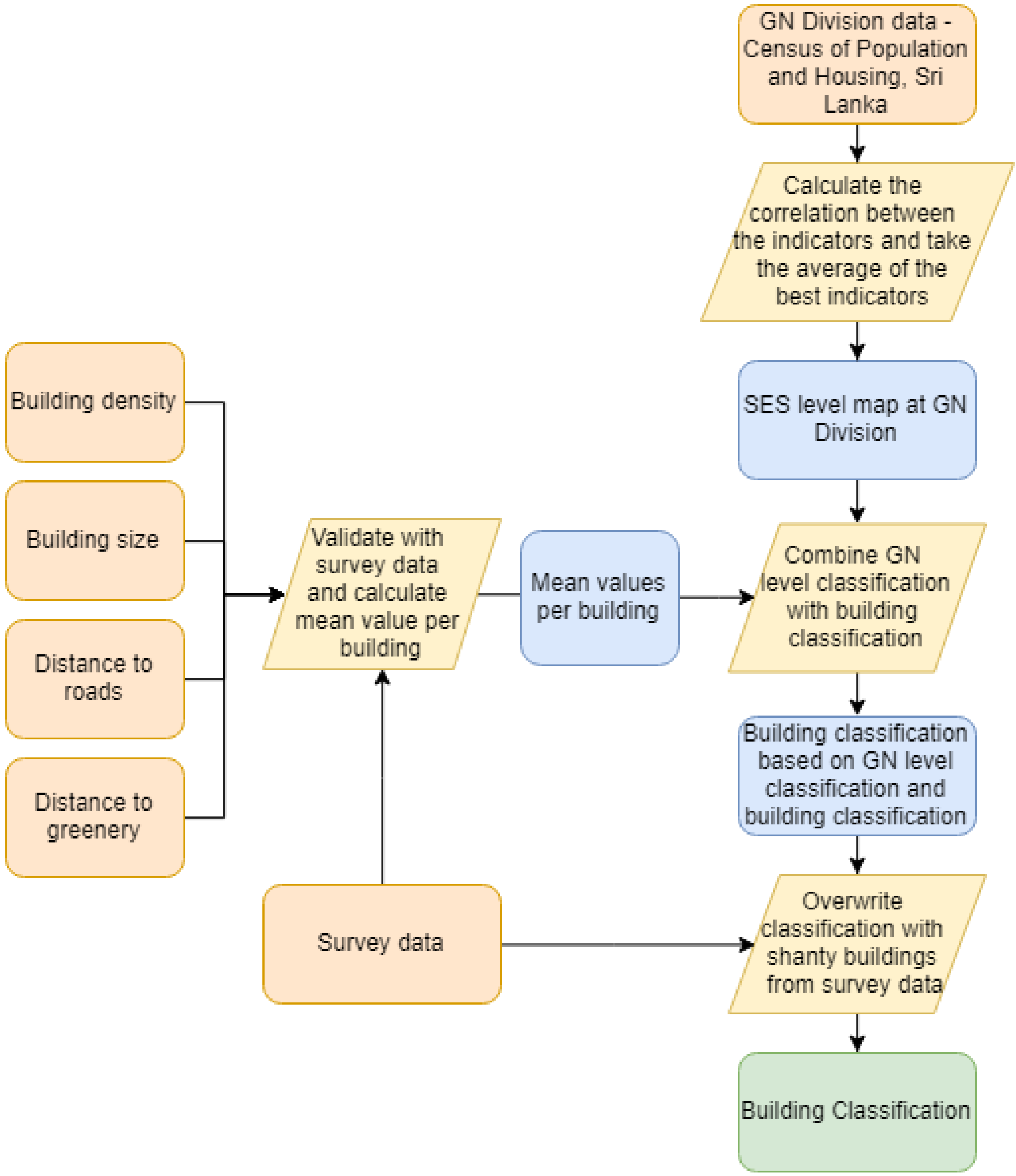


Figure 1. Conceptual framework

2.3 Available data

The conceptual framework required both open and survey data sources. The open data sources used are; census data from the department of statistics in Colombo, Open Street Map and satellite imagery. An overview of the datasets used is shown in table 3.

Survey data at building level, which is measured through field observations, is available in a shapefile. This contains information about the condition of the building, rated in ‘poor’, ‘moderate’ and ‘good’. The ‘poor’ condition buildings are defined as ‘shanties’, and covers approximately 11,3% of the total buildings. The ‘moderate’ condition is labelled to approximately 12% of buildings, while the ‘good’ condition covers 76,7%, Although the detailed spatial availability of this data is a big advantage, the data seems to underrepresent the number of people living in poorer conditions. To test this, and to get a wider distribution of the different categories, second data sources are used, looking at indicators of SES status at local GN level and open data from satellite images and open street map.

Census data at GN level. The department of census and statistics in Sri Lanka has conducted a census survey among the population in 2012. The data contains information about the inhabitants (e.g. age, economic activity and educational attainment) and about the housing of the inhabitants (e.g. construction material, type of roof, source of drinking water etc.). A drawback is that the census data is only available at GN level and not at building level. However, since the census data contains a lot of residential indicators (mentioned in table 2) it is used to create a SES distribution map at GN level, which will be used to determine the locations containing the highest differentiation between low and high. These districts will be further analyzed at building level using more detailed information from Open Streep Map and satellite imagery.

Open Street Map is used to get shapefiles of the roads and waterways in Colombo. These can be used for determining the distance to highways and rivers, which are indicators mentioned by Davis, 2006.

Open satellite data sources were examined to find the best resolution imagery for Colombo, Sri Lanka. However, it turned out that even the sentinel 2 satellite imagery did not contain the required resolution for classifying a district at building level. For example, Sentinel 2 contained a 60 m spectral resolution. While most articles which define SES status use 0.6 m resolution e.g (Kohli, Sliuzas, & Stein, 2016).

To still get a high resolution imagery, snapshots were taken from google earth. The snapshots of google earth were then georeferenced in ArcMap.

The satellite images can be used to detect the amount of greenery per GN district, because an area containing a lot of greenery is more likely to have a higher SES status than an area which contains no or less greenery (Davis, 2006).

Table 3. Data sources

Title	Source	Year	Scale	Format	Fit for purpose
Satellite imagery – Google Earth	Google	2018	65 cm spatial resolution	.png	To determine the distance to greenery
Open Street Map (for roads and buildings)	OSM	-	-	.shp	To determine the distance from the buildings to the roads as indicator
Census of Population and Housing	Department of Census and Statistics, Sri Lanka	2012	DN level	.pfd	For creating a SES classification at GN level
Field observation data	Students in Colombo, Sri Lanka	2016	Building level	.shp	For validation of SES differentiation.

3. Results

3.1 GN Classification

The first step in making a SES classification at GN level was to define the indicators. Several indicators; level of education of adult household members, (Falkingham & Namazie, 2002), type of cooking fuel (Rutstein & Johnson, 2004) source of drinking water (Rutstein & Johnson, 2004) roof material, wall material, floor material and households by tenure (Habitat, 2003) were selected by literature research and were validated by experts in Colombo with local knowledge on the living conditions of the inhabitants.

Each indicator was then subdivided into three classes: low, moderate and high. Each indicator consisted of several sub-indicators, that had to be classified in either 'low', 'moderate', or 'high'. This clustering has been based on the original order in the census data and the validation of the experts in Colombo. The classification table can be found at Appendix A.

To further validate the selection of the indicators and clustering of sub-indicators, the Pearson correlation method was used to calculate the correlation. The resulting coefficient is a measure of the strength of the linear relationship between two variables.

The correlation between the indicators is calculated for each category (low, moderate and high). This results into three correlation tables, which were merged to one table by taking the mean values of the three tables. The resulting table (table 4) shows that three of the six indicators contain a high correlation with each other. These indicators are source of drinking water, construction material floor and education.

Table 4. Pearson correlation between indicators

	Source of Drinking Water	*Households by Tenure*	*Construction Material Floor*	*Construction Material Roof*	*Construction Material Wall*	*Type of Structure*	*Education*
Source of Drinking Water	1						
Households by Tenure	0.06	1					
Construction Material Floor	0.59	-0.058	1				
Construction Material Roof	0.12	0.32	0.11	1			
Construcion Material Wall	0.17	0.13	0.15	0.15	1		
Type of Structure	0.3	0.24	0.38	0.23	-0.15	1	
Education	0.48	0.02	0.78	0.20	0.12	0.44	1

However the correlation between the other indicators turned out to be very low. It was first tested whether this would change if the clustering of the sub-indicators changes (e.g. assigning a roof material to low instead of moderate). However the correlation remained very low, therefore these indicators were excluded as SES indicator.

The second step was to combine the different SES indicators by calculating the mean value of each of the categories; low, moderate and high per GN district. This results in three maps showing the distribution for each of the three categories low, moderate and high. The distribution of 'low' and 'high' is shown in figure 2, the distribution of the moderate category can be found in Appendix B.1. The classification method used is natural breaks (jenks), which is based on natural grouping inherent in the data. The class breaks are defined by the method, which seeks to reduce the variance within classes and maximize the variance between classes.

SES Distribution at GN level

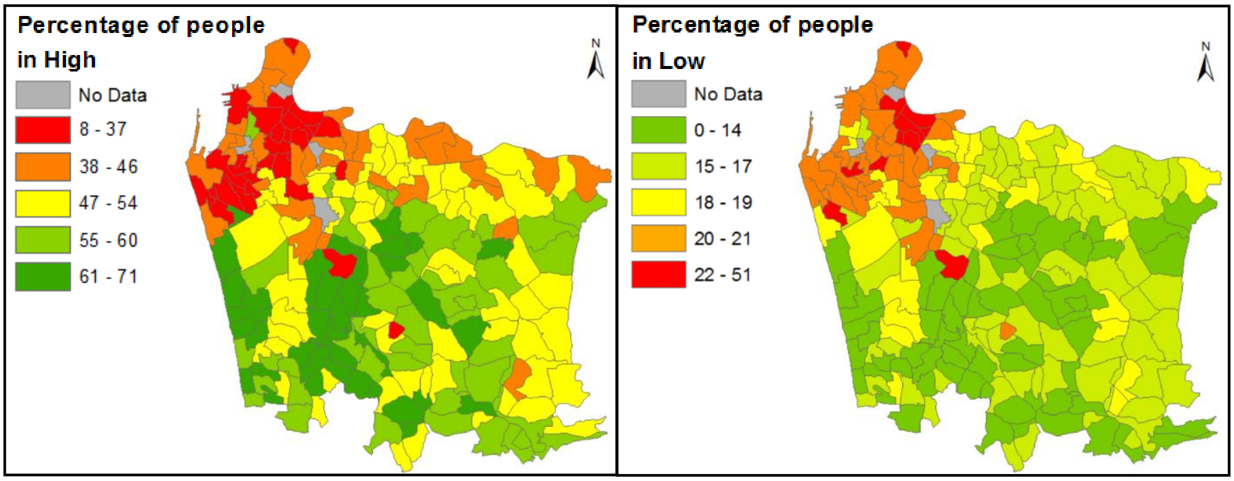


Figure 2. Distribution of low and high by source of drinking water, education and floor material

As was to be expected, the low and high maps show very similar results, which is explainable as a district with a high score in the category low will automatically have a low score in the category high (unless the moderate category is very big). The GN districts which do not have the same classification in low and high are more likely to contain a high distribution between low and high. The moderate map merely indicates the number of people in the middle category, and therefore does not have any values/colour codes of 'bad' (red) or 'good' (green) assigned to it.

The third step was to select two districts, the first one based on the comparison between the SES map at GN level and the survey building level classification, and the second one based on the differentiation within the district. Both districts should contain a substantial flood risk.

Figure 3 shows the GN classification map and the survey data classification. The map containing the flooding extent can be found in Appendix B.2

Comparison between GN level classification and survey classification

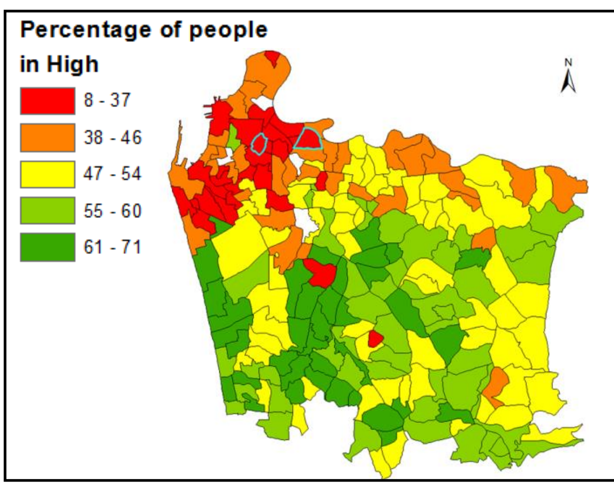


Figure 3a. Mean 'High' Values of Source of drinking water, Floor material and Education per GN District, selected areas: GrandPass High (left), Kotuvila (right)

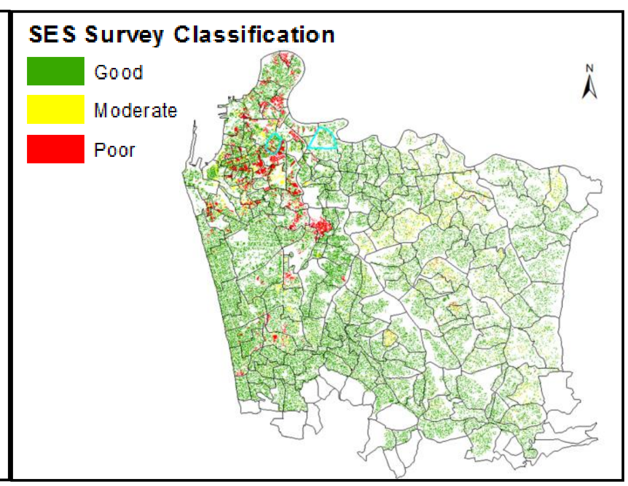


Figure 3b. SES Classification at Building level by Survey Data, selected areas: GrandPass High (left), Kotuvila (right)

Because of the similarity of the low and high map (figure 2), the 'high' map is shown to display the similarities with the survey data map. Generally, the areas classified as poor in the survey data map are located in a red coloured district at the GN level map, which means they contain the lowest percentage of people in ' high' compared to the other GN districts.

Outstanding is the orange area at the north in the middle of the map. This is classified as green in the survey data map. This is explainable, because this area contains a lot of social housing, which means that the condition of the buildings is good, but the people living in the buildings are actually very poor. The GN classification method used level of education as indicator which is resident related indicator instead of a building characteristic.

The first district chosen, Kotuvila, lies near this area (the selected district on the right in figure 3a). Kotuvila is classified as 'high' on the survey data map but as 'red' at the GN-level map and contains a substantial flood risk.

When looking at the GN-level classification maps a second GN district is selected which contains a high percentage of people in the low category and a high percentage of people in the high category. The survey data is used to validate if the GN district actually contains shanty building. A GN district which meets these conditions and which contains a substantial flood risk is the 'GrandPass High' district (the district on the left in figure 3a).

3.2 Building-level classification

To determine the locations of the buildings, the indicators; building density, building size, distance to greenery and distance to the nearest road were analyzed.

Because we expected the shanty buildings in the survey data to be correctly classified, the survey data is shown in figure 4.

Survey Classification

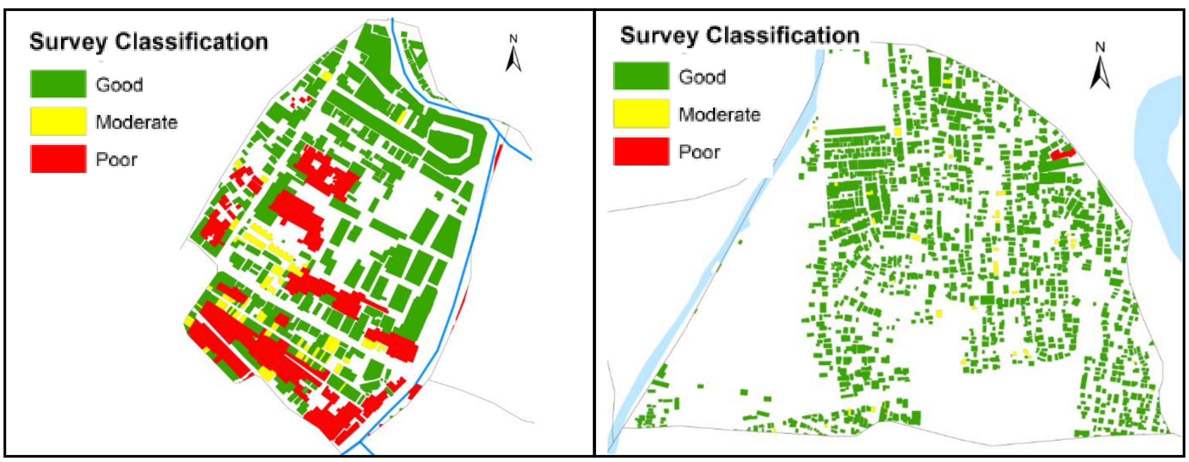


Figure 4a. Survey classification of GrandPass High

Figure 4b. Survey classification of Kotuvila

Building density is an often used indicator for defining the SES distribution. Areas with a lower SES status generally have high roof coverages with no spaces or vegetation (Davis, 2006).

When comparing the resulting building density map for GrandPass North (figure 5a) with the Survey classification in figure 4a, it turns out that the differentiation between the categories is quite similar. Kotuvila shows more differences between the survey data and the density map. However, when looking at google street view, the area classified as 'low' in the density map actually is a very poor area. Which concludes that the density classification may be more accurate than the survey data classification. The white space on the left of the buildings is occupied by oil industry. It would be a logical conclusion that this is not a very nice area to live, which indicates that there is a high change on slum settlements. A more detailed analyses about Kotuvila can be found in Appendix C.1

Building density

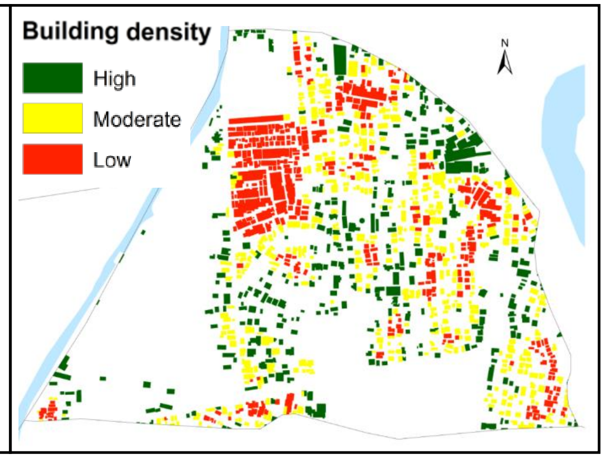
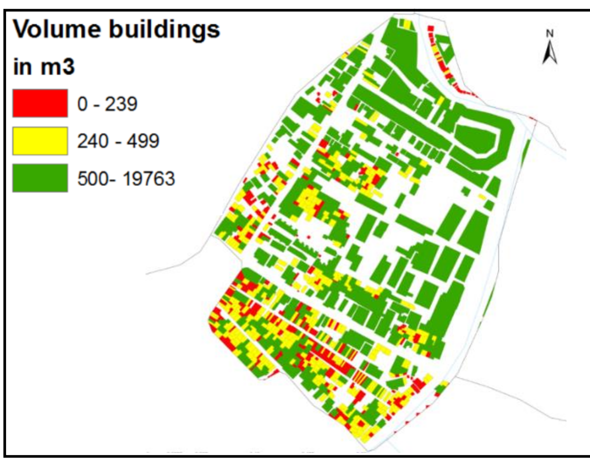


Figure 5a. Building density of GrandPass North *Figure 5b. Building density of Kotuvila*

Just like building footprint density is **building size** also an import indicator for defining the SES differentiation. Small buildings have a high probability of indicating a shabby building i.e. a building located in a slum (Kohli et al., 2012).

When comparing figure 4 and 6, it shows that the building size map of GrandPass North does actually match the Survey classification map quite well. The buildings with an average and low volume (yellow and red buildings) are classified as low in the Survey classification. The volume map of Kotuvila shows the same pattern as the density map in figure 5b, most areas categorized as low are located in the North-West of the map, where a lot of small buildings contain a high building density.

Building size



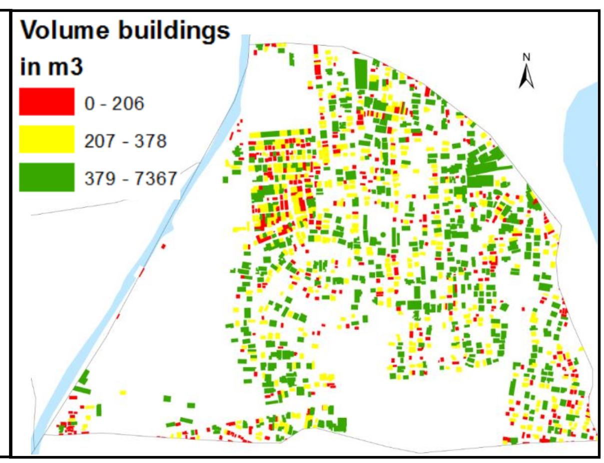


Figure 6a. Building sizes in GrandPass North *Figure 6b. Building sizes in Kotuvila*

(Sliuzas et al., 2008) concludes that small slums lack the access to roads and cluster along public infrastructure, i.e. highways, canals and railways. While the areas chosen do not contain much canals, railways and highways, the accessibility to the roads is mapped by calculating the **distance to the nearest road** for each building.

The resulting map, shown in figure 7, does not exactly match the Survey classification map in figure 4, but the pattern is recognizable. The red areas in the South-West of GrandPass North are also classified as 'shanty' in the Survey classification. The red area in the North-West of Kotuvila is also classified as 'low' by the density and building size indicator (figure 5b and 6b). Therefore this is also considered a good indicator for defing the SES distribution.

Distance to nearest roads

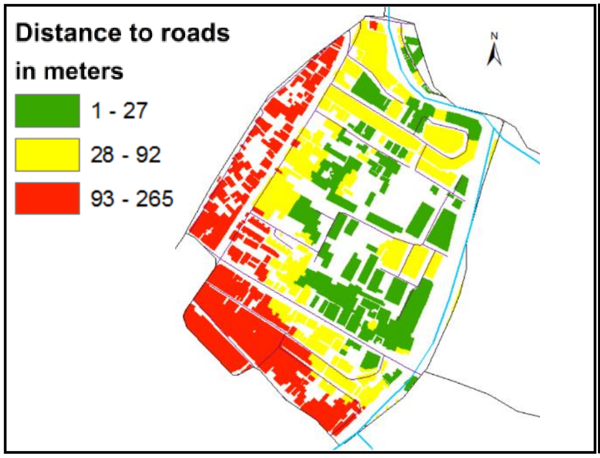


Figure 7a: Distance to nearest road in GrandPass North

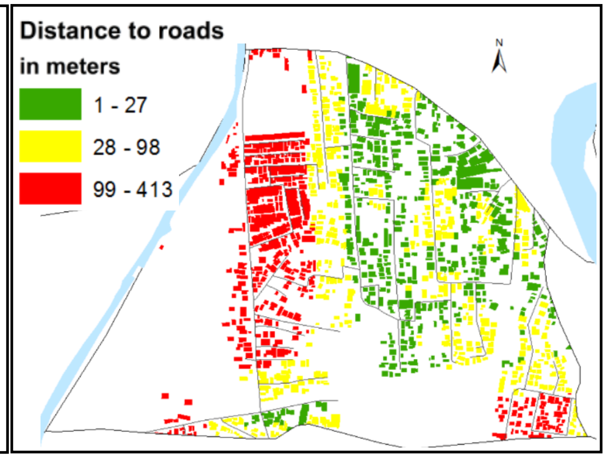


Figure 7b: Distance to nearest road in Kotuvila

The building level indicators are divided into three categories based on the quintile classification. This means that each class contains approximately the same class width, which gives the colours on the map about equal area. This results in a clear distribution between the colours on the map. Three categories are chosen because the Survey classification also contains three categories. This makes it easier to compare the maps with each other. Appendix C.2 gives a short summery of the tools in ArcMap used to create the resulting maps.

It turned out that distance to greenery was not a good SES indicator because of the large amount of wetlands in Colombo. The center of Colombo is build up area but the outskirts of the center are located in the wetlands of Colombo. This means that, even in the poorer neighbourhoods, an abundance of green is present. Therefore this indicator is not used for determining the SES differentiation. A more detailed analysis of distance to greenery can be found in appendix C.3

3.3 GN level and building level combined

To combine the different indicators with each other, the amount of people per category at GN level need to be converted to the amount of households per category. To achieve this, an assumption about the average household size is needed. The average household size estimate originates from Michael Bauer Research, which is made available by esri in a shapefile. The shapefile consists of subdistricts which all hold a minimum and maximum household value in each subdistrict in Sri Lanka. The minimum household value is used for the category high, and the maximum household value for the category low. The average of the two values is assigned to the moderate category. The amount of households in each category is calculated by dividing the amount of people by the average household sizes. The amount of households is then converted to percentages. With the percentage of households in each category, the amount of buildings in each category is calculated.

The next step was to convert the indicator values per building to values between 0 and 1. The three indicators used; building density, building volume and distance to nearest roads, are then combined into one resulting index by taking the average value.

The next step is classify the X highest index values as low (high index value means a low SES status), the X lowest index values as high and the remaining values as moderate, where X is the amount of buildings.

A subsequent step is taken to exclude all non-residential functions For example, 228 residential buildings are categorized as low in table 5, the index values are then sorted from high to low and the 228 residential buildings with the highest values (a high value means a low SES status) are classified as 'low'. The same method is used to classify the buildings in GrandPass North.

Table 5. GN Classification results for Kotuvila

Kotuvila	*Total*	*Low*	*Moderate*	*High*
Population	5083	1127	2173	1781
Household average		6,8	5,55	4,3
Amount of households	971	166	392	414
Households %		17%	40%	43
Amount of buildings	1447	247	583	617
Amount of buildings residential	1338	228	539	571

The resulting building classification is then combined with the already classified shanty buildings. Kotuvila did not have any shanty buildings classified in the survey classification, so only the map of GrandPass North changed. The resulting maps are shown in figure 8a and 8b.

SES Classification GrandPass North and Kotuvila

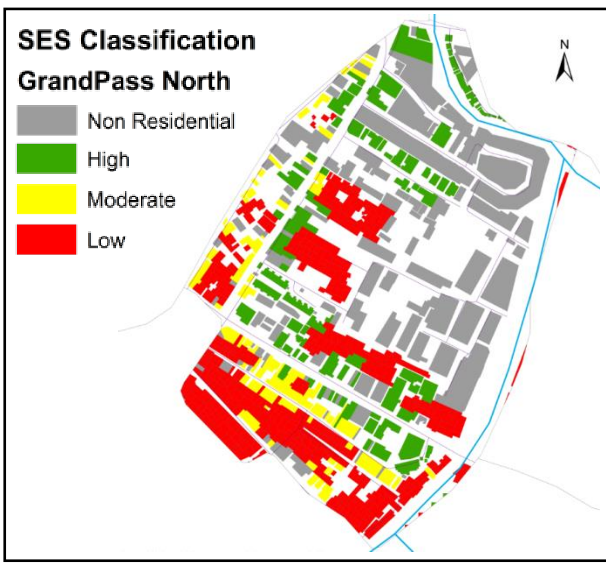


Figure 8a. Resulting SES Classification of GrandPass North

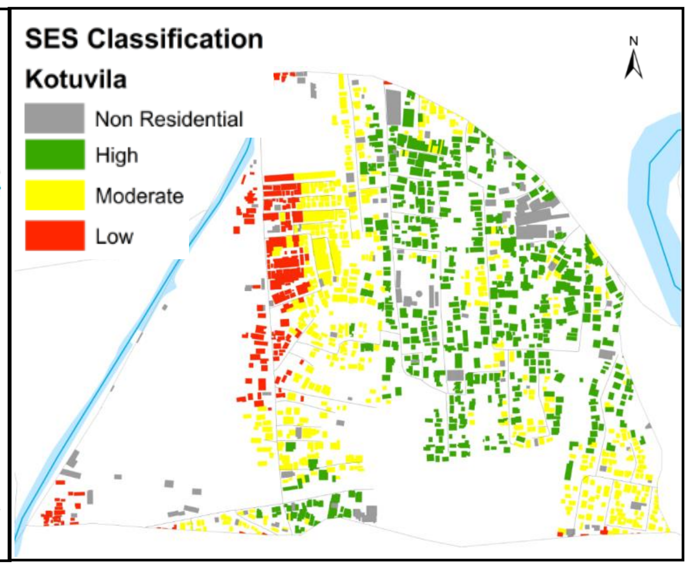


Figure 8b. Resulting SES Classification of Kotuvila

The SES Classification of GrandPass North was expected to contain more buildings classified as 'low'. However, this is not the case because of the large amount of non-residential functions in this area. The SES classification of Kotuvila meets the expectations better. The areas classified as 'low' are surrounding the oil industry which is not a nice area to live, which indicates that there is a high change on slum settlements. The Survey classification did not show this area as 'low' so the developed classification method may be more accurate.

4. Discussion

The maps created by the developed method show exactly the expected SES classification . For example, the map about Kotuvila shows that there are more buildings classified as low than the survey classification shows, which means that the GN level indicators; level of education (Falkingham & Namazie, 2002), source of drinking water (Rutstein & Johnson, 2004) and the construction material of the floor (Habitat, 2003) were good indicators for defining the SES distribution at GN level. Besides, building density (Kohli et al., 2012), building size (Kohli et al., 2012) and distance to the nearest road (Sliuzas et al., 2008) were derived from literature as indicators for defining the SES distribution at building level. The resulting map, which combines these indicators, does meet the expectations about the location of the ‘low’, ‘moderate’, and ‘high’ category. Which leads to the conclusion that these indicators also give a good indication about the SES distribution.

However, this method contains some assumptions which may not be perfectly correct, for example the adaption of the average household size. Different household sizes lead to different amounts of households in the low, moderate and high category, which results in different amounts of buildings classified as low, moderate and high. This affects the resulting map.

Uncertainty also occurs in the selection of the indicators. Both dense formal neighbourhoods and low density slums are sufficiently common worldwide to ensure ambiguity. For the case of Colombo, it turned out that a high building density matches the buildings classified as ‘low’. However, in other countries this may not be the case.

Scaling the method up for the other GN districts in Colombo is possible. However, the GN districts selected in this article contained no more than 1 or two stories, other districts may contain bigger buildings with more households living in it. This should be taken into account when scaling it to the whole of Colombo.

Another limitation for scaling the method up is that excel is used to determine the building classification e.g. select the highest X% values and classify it as ‘low’. This is done manually, which is a lot of work when applying it for each GN district. However, this problem could possibly be solved by writing a python script which automates this process.

5. Conclusion

The advantage of the developed method is that this way the census information about the population is taken into account by determining the amount of people in low, moderate and high and the building characteristics are taken into account by determining the location of each category. So both the resident characteristics at GN level and physical characteristics at building level are used to make the SES classification. The disadvantage of the method is that is that some indicators should be adapted when using it for different GN divisions. Besides, when using this method for the whole of Colombo, it will be very time consuming. This is due to the combination of the GN level classification and the building level classification, which is done manually using excel. However, this problem could possibly be solved by writing a python script which automates this process. Altogether, the conclusion can be drawn that the method is suitable, but especially for specific districts from which the SES indicators are known.

References

Baker, E. H. (2014, February 21). Socioeconomic Status, Definition. *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society*. https://doi.org/doi:10.1002/9781118410868.wbehibs395

Davis, M. (2006). *Planet of slums*. London; New York: Verso.

Falkingham, J., & Namazie, C. (2002). Measuring health and poverty: a review of approaches to identifying the poor. *Health Systems Resource Centre*, *44*(0), 70. https://doi.org/10.1111/j.1467-7660.2010.01678.x

Friesen, J., Taubenböck, H., Wurm, M., & Pelz, P. F. (2018). The similar size of slums. *Habitat International*, *73*(September 2017), 79–88. https://doi.org/10.1016/j.habitatint.2018.02.002

Graesser, J., Cheriyadat, A., Vatsavai, R. R., Chandola, V., Long, J., & Bright, E. (2012). Image Based Characterization of Formal and Informal Neighborhoods in an Urban Landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(4), 1164–1176. https://doi.org/10.1109/JSTARS.2012.2190383

Habitat, U. N. (2003). *Slums of the World: the face of urban poverty in the new millennium*. *UN-Habitat, Nairobi*.

Kind, J., Wouter Botzen, W. J., & Aerts, J. C. J. H. (2017). Accounting for risk aversion, income distribution and social welfare in cost-benefit analysis for flood risk management. *Wiley Interdisciplinary Reviews: Climate Change*, *8*(2), 1–20. https://doi.org/10.1002/wcc.446

Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, *36*(2), 154–163. https://doi.org/10.1016/j.compenvurbsys.2011.11.001

Kohli, D., Sliuzas, R., & Stein, A. (2016). Urban slum detection using texture and spatial metrics derived from satellite imagery. *Journal of Spatial Science*, *61*(2), 405–426. https://doi.org/10.1080/14498596.2016.1138247

Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space-15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6). https://doi.org/10.3390/rs8060455

Kuffer, M., Pfeffer, K., Sliuzas, R., Baud, I., & van Maarseveen, M. (2017). Capturing the diversity of deprived areas with image-based features: The case of Mumbai. *Remote Sensing*, *9*(4). https://doi.org/10.3390/rs9040384

Motholo, G. L. (2014). Inferring urban household socio-economic conditions in Mafikeng, South Africa, using high spatial resolution satellite imagery AU - Munyati, Christopher. *Urban, Planning and Transport Research*, *2*(1), 57–71. https://doi.org/10.1080/21650020.2014.901158

Shea Oscar Rutstein, & Kiersten Johnson. (2004). The DHS Wealth Index. *DHS Comparative Reports No. 6*, (December). Retrieved from http://www.measuredhs.com/pubs/pdf/CR6/CR6.pdf

Sliuzas, R., Kuffer, M., & Planning, R. (2008). Analysing the Spatial Heterogeneity of Poverty Using Remote Sensing : Typology of Poverty Areas Using Selected. *Population (English Edition)*, (i), 158–167.

Williams, N., Quincey, D., & Stillwell, J. (2016). Automatic Classification of Roof Objects From Aerial Imagery of Informal Settlements in Johannesburg. *Applied Spatial Analysis and Policy*, *9*(2), 269–281. https://doi.org/10.1007/s12061-015-9158-y

Wong, D. W. (2013). Exploring structural differences between rural and urban informal settlements from imagery: the basureros of Cobán AU - Owen, Karen K. *Geocarto International*, *28*(7), 562–581. https://doi.org/10.1080/10106049.2012.734533

Zhang, Y., Li, Q., Huang, H., Wu, W., Du, X., & Wang, H. (2017). The combined use of remote sensing and social sensing data in fine-grained urban land use mapping: A case study in Beijing, China. *Remote Sensing*, *9*(9). https://doi.org/10.3390/rs9090865

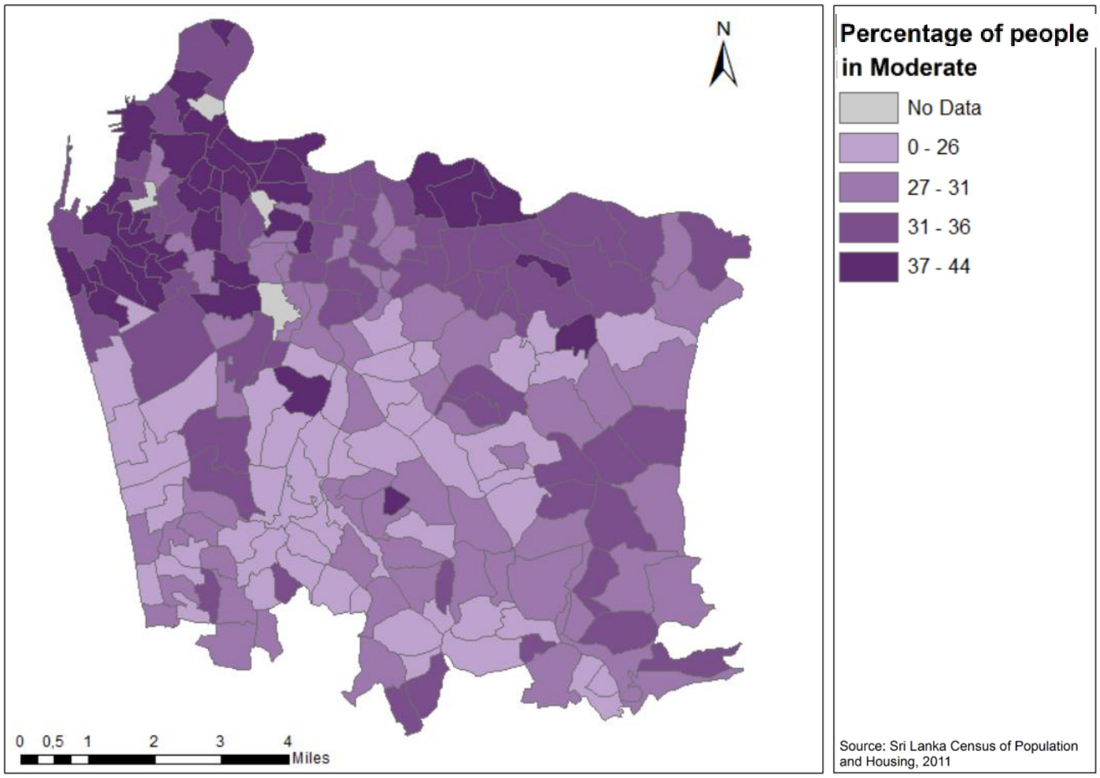
Appendices

A. Table for GN Level Classification

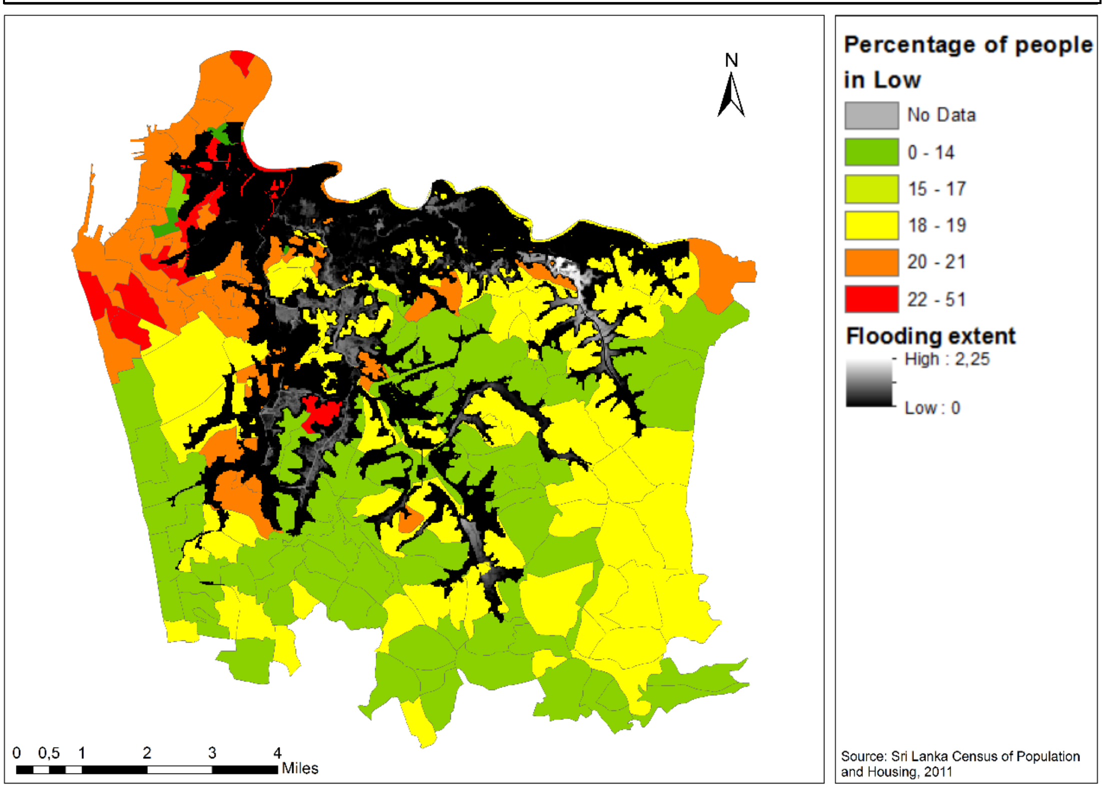
	Low	**Moderate**	**High**
Source of drinking water	Unprotected well Bowser River/Tank/Stream Rain water Bottled water Protected well outside premises	Tap within premises but outside unit (main line) Tap outside premises (main line) Tube well	Tube well
Education	No schooling Primary	G.C.E. (O/L)-Secondary	G.C.E. (A/L) Degree and above
Households by tenure	Rent free occupied Encroached Other	Owned by a household member Rent/Lease Government owned Rent/Lease Privately owned Rent free occupied	Owned by a household member
Construction Material Floor	Cement Mud Wood Other	Cement Concrete	Tile/Granite/Terrazo Process Wood
Construction material roof	Cadjan/Palmyrah/Straw Takaran sheets Other	Tile Asbestos	Tile Asbestos Concrete
Construction material wall	Cabook Soil bricks Mud Cadjan/Palmyrah Plank/Metal Sheet Other	Brick Cement block	Brick Cement block/Stone Stone
Type of Structure	Row/Line room Hut/Shanty	Single 1 story Attached house/Annex Low cost Flat Twin house	Single 1 story Single 2 story Flat/Apartments/Condominium Twin house

B. SES Classification Maps

B.1 SES Differentiation of the Category: Moderate



B.2 SES Classification of the ‘Low’ Category combined with flooding extent



C. Building level indicators

C.1 Building density validation

The map below shows the density map of Kotuvila divided into 5 categories, based on the quintile classification. This shows a more specific distribution of the building density in the area. Google Street View images are used to validate each area. As shown in the pictures, the red buildings in the map are mostly small and poor buildings, the bright green areas are most often large buildings with a big garden, and the areas between, coloured orange, yellow and lightgreen are mostly the ‘ moderate’ category, which contains one or two story buildings with an average sized garden.

SES distribution Kotuvila

Poor area near oil industry[1]

Rich/Moderate area near river[2]

Figure 9: Density map of Kotuvila

Moderate area with one and two stories[3]

[1] Google street view image, consulted on 20 January 2018
[2] Google street view image, consulted on 20 January 2018
[3] Google street view image, consulted on 20 January 2018

C.2 Classification in ArcMap

<u>Building density</u>

The first step for making the density classification was to convert each building to a point. Then the point density was calculated using ArcMap, which was then converted to building level using the zonal statistics tool which took the mean density value per building.

<u>Building size</u>

To determine the building size, the building footprint area is calculated. While information of the amount of stories per building is available, the building area in m2 is multiplied by the height of the building. To achieve this, the assumption is made that one story is approximately 2.8 meter.

<u>Distance to Roads</u>

The roads from Open Street Map where downloaded and converted to a shapefile format to add the layer in ArcMap. The distance to the roads is calculated using the Euclidean distance tool. This results in a map where each point contains the distance to the nearest road. The zonal statistics tool is used to average the distance values per building.

C.3 Distance to greenery

Open data in the form of satellite images is for Colombo, Sri Lanka available by sentinel-2 at 60 m spatial resolution. This resolution is not high enough to make a good distinction between the buildings and thus for determining the SES differentiation. Kohli et al., (2016), for example, used a cloud-free pan-sharpened Quickbird satellite image of 0.6 m resolution to defect slums in urban areas.

To get high resolution imagery, google earth is used. Most of the VHR images of google earth are from DigitalGlobe Quickbird, which is roughly 65 cm sharpened. The snapshots of google earth were georeferenced in ArcMap and then combined using the 'mosaic to new raster' tool.

The supervised classification method is used to classify the greenery in each GN district. The classes needed for identifying greenery were known in advance, which makes this method most likely to use. The Euclidean distance tool was used for determining the distance to the greenery, subsequently the zonal statistics tool was used to determine the distance to greenery per building. Figure 10 shows the resulting maps. When comparing these maps to figure 4 in the report it shows that there are almost no similarities between the 'low' classified buildings and the shanty buildings. This is explainable by the large amount of greenery in Colombo.

Distance to Greenery in GrandPass North and Kotuvila

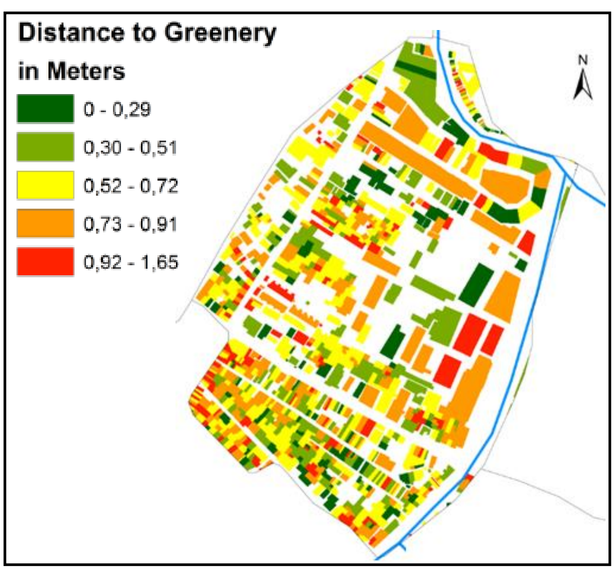


Figure 10a: Distance to greenery in GrandPass North

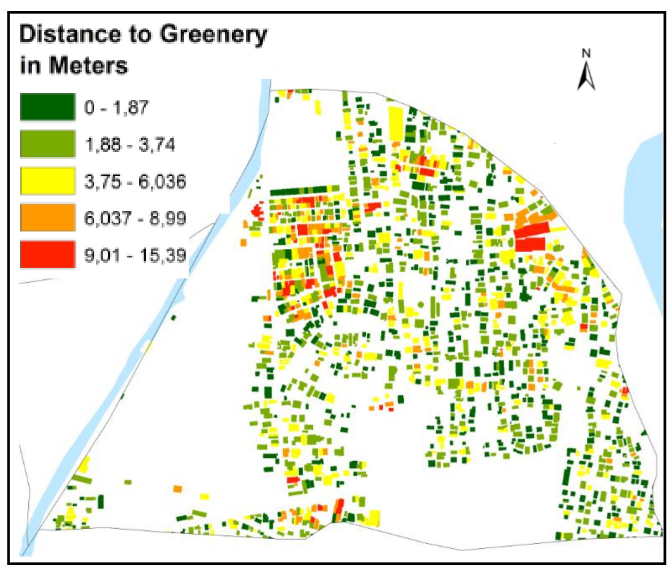


Figure 10b: Distance to greenery in Kotuvila