

Escuela de Economía y Finanzas

Documentos de trabajo

Economía y Finanzas

Centro de Investigación
Económicas y Financieras

No. 13-23
2013

Quantifying slumness with remote sensing data

Duque, Juan C.; Patiño, Jorge; Ruiz, Luis A.; Pardo-Pascual, Josep E.



Quantifying slumness with remote sensing data

*Juan C. Duque*¹
*Jorge E. Patino*²
*Luis A. Ruiz*³
*Josep E. Pardo-Pascual*³

June 26 2013

Abstract

The presence of slums in a city is an indicator of poverty and its proper delimitation is a matter of interest for researchers and policy makers. Socio-economic data from surveys and censuses are the primary source of information to identify and quantify slumness within a city or a town. One problem of using survey data for quantifying slumness is that this type of data is usually collected every ten years and is an expensive and time consuming process. Based on the premise that the physical appearance of an urban settlement is a reflection of the society that created it and on the assumption that people living in urban areas with similar physical housing conditions will have similar social and demographic characteristics (Jain, 2008; Taubenböck et al., 2009b); this paper uses data from Medellin City, Colombia, to estimate slum index using solely remote sensing data from an orthorectified, pan-sharpened, natural color Quickbird scene. For Medellin city, the percentage of clay roofs cover and the mean swimming pool density at the analytical region level can explain up to 59% of the variability in the slum index. Structure and texture measures are useful to characterize the differences in the homogeneity of the spatial pattern of the urban layout and they improve the explanatory power of the statistical models when taken into account. When no other information is used, they can explain up to 30% of the variability of the slum index. The results of this research are encouraging and many researchers, urban planners and policy makers could benefit from this rapid and low cost approach to characterize the intra-urban variations of slumness in cities with sparse data or no data at all.

Keywords

Regional Science, Remote Sensing, Slum, GEOBIA

JEL Clasification

C8 Data Collection and Data Estimation Methodology, Computer Programs.
R14 Land Use Patterns.

¹ Research in Spatial Economics (RISE-group), Department of Economics, EAFIT University,
e-mail: jduquec1@eafit.edu.co
Phone: (574) 261-9354, Fax: (574) 261-9294

² Research in Spatial Economics (RISE-group), Department of Economics, EAFIT University

³ Geo-Environmental Cartography and Remote Sensing Group, Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València

* The authors wish to thank EAFIT University and Medellin City Hall EnlazaMundos program for their financial support. The usual disclaimer applies.

1 Introduction

Today the world population is mostly urban and information about urban areas, their internal differences and their dynamics is very important to the preservation of certain standards of living (Phinn et al., 2002). According to United Nations more than half of the world population lives in urban areas: 78% in the more developed countries and 47% in the less developed regions of the planet (United Nations, 2012). The percentage of urban dwellers increased from 43% in 1990 to 52% in 2011 and it is expected to grow to 67% in 2050 (United Nations, 2007, 2008, 2012). More important, all the population growth from 2011 to 2050 is expected to be absorbed by urban areas, and most of this growth will take place in the cities of the less developed regions of the world (United Nations, 2012). The variations in poverty and health within urban areas in developing countries can be larger than the differences between urban and rural areas (Montgomery and Hewett, 2005; Stow et al., 2007); thus inequities within cities and metropolitan areas are of major concern and the projected population growth confirms that cities and urban settlements must be studied with maximum detail. This analysis becomes even more important when urban growth occurs in a disordered way due to rural urban migrations to unplanned urban spaces (Duque et al., 2012b) as we see it in most of the cities of the less developed countries in the world.

Medellin city is the second largest city in Colombia with a population of 2.4 million inhabitants and it has grown rapidly in the last decades. Political violence, guerrillas and paramilitary presence, and drug war in the rural areas of the country have caused important migrations of farmers and peasants from rural areas to the cities, which impacted urban population growth rates since the mid 20th century. In Medellin this demographic growth has been concentrated in the northern periphery of the city in unplanned spaces, exceeded the capacity of the local government to deliver services and infrastructure and caused the appearance of slums (Duque et al., 2012b).

According to UN-Habitat, a slum household is a group of individuals living under the same roof in an urban area that lacks one or more of the following: durable housing of a permanent nature, sufficient living space (not more than three people sharing the same room), easy access to safe water in sufficient amounts at an affordable price, access to adequate sanitation in the form of a private or public toilet shared by a reasonable number of people, and security of tenure.¹ The presence of slums in a city is an indicator of poverty and deprivation, and United Nations (2011) estimates that about one-third of the urban population in the less developed countries lives in slums.

The delineation of slums in Medellin city has the same caveats as in many cities of the developing countries: it is often based on subjective aspects because socio-economic surveys are very expensive, time consuming, and are only statistically significant for too large spatial units to be useful for urban planning and policy making. The local government makes a socio-economic survey every year that costs approximately 450.000 USD and takes almost a whole year for getting aggregated results for city communes.² An additional limitation of field surveys is that respondents in deprived areas often give erroneous or no information for fear that their responses could be used against them (Galeon, 2008; Owen and Wong, 2013). However, the proper identification of slum areas is a central issue for policy makers since it can help to prevent poverty traps and crime nests (Duque et al., 2012b), and to allocate public investments where most needed within the city.

Since the slum index, as defined by UN-Habitat, is based on physical aspects of dwelling units, it is feasible to think that remote sensing can deliver proxy measures of it with sufficient accuracy to delineate the slum areas within a city. Indeed, the relationships between urban land cover and environmental factors and socio-economic variables have been explored since the 1950s with the use of aerial photography (Green, 1956, 1957), and have been established in recent years in different cities around the world using satellite remote sensing (Patino and Duque, 2013). These relationships are based on the premise that the physical appearance of a human settlement is a

¹ http://www.unhabitat.org/documents/media_centre/sowcr2006/SOWCR%205.pdf accessed 14 March 2013.

² A commune is an administrative unit that can be larger than a small town, with a total population ranging from 85.000 to 195.000 inhabitants (DANE and Municipio de Medellín, 2010).

reflection of the society that created it and on the assumption that people living in urban areas with similar physical housing conditions will have similar social and demographic characteristics (Jain, 2008; Taubenböck et al., 2009a).

The main advantage of using remote sensing as data source for slum index quantification in a city is that this kind of data can be obtained quicker and for a fraction of the cost than censuses, surveys and *in situ* measurements and for time periods much shorter than those of field surveys and census data, which in most countries are collected every 10 to 12 years (Patino and Duque, 2013). On the other hand, the processing time required since image acquisition up to the obtaining of results can be considerably shorter when compared to those of censuses and field surveys; and the repetitive nature of most satellite remote sensing systems opens the possibility of updating slum index city maps in an objectively way and in a regular basis with little human interaction once a statistical model is built and fitted to local conditions. Finally, the remote sensing approach to slum detection allows comparative analysis of multiple locations as long as environmental conditions, like weather and the availability building materials, be similar between those locations.

Previous works have explored the use of remote sensing as a proxy for socio-economic class differentiation within a city (Avelar et al., 2009; Tapiador et al., 2011), and several works have demonstrated the usefulness of remote sensing to distinguish a slum from its surrounding neighborhoods in the last decade (Barros, 2008; Hofmann et al., 2008; Kit et al., 2012, 2013; Kohli et al., 2012; Netzband et al., 2009; Niebergall et al., 2007; Owen and Wong, 2013; Rhinane, 2011; Stow et al., 2007; Shekhar, 2012; Stoler et al., 2012; Weeks et al., 2007). Most previous works of slum detection from remotely sensed data have used high and very high spatial resolution (VHR) imagery from Quickbird, Ikonos, and SPOT 5 sensors, and some approaches have used complementary socio-economic data on housing size and value, while others have used only textural and fractal dimensions derived from satellite imagery to implement statistical regressions (Patino and Duque, 2013).

Avelar et al. (2009) and Tapiador et al. (2011) used Quickbird imagery to differentiate socio-economic classes in Lima, Peru, using per-pixel supervised classification techniques to identify urban features such as green areas, water bodies, paved streets, swimming pools, and different types of buildings, that were later used as input for socio-economic class asignation. Object-oriented classification schemes with VHR imagery have been tested in different cities around the world for slum detection: Delhi, India (Niebergall et al., 2007), Padang, Indonesia (Taubenböck et al., 2009a), Casablanca, Morocco (Rhinane, 2011), and Sao Paulo and Rio de Janeiro, Brazil (Hofmann et al., 2008; Kux et al., 2010; Novack and Kux, 2010), and all of them reported encouraging results. Weeks et al. (2007), Stow et al. (2007), and Stoler et al. (2012) tested the Vegetation-Impervious Surface-Soil (VIS) model of the urban scene devised by Ridd (1995)³ and the use of texture measures from medium to very high spatial resolution satellite imagery to develop spatial econometric models to predict the slum index as a function of remote sensing derived variables with good results (pseudo R-squared of 0.401 in the best model). Barros (2008) and Kit et al. (2012) used an approach based on lacunarity texture analysis,⁴ to identify slums in Recife and Campinas, Brazil and Hyderabad, India, respectively, and both reported good results. Owen and Wong (2013) used a Quickbird image as well as digital elevation data to implement a model for slum delineation in Guatemala City based on image and DEM derived indicators for road coverage, road connectivity, settlement texture and vegetation, and they reported very good results.

³ Ridd (1995) proposed a model for urban ecosystem analysis that is very well suited to medium spatial resolution images and addresses the problem of mixed pixels. This model decomposes the urban scene into image fractions of vegetation, impervious surfaces and soil (VIS), and then urban land cover is classified according to the fractions for each pixel in a similar way as the tertiary sand-silt-clay diagram is used to quantify soil texture composition in the Earth sciences. This model is usually implemented through Spectral Mixture Analysis techniques (Lu and Weng, 2004; Rashed et al., 2005; Setiawan et al., 2006).

⁴ Lacunarity is a fractal dimension that accounts for the distribution of empty spaces (lacunas) in an image (Barros, 2008).

The objective of this paper is to illustrate how remote sensing derived variables with physical meaning can be used as a reliable proxy for slum index estimation within cities and urban agglomerations, using Medellin, Colombia as study case. We use a robust approach in which we pay careful attention to the presence of spatial effects in the statistical regressions. We use state of the art geocomputation algorithms for spatially constrained clustering to design meaningful spatial units (analytical regions) that allow for a more accurate calculation of the slum index from socioeconomic survey data. Data from an orthorectified, pan-sharpened, natural color Quickbird scene, that is similar to the imagery already available through internet services like Google Earth and Microsoft Bing, is used as input to the process. We apply basic image processing and a robust supervised image classification procedure to derive a set of basic land cover variables; and geographic based image analysis (GEOBIA) techniques to compute another set of variables from the satellite image at the analytical region level. This approach is promising in that the remote sensing data could be acquired at a very low cost and is processed with free software tools like ClusterPy and FETEX 2.0; thus, it could open the possibility of using remote sensing data for analyzing intra-urban poverty patterns in a fast, repetitive, inexpensive and reliable way.

The structure of this chapter is as follows: section 2 presents an outline of the proposed model from a theoretical point of view and the discussion of what type of environmental information can be measured from remote sensing as a proxy for slum index. A short discussion about the spatial unit of analysis is made in section 3. Section 4 presents the remote sensing data and the variables derived from it. Results and its discussion are presented in section 5. Section 6 presents the main conclusions and public policy implications of this line of research for local government and authorities.

2 Model

We develop a model for slum index estimation based solely on remote sensing data from a VHR color image in order to assess the performance of the remote sensing derived variables for this purpose. Although there exists VHR imagery with higher spectral resolution, we deliberately chose to work with an image whose quality is comparable to any aerial color photography to facilitate the replicability of this approach in other urban areas around the world. The value of the variables that compose the slum index cannot be measured directly from remote sensing data, but the spatial characteristics of urban land cover elements can be quantified to serve as a proxy for the slum index (Weeks et al., 2007). These characteristics could provide a framework for the analysis of intra-urban variations of poverty so that urban planners and policy makers are able to identify which areas are most in need of attention (Stoler et al., 2012).

Weeks et al. (2007); Stow et al. (2007) and Stoler et al. (2012) developed spatial econometric models to estimate the slum index in Accra, Ghana, as a function of within neighbourhoods measures of vegetation, impervious surface and soil that were obtained from applying spectral mixture analysis (SMA) on remote sensing imagery. Owen and Wong (2013) made a review of indicators derived from remote sensing imagery and digital elevation data that have been reported as useful for slum delineation. Using information of Guatemala City, they found that the variables that best contributed to separate informal from formal settlements are *entropy of roads*, *vegetation patch size*, *vegetation compactness*, *profile convexity*, *road density* and *soil coverage*. Thus, according to these previous studies, a model for slum index estimation from remote sensing imagery should include information not only about land cover materials but also about the spatial pattern of the urban fabric.

From the urbanism point of view, Medellin City is a good example of an average Latin American city: it has many contrasts between rich and poor areas, and there are some land cover features related to poverty or slumness that are also present in most Latin American cities with similar weather conditions (Patino, 2010). We use basic image processing and a recently released software

for object-based image analysis called FETEX 2.0,⁵ to generate a wide set of variables that allow us to quantify some physical aspects of the urban settings that give information about rooftops materials, the presence or absence of some features that inform about economic affluence, and the spatial arrangement of dwelling units or the urban layout. Then we assess which variables best explain the variability of slum index using ordinary least squares regression (OLS) analysis, while controlling for multicollinearity and the presence of spatial effects.

Since the dependent variable (slum index) is calculated from survey data, we pay special attention to the spatial unit of analysis by using the analytical regions defined by Duque et al. (2012b). These units were specifically designed to study the intra-urban variations of poverty in Medellín City. As highlighted by Weeks et al. (2007), a proper definition of the spatial unit of analysis increases the accuracy of spatial delimitation of slums, and also avoids the potential presence of spurious spatial autocorrelation in the econometric analysis.

Figure 1 outlines the methodology of image processing and neighbourhood delineation that were followed to develop an econometric model for slum index estimation from VHR remote sensing data. Dotted frames are the basic steps applied by Duque et al. (2012b) to define analytical regions using CluserPy⁶ with survey data. Basic image processing techniques are applied to the input orthorectified VHR color image for two purposes: to enhance the spectral information before image classification and to summarize the spectral information of the three color bands (red, green and blue) into one single band. Band ratios *red/blue* and *green/red* are applied to enhance the spectral information for a better differentiation of reddish and green color surfaces respectively. Principal component analysis is performed to the three color bands to summarize the spectral information contained in the three bands into one single band (the first principal component explains the 98% of the spectral variability contained in the three color bands). Clean water surfaces and swimming pools have high values in the *green/red* band ratio, so we use it to perform swimming pool detection by thresholding and further extraction of swimming pool centroids. These centroids are then used to make a map of swimming pool density by square kilometer and then we calculate the mean swimming pool density within analytical regions. The first principal component (PC 1) and both band ratios are used as input to a supervised classification to differentiate basic land cover types of the urban scene: clay roofs, soil, vegetation and impervious surfaces. Then we calculate the percentage of each land cover type within analytical regions. On the other hand, we use PC 1 and the analytical regions as input to the calculation of structural and texture variables using FETEX 2.0. We then integrate all these remote sensing derived variables into one single database at the analytical region level with the slum index calculated from survey data in order to perform statistical analysis for model specification.

3 Spatial unit of analysis

The urban area of Medellín city has two levels of administrative spatial units: communes (16) and neighborhoods (243). The slum index is usually reported at the commune level from socio-economic data available in the Quality of Life Survey, which sampling process is designed to be representative at this spatial scale. There are two main disadvantages of using communes as spatial units of analysis. First, these units are too large to study the spatial pattern of intra-urban poverty levels; and second, the statistical inference based on large administrative units may be affected by aggregation problems such as the ecological fallacy (Robinson, 1950), the modifiable areal unit problem (Openshaw, 1977b,a; Openshaw and Taylor, 1981; Openshaw, 1984; Arbia, 1989), and aggregation bias (Fotheringham and Wong, 1991; Amrhein and Flowerdew, 1992; Paelinck and Klaassen, 1979; Paelinck, 2000). At the neighborhood level the average number of surveyed households, in 2007, is 84 ± 57 , with values ranging from 3 to 296, and 15.64% of the neighbor-

⁵ FETEX 2.0 is a free software tool that runs over ENVI and IDL (Ruiz et al., 2011).

⁶ For a detailed version of this process see figure 12.3 on page 245 of Duque et al. (2012b).

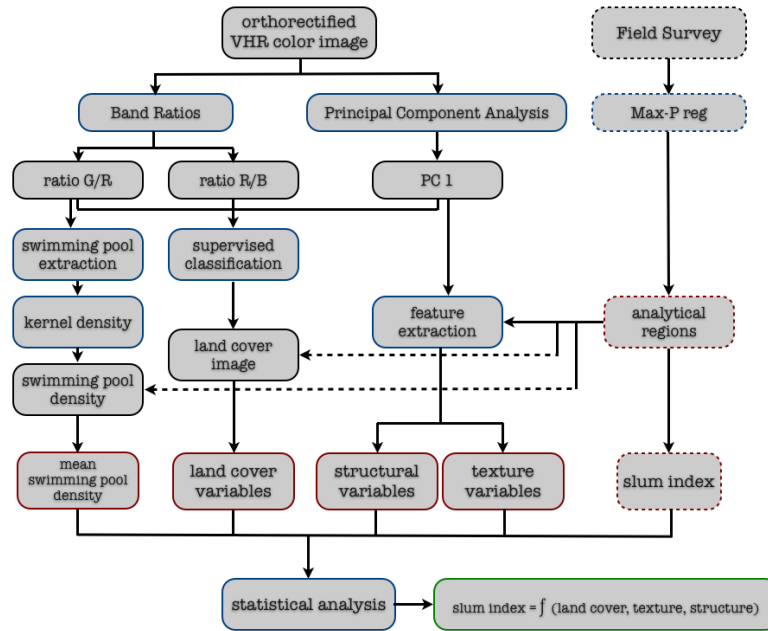


Figure 1 Steps for model specification of slum index as a function of remote sensing derived variables from a VHR image of the urban scene.

hoods contain less than 30 surveyed households. Thus, working with administrative neighborhoods can involve three problems: the lack of statistical validity, the small numbers problem for rates calculation (Diehr, 1984), and the potential appearance of spurious spatial autocorrelation (Weeks et al., 2007).

An alternative to deal with the problems described above is to use analytical regions, which are spatial units that fulfill specific criteria (e.g., size, shape, attribute homogeneity, among others) relevant to the phenomena under study (Duque et al., 2006). Duque et al. (2012b) designed analytical regions for the study of intra-urban poverty variations in Medellin city. Each analytical region was required to meet the following two conditions: it must be homogeneous in terms of the socio-economic characteristics and it must have a minimum of 100 surveyed households to ensure statistical validity. These units were delineated by applying the Max-P-Regions algorithm (Duque et al., 2012a, 2011), which grouped the 243 city neighborhoods into 139 analytical regions (figure 2).

4 Data

Medellin City is located at -75.58° of Longitude and 6.22° of Latitude in the northwest region of Colombia on the Central Andean mountain range at an average height of 1,500 meters above sea level (figure 3). The city extends over a narrow valley crossed by Medellin River from South to North. Initially the city grew over the alluvial plain of the river, but in the last decades it spread towards the western and eastern slopes of the valley in an unplanned way. The average annual tem-

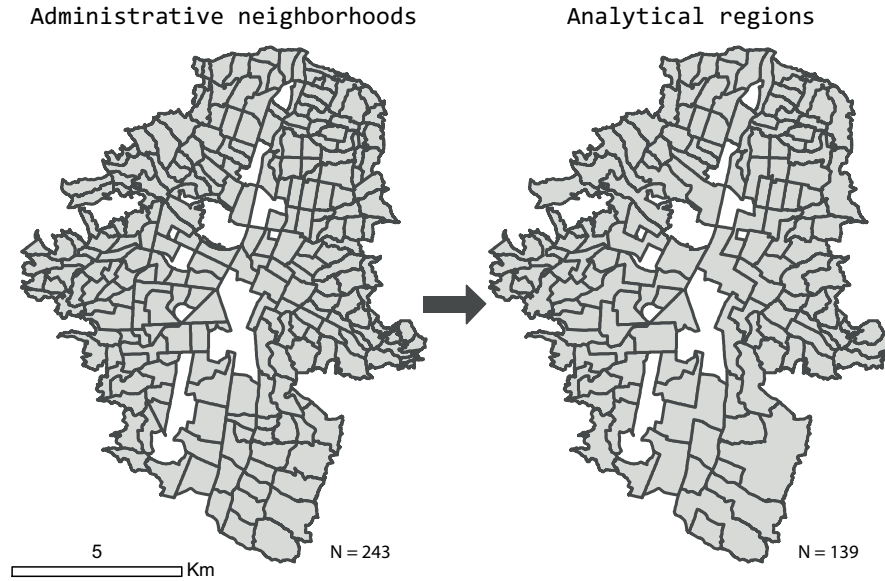


Figure 2 Medellín administrative neighborhoods and the resulting analytical regions from [Duque et al. \(2012b\)](#).

perature is around 22 °C, with minimal variations due to its proximity to the Equator. The city has 2.4 million inhabitants, with 93.4% of them considering themselves as white descendants (which include those of mixed Spanish and indigenous Amerindian descent), 6.5% Afro-descendant and only 0.1% indigenous Amerindian ([DANE, 2010](#)).

The slum index was computed using data from the 2007 Quality of Life Survey of Medellín. This survey is formed by 184 questions on 9 dimensions: housing, households, demography, education, social security, income and employment, social participation, gender and family violence, and nutrition. The sample include 21,861 households that represent 79,912 persons ([Duque et al., 2012b](#)). The survey considers a wide list of questions related with poverty, from which [Duque et al. \(2012b\)](#) selected the five aspects considered by UN-Habitat that make up the slum index and they constructed binary variables indicating whether or not the housing unit fulfill each aspect. [Table 1](#) describes the variables used to calculate slum index and the proportion of housing units in the city under those conditions.

Table 1 Variables used to calculate slum index from the 2007 Quality of Life Survey of Medellín and proportion of housing units under those conditions. Modified after [Duque et al. \(2012b\)](#).

Variable	Description	Dimension	Housing units (%)
Material	Walls are not of durable material	Housing	0.14
Overcrowding	Three or more persons per room	Households	16.77
Water	Absence of piped water	Housing	0.01
Toilet	Toilet not connected to a sewer	Housing	3.77
Ownership	Residents are not the owners	Households	35.46

A VHR color image was used as input data to image processing in order to build a database of a wide set of image derived variables to be used as input to the econometric analysis. The image is a

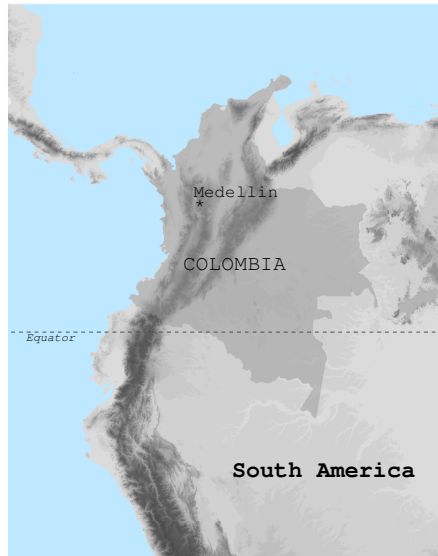


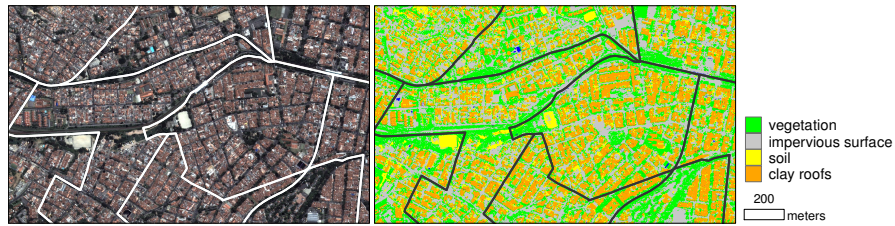
Figure 3 Location of Medellín City in Colombia and South America.

Quickbird, orthorectified, pan-sharpened, natural color scene of Medellín city that was acquired in May 2008. Image preprocessing included the geometric rectification using the rational polynomial coefficients (RCP) provided with the image, and pan-sharpening using the panchromatic band to produce a natural color scene with 0.6 meters of pixel size. As mentioned above, we applied Principal Components Analysis to summarize the spectral information contained in the three bands (red, green and blue) into one single band (the first principal component, PC 1), and the band ratios *red/blue* and *green/red* to enhance the spectral information for a better differentiation of reddish and green color surfaces. Image derived variables are grouped in three sets: a set of land cover features, a set of structural features and a set of texture features. Land cover features were calculated after swimming pool extraction and image classification into land cover classes, while structural and texture features were calculated directly using FETEX 2.0 software.

Land cover features were obtained through a supervised classification using the maximum likelihood algorithm in ENVI with the composite image of PC 1 and the *red/blue* and *green/red* band ratios. The thematic accuracy of the classification process was assessed using a traditional point-based technique with a reference dataset of randomly selected points (a wide sample of ground truth points were collected and randomly divided into two datasets: 80% for classification purposes and the other 20% for the accuracy assessment of the classification results). Overall accuracy of this classification was 95% and the *kappa* statistic was 0.94. Table 2 shows the confusion matrix for the classification results. The producer's accuracy for vegetation, impervious surface, soil and clay roofs classes is over 93%, while user's accuracy is over 91% for the vegetation class, and over 98% for impervious surface and clay roofs classes, and about 77% for the soil class. Then we calculate the percentage cover of vegetation, impervious surface, soil and clay roofs classes within analytical regions to build the following variables: *percent of vegetation*, *percent of impervious surface*, *percent of soil* and *percent of clay roofs* (figure 4). Swimming pool locations were extracted by thresholding the *green/red* band ratio and a raster map of swimming pool density by square kilometer was built using the Kernel Density function in ArcGIS. Then the *mean swimming pool density* within analytical regions was computed.

Table 2 Accuracy assessment of image classification results.

Ground truth	Vegetation	Impervious surf.	Water	Shadow	Soil	Clay roofs	Total
Classified as							
Vegetation	712	24	0	26	0	14	776
Impervious surface	10	1356	6	1	9	0	1382
Water	0	0	400	0	0	0	400
Shadow	6	0	18	165	0	0	189
Soil	10	29	9	0	186	8	242
Clay roofs	0	0	0	0	5	708	713
Total	738	1409	433	192	200	730	
Producer's accuracy (%)	96.48	96.24	92.38	85.94	93.00	96.99	
User's accuracy (%)	91.75	98.12	100.00	87.30	76.86	99.29	
Overall classification accuracy (%)	95.27						
Kappa value	0.94						

**Figure 4** Example of the input VHR color image and classified land cover types within analytical regions.

We use structural and texture features in order to characterize the spatial pattern of the urban layout. The urban layout in slum-like neighborhoods often show a more organic, crowded and cluttered pattern than that in more formal and wealthy neighborhoods, thus the spatial pattern of the urban fabric may be an indicator of poverty in this city (Patino, 2010), and we expect that some of the variables used to characterize the spatial pattern of the urban layout may be related to the slum index at the analytical region level. The PC 1 band and the analytical regions polygons were used as input in FETEX 2.0 software to calculate these variables. Structural features provide information of the spatial arrangement of elements inside the polygons, in terms of randomness or regularity of the distribution of the elements (Balaguer et al., 2010; Balaguer-Beser et al., 2013; Ruiz et al., 2011). For this research we computed a set of structural features of the analytic regions using FETEX 2.0 and the experimental semivariogram approach. According to Ruiz et al. (2011), the semivariogram quantifies the spatial associations of the values of a variable, measures the degree of spatial correlation between different pixels in an image and it is a suitable tool for the characterization of regular patterns. FETEX 2.0 obtains the experimental semivariogram of each polygon by computing the mean of the semivariogram calculated in six different directions, from 0° to 150° with step increments of 30° , and then each semivariogram curve is smoothed using a Gaussian filter in order to reduce experimental fluctuations (Ruiz et al., 2011). Structural features extracted from the semivariogram are based on the zonal analysis defined by a set of singular points

on the semivariogram, such as the first maximum, the first minimum, and the second maximum (Ruiz et al., 2011).⁷ The structural features calculated for each analytic region are listed in table 3.

Texture features inform about the spatial distribution of intensity values in the image and provide information about contrast, uniformity, rugosity, etc. (Ruiz et al., 2011). FETEX 2.0 calculates the Grey Level Co-occurrence Matrix (GLCM) and the histogram of pixel values inside each polygon for texture feature extraction. The *kurtosis* and *skewness* features are based in the histogram of the pixel values inside the polygon. The GLCM describes the co-occurrences of the pixel values that are separated at a distance of one pixel inside the polygon and it is calculated considering the average value of four principal orientations: 0°, 45°, 90° and 135°, in order to avoid the influence of the orientation of the elements inside the polygon (Ruiz et al., 2011). The GLCM is used in FETEX 2.0 to calculate a set of variables proposed by Haralick et al. (1973) that is widely used in image processing: *uniformity*, *entropy*, *contrast*, *inverse difference moment (IDM)*, *covariance*, *variance*, and *correlation*. The *edgeness factor* is another useful feature that represents the density of edges present in a neighbourhood and the *mean* and the *standard deviation* of the *edgeness factor* (MEAN_EDG, STDEV_EDG) are also computed within this set of texture features in FETEX 2.0 (Ruiz et al., 2011). Table 3 shows the complete list of remote sensing variables used in the econometric analysis.

Table 3 Remote sensing derived variables and summary statistics (all variables by analytical region).

Group	Variable name	Description	Mean	STDEV
Land cover features	S_POOL_DENS	Mean swimming pool density	5.04	6.50
	VEG_P	Percentage of vegetation cover	36.66	13.19
	IMP_SURF_P	Percentage of impervious surface cover	54.09	12.75
	SOIL_P	Percentage of bare soil cover	8.04	2.60
	CLAY_ROOFS_P	Percentage of clay roof cover	13.22	9.24
Structural features	RVF	Ratio variance at first lag	12.23	2.80
	RSF	Ratio between semivariance values at second and first lag	1.80	0.02
	FDO	First derivative near the origin	116.93	20.46
	SDT	Second derivative at third lag	15.15	3.18
	MFM	Mean of the semivariogram values up to the first maximum	822.65	132.71
	VMF	Variance of the semivariogram values up to the first maximum	1.44	0.22
	DMF	Difference between the mean of the semivariogram values up to the first maximum and the semivariance at first lag	676.61	108.08
	RMM	Ratio between the semivariance at first local maximum and the mean semivariogram values up to this maximum	1.41	0.04
	SDF	Second order difference between first lag and first maximum	-338.46	73.52
	AFM	Area between the semivariogram value in the first lag and the semivariogram function until the first maximum	5236.57	890.74
Texture features	SKENNESS	Skewness value of the histogram	-0.75	0.26
	KURTOSIS	Kurtosis value of the histogram	0.33	0.69
	UNIFOR	GLCM uniformity	0.01	0.002
	ENTROP	GLCM entropy	2.61	0.07
	CONTRAS	GLCM contrast	15.23	2.73
	IDM	GLCM inverse difference moment	0.43	0.04
	COVAR	GLCM covariance	97.27	14.35
	VARIAN	GLCM variance	104.89	14.41
	CORRELAT	GLCM correlation	0.93	0.02
	MEAN_EDG	Mean of the edgeness factor	8.37	1.02
STDEV_EDG	Standard deviation of the edgeness factor	7.42	0.54	

5 Results and discussion

As we intended to build a model for slum index as a function of remote sensing variables, we looked first to the statistical correlation between the slum index and each of the image derived variables using Pearson's correlation coefficients. We selected the variables that have a Pearson's correlation coefficient greater than 0.5 and statistically significant as candidates to be included as dependent variables in OLS regressions. Figure 5 shows a correlation matrix of slum index and the selected variables generated in R software (R Core Team, 2013). This matrix is used for two purposes: to inform about the relationship between slum index and each independent variables,

⁷ For a full description of these features see Balaguer et al. (2010); Balaguer-Beser et al. (2013) and Ruiz et al. (2011).

and to check for correlations among independent variables that could lead to multicollinearity in the OLS regressions.

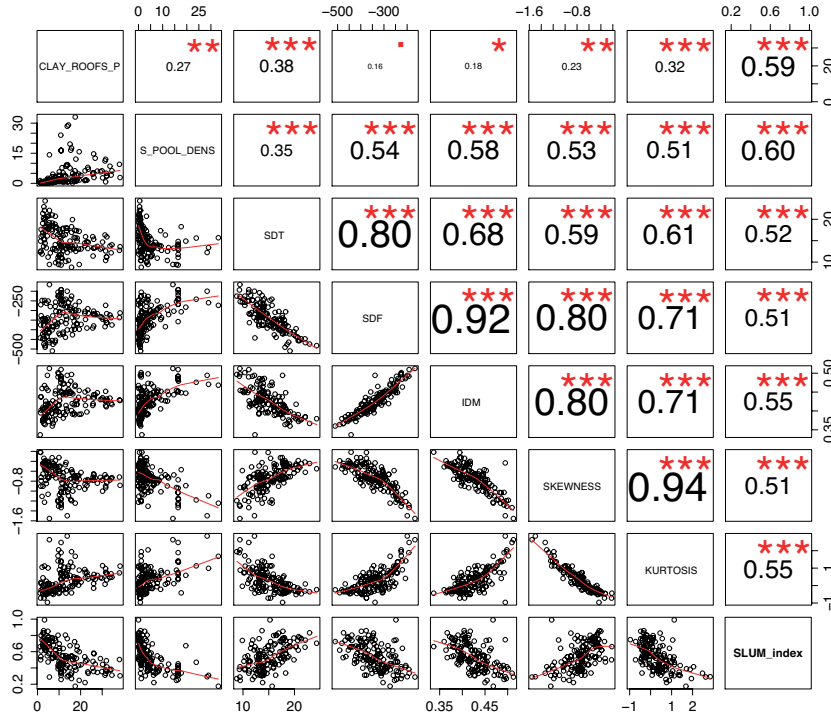


Figure 5 Graph of the correlation matrix of the selected remote sensing variables and Slum_index. Upper right part shows the absolute value of the Pearson’s correlation coefficient and significance levels as symbols in red: 0.1%: ***, 1%: **, 5%: *, 10%: . Lower left part shows bivariate scatter plots with a fitted line in red.

The last row in figure 5 shows bivariate scatter plots of remote sensing derived variables (plotted in the X-axis) vs. slum index (plotted in the Y-axis). The last column shows the absolute value of the Pearson’s correlation coefficient of remote sensing derived variables with slum index. The direction of the relationships with land cover variables clay roof percentage cover and mean swimming pool density is as expected in both cases. The slum index goes down when there are more clay roof surfaces within the analytical region, and also when the swimming pool density is higher. Both of these features are related to wealthier neighborhoods in Medellin city: clay tile roofs are more expensive to install and maintain than other roofing materials such as industrial roof tiles made of zinc or asbestos. Historically, roofing materials and styles have been seen as a reflection of the wealth and power of the dwelling owners since ancient times (Fiumi, 2012). Swimming pools are also expensive to build and they require maintenance and water quality checks in a regular basis, thus their presence indicates that their owners can afford them. Moreover, the presence of swimming pools in urban areas has been reported as related to higher socio-economic classes in other cities around the world such as Lima (Peru), Cairo (Egypt), and Rio de Janeiro (Brazil) (Tapiador et al., 2011).

The structural variables (SDT and SDF) have an inverse behavior in relation to slum index. On one hand, SDT provides information about the concavity or convexity level of the semivariogram at short distances (few meters), corresponding with the heterogeneity of the image objects (Balaguer-Beser et al., 2013). When applied to urban neighborhoods, the value of SDT increases as the image objects become more heterogeneous at short distances. This finding is consistent with previous results reported for agricultural plots (Balaguer-Beser et al., 2013). In relation to slum index, the higher the value of slum index, the higher the value of SDT; which indicates that high heterogeneity (or less homogeneity) at short distances is related with higher slum index values at the analytical region level. On the other hand, SDF provides information about the semivariogram curvature in the interval between the semivariogram's first lag and its first maximum and it represents the low frequency values in the image (Balaguer et al., 2010). It captures information about the spatial arrangement of the elements at short distances and it provides another way to quantify the homogeneity or heterogeneity of the spatial pattern inside the analytical regions. In agricultural plots, the value of SDF increases as the homogeneity at short distances also increases (Balaguer-Beser et al., 2013), showing an inverse behavior compared to SDT. We observed the same behavior of these two variables in urban settings as was previously reported for agricultural plots. With respect to slum index, the higher the value of SDF, the lower the slum index value.

The selected texture variables (IDM, skewness and kurtosis) provide another way to characterize the homogeneity or heterogeneity of the urban layout. These variables along with the selected structural variables all provide different ways to quantify homogeneity or heterogeneity inside the analytical regions. The values of IDM and kurtosis increase as the homogeneity inside analytical regions increase, and the opposite occurs for skewness. In general, the homogeneity of the urban layout has an inverse relation with slum index: more homogeneity at short distances is characteristic of analytical regions with low slum index, and the less homogeneity, the higher the slum index.

The value of either this five remote sensing derived variables (SDT, SDF, IDM, skewness and kurtosis) may be seen as an indicator of household size and crowdedness: the homogeneity at short distances inside the analytical regions is higher in settings where the buildings and households are well separated from each other, with vegetation or other rather homogeneous cover among them; and the homogeneity decreases as the building sizes are lower, their rooftops are diverse in materials, and they are closer to each other with no other land cover among them. Figure 6 shows square image tiles of 500 meters and its corresponding values for each remote sensing derived variable. The image tiles are organized according to increasing values of each variable from left to right in order to show the differences in the urban scene as the variable's value increases.

A number of statistical models were run to test the relationship between slum index and the selected remote sensing variables. Ordinary Least Squares (OLS) regressions were used with slum index as a dependent variable and the remote sensing variables as independent variables. Table 4 shows the OLS models tested in the model specification process. Six models were run: one with slum index as function of the selected land cover variables (M1) and the other five models with one added variable at a time from the other groups (texture and structure, models M2 to M6). We did not include different structure and texture variables in the same model in order to avoid multicollinearity, because it is clear from figure 5 that they are highly correlated among them. The first OLS model (M1) shows that swimming pool density and clay roofs percentage cover together account for 55% of the variability in the slum index.



Figure 6 Square image tiles showing urban areas (500 x 500 meters) that have different values of the remote sensing variables selected in model specification. From top to bottom: *clay roofs percentage cover*, *swimming pool density*, *SDT*, *SDF*, *IDM*, *skewness* and *kurtosis*. Tiles organized according to increasing values of each variable from left to right.

Table 4 Multivariate Ordinary Least Squares models of Slum index as a function of remote sensing variables. Spatial unit: analytical regions. N = 139. Contiguity matrix: Euclidean distance weight matrix with threshold at 1,700 m.

Exogenous variables	M1	M2	M3	M4	M5	M6
Constant	0.7060926***	0.5038889***	0.4903547***	1.2441700***	0.7898861***	0.6996642***
CLAY_ROOFS_P	-0.0081365***	-0.0069110***	-0.0080638***	-0.0079703***	-0.0077744***	-0.0073246***
S_POOL_DENS	-0.0120699***	-0.0104887***	-0.0085596***	-0.0079300***	-0.0093146***	-0.0094476***
SDT		0.0117542***				
SDF			-0.0005823***			
IDM				-1.3148930***		
SKEWNESS					0.1362412***	
KURTOSIS						-0.0537139***
Adjusted R^2	0.55	0.59	0.60	0.60	0.58	0.58
Multicollinearity Condition Number	3.64	14.96	14.93	37.00	8.76	4.09
Breusch-Pagan test	1.24	4.23	3.80	4.04	5.70	1.80
Lagrange Multiplier - lag test	12.45***	4.05**	5.39**	5.19**	8.77***	8.36***
Robust LM - lag test	10.93***	1.95	2.62*	3.49*	4.46**	4.70**
Lagrange Multiplier - error test	2.51	2.12	2.84*	1.72	4.36**	3.66**
Robust LM - error test	0.99	0.02	0.08	0.03	0.05	0.00

Note: Statistical significance is at the 1, 5, and 10% level as indicated by ***, **, and * respectively.

The second group tested for model specification was the selected variables from the structural features group: SDT and SDF. There is high correlation between these two variables, hence including both of them in the same model increase the multicollinearity in the regression, so we added SDT and SDF separately to the previous model specification. Including one of these structural variables increases the explanatory power of the models to 60% (table 4, M2 and M3). As mentioned above, we observed that the more heterogeneity at short distances, the higher the slum index, and this was corroborated by the sign of the coefficients in these OLS models.

The selected texture variables all measure the same aspect as the structural variables (homogeneity), and we tested whether or not each of them improved the model's explanatory power when they are taken into account together with land cover variables (table 4, M4 to M6). The results indicate that the inclusion of any of them improves the OLS regressions up to 5%, increasing the adjusted R^2 from 0.55 (M1) to 0.60 in the case of IDM (M4), and to 0.58 for skewness (M5), and kurtosis (M6). But multicollinearity becomes serious in the case of IDM, indicating that the information it provides was already captured by the land cover variables. From figure 5 it is clear that IDM is positively correlated with swimming pool density (with a Pearson's coefficient of 0.58).

We run different tests to find out if these regressions were robust to the potential presence of spatial effects: heteroskedasticity and spatial autocorrelation. The results of the Breusch-Pagan tests indicate that there isn't overwhelming evidence of heteroskedasticity in these models. In order to test for the presence of spatial autocorrelation, we first used the use the Rook and Queen specifications of contiguity matrices but they both dropped confusing results regarding the type of spatial dependence present in these regressions, so we moved to the specification of a neighbor weight matrix based on the euclidean distance proximity, which means that all neighbors within a specified distance of a given polygon would be considered neighbors. To determine an appropriate distance threshold we followed the strategy outlined by Troy et al. (2012) and built the experimental semivariogram of the endogenous variable in order to get an idea of the distance at which spatial autocorrelation decays. The semivariogram showed a terrace-type pattern with a sill at about 2 Km, which indicates that the autocorrelation of observation pairs levels off at that distance. We then compared some fixed distance bands for the neighbor matrix above and below 2 Km and found that 1.7 Km resulted in the best R -squared models and also in best consistency among spatial dependence tests for all models. The results of the Lagrange Multiplier tests indicated that we needed to include a spatial lag of the endogenous variable in order to account for spatial autocorrelation in the statistical regressions, and we feel justified to use a spatial lag model since it has been established in the literature that poverty is a phenomena that can spill over into adjacent regions (Duncan et al., 2012; Holt, 2007; Okwi et al., 2007; Orford, 2004; Sowunmi et al., 2012; Voss et al., 2006).

Table 5 shows the spatially adjusted regressions of Slum index as a function of remote sensing derived variables for all previous model specifications except for M4, which was removed from this analysis due to the serious presence of multicollinearity. The Likelihood Ratio tests indicate that there is an improvement of all spatial models over their non-spatial counterparts. Three significant values for the spatial lag coefficient (W_Slum_index) and the Anselin-Kelejian tests show that there is no spatial autocorrelation in the residuals of these models.

Table 5 Spatial Lag models of Slum index as a function of remote sensing variables. Spatial unit: analytical regions. N = 139. Contiguity matrix: Euclidean distance weight matrix with threshold at 1,700 m.

Exogenous variables	SL_M1	SL_M2	SL_M3	SL_M5	SL_M6
Constant	0.4836126***	0.4093310***	0.3731374***	0.5881738***	0.5174431***
CLAY_ROOFS_P	-0.0064210***	-0.0061008***	-0.0068782***	-0.0063887***	-0.0060473***
S.POOL_DENS	-0.0089232***	-0.0088408***	-0.0069830***	-0.0071249***	-0.0072947***
SDT		0.0089001***			
SDF			-0.0004792***		
SKEWNESS				0.1133092***	
KURTOSIS					-0.0447084***
W_Slum_index	0.3445581***	0.2224816**	0.2407183**	0.2905516***	0.2838776***
Pseudo- R^2	0.59	0.61	0.62	0.61	0.61
Likelihood Ratio test	10.42***	3.69**	4.88**	7.43***	7.08***
Anselin-Kelejian test	1.96	0.02	0.03	0.16	0.32

Note: Statistical significance is at the 1, 5, and 10% level as indicated by ***, **, and * respectively.

The results of these analysis indicate that the most important predictors of slum index at the analytic region level in Medellin city are the selected variables from the land cover set: clay roofs percentage cover and swimming pool density. The model is improved further up to 3% by adding only a single variable that accounts for homogeneity within the analytical regions. That variable could be from the structural variable set (SDT or SDF, table 4, M2 and M3, and table 5, SL_M2 and SL_M3), or from the texture variable set (skewness or kurtosis, M5 and M6, and SL_M5 and SL_M6), which were introduced to account for the differences in the spatial pattern of the urban layout. Structural and texture variables are highly correlated among them, and they slightly improve the explanatory power of the OLS and the spatially adjusted regression models of slum index, so any of them alone should be used to account for the differences in the spatial pattern of the urban layout at the analytical region level. In the case of Medellin city, the variable that best performs in the models is SDF, while the other three, SDT, skewness, and kurtosis, all perform in a similar fashion as indicated by similar values of Pseudo- R^2 .

The used land cover variables, i.e., clay roofs percentage cover and mean swimming pool density, are closely related to the geographical location of the city and these variables are more expensive to obtain than structural and texture variables, since they require the previous classification of the urban scene into different land cover types in the case of clay roofs percentage cover, and the extraction of swimming pools centroids and posterior computation of a density map in the case of mean swimming pools density. The presence and abundance of orange and reddish clay tile roofs in a city is closely related to its environmental conditions since it depends on the availability and mineral types of the clays and sands located in the proximity of the city. The abundance of swimming pools in a city is also related to the environmental conditions of the city since this urban feature is more common in hot and warm weather sites than in cold ones. Moreover, the cities where this feature has been reported to be related to higher socio-economic status are all located near or within the world's tropical zone: Lima, Peru; Rio de Janeiro, Brazil; and Cairo, Egypt (Tapiador et al., 2011); Sao Paulo, Brazil (Novack and Kux, 2010); and Medellin, Colombia (this research).

On the contrary, the descriptors of the urban layout, i.e., the structure and texture variables, are easily computed from VHR imagery and polygon boundary data and the use of texture measures

for slum detection and mapping has been reported in different cities around the world: Accra, Ghana (Weeks et al., 2007); Campinas, Brazil (Barros and Sobreira, 2005); Casablanca, Morocco (Rhinane, 2011); Delhi, India (Niebergall et al., 2007); Guatemala, Guatemala (Owen and Wong, 2013); Hyderabad, India (Kit et al., 2012); Pune, India (Shekhar, 2012); and Rio de Janeiro, Brazil (Barros and Sobreira, 2005). We tested univariate regressions in order to get a better understanding of the explanatory power of each selected structure and texture variables alone for slum index (table 6). We didn't test for composite models combining these variables to avoid multicollinearity. When using one of these variables alone, each of them explain from 26% up to 30% of the variability of the slum index in Medellin city.

Table 6 Univariate Ordinary Least Squares models of Slum index as a function of structure and texture variables. Spatial unit: analytical regions. N = 139. Contiguity matrix: Euclidean distance weight matrix with threshold at 1,700 m.

Exogenous variables	M7	M8	M9	M10	M11
Constant	0.1306056**	0.1492380***	1.6155320***	0.7821080***	0.5804296***
SDF	0.0268777***				
SDF		-0.0011478***			
IDM			-2.5256920***		
SKEWNESS				0.3249882***	
KURTOSIS					-0.1310201***
Adjusted R ²	0.27	0.26	0.30	0.26	0.30
Breusch-Pagan test	2.19	2.53	2.00	2.16	0.41
Lagrange Multiplier - lag test	53.20***	66.79***	58.64***	76.16***	66.03***
Robust LM - lag test	11.74***	15.01***	22.23***	5.24**	7.25***
Lagrange Multiplier - error test	42.16***	52.46***	36.44***	79.30***	65.09***
Robust LM - error test	0.69	0.69	0.03	8.38***	6.32**

Note: Statistical significance is at the 1, 5, and 10% level as indicated by ***, **, and * respectively.

Again, the results of the Breusch-Pagan test indicated that there isn't evidence of heteroskedasticity in these regressions (M7 to M11), but the Lagrange Multiplier tests pointed to the presence of spatial autocorrelation that should be addressed through spatially adjusted regressions. Table 7 shows the respective spatial lag models for each structure and texture variables (SL_M7 to SL_M11). The results of the Likelihood Ratio tests pointed again to the superiority of the spatially adjusted regressions over their non-spatial counterparts, and the Pseudo-R² values indicate that there is a significant improvement of the explanatory power of these models when spatial effects are taken into account.

Table 7 Spatial Lag univariate models of Slum index as a function of structure and texture variables. Spatial unit: analytical regions. N = 139. Contiguity matrix: Euclidean distance weight matrix with threshold at 1,700 m.

Exogenous variables	SL_M7	SL_M8	SL_M9	SL_M10	SL_M11
Constant	0.0034071	-0.0139045	0.8054779***	0.3233062***	0.2265295***
SDF	0.0126591***				
SDF		-0.0006034***			
IDM			-1.4078950***		
SKEWNESS				0.1810804***	
KURTOSIS					-0.0757800***
W_Slum_index	0.6416619***	0.6506931***	0.6238322***	0.6566835***	0.6291649***
Pseudo-R ²	0.47	0.49	0.50	0.50	0.51
Likelihood Ratio test	36.20***	41.53***	38.00***	44.56***	40.41***

Note: Statistical significance is at the 1, 5, and 10% level as indicated by ***, **, and * respectively.

To assess the performance of this approach, we took the best performing spatial regression (SL_M3, from table 5) and we built a box map of the residuals in order to examine the location of the analytical regions with the greatest errors (figure 7). The box map shows that there are no spatial pattern in this residuals, as indicated by the Anselin-Kelejian test reported in table 5, and clearly shows two upper outliers where the estimated slum index is lower than the reference slum index computed from socioeconomic data. The greatest error occurs in an analytical region at the northeast border of the city (ID = 94), where the reference slum index is 0.99 but the estimated slum index is 0.62. This region is composed mainly of small, non-durable households, and it can be seen that the region boundary encloses not only the built-up areas, but also big green areas corresponding to vacant lots located in the urban-rural border that aren't yet occupied nor belong to the urban fabric. These green areas contribute to rise the homogeneity inside the region, which is captured in the regression by the structure variable SDF, hence lowering the estimated slum index. The second greatest error occurs in an analytical region located in the west part of the city (ID = 122, reference slum index 0.67, estimated 0.40). Careful inspection of this region shows that this error could be an artifact of the regionalization process, since there are high standard houses at the center of this region with almost one swimming pool every two houses, while the upper left and lower right areas within this region have lower standard housing with smaller buildings and more heterogeneity in the short distances. These errors indicate that one must pay careful to the delineation of the spatial unit of analysis when extracting remote sensing variables, thus administrative units and the resulting analytical regions of the regionalization process sometimes may include green spaces or undeveloped areas that don't belong to the urban fabric.

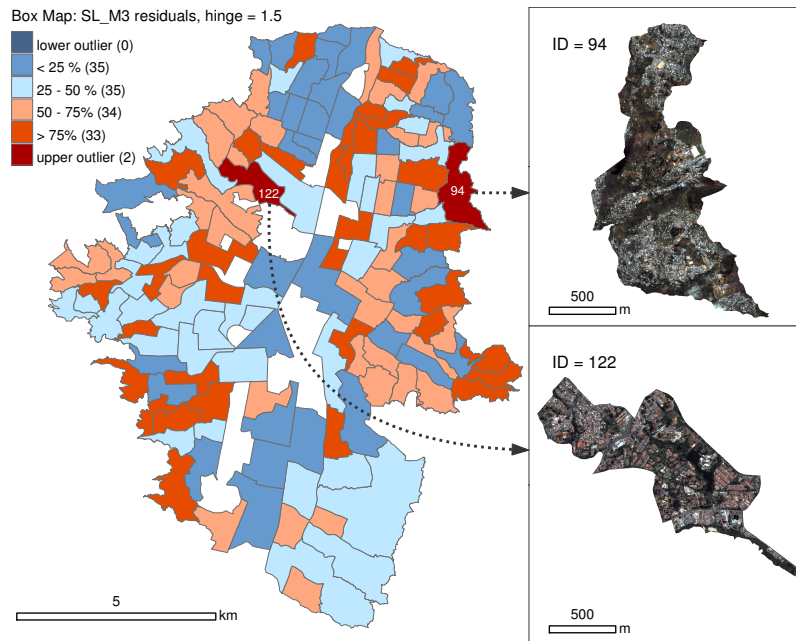


Figure 7 Box map of SL_M3 residuals (upper part) and image details of analytical regions ID = 94 and ID = 122 (lower part).

6 Conclusions

In this paper we used remote sensing data to quantify the slum index in Medellin, Colombia, based in the premise that the physical appearance of an urban settlement is the result of the social and economical conditions and interactions of their inhabitants and on the assumption that people living in urban areas with similar physical housing conditions will have similar social and demographic characteristics. The usefulness of remote sensing data to estimate the slum index in a city had been previously tested in Accra, Ghana; and the findings of the present research corroborate those results in a city whose geographical conditions are rather different: Medellin, Colombia. While the most important remote sensing predictor of slum index in Accra is the amount of vegetation (Weeks et al., 2007; Stoler et al., 2012), the same variable in Medellin does not perform in the same fashion thus it shows a very low Pearson's correlation coefficient in relation with the slum index (0.14). This may be explained by the differences in climate conditions that influences the vegetation abundance in each location, as well as by the cultural differences of the inhabitants of this two cities.

The best performing variables for slum index quantification at the analytical region level from remote sensing data in Medellin are the percentage of clay roofs cover and the mean swimming pools density, which together can explain up to 55% of the variability in the slum index (table 4), and their explanatory power increases to 59% when the spatial effects are taken into account (table 5). These variables are closely related to the geographical location of the city in terms of the availability of building materials, that depend of the existent lithology in the proximity of city, and of the city's climate that influences the abundance of swimming pools. The models developed so far for Medellin city could be transferred to other cities in the world as long as they have comparable environmental conditions: wide availability of orange and reddish clays as building materials, and similar warm to hot climate.

Another finding of this research is that the structure variables (SDT and SDF), calculated from the experimental semivariogram of the intensity values of the image, can be used to characterize the spatial pattern of the urban layout, although these variables were specifically developed to characterize agricultural plots. The texture variables (IDM, skewness, and kurtosis) are useful as well to quantify the spatial pattern of the urban layout of the neighborhoods. These structure and texture variables are highly correlated to the slum index (figures 5 and 6), and all explain similar aspects of homogeneity (or heterogeneity) at short distances inside the analytical regions, as was indicated by the high and significant Pearson's correlation coefficients amongst them. This research also verified that these structure and texture variables add to the explanatory power of the models up to 3%. And when no other information is available, they alone can explain up to 30% of the variability of the slum index. The homogeneity (or heterogeneity) at short distances can be seen as an indicator of dwelling size and crowdedness, which are similar characteristics of slums all over the world, thus these structure and texture measures could be used as well for rapid slum index estimation world wide when no other information is available. The examination of the residuals box map for the best performing model (SL_M3) indicates that the structure and texture variables are very sensitive to the presence of homogeneous land covers, such as bare soil, grass or dense forest, and that the model could perform even better if the analytic regions boundaries were delineated including only the built-up areas and their related open spaces, and excluding green adjacent areas that aren't part of the neighborhoods.

The presence of spatial autocorrelation in regard to the slum index in Medellin city indicate that poverty in this city could be tackled by a policy of strategically located investments and programs to booster economic development in some neighborhoods of the city, since their outcomes can be expected to affect not only the neighborhood where the investment is going to be located itself, but the surrounding neighborhoods as well. From this point of view, the city public investments made in the last decade, such as the public library parks and high quality schools, seem to be an appropriate strategy to reduce poverty in this city and should be encouraged for continuing and improving.

Finally, we highlight the fact that the input remote sensing data used in this research is rather simple: the most important requirement is the high spatial resolution in order to capture the differences in the homogeneity of the urban layout. We didn't use near-infrared spectral information nor other special spectral features of the satellite imagery and yet we could classify the input image into basic land cover types with an overall accuracy of 95%. The structure and texture variables were calculated from the first principal component band after the processing of the red, green and blue bands of the color image. Any aerial color image with very high spatial resolution could serve for this purpose. This opens up the possibility for rapid slum index estimation and intra-urban poverty pattern analysis from remote sensing data at a very low cost, since similar imagery to the used in this research could be easily obtained from internet services like Google Earth and Microsoft Bing Imagery. Urban planners and policy makers from many cities in the less developed countries of the world that have sparse data or no data at all can benefit with this rapid and low-cost approach for intra-urban poverty mapping and smart resource allocation. Future research should be made in order to develop applications that can make use of those internet imagery services without the need to download the imagery to the researcher's computer.

References

- Amrhein, C. G. and Flowerdew, R. (1992). The effect of data aggregation on a Poisson regression model of Canadian migration. *Environment and Planning A*, 24(10):1381–1391.
- Arbia, G. (1989). *Spatial data configuration in statistical analysis of regional economic and related problems*. Advanced Studies in Theoretical and Applied Econometrics. Kluwer Academic Publishers, Dordrecht, The Netherlands, advance st edition.
- Avelar, S., Zah, R., and Tavares-Correa, C. (2009). Linking socioeconomic classes and land cover data in Lima, Peru: Assessment through the application of remote sensing and GIS. *International Journal of Applied Earth Observation and Geoinformation*, 11(1):27–37.
- Balaguer, A., Ruiz, L. A., Hermosilla, T., and Recio, J. A. (2010). Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. *Computers & Geosciences*, 36(2):231–240.
- Balaguer-Beser, A., Ruiz, L. A., Hermosilla, T., and Recio, J. A. (2013). Using semivariogram indices to analyse heterogeneity in spatial patterns in remotely sensed images. *Computers and Geosciences*, 50:115–127.
- Barros, M. (2008). Slums detection through lacunarity-based texture analysis of remote sensing images. In *Expert group meeting on slum mapping*. ITC, The Netherlands, Enschede. ITC, UN-HABITAT, CIESIN.
- Barros, M. and Sobreira, F. (2005). Analysing spatial patterns in slums: A multiscale approach. In *Congresso Planejamento Urbano Regional Integrado Sustentável PLURIS*, Sao Carlos.
- DANE (2010). Boletín Censo general 2005. Perfil Medellín, Antioquia. Technical report, DANE, Bogota, Colombia.
- DANE and Municipio de Medellín (2010). Proyecciones de población Municipio de Medellín por comunas y corregimientos. Años 1993, 2005 - 2015. Technical report, DANE & Alcaldía de Medellín, Bogota, Colombia.
- Diehr, P. (1984). Small area statistics: large statistical problems. *American Journal of Public Health*, 74(4):313–4.
- Duncan, D. T., Aldstadt, J., Whalen, J., White, K., Castro, M. C., and Williams, D. R. (2012). Space, race, and poverty: Spatial inequalities in walkable neighborhood amenities? *Demographic Research*, 26(17):409–448.
- Duque, J. C., Anselin, L., and Rey, S. J. (2012a). The max-p-regions problem. *Journal of Regional Science*, 52(3):397–419.

- Duque, J. C., Artís, M., and Ramos, R. (2006). The ecological fallacy in a time series context: evidence from Spanish regional unemployment rates. *Journal of Geographical Systems*, 8(4):391–410.
- Duque, J. C., Dev, B., Betancourt, A., and Franco, J. L. (2011). ClusterPy: Library of spatially constrained clustering algorithms, Version 0.9.9.
- Duque, J. C., Royuela, V., and Noreña, M. (2012b). A stepwise procedure to determine a suitable scale for the spatial delineation of urban slums. In Fernandez, E. and Rubiera Morollón, F., editors, *Defining the spatial scale in modern regional analysis. Advances in Spatial Science*, chapter ??, pages xxx – xxx. Springer-Verlag, Berlin Heidelberg.
- Fiumi, L. (2012). Surveying the roofs of Rome. *Journal of Cultural Heritage*, 13:304–313.
- Fotheringham, A. S. and Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, 23(7):1025–1044.
- Galeon, F. A. (2008). Estimation of population in informal settlement communities using high resolution satellite image. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(Part B4):1377–1382.
- Green, N. E. (1956). Scale analysis of urban structures: A study of Birmingham, Alabama. *American Sociological Review*, 21(1):8–13.
- Green, N. E. (1957). Aerial photographic interpretation and the social structure of the city. *Photogrammetric Engineering*, 23:89–99.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621.
- Hofmann, P., Strobl, J., Blaschke, T., and Kux, H. J. H. (2008). Detecting informal settlements from QuickBird data in Rio de Janeiro using an object-based approach. In Blaschke, T., Lang, S., and Hay, G. J., editors, *Object-based image analysis - Spatial concepts for knowledge-driven remote sensing applications*, chapter 6.1, pages 531–553. Springer, New York.
- Holt, J. B. (2007). The topography of poverty in the United States: a spatial analysis using county-level data from the Community Health Status Indicators project. *Preventing chronic disease*, 4(4):1–9.
- Jain, S. (2008). Remote sensing application for property tax evaluation. *International Journal of Applied Earth Observation and Geoinformation*, 10(1):109–121.
- Kit, O., Lüdeke, M., and Reckien, D. (2012). Texture-based identification of urban slums in Hyderabad, India using remote sensing data. *Applied Geography*, 32(2):660–667.
- Kit, O., Lüdeke, M., and Reckien, D. (2013). Defining the bull’s eye: satellite imagery-assisted slum population assessment in Hyderabad, India. *Urban Geography*, 0(0):1–12.
- Kohli, D., Sliuzas, R., Kerle, N., and Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, 36(2):154–163.
- Kux, H. J. H., Novack, T., Ferreira, R., and Oliveira, D. A. (2010). Urban land cover classification using optical VHR data and the knowledge-based system InterIMAGE. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII(4-C7).
- Lu, D. and Weng, Q. (2004). Spectral mixture analysis of the urban landscape in Indianapolis with Landsat ETM+ imagery. *Photogrammetric Engineering and Remote Sensing*, 70(9):1053–1062.
- Montgomery, M. R. and Hewett, P. C. (2005). Urban poverty and health in developing countries : Household and neighbourhood effects. *Demography*, 43(3):397–425.
- Netzband, M., Banzhaf, E., Höfer, R., and Hannemann, K. (2009). Identifying the poor in cities: how can remote sensing help to profile slums in fast growing cities and megacities? *IHDP Update*, pages 22–28.
- Niebergall, S., Loew, A., and Mauser, W. (2007). Object-oriented analysis of very high-resolution QuickBird data for mega city research in Delhi, India. In IEEE, editor, *2007 Urban Remote Sensing Joint Event*, pages 0–7, Paris, France. IEEE.
- Novack, T. and Kux, H. J. (2010). Urban land cover and land use classification of an informal settlement area using the open-source knowledge-based system InterIMAGE. *Journal of Spatial Science*, 55(1):23–41.

- Okwi, P. O., Ndeng'e, G., Kristjanson, P., Arunga, M., Notenbaert, A., Omolo, A., Henninger, N., Benson, T., Kariuki, P., and Owuor, J. (2007). Spatial determinants of poverty in rural Kenya. *Proceedings of the National Academy of Sciences of the United States of America*, 104(43):16769–16774.
- Openshaw, S. (1977a). A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the Institute of British Geographers*, 2(4):459–472.
- Openshaw, S. (1977b). Optimal zoning systems for spatial interaction models. *Environment and Planning A*, 9(2):169–184.
- Openshaw, S. (1984). *The modifiable areal unit problem*, CATMOG 38. Geo Books, Norwich, vol. 38. edition.
- Openshaw, S. and Taylor, P. J. (1981). The modifiable areal unit problem. In Wrigley, N. and Bennett, R. J., editors, *Quantitative Geography*, pages 60–70. Routledge and Kegan Paul, Henley-on-Thames, Oxon.
- Orford, S. (2004). Identifying and comparing changes in the spatial concentrations of urban poverty and affluence: a case study of inner London. *Computers, Environment and Urban Systems*, 28(6):701–717.
- Owen, K. K. and Wong, D. W. (2013). An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics. *Applied Geography*, 38:107–118.
- Paelinck, J. H. (2000). On aggregation in spatial econometric modelling. *Journal of Geographical Systems*, 2(2):157–165.
- Paelinck, J. H. P. and Klaassen, H. (1979). *Spatial econometrics*. Saxon House, Farnborough, UK.
- Patino, J. E. (2010). Potencial de las imágenes satelitales como herramienta de planificación urbana y ambiental en el Valle de Aburrá. In Hermelin Arbaux, M., Echeverry Restrepo, A., and Giraldo Ramirez, J., editors, *Medellín, Medio Ambiente, Urbanismo y Sociedad*, chapter 1, pages 112–127. Fondo Editorial EAFIT, Medellin, 1st edition.
- Patino, J. E. and Duque, J. C. (2013). A review of regional science applications of satellite remote sensing in urban settings. *Computers, Environment and Urban Systems*, 37:1–17.
- Phinn, S. R., Stanford, M., Scarth, P., Murray, A. T., and Shyy, P. T. (2002). Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques. *International Journal of Remote Sensing*, 23(20):4131–4153.
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rashed, T., Weeks, J. R., Stow, D., and Fugate, D. (2005). Measuring temporal compositions of urban morphology through spectral mixture analysis: toward a soft approach to change analysis in crowded cities. *International Journal of Remote Sensing*, 26(4):699–718.
- Rhinane, H. (2011). Detecting slums from SPOT data in Casablanca Morocco using an object based approach. *Journal of Geographic Information System*, 03(03):209–216.
- Ridd, M. K. (1995). Exploring a VIS (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. *International Journal of Remote Sensing*, 16(12):2165–2185.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3):351–357.
- Ruiz, L. A., Recio, J. A., Fernández-Sarría, A., and Hermosilla, T. (2011). A feature extraction software tool for agricultural object-based image analysis. *Computers and Electronics in Agriculture*, 76(2):284–296.
- Setiawan, H., Mathieu, R., and Thompson-Fawcett, M. (2006). Assessing the applicability of the VIS model to map urban land use in the developing world: Case study of Yogyakarta, Indonesia. *Computers, Environment and Urban Systems*, 30(4):503–522.
- Shekhar, S. (2012). Detecting slums from Quick Bird data in Pune using an object oriented approach. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX(B8):519–524.

- Sowunmi, F. A., Akinyosoje, V. O., Okoruwa, V. O., and Omonona, B. T. (2012). The Landscape of Poverty in Nigeria: A Spatial Analysis Using Senatorial Districts- level Data. *American Journal of Economics*, 2(5):61–74.
- Stoler, J., Daniels, D., Weeks, J. R., Stow, D. a., Coulter, L. L., and Finch, B. K. (2012). Assessing the utility of satellite imagery with differing spatial resolutions for deriving proxy measures of slum presence in Accra, Ghana. *GIScience & Remote Sensing*, 49(1):31–52.
- Stow, D., Lopez, A., Lippitt, C., Hinton, S., and Weeks, J. R. (2007). Object-based classification of residential land use within Accra, Ghana based on QuickBird satellite data. *International Journal of Remote Sensing*, 28(22):5167–5173.
- Tapiador, F. J., Avelar, S., Tavares-Corrêa, C., and Zah, R. (2011). Deriving fine-scale socioeconomic information of urban areas using very high-resolution satellite imagery. *International Journal of Remote Sensing*, 33(21):6437–6456.
- Taubenböck, H., Wegmann, M., Roth, A., Mehl, H., and Dech, S. (2009a). Urbanization in India Spatiotemporal analysis using remote sensing data. *Computers, Environment and Urban Systems*, 33(3):179–188.
- Taubenböck, H., Wurm, M., Setiadi, N., Gebert, N., Roth, A., Strunz, G., Birkmann, J., and Dech, S. (2009b). Integrating remote sensing and social science - The correlation of urban morphology with socioeconomic parameters. In IEEE, editor, *2009 Urban Remote Sensing Joint Event*, London.
- Troy, A., Grove, J. M., and O’Neil-Dunne, J. (2012). The relationship between tree canopy and crime rates across an urbanrural gradient in the greater Baltimore region. *Landscape and Urban Planning*, 106:262–270.
- United Nations (2007). World urbanization prospects: The 2007 revision.
- United Nations (2008). Urban population, development and the environment 2007.
- United Nations (2011). The millennium development goals report 2011. Technical report, United Nations, New York.
- United Nations (2012). World Urbanization Prospects: The 2011 Revision. Technical report, United Nations Department of Economic and Social Affairs/Population Division, New York.
- Voss, P. R., Long, D. D., Hammer, R. B., and Friedman, S. (2006). County child poverty rates in the US: a spatial regression approach. *Population Research and Policy Review*, 25(4):369–391.
- Weeks, J. R., Hill, A. G., Stow, D., Getis, A., and Fugate, D. (2007). Can we spot a neighborhood from the air? Defining neighborhood structure in Accra, Ghana. *GeoJournal*, 69(1-2):9–22.