UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DEPARTMENT OF CARTOGRAPHIC
ENGINEERING, GEODESY AND
PHOTOGRAMMETRY

CITYSCAPE, POVERTY AND CRIME: A QUANTITATIVE ASSESSMENT USING VHR IMAGERY

PhD DISSERTATION

Jorge Eduardo Patiño Quinchía

Advisors:

Josep Eliseu Pardo Pascual, Universitat Politècnica de València
Luis Ángel Ruiz Fernández, Universitat Politècnica de València
Juan Carlos Duque Cardona, EAFIT University

March 2015

© copyright disclaimer

Acknowledgements

The author thanks the guidance and assistance of all his advisors, professors Josep E. Pardo, Luis Ángel Ruiz, and Juan C. Duque, who also asked the seminal question that led to this research. The author also thanks Alejandro Betancourt of RiSE Group at EAFIT University, and Jaime Almonacid, José Luis Gil and Txomin Hermosilla of Geo-Environmental Cartography and Remote Sensing Group at Universitat Politècnica de València for their friendship and their invaluable help with the learning of new tools needed to develop this work.

This research was made possible by funding of Medellin City Hall Enlaza-Mundos program, 2010-2 Call, and EAFIT University, Colombia, research projects 342-000033 (2012), 435-000060 (2013), and 513-000084 (2014). The author also thanks Professor John Weeks, director of the International Population Center at San Diego State University and professors Hermilson Velázquez, Andrés Ramírez Hassan and Gustavo Canavire at EAFIT University for their insightful observations and suggestions during the different stages of this work.

Finally but most important, infinite thanks to Ani, who suffered the author's thesis mood during this time and chose to support him ever!

Abstract

The first part of this work reviews the potential applications of satellite remote sensing to regional science research in urban settings. Regional science is the study of social problems that have a spatial dimension. The availability of satellite remote sensing data has increased significantly in the last two decades, and these data constitute a useful data source for mapping the composition of urban settings and analyzing changes over time. The increasing spatial resolution of commercial satellite imagery has influenced the emergence of new research and applications of regional science in urban settlements because it is now possible to identify individual objects of the urban fabric. The most common applications found in the literature are the detection of urban deprivation hot spots, quality of life index assessment, urban growth analysis, house value estimation, urban population estimation, urban social vulnerability assessment, and the variability of intra-urban crime rates. The satellite remote sensing imagery used in these applications has medium, high or very high spatial resolution, such as images from Landsat MSS, Landsat TM and ETM+, SPOT, ASTER, IRS, Ikonos and QuickBird. Consistent relationships between socio-economic variables derived from censuses and field surveys and proxy variables of vegetation coverage measured from satellite remote sensing data have been found in several cities in the US. Different approaches and techniques have been applied successfully around the world, but local research is always needed to account for the unique elements of each place. Spectral mixture analysis, object-oriented classifications and image texture measures are some of the techniques of image processing that have been implemented with good results. Many regional scientists remain skeptical that satellite remote sensing will produce useful information for their work. More local research is needed to demonstrate the real potential and utility of satellite remote sensing for regional science in urban environments.

The second part of this work focuses on data manipulation and extraction of urban fabric descriptors from a very high spatial resolution (VHR) image, and the integration with socioeconomic data at object level. We extract information on land

cover composition using per-pixel classification and on urban texture and structure using an automated tool for texture and structure feature extraction at object level. We use data from Medellin (Colombia), which is the second larger city in the country, has been one of the most violent cities in the world in past decades and is still one of the most socioeconomically divergent. This city is a useful location for conducting intra-urban variability studies because it has experienced high population growth rates since the 1950s, and the unplanned urban growth in some parts of the city resulted in a high degree of spatial heterogeneity in both the socioeconomic and physical characteristics of its neighborhoods.

The third part of this work contributes empirical evidence about the usefulness of remote sensing imagery to quantify the degree of poverty at the intra-urban scale. This concept is based on two premises: first, that the physical appearance of an urban settlement is a reflection of the society; and second, that the people who reside in urban areas with similar physical housing conditions have similar social and demographic characteristics. We evaluate the potential of the image-derived urban fabric descriptors to explain a measure of poverty known as the Slum Index. We found that these variables explain up to 59% of the variability in the Slum Index. Similar approaches could be used to lower the cost of socioeconomic surveys by developing an econometric model from a sample and applying that model to the rest of the city and to perform intercensal or intersurvey estimates of intra-urban Slum Index maps.

The last part of this work analyzes the relation between the urban layout and crime. The link between place and crime is at the base of social ecology theories of crime that focus in the relationship of the characteristics of geographical areas and crime rates. The broken windows theory states that visible cues of physical and social disorder in a neighborhood can lead to an increase in more serious crime. The crime prevention through environmental design (CPTED) planning approach seeks to deter criminal behavior by creating defensible spaces. Based on the premise that a settlement's appearance is a reflection of the society, we ask whether a neighborhood's design has a quantifiable imprint when seen from space using urban fabric descriptors computed from VHR imagery. We tested which land cover, structure and texture descriptors were significantly related to intra-urban homicide rates in Medellin, Colombia, while controlling for socioeconomic confounders. The percentage of impervious surfaces other than clay roofs, the fraction of clay roofs to impervious surfaces, two structure descriptors related to the homogeneity of the urban layout, and the uniformity texture descriptor were all statistically significant. Areas with higher homicide rates tended to have higher local variation and less general homogeneity; that is, the urban layouts were more crowded and cluttered, with small dwellings with different roofing materials located in close proximity to one another, and these regions

often lacked other homogeneous surfaces such as open green spaces, wide roads, or large facilities. These results seem to be in agreement with the broken windows theory and CPTED in the sense that more heterogeneous and disordered urban layouts are associated with higher homicide rates.

Resumen

La primera parte de este trabajo presenta una revisión de las potenciales aplicaciones de la teledetección satelital en la investigación de ciencia regional en entornos urbanos. La ciencia regional es el estudio de los problemas sociales que tienen una dimensión espacial. La disponibilidad de datos de percepción remota satelital se ha incrementado significativamente en las dos últimas décadas, y éstos datos son una fuente de información útil para el mapeo de la composición de los entornos urbanos y para analizar sus cambios en el tiempo. La resolución espacial de las imágenes satelitales comerciales también se ha venido aumentando, y esto a su vez ha influenciado el surgimiento de nuevas investigaciones y aplicaciones de ciencia regional en asentamientos urbanos ya que ahora es posible identificar los objetos que componen el tejido urbano. Las aplicaciones más comunes encontradas en la literatura son la detección de *hot spots* de pobreza urbana, la evaluación de índices de calidad de vida, el análisis del crecimiento urbano, la estimación de valores de vivienda, la estimación de población urbana, la evaluación de la vulnerabilidad social y las variaciones intra-urbanas en tasas de crimen. Las imágenes satelitales usadas en esas aplicaciones tienen resolución espacial media, alta o muy alta, tales como las imágenes de Landsat MSS, Landsat TM y ETM+, SPOT, ASTER, IRS, Ikonos y Quickbird. Se han encontrado relaciones consistentes entre variables socio-económicas obtenidas de censos y encuestas y variables proxy de la cobertura de vegetación medidas con imágenes satelitales en varias ciudades de Estados Unidos. Técnicas y aproximaciones diferentes se han aplicado exitosamente alrededor del mundo, pero siempre es necesaria la investigación local para dar cuenta de los elementos únicos de cada lugar. Algunas de las técnicas que se han implementado y obtenido buenos resultados son el análisis de mezcla espectral, las clasificaciones orientadas a objetos y las medidas de textura de la imagen. Muchos científicos regionales aún son escépticos con respecto a que la teledetección satelital puede producir información útil para su trabajo. Es necesario más investigación local para demostrar el potencial y la utilidad real de la percepción remota para la ciencia regional en ambientes urbanos.

La segunda parte de este trabajo se enfoca en la manipulación de datos y la extracción de descriptores del tejido urbano a partir de una imagen de muy alta resolución espacial, así como de la integración con datos socioeconómicos a escala de objeto. Extraemos información de la composición de las coberturas del suelo usando una clasificación por pixel y de la textura y estructura urbana usando una herramienta automática para la extracción de variables de textura y estructura a escala de objeto. Usamos datos de Medellín (Colombia), que es la segunda ciudad más grande en el país, ha sido una de las ciudades más violentas del mundo en décadas pasadas y es todavía una de las ciudades con más desigualdades socioeconómicas. Esta ciudad es un sitio útil para llevar a cabo estudios de variaciones intra-urbanas porque ha experimentado altas tasas de crecimiento poblacional desde la década de 1950 y el crecimiento urbano no planeado en algunas partes de la ciudad produjo un alto de grado de heterogeneidad espacial en las características físicas y socioeconómicas de sus barrios.

La tercera parte de este trabajo aporta evidencia empírica acerca de la utilidad de las imágenes satelitales para cuantificar el grado de pobreza a escala intra-urbana. Este concepto se basa en dos premisas: primero, que la apariencia física de un asentamiento urbano es un reflejo de la sociedad que lo habita; y segundo, que las personas que residen en áreas urbanas con condiciones físicas de vivienda parecidas tienen también características sociales y demográficas similares. Evaluamos el potencial de los descriptores del tejido urbano extraídos de la imagen para explicar una medida de pobreza conocida como el índice *Slum*. Encontramos que esas variables explican hasta un 59% de la variabilidad en el índice *Slum*. Aproximaciones similares a esta podrían usarse para disminuir el costo de encuestas socioeconómicas por medio del desarrollo de un modelo econométrico usando una muestra y luego aplicando el modelo al resto de la ciudad, y para elaborar estimaciones inter-censales o inter-encuestas de mapas intra-urbanos del índice *Slum*.

La última parte del trabajo analiza la relación entre el trazado urbano y crimen. El enlace entre el lugar y el crimen está en la base de las teorías socio-ecológicas de crimen que se enfocan en la relación de las características de las áreas geográficas y las tasas de crimen. La teoría de las ventanas rotas afirma que las evidencias visibles de desorden físico y social en un barrio pueden llevar al incremento de crímenes más serios. La aproximación de prevención del crimen por medio del diseño ambiental urbano (CPTED, por sus siglas en inglés) busca desalentar el comportamiento criminal por medio de la creación de espacios defendibles. Con base en la premisa de que la apariencia de un asentamiento es un reflejo de la sociedad, nos preguntamos si el diseño del barrio tiene un impacto cuantificable cuando se observa desde el espacio usando descriptores del tejido urbano obtenidos de imágenes de muy alta resolución.

Para esto probamos cuáles de las variables de coberturas del suelo, estructura y textura están relacionadas significativamente con tasas intra-urbanas de homicidio en Medellín, Colombia, al tiempo que controlamos por factores socioeconómicos. El porcentaje de superficies impermeables diferentes a los techos de arcilla, la fracción de techos de arcilla sobre las superficies impermeables, dos variables de estructura relacionadas con la homogeneidad del trazado urbano y la variable de textura uniformidad resultaron estadísticamente significativas. Las áreas con tasas de homicidio más altas tienden a tener mayor variación local y menor homogeneidad general; esto es, los trazados urbanos son más desordenados y hacinados, con pequeñas viviendas que tienen materiales diferentes en sus techos localizadas muy cerca unas de otras, y estas áreas carecen a menudo de otras superficies homogéneas tales como espacios verdes abiertos, vías amplias y grandes construcciones industriales o institucionales. Estos resultados parecen estar en acuerdo con la teoría de las ventanas rotas y CPTED en el sentido de que los trazados urbanos más desordenados y heterogéneos están asociados con tasas de homicidio más altas.

Resum

La primera part del treball presenta una revisió de les potencials aplicacions de la teledetecció espacial a la investigació en ciència regional en entorns urbans. La ciència regional és l'estudi dels problemes socials que hi tenen una dimensió espacial. La disponibilitat de dades de percepció remota des de satèl·lits s'ha incrementat significativament a les dues últimes dècades i, en aquest cas, són una font d'informació útil per al cartografiat de la composició del entorns urbans i per analitzar-hi els seus canvis al llarg del temps. La resolució espacial de les imatges de satèl·lit comercials també han anat augmentant i això, ha influït en l'aparició de noves investigacions i aplicacions a la ciència regional en assentaments urbans ja que ara hi és possible identificar els objectes que composen el teixit urbà. Les aplicacions més comunes trobades a la literatura són la detecció de punts calents de pobresa urbana, l'avaluació dels índex de qualitat de vida, les anàlisis de creixement urbà, l'avaluació de la vulnerabilitat social i les variacions intraurbanes de les taxes de crims. Les imatges de satèl·lit emprades en aquestes aplicacions tenen resolució espacial mitjana, alta o molt alta, com ara les imatges Landsat MSS, Landsat TM i ETM+, SPOT, ASTER, IRS, Ikonos y Quickbird. S'han torbat relacions consistents entre variables socioeconòmiques obtingudes de censos i enquestes i variables aproximades de la cobertura de vegetació mesurades amb imatges de satèl·lit de varies ciutats del Estats Units. Diferents tècniques i aproximacions metodològiques s'han aplicat en diferents indrets del món però sempre és necessària la investigació local per considerar els elements únics de cada lloc. Algunes de les tècniques que s'han implementat i han donat bons resultats són l'anàlisi de mescla espectral, les classificacions orientades a objecte i les mesures de textura de les imatges. Molts científics regionals però encara en són d'escèptics a que la teledetecció des de satèl·lit puga produir informació útil per al seu treball. És necessari més investigació local per demostrar el potencial i la utilitat real de la percepció remota per a la ciència regional en ambients urbans.

La segona part del treball es dirigeix cap a la manipulació de dades i la extracció de descriptors del teixit urbà a partir de una imatge de molt alta resolució espacial,

així com la integració amb dades socioeconòmiques a escala d'objecte. Extraiem informació de la composició de les cobertures del sòl emprant una classificació per píxel, i de la textura i estructura urbana emprant una eina automàtica per a l'extracció de variables de textura i d'estructura a escala d'objecte. Utilitzem dades de Medellín (Colòmbia), que és la segona ciutat més gran del país, i ha estat una de les ciutats més violentes del món en dècades passades i, encara avui en dia, és una de les ciutats amb més desigualtats socioeconòmiques. Aquesta ciutat és un lloc útil per a dur endavant estudis de variacions intraurbanes perquè ha experimentat altes taxes de creixement poblacional des de la dècada dels cinquantes del segle passat i el creixement urbà no planejat en algunes parts de la ciutat va produir un alta grau de heterogeneïtat espacial en les característiques físiques i socioeconòmiques dels seus barris.

La tercera part d'aquest treball aporta evidència empírica sobre la utilitat de les imatges de satèl·lit per quantificar el grau de pobresa a escala intraurbana. Aquest concepte es bassa en dues premisses: primer, que l'aparença física d'un assentament urbà n'és un reflex de la societat que l'habita; i segon, que les persones que resideixen en àrees urbanes amb condicions físiques de vivenda paregudes tenen també característiques socials i demogràfiques similars. Avaluem el potencial dels descriptors del teixit urbà extrets de la imatge per explicar una mesura de pobresa coneguda com index *Slum*. Trobem que aquestes variables expliquen fins un 59% de la variabilitat de l'índex *Slum*. Aproximacions semblants a aquesta es podrien emprar per a disminuir el cost de les enquestes socioeconòmiques mitjançant el desenvolupament d'un model economètric utilitzant una mostra i després aplicant el model a la resta de la ciutat, i per elaborar estimacions inter-censals o inter-enquestes de mapes intraurbans de l'índex *Slum*.

La darrera part del treball analitza la relació entre el traçat urbà i el crim. L'enllaç entre el lloc i el crim està a la base de les teories socio-ecològiques del crim que es centren en la relació de les característiques de les àrees geogràfiques i les taxes de crims. La teoria de les finestres trencades afirma que les evidències visibles de desordre físic i social d'un barri pot portar a l'augment de crims més greus. L'aproximació de prevenció del crim mitjançant el disseny ambiental urbà (CPTED, per les seues sigles en anglès) cerca desanimar el comportament criminal mitjançant la creació d'espais defensables. Basant-se en la premissa de que l'aparença d'un assentament n'és el reflex de la societat, ens hi preguntem si el disseny del barri té un impacte quantificable quan s'observa des de el espai, utilitzant descriptors del teixit urbà obtinguts de imatges de molt alta resolució. Per açò provem quines de les variables de cobertures del sòl, estructura i textura estan relacionades significativament amb taxes intraurbanes d'homicidi a Medellín, Colòmbia, al temps que controlem els factors socioeconòmics. Han resultat estadísticament significatius el percentatge de superfícies impermeables

diferents a les teulades de argila, la fracció de teulades d'argila sobre les superfícies impermeables, dues variables d'estructura relacionades amb la homogeneïtat del traçat urbà i la variable de textura de uniformitat. Les àrees amb taxes d'homicidi més altes tendeixen a presentar una major variació local i una menor homogeneïtat general; és a dir, el traçats urbans són més desordenats i amuntonats, amb petites vivendes que tenen materials diferents a les seues teulades localitzades molt prop unes d'altres, i aquestes àrees manquen sovint d'altres superfícies homogènies, com ara espais verds oberts, vies amplies i grans construccions industrials o institucionals. Aquests resultats pareixen estar-hi d'acord amb la teoria de les finestres trencades i CPTED en el sentit de que els traçats urbans més desordenats i heterogenis estan associats amb taxes d'homicides més altes.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter is an edited compilation of some sections of the paper 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. doi: 10.1016/j.compenvurbsys.2012.06.003.

1.1 Background: regional science and remote sensing

Regional science can be defined as a discipline concerned with the study of social, economic, political and behavioral phenomena that have regional or spatial dimensions, using combinations of analytical and empirical research methods (Isard, 1975; Isserman, 2004). The spatial aspect involved in regional science research is what makes remote sensing appealing for extracting spatial data that can be related to the phenomena under analysis. Aerial photography has been used as input data in regional science applications in urban settlements since the late 1950s. The use of this type of remotely sensed data is supported by the hypothesis that the surface appearance of a settlement is the result of the human population's social and cultural behavior and interaction with the environment, which leave their mark on the landscape. Remotely sensed data from satellites can also be used to measure the context of social phenomena, to gather additional contextual data on the environment in which people live, and to measure the environmental consequences of social processes (Rindfuss and Stern, 1998). Lo and Faber (1997) state that it is through greenness (i.e., the amount of green vegetation) that remotely sensed environmental data can be combined with socioeconomic census data and list early examples of the use of remotely sensed data for the social analysis of cities. Topics covered in those early studies included social structure and residential desirability (Green, 1957; Monier and Green,

1957), urban poverty (Metivier and McCoy, 1971; Mumbower and Donoghue, 1967), and urban quality of life (Weber and Hirsch, 1992). Mullens and Senger (1969) and Miller and Winer (1984) are also pioneering works that reported relationships between aerial remotely sensed data and demographic, social and economic characteristics of neighborhoods in a city. The application of empirical models to estimate biophysical, demographic and socioeconomic variables has been reported as one of the top five recurrent research themes in applications of remote sensing to urban environments (Phinn et al., 2002).

The availability of remotely sensed data has increased significantly as its costs have decreased in the last two decades. These data provide a useful method for mapping the compositions of cities and analyzing changes over time (Weng and Quattrochi, 2006). The unique characteristics of remotely sensed data, such as repeat cycle and wide area coverage, provide means for exploring and testing hypotheses and models about urban areas and for constructing new theories that can help policy makers to analyze and respond to problems that involve urbanization processes (Rashed et al., 2005). The increasing spatial resolution of commercial satellite imagery has been crucial to the emergence of new studies and applications related to urban settlements because it is now possible to identify the individual objects of the urban fabric, such as individual buildings and details of road networks and open spaces (Sliuzas et al., 2010).[1]

The increased availability of high-resolution remote sensing data can also be considered a response to the growing need for high spatial and temporal resolution data on urban agglomerations.[2] The worlds population is now mostly urban, and information about urban settings, their internal compositions and their dynamics is very important to the preservation of certain standards of living (Phinn et al., 2002). Processes taking place in urban areas are among the main drivers of land change on local to global scales (Herold, 2009). According to the United Nations, the 20th century witnessed the urbanization of the world's population, with the percentage of urban dwellers increasing from 13% in 1900 to 49% in 2005, and this figure is expected to grow to 60% in 2030 (United Nations, 2007, 2008). In 2007, the populations of the Americas, Europe and Oceania were over 70% urban, and those of Asia and Africa were approximately 40% urban. Every year, a larger absolute number of persons is added to the world's urban population; this situation is believed to be more significant in the less developed regions of the world, whose population growth in urban areas is projected to account for almost all of the world population growth between

[1] The complete Landsat archive is freely available through the University of Maryland Global Land Cover Facility (GLCF) and the United States Geological Survey (USGS) Internet sites.

[2] High-resolution population data in gridded format for 50 metropolitan statistical areas (MSAs) in the US have recently been released into the public domain.

2005 and 2030 (United Nations, 2008). This situation brings attention to the existing inequities within urban settings. In developing countries, variations in poverty and health within urban areas can be larger than the differences between urban and rural settlements (Montgomery and Hewett, 2005; Stow et al., 2007); hence, urban settings need to be studied and monitored at very high spatial and temporal resolutions.

Remote sensing is widely known among urban planners, city planners and policy makers as a useful tool for extracting biophysical information about the urban environment, including land-cover and land-use mapping, urban morphology description and analysis, vegetation distribution and characterization, hydrography and disaster relief. This tool is also widely used in the field of natural resource exploration and management. However, little is known about the detection of the subtle relationships between the physical appearance of the urban landscape and the socioeconomic conditions of the population. The data that are currently available from Earth observation systems present an opportunity to collect information about urban settlements at several scales and on several dimensions (Netzband and Jürgens, 2010), and urban population growth and problems will increase in relevance in the coming decades (Sembler, 2006; Stow et al., 2007; Weeks et al., 2004). Therefore, it is important to demonstrate how remote sensing tools can contribute useful information to the study of cities and urban settlements.

Remote sensing is used to gather information from distant objects by measuring the radiation that they reflect or emit. Remote sensing began with the development of photography in early 1800s by Louis Daguerre. The first aerial photographs were taken in the 1860s by Felix Tournachon in France using a camera mounted on a balloon (de Sherbinin et al., 2002; Short, 2010). Cameras mounted on planes were first used for military reconnaissance in World War I and II. The military need for separating real vegetation from camouflage resulted in the development of remote sensing beyond the human eye's visible range with the introduction of infrared wavelength sensors. After World War II, civilian applications of airborne remote sensing were developed for hazard mapping, vegetation mapping and planning. Space-based remote sensing started in the late 1950s with the launch of the first military intelligence satellite. A few years later, the first US meteorological satellite was launched, which was designed to aid in the production of generalized weather maps. The first Earth observation satellite was launched a decade later in 1972 and is known today as Landsat (de Sherbinin et al., 2002).

Although aerial photography has been used as a tool for urban analysis since the late 1950s, the focus of remote sensing research has shifted to the use of imagery acquired by Earth-orbiting satellite sensors as a result of the lower costs and frequency of updates of this imagery (Donnay et al., 2001). The earliest launched satellites are

known as first-generation sensors and were able to produce digital images of the Earth's surface with relatively moderate spatial resolution (i.e., 80m of pixel size for Landsat MSS – Multispectral Scanning System) that were used primarily for regional scale studies. Second-generation satellites, such as Landsat Thematic Mapper and SPOT-HRV, increased the spatial resolution to 30 and 10 m, respectively, and enabled more detailed studies of urban systems. Third-generation satellites with very high spatial resolution (5–0.5 m), such as Ikonos and Quickbird, were launched in the last decade and have stimulated the development of newer detailed scale applications related to urban settlements, as anticipated by Donnay et al. (2001).

Sensors on board satellites used to gather data are of one of two types: active or passive. Both types collect electromagnetic radiation from the Earth's surface. Active sensors have their own energy source and emit a signal that travels through the atmosphere, reflects on the Earth's surface and returns to the sensor, which measures the signal's travel time and strength. Synthetic Aperture Radar (SAR) is an example of an active sensor that uses long-wavelength signals and thus can penetrate clouds or bad weather conditions. Passive sensors do not have their own energy source and usually record the radiation from the Sun that is reflected from the Earth's surface. Photographic cameras and multispectral scanners are passive sensors often used in satellite remote sensing.

The visible part of the electromagnetic spectrum is very small, and most satellite systems have been designed to be sensitive to other portions of the spectrum as well. This characteristic enables remote sensing analysts to see portions of the spectrum that the human eye cannot detect, thereby enhancing their ability to identify different surface materials. The spectral properties of a sensor are defined by the number, placement and width of bands within the electromagnetic spectrum that it is able to record. Panchromatic sensors measure reflected radiation in a single portion, usually located in the visible or infrared part of the electromagnetic spectrum, whereas multispectral sensors collect radiation in discrete parts of the spectrum, which are recorded as separate images called bands or channels. Today, most high-resolution satellite remote sensing systems are composed of a panchromatic sensor and a multispectral sensor. Each sensor usually has a different spatial resolution, with the resolution of the panchromatic sensor being higher than that of the multispectral sensor.

Satellite systems can be placed in two types of orbits around the Earth. A geostationary orbit is obtained when the satellite orbits at a very high altitude and at the same speed as the Earth's rotation, thus remaining in a stationary position relative to the Earth and directed to the same portion of the Earth's surface. Satellites in this type of orbit are limited to a spatial resolution of 1–10 square kilometers and are used for weather and climate data collection and communications. A polar orbit is

closer to the Earth's surface, between 700 and 1000 kilometers; passes through the poles of the Earth, describing a steep inclination relative to the Equator; and orbits in the opposite direction to the Earth's rotation. Satellites in this type of orbit can cycle around the Earth in 100–120 min, several times per day, and return to the same position after 2 weeks or more, allowing them to gather data on the same location over time. The time that elapses before a satellite returns to the same position is called the temporal resolution of the remote sensing system, and it has been defined as the capability for acquiring repetitive imagery over a certain time interval (Fugate et al., 2010). Satellites in this type of orbit can obtain images with higher spatial resolution, ranging from 1 to 200m (de Sherbinin et al., 2002).

Thus far, most regional science research in urban settings has used imagery from polar-orbiting satellites with passive sensors with medium, high or very high spatial resolution that provide good spectral resolution with multiple bands in the visible portion of the electromagnetic spectrum, at least one band located in the infrared portion of the spectrum and a panchromatic band. Table 1.1 lists the satellite remote sensing systems that are frequently used in urban regional science applications.[3] Figure 1.1 illustrates the differences in spatial resolution of imagery from four different sensors by showing the same area of a city.

The two most recent reviews of the use of remote sensing imagery in research related to socioeconomic issues in urban settings are Jensen and Cowen (1999) and Miller and Small (2003). Jensen and Cowen (1999) focused on the technical requirements of remotely sensed data to extract information related to urban and suburban infrastructure and socioeconomic attributes. They reviewed only population estimation and quality of life indicators. They stated that one of the most important requirements for detecting those features of interest in the image is the spatial resolution of the remotely sensed data. In the last decade, we have witnessed a large increase in the use of very high-resolution space-borne sensors and programs and the introduction of "government-wide" data purchases, which, in turn, have resulted in an increase in the availability of imagery at very high spatial resolutions. Thus, the spatial resolution requirement has been met, and the temporal resolution issue is becoming more of a budget issue than a technical one. These technological advancements have created a clear opportunity for the regional science community to begin to explore and use remotely sensed data in their daily work.

Miller and Small (2003) reviewed the potential applications of remote sensing in urban environmental research and policy. They showed that remotely sensed data could be used to obtain internally consistent measurements of physical properties

[3]For a detailed list of satellite platforms, the reader can review the ITC's database of Satellites and Sensors at http://www.itc.nl/research/products/sensordb/AllSatellites.aspx .

Table 1.1: Most often used satellite remote sensing systems in regional science applications in urban settlements.

System	Spectral resolution	Spatial resolution (pixel size - meters)	Temporal resolution (days)	Archive since
Landsat MSS	3 bands visible 1 band infrared 1 band thermal infrared	80	18	1972
Landsat TM	3 bands visible 3 bands infrared 1 band thermal infrared	30 - visible and infrared 60 - thermal infrared	16	1986
Landsat ETM+	3 bands visible 3 bands infrared 2 bands thermal infrared 1 band panchromatic	30 - visible and infrared 60 - thermal infrared 15 - panchromatic	16	1999
Landsat 8	4 bands visible 4 bands infrared 2 bands thermal infrared 1 band panchromatic	30 - visible and infrared 100 - thermal (resampled to 30) 15 - panchromatic	16	2013
SPOT 1 SPOT 2 SPOT 3	2 bands visible 1 band infrared 1 band panchromatic	20 - visible and infrared 10 - panchromatic	26	1986
SPOT 4	2 bands visible 2 bands infrared 1 band panchromatic	20 - visible and infrared 10 - panchromatic	2-3	1998
SPOT 5	2 bands visible 2 bands infrared 1 band panchromatic	20 - mid infrared 10 - visible and near infrared 2.5-5 - panchromatic	2-3	2002
SPOT 6	3 bands visible 1 band infrared 1 band panchromatic	6 - visible and near infrared 1.5 panchromatic	3	2012
ASTER	3 bands visible 6 bands infrared 5 bands thermal infrared	15 - visible 30 - infrared 90 - thermal infrared	16	1999
IRS-C1	2 bands visible 2 bands infrared 1 band panchromatic	23.5 - multispectral 5 - panchromatic	5-24	1995
Ikonos	3 bands visible 1 band infrared 1 panchromatic	4 - multispectral 0.8 - panchromatic	1.5-2.9	1999
Quickbird	3 bands visible 1 band infrared 1 band panchromatic	2.4 - multispectral 0.6 - panchromatic	1-3.5	2001
Worldview-1	1 band panchromatic	0.5	1.7	2007
Worldview-2	6 bands visible 2 bands infrared	1.85 - multispectral 0.5 - panchromatic	1.1	2009
Geoeye-1	3 bands visible 1 band infrared 1 band panchromatic	1.65 - multispectral 0.4 - panchromatic	3	2008

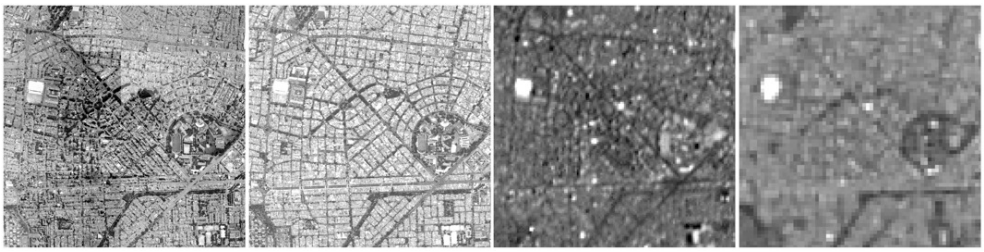

Figure 1.1: Different images for an area of 2 km x 2 km of Medellin, Colombia. From left to right: digital orthophoto (pixel size: 0.125 m); Quickbird (0.6 m), Landsat ETM+ panchromatic band (15 m), and Landsat TM (30 m).

at a lower cost than that of in situ measurements. The use of remote sensing data is usually more suitable for measuring and monitoring urban environmental conditions than for urban planning purposes because in the latter case, governmental and private sector data are more easily obtained. However, new developments and applications have taken advantage of the consistency of the remote sensing data to study the spatio-temporal dynamics of urbanization, suburbanization, land-cover or land-use changes, and urban morphology, which are topics of interest for urban planning (Griffiths et al., 2010; Lu and Weng, 2004; Rashed et al., 2001; Taubenböck et al., 2009a; Yin et al., 2005). Since Miller and Small (2003), advances in remote sensing applications for regional science in urban settlements have occurred in the areas of crime and nighttime lighting relationships (Weeks, 2003); urban land cover and socioeconomic change relationships (Mennis and Liu, 2005); population density estimation (Liu et al., 2006); house value modeling (Yu and Wu, 2006); socioeconomic status mapping (Avelar et al., 2009; Stow et al., 2007); population estimation in informal settlements (Galeon, 2008); social vulnerability to landslides (Ebert and Kerle, 2008); land surface temperature relationship with socioeconomic parameters (Rajasekar and Weng, 2009); urban morphology and socioeconomic parameters (Taubenböck et al., 2009b); and spatio-temporal analysis of urban sprawl (Taubenböck et al., 2009a). Multidisciplinary work is often used in these applications, with experts in remote sensing working together with architects, urban planners, sociologists and geographers to extract useful information from satellite imagery.

1.2 Research justification

Although remote sensing imagery cannot replace socioeconomic surveys and censuses, it can improve our knowledge about the social setting of an urban settlement in several ways. Aerial and satellite imagery can be used to improve explanations about

the interactions between people and place, to quantitatively assess the cityscape characteristics and their relationships with the social and economic conditions of urban dwellers, and could be used as an alternative information source for proxies of socioeconomic variables when no other information is available. The development of quantitative methods of image information extraction for regional science is of particular interest in developing countries that lack robust statistical offices, and to world wide organizations like United Nations that struggle to obtain data in the less developed regions of the world. The availability of socioeconomic data in developing countries is usually sparse, and the census frequency can be too low due to budget constraints. In those instances, the use of remote sensing could lower the costs of socioeconomic surveys, could be used to perform intercensal or intersurvey estimates of some socioeconomic variables, or could be used to quantify some physical features of the urban fabric at a lower cost than field surveys and measurements.

But the remote sensing approach has limitations. They are related to the availability of means and to the specific techniques being developed. Although almost the whole surface of the planet is being recorded regularly from different satellite remote sensing platforms, the cost of the satellite imagery with enough spatial resolution to study intra-urban differences can be another constraint because of budget limitations. Recent technological advances in platforms and sensors, like remote controlled drones and unmanned aircraft systems, are lowering the cost of very high-resolution aerial imagery. Some Internet services like Google Earth and Microsoft Bing can deliver very high-resolution satellite and aerial imagery at no cost for researchers, but this imagery lacks near infrared information, which is very useful to identify vegetation and water bodies. However, it is important to explore the usefulness of these less sophisticated imagery sources because they can be the unique available source of remote sensing data in some regions of the world.

Some of the techniques being developed exploit the particular conditions of the places where they are used, such as the quantification of green spaces, the presence of swimming pools, and the usage of specific roofing materials that are related to the wealth of dweller owners. Those characteristics are closely related to the environmental setting of the urban settlement: the presence of vegetation is related to the availability of water, open and visible swimming pools are very common in hot places but very rare in cold ones, and roofing materials are related to the availability of building materials near or within the cities and to the cultural heritage of the society. As these features are dependent of the city's particular conditions, they can inform about one situation in one city and about the opposite in another. The conclusions drawn with the use of these features will be valid only for the specific places where they were tested. But the application of these techniques in different places around

the world can help to identify patterns that could allow the development of more sophisticated information extraction methods that account for cultural and geographical differences.

Other characteristics of urban settlements that can be quantified are more general. The disorder in the urban layout, the organic pattern in urban settings where self-build processes are the norm, and the arrangement of different buildings in close proximity with different forms and materials, are all examples of features that can be related to poor urban settings almost world wide. Thus the use of textural and structure analysis techniques that quantify the diversity of landscape elements and the spatial pattern of the urban layout offers more potential in terms of usability and applicability; and recently developed automated image feature extraction tools at object level are of particular interest for this task.

Contributing to the research community with empirical evidence about the usefulness of non-sophisticated, very high-resolution imagery for regional science applications in urban settings is very important to show how, and to what point, this information source can be used as an alternative or complementary data source when other socioeconomic information is scarce or too expensive to obtain.

1.3 Aim and objectives

This research aims to bring closer the remote sensing world to the regional scientists by providing new evidence about the usefulness of very high-resolution images as an alternative and complementary data source for the study of socioeconomic phenomena. In particular, this work intends to show the usefulness of very high-resolution imagery similar to the imagery already available in some Internet services -without near infrared information nor three-dimensional data- to extract descriptors of the urban fabric that can enhance our knowledge about social aspects in urban settings. In this way, it could benefit regional scientists that work in cities and towns with very limited resources and sparse data by showing how to extract and use quantitative information derived from the visible spectrum bands of very high-resolution imagery using available off-the-shelf software and tools. To do that, this work illustrates the use of the image-derived features in two different quantitative applications of regional science in an urban setting.

1.3.1 Objectives

To face the general aim of this research, the following specific objectives have been proposed:

- To revise the usefulness of recently developed methods that use remote sensing data to obtain or complement socioeconomic information from urban settings. This review is intended to identify the regional science applications in urban settings which can benefit from the use of remote sensing data as a complementary or as an alternative source of information.

- To integrate spatial data sets of socioeconomic variables, obtained from surveys and censuses, with urban fabric descriptors extracted from a very high-resolution image to enhance the spatial econometric modeling of socioeconomic phenomena at intra-urban level. This implies the use of state-of-the-art tools for image feature extraction that work at object level to ease the integration of this information with the socioeconomic data, which is usually available at predefined spatial units.

- To estimate an intra-urban poverty index using only remote sensing derived metrics to quantify their explanatory power and to assess their usefulness when no other information is available. This application is intended to show the usefulness of the remote sensing metrics as an alternative data source for researchers in settings with very limited access to quantitative socioeconomic data and with very low financial resources.

- To quantitatively verify the influence of the urban layout in crime using remote sensing variables while controlling for potential socioeconomic confounders. This application is intended to show the usefulness of the remote sensing variables as a complementary data source to quantify some characteristics of the built environment that could influence human behavior, and that are very expensive and time consuming to quantify using traditional methods such as surveys and field appraisals.

1.4 Document structure

This work explores the potential of satellite remote sensing as a tool for regional scientists who are interested in intra-urban variations of socioeconomic phenomena. It addresses the specific topic of satellite remote sensing tools applied to regional science research in urban settings. The main components of this research are a review of the state of the art and two different quantitative applications, and there is an additional chapter to show the reader the general workflow for data integration and quantitative analysis.

The literature review, presented in chapter 2, shows recent developments in this

field. Through this review the reader will be introduced to studies that have used remote sensing as input data to analyze a variety of urban issues such as poverty, population dynamics, quality of life, socioeconomic status, and crime among others. It shows what remote sensing can do for regional science and how, and finishes summarizing the main types of data being used, and the methods and approaches to problem solving in this field.

A general work flow for data integration and quantitative analysis is presented in chapter 3. It describes the main features of the study site and the used data sources in the following specific applications. It also addresses the issue of the selection of a proper spatial unit of analysis and the extraction of remote sensing and socioeconomic variables at this spatial scale to allow the econometric modeling.

Two specific applications, presented in chapters 4 and 5, contribute to the growing literature in this field and show examples about the use of a non-sophisticated very high-resolution imagery for the quantitative analysis of two different socioeconomic phenomena at intra-urban scale. Chapter 4 shows the potential of remote sensing as an alternative data source for intra-urban poverty mapping. It analyzes the potential of image-derived variables to estimate an intra-urban poverty index.

Chapter 5 shows an example of the use of image-derived metrics as a complementary data source for the analysis of urban crime. It starts with the analysis of intra-urban homicide rates in a city and its potential socioeconomic drivers, and then it focuses in the usefulness of the remote sensing-derived variables to help explain the relationship between this type of crime and the urban layout structure and appearance when seen from the space.

Chapter 6 presents the general conclusions of this work.

Chapter 2

Satellite remote sensing and regional science in urban settings

This chapter is an edited compilation of some sections of the paper 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. doi: 10.1016/j.compenvurbsys.2012.06.003

Although the variables of interest for regional scientists are not directly measured from the air, remote sensing can measure the context of social phenomena and their effects on the land surface (Rindfuss and Stern, 1998). The unique characteristics of remotely sensed data, such as wide area coverage and repeat cycle, provide a means for exploring and testing hypotheses and models about urban areas and for constructing new theories that can help in the analysis and response by policy makers of problems that involve urban processes (Rashed et al., 2005). The description of patterns of the urban landscape is a fundamental question in urban analysis (Rashed et al., 2001). This topic has been addressed through land-use/land-cover mapping, in which the landscape fabric is viewed as an arrangement of discrete pieces of homogeneous landscape, each one with a different type of land cover and function. In many urban applications, land-cover/land-use maps generated by the thematic classification of satellite images can be considered the starting point for further analyses (Donnay et al., 2001).

The relationships between urban land cover and other environmental factors and the variables that describe socioeconomic conditions have been explored since the late

1950s with the use of aerial photography (Green, 1956, 1957). Research in the last three decades has explored the use of satellite remote sensing to characterize these relationships at a lower cost. In the urban environment, these relationships are based on the concept 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).

Although Jensen and Cowen (1999) stated that the most important requirement for properly identifying urban socioeconomic attributes using satellite remote sensing is the spatial resolution, several works have used satellite remote sensing at a medium resolution, such as Landsat MSS, TM and ETM+ imagery, to explore the relationships between land cover and socioeconomic data (Emmanuel, 1997; Forster, 1983; Jenerette et al., 2007; Mennis, 2006). Because of the repeat cycles and length of archive of these medium spatial resolution systems, satellite imagery is available dating back to the 1970s, and the imagery is well suited to studying and monitoring urban changes and growth trends (Pham et al., 2011; Rashed et al., 2005; Sabet et al., 2011; Van de Voorde et al., 2011; Weng, 2012). Forster (1983) studied Landsat MSS images with pixel size of 80m as a proxy for some socioeconomic variables in Sydney, Australia. This study attempted to determine the reflectance of urban residential surfaces using Landsat MSS (multi-spectral scanner) data in an effort to develop equations to predict the percentages of different surfaces contributing to the total reflectance of a pixel and to determine the relationships between land-cover percentages and average residential home value as inputs to measure urban residential quality. The results were encouraging, and it was stated at that time that the potential of satellite remote sensing as a tool for urban analysis would be enhanced when newer satellite systems with higher spatial resolution became available.

Landsat and ASTER images have also been tested in research on relationships between urban land-cover and socioeconomic data and their changes over a time span of decades in several cities in the US (Emmanuel, 1997; Jenerette et al., 2007; Mennis, 2006; Mennis and Liu, 2005; Rajasekar and Weng, 2009). Emmanuel (1997) found a significant and positive relationship between increases in urban vegetation measured from Landsat images from 1975 and 1992 and demographic factors associated with urban decay in Detroit, Michigan. It was concluded that vegetation trends in Detroit could be used as indicators of urban socioeconomic changes. Mennis (2006) explored the relationship between urban vegetation and socioeconomic conditions in Denver, Colorado. Land-cover data derived from aerial photography and vegetation data derived from Landsat ETM+ imagery were integrated using a geographic information system (GIS) with census data at the census-tract level. Multivariate statistics and

choropleth mapping were used to explore relationships among variables, and association rule mining was used to explore other non-linear relationships.[1] The results from multivariate regression analyses showed that the amount of vegetation was positive correlated with a set of socioeconomic variables that included median income, educational attainment, number of rooms and housing value and negative correlations with variables describing population density, commercial density and median year in which the housing unit was built. Results from association rule mining generally confirmed and improved the understanding of the multivariate regression results. These findings demonstrated the utility of integrating remote sensing and socioeconomic data for the study of interactions of urban ecological and social systems. Mennis and Liu (2005) used similar techniques with GIS data to analyze urban land-cover and socioeconomic changes in the same city from 1970 to 1990.

Jenerette et al. (2007) investigated regional relationships between surface temperature, vegetation conditions and human settlement patterns in the region of Phoenix, Arizona. Census data and a Landsat ETM+ image were used as inputs in this research. Surface temperature and vegetation data were derived from Landsat images, while social variables were derived from the 2000 US Census and included median household income, percent Hispanic population, median age of housing and population density. GIS was used to integrate these variables at the census tract level, and a path analysis multivariate model was used to measure the interaction between them. They found a strong correlation between vegetation and surface temperature, as expected. The correlations of temperature with median household income and with the percentage of Hispanic population were the strongest among all social variables. They also found that residential segregation of neighborhoods by socioeconomic status influenced the vegetation and surface temperature patterns in the city.

Rajasekar and Weng (2009) used data mining techniques to explore the relationships between urban surface temperature and various biophysical and social parameters by integrating an ASTER satellite image and census data for Marion County, Indiana. The multispectral nature of ASTER images, with several spectral bands in the near infrared and thermal infrared parts of the spectrum, allowed the derivation of land surface temperature data, land-use/ land-cover data and vegetation data (estimated with a scaled normalized difference vegetation index – NDVI). Population density was obtained from census data, and GIS land-use zoning and transportation buffer zone maps of Marion County were also included in the analysis. The model was developed using association rule mining and yielded interesting rules that may help researchers, planners and environmental managers to understand the relationships between biophysical and social conditions in that county.

[1] For more information on this technique, see Zhang and Zhang (2002).

In recent years, very high spatial resolution satellite images have also been tested. Avelar et al. (2009) explored the relationship between land cover in Lima, Peru, and the distribution of socioeconomic classes using a very high spatial resolution Quickbird image, census data and field data on socioeconomic classes. Image data were classified using conventional supervised classification techniques to identify land-cover features such as green areas, water bodies, paved streets, bare soils, high-quality buildings and low- quality buildings. They used a regular square grid with a cell size of 1 km to store the percentages of each land-cover feature within each cell. Then, the combination of urban features was analyzed to obtain a socioeconomic class for each cell based on membership rules for five socioeconomic classes. The proposed method allowed the rapid assessment of socioeconomic classes in large cities like Lima; however, the accuracy assessment showed that the method must be improved further to become fully operational and that a cell size of 1 km is too coarse for urban analysis. Improving on this work, Tapiador et al. (2011) developed a different approach to deriving the socioeconomic class from very high-resolution satellite imagery in the same city. This approach is based on the assumption that there is a relationship between the socioeconomic classes and the urban morphology, described in terms of the availability of green areas, sport facilities, private swimming pools and pavement in various conditions. A standard image classification process was applied to plot the components of the urban morphology, and a neural network classification process was used to assign social classes to each pixel in the image. The results were not as good as expected, but the researchers stated that the method could be used to identify deficiencies in urban services, to monitor social policies and as complementary information for decision making in urban planning. They also tested the applicability of this approach in Rio de Janeiro, Brazil, and Cairo, Egypt, and found that the method could be applied with few modifications.

Night-time satellite imagery on a regional scale from the Defense Meteorological Satellite Program (DMSP) has been used to estimate economic activity and poverty and to map urbanization dynamics at a global scale (Doll, 2008; Doll et al., 2000; Elvidge et al., 2001, 2009; Zhang and Seto, 2011). Recently, high-resolution night-time imagery has been tested in Israel as an indicator of the demographic and socioeconomic properties of urban areas at a local scale (Levin and Duke, 2012). Table 2.1 summarizes reported relationships among proxy variables of vegetation derived from satellite remote sensing and socioeconomic variables measured in census and field surveys for several cities in the US. Table 2.2 presents similar information for surface temperature rather than vegetation.

The results of the aforementioned studies indicate that relationships among land-cover and socioeconomic variables can be established in different cities around the

Table 2.1: Reported relationships among vegetation and socioeconomic variables, (+) positive correlation, (-) negative correlation.

Reported in	Mennis (2006)	Lo (1997) and Lo and Faber (1997)	Jenerette et al. (2007)	Li and Weng (2007)	Jensen et al. (2004)
City	Denver, CO	Athens-Clarke, GA	Phoenix, AZ	Marion County, IN	Terre Haute, IN
Image	Landsat ETM+	Landsat TM	Landsat ETM+	Landsat ETM+	ASTER
Vegetation proxy	NDVI	NDVI	SAVI*	NDVI	Leaf Area Index
Methods	Association Rule Mining	Principal Component Analysis	Path Analysis	Factor Analysis Regression Analysis	Regression Analysis
Socioeconomic variable					
Residential density	+				
Commercial density	-				
Population density	-	-	-	-	-
Income	+	+	+	+	+
Percent minority / % Hispanic	-		-		
Educational attainment	+	+		+	
Number of rooms	+			+	
Home year built	-		-		
Home value	+	+		+	+
Percent of urban use		-			
Poverty				-	
Unemployment rate				-	

* SAVI: Soil Adjusted Vegetation Index.

world using satellite remote sensing, but local research is always needed, as the particular conditions of each city must be addressed and contrasted with socioeconomic data to find useful correlations ((Besussi et al., 2010). The most obvious advantage of establishing these relationships in a city is that they allow social variables to be estimated quicker and at a lower cost than *in situ* measurements and for dates other than those of field surveys and census data, which in most countries are collected every 10 or 12 years. According to Phinn et al. (2002), five recurrent research themes in applications of remotely sensed data to urban environments are (i) land-use/ land-cover mapping; (ii) assessment of the usefulness of texture measures to aid in separating urban land-cover and land-use types; (iii) impervious surface mapping for input to energy and moisture flux models; (iv) land-use/land-cover change mapping; and (v) application of empirical models to estimate biophysical, demographic and socioeconomic variables. The last theme is of special interest for regional scientists, urban planners and policy makers. A sample of the applications of satellite remote sensing to the specific field of regional science in urban settings is described next.

2.1 Slum detection (deprivation hot-spots)

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

Table 2.2: Reported relationships among surface temperature and socioeconomic variables, (+) positive correlation, (-) negative correlation.

Reported in	Lo (1997) and Lo and Faber (1997)	Jenerette et al. (2007)	Li and Weng (2007)
City	Athens-Clarke, GA	Phoenix, AZ	Marion County, IN
Image	Landsat TM	Landsat ETM+	Landsat ETM+
Methods	Principal Component Analysis	Path Analysis	Factor Analysis Regression Analysis
Socioeconomic variable			
Population density	+	+	+
Income	-	-	-
Percent minority / % Hispanic		+	
Educational attainment	-		-
Number of rooms			-
Home year built		-	
Home value	-		-
Percent of urban use	+		
Poverty			+
Unemployment rate			+

to adequate sanitation in the form of a private or public toilet shared by a reasonable number of people, and security of tenure.[2] The presence of slums in a city is an indicator of deprivation and poverty, and it is estimated that approximately one-third of the urban population in the developing world lives in slums (United Nations, 2011); hence, researchers are interested in the proper identification and characterization of slum neighborhoods.

The usefulness of satellite remote sensing for distinguishing a slum from its surrounding neighborhoods has been addressed in the last decade (Barros, 2008; Hofmann et al., 2008; Kohli et al., 2012; Rhinane, 2011; Stow et al., 2007; Weeks et al., 2007). The degree of slumness of neighborhoods and places within a city has been estimated using remotely sensed data from very high spatial resolution platforms such as Quickbird, Ikonos and SPOT 5. Some approaches have used complementary socioeconomic data on housing size and value, whereas others have used only texture and fractal dimensions derived from satellite imagery to develop statistical regressions. A very high resolution Quickbird satellite image with a pixel size of 0.6 m has been tested with three different strategies in the city of Accra, Ghana (Weeks et al., 2007). The first strategy was based on the Vegetation-Impervious surface- Soil (VIS) model of the urban scene devised by Ridd (1995)[3] and the use of image texture measures to infer

[2]http://www.unhabitat.org/documents/media_centre/sowcr2006/SOWCR%205.pdf accessed 31 July 2011.

[3]Ridd (1995) proposed a model for urban ecosystem analysis that is very well suited to medium

land use from land cover data derived from satellite imagery (Weeks et al., 2007). A slum index map of the city was generated from census data to assess the usefulness of the remotely sensed data to perform the task. In this approach, it was found that the variability of a neighborhood's slum index can be predicted from remotely sensed data and that the amount of vegetation is the most important predictor, along with the image texture measures. By using these measures derived from remotely sensed images, it was possible to identify the parts of the city of Accra where 6 of the 10 worst slums are located. Stow et al. (2007) continued the work of Weeks et al. (2007) by using a more sophisticated classification scheme with the same satellite image in the same city. The image data were classified into non-residential, low socio-economic residential and high socio-economic residential areas using two different object-oriented image classification strategies, and their results were compared against the slum index map generated from census data. One strategy was based on spatial frequency characteristics of multispectral data for pixels within neighborhood-size segments, and the other was based on the proportions of vegetation, impervious surfaces and soil sub-objects (as in the VIS model) within similar segments. Both strategies showed similar accuracy, although the spatial patterns of land-use types were different, and a closer agreement with the map generated from census data was observed for the VIS land-cover sub-objects strategy.

Very high spatial resolution Quickbird images have also been tested for slum detection with object-oriented classification schemes in Sao Paulo and Rio de Janeiro, Brazil, with encouraging results (Hofmann et al., 2008; Kux et al., 2010; Novack and Kux, 2010). Barros (2008) tested both Quickbird and Ikonos images of Recife and Campinas, Brazil, with an approach based on lacunarity texture analysis.[4] The slum locations were identified first by calculating an inhabitability index from census data, and image samples for those sites were then tested. The method proved to be useful for distinguishing images from slum and non-slum areas, and a strong correlation between lacunarity and inhabitability was found for both tested image types. It was also reported that this approach does not require intense field surveys, that it can be replicated and that it allows comparative analysis of multiple locations. Taubenböck et al. (2009b) used an Ikonos image with an automatic object-oriented methodology and a semantic classification to classify the urban area of Padang, Indonesia, into

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 (V-I-S), 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).

suburbs, slums, low-class, middle-class and high-class areas. SPOT 5 images with 2.5-m spatial resolution have also proven useful to identify and quantify slums in Casablanca, Morocco, again using an object-oriented classification approach in an attempt to facilitate the monitoring and mapping tasks of Cities without Slums, a Moroccan government program launched in 2005 (Rhinane, 2011). Lacunarity texture measures have also been used successfully to identify urban slums in Hyderabad, India, using a Quickbird image Kit et al. (2012).

Very high spatial resolution satellite images have been tested successfully in several cities around the world, and it can be inferred that the situation will be similar for other cities in similar geographical contexts. The UN-Habitat objective definition of a slum is intended to be valid worldwide, and there is evidence of the usefulness of satellite remote sensing in cities with very different environmental conditions. However, there is still a research gap in slum detection from remote sensing that must be filled to develop a method applicable worldwide that takes advantage of the similar appearances of slums all over the world, such as crowding, small dwelling size and precarious and irregular street network patterns. Image texture measures, mathematical morphology and fractal dimensions such as lacunarity that use generalized urban morphology instead of local features are believed to be useful for this purpose. The work of Kohli et al. (2012) attempts to fill this gap by developing an ontological framework to conceptualize slums and to contribute to their detection and classification using very high resolution imagery, though this approach still requires adaptation to local conditions.

2.2 Quality of Life index assessment

An important topic for urban policy makers is the objective measurement of the quality of life in urban areas through an index that takes into account not only socioeconomic information about people but also information on the geographical context and environmental conditions of urban areas. An evaluation of the quality of life of the urban population is important to support decision making and sustainable urban management and planning and to measure policy outcomes (Craglia et al., 2004; Stathopoulou and Cartalis, 2006).[5] This approach is intended to assess several dimensions of quality of life: the social and the economic dimensions are measured from census data, and the environmental quality of the places inhabited is measured from remotely sensed data (Forster, 1983; Jensen et al., 2004; Lo, 1997; Lo and Faber, 1997; Weber and Hirsch, 1992). Various indices of quality of life, residential quality,

[5]For more information on quality of life concepts and definitions, see Van Kamp et al. (2003); Pacione (2003).

attractiveness and a housing index have been estimated using remotely sensed data from Landsat MSS, TM, ETM+, and SPOT sensors and complementary data from censuses and surveys on housing size and value, population density, income and education. A description of different strategies to calculate a quality of life index with remote sensing data follows.

Forster (1983) shows how to derive a general residential quality index from satellite remote sensing data for Sydney, Australia. A correlation between average housing value and Landsat-derived reflectance data was demonstrated first. As in traditional hedonic models for house pricing, it was assumed that housing size is a surrogate of housing value and that a measure of housing quality and social environment may be positively related to a residential quality index. Vegetation content was assumed to be positively related as well, and other surfaces, such as roads and non-residential buildings, were assumed to have a negative impact on residential quality due to high noise and pollutant levels expected for those urban functions. Equations for residential quality index estimations from Landsat MSS reflectance data were developed using multiple regression analysis, which yielded a multiple correlation coefficient of 0.87. Although a residential quality map was not published, the analysis indicated that the major influence on residential quality came from vegetation content estimation from the infrared/visible bands ratio of the satellite image. Following similar principles, Weber and Hirsch (1992) calculated a quality of life index for Strasbourg, France, using census data and the quality of the urban landscape as characterized and mapped from SPOT satellite images. Three different indices were developed and mapped from mixed data: a housing index, an attractiveness index and a quality index. The results produced realistic spatial distributions but were only weakly related to the census data that were usually used to define the urban landscape quality alone; these results thus demonstrate the importance of including environmental factors in the quantification of quality of life indices.

Lo (1997) and Lo and Faber (1997) used Landsat TM images combined with census data to assess the quality of life in Athens-Clarke County, Georgia, using image processing techniques and information overlay in GIS. The environmental variables extracted from the image data were land-use/land-cover as a percentage of urban use, vegetation as a normalized difference vegetation index (NDVI), and apparent surface temperature. The socioeconomic variables extracted from the census data included population density, per capita income, median home value and percentage of college graduates. Two approaches were used to integrate the layers of data: principal components analysis (PCA) and GIS overlay. PCA showed a strong relation between biophysical data and socioeconomic data, with NDVI showing strong negative correlations with land surface temperature and percentage of urban cover and positive

correlations with per capita income, median home value and education. The first component was assumed to be an index of quality of life because it explained 54% of the variance and included both socioeconomic and environmental variables. It was aggregated to the census block group level to produce a map of quality of life showing intra-urban variations. In the GIS overlay approach, all variables were ranked in terms of desirability for quality of life and combined at the census block group level to obtain a quality of life score map of the county. Both approaches produced very similar results, and both methods have the advantage that they can be easily performed with commercial remote sensing and GIS software.

Stathopoulou and Cartalis (2006) used a similar approach to assess quality of life in Athens, Greece, as did Li and Weng (2006) and Li and Weng (2007) for Indianapolis, Indiana. Stathopoulou and Cartalis (2006) identified not only the areas of the city with lower quality of life but also the variables that caused quality of life to be lowered in each area.Li and Weng (2006) developed a methodology to model quality of life in Marion County, Indiana, and validated it by applying it to Monroe and Vigo Counties in Indiana. It was concluded that the model may be useful for predicting quality of life and analyzing spatial variations within a specific region. Li and Weng (2007) also reported strong positive correlations of green vegetation with income, house value and education and negative correlations with temperature, impervious surface and population density for Indianapolis. Nichol and Wong (2006) used the same approach as Lo (1997) and Lo and Faber (1997) to assess urban environmental quality in the Kowloon Peninsula, Hong Kong. Field measurements were integrated with environmental parameters derived from Landsat ETM+ and Ikonos images. Although no accuracy assessment was made, they reported strong correlations between the results obtained from PCA and GIS overlay methods. Interest in quality of life studies has recently increased. Hundreds of communities all over the world have conducted quality of life indicator research, and there are still important differences of opinion on the indicators of and contributors to quality of life and on how to use the information that indicators provide (Li and Weng, 2006, 2007). Moreover, each city has its own mix of environment and population attributes, which means that something that is found to be true for one city might be false for another. As with the relationships between land-cover and socioeconomic variables, local research and knowledge must be considered for the results to be useful in policy making, urban planning and management.

2.3 Housing value estimation

Housing value estimation with satellite remote sensing has been tested in several cities around the world. House prices have been estimated using data from Landsat MSS, TM, ETM+ and ASTER sensors as well as complementary data from censuses and surveys of house structural attributes, population density, income, and distance to a central business district. Forster (1983) reported that the most significant variable for predicting average house value in the city of Sydney, Australia, was the number of rooms. House size was also believed to be a good predictor, and regression models for house value estimation from house size derived from a Landsat MSS image were developed. The results of these models were equivalent to the results obtained using known housing characteristics from census data. In a second step, Forster (1983) tested reflectance values and distance to a central business district directly in a regression equation with average house value. The results indicated that the average house value in Sydney can be predicted from Landsat-derived reflectance and that these results could be applicable in other cities with similar climate, residential density, population, size and urban morphology.

Several studies in US cities have found positive relationships between vegetation measures from remote sensing data and housing value, among other variables, using a variety of methods and approaches(Jensen et al., 2004; Li and Weng, 2007; Lo, 1997; Lo and Faber, 1997; Mennis, 2006). Jensen et al. (2004) used remote sensing and GIS to examine the relationships between a biophysical variable and standard socioeconomic variables. They quantified the urban forest by measuring the Leaf Area Index (LAI, m^2 of leaves per m^2 of ground) of the vegetation in Terre Haute, Indiana, in 143 random sampling locations, which were later extended to the entire area of study using data from an ASTER satellite image and an artificial neural network. The result was integrated with census data using ordinary least squares regression techniques to establish relationships between urban LAI and population density, median income and median housing value. Urban LAI was positively related to the observed median home value and median income; the researchers stated that this relationship could be used to focus urban planning and management to ensure future investments are made where they are most needed. They went further in their conclusion to state that urban amenity variables could be used to explain and investigate the uneven distribution of urban resources.

Yu and Wu (2006) used a Landsat ETM+ image for house value modeling in Milwaukee, Wisconsin. They integrated environmental characteristics derived from the remotely sensed image with house structural attributes to model house values using two different techniques: global ordinary least squares regression and a regression

tree approach. The environmental characteristics derived from the image were the fractions of vegetation, impervious surface and soil, which were measured through the normalized spectral mixture analysis developed by Wu (2004). The results showed that environmental characteristics have a strong influence on house values and that the performance of models is improved significantly when environmental characteristics are taken into account. The regression tree approach performed better than ordinary least squares regression, but both techniques yielded correlation coefficients above 0.9 between the predicted and observed house values. All of the aforementioned studies have used coarse (Landsat MSS) and medium (Landsat TM and ETM+) spatial resolution imagery. The use of very high spatial resolution imagery in house value models deserves further research because it might provide better information for understanding urban housing markets (Yu and Wu, 2006). House size and house density are believed to be good predictors of house value, and these variables could be derived from very high spatial resolution imagery in addition to other environmental factors that could be integrated into house value models. Semantic and object-oriented image classification techniques are believed to be the most promising way to derive such information. This approach also offers potential for identifying housing submarket partitions on the basis of the differentiation of homogeneous urban morphology zones within a city. Taubenböck et al. (2009b) tested the mentioned approach with an Ikonos image for classifying the urban area of Padang, Indonesia, into suburbs, slums, low-class, middle-class and high-class areas. They investigated the correlation of these classes with mean income and mean value of property, derived from a field survey, and found that the correlation is as expected but that location also played an important role in the value of property parameter, with the highest values located in the central area of the city. As with many other remote sensing applications, further research is needed to determine whether this correlation is valid in other urban environments around the world.

2.4 Urban growth

Remote sensing can play an important role in policy assessment related to urban growth, population dynamics and environmental issues within cities (Bhatta et al., 2010a; Miller and Small, 2003). Quality of life, as described in section 2.2, is one of several topics that have been addressed to improve urban policy making by monitoring the effects and consequences of policies on the ground(Pacione, 2003). Knowledge of actual urban growth trends is another topic of great importance to city planners, environment managers, and policy makers. There are considerable differences between the growth that is planned by a city government and the actual growth trends that

take place on the ground in many cities in the developing world; therefore, monitoring and understanding those trends is important for understanding how to take proper action on them. Satellite remote sensing offers a reliable source of information about urban growth and sprawl processes because of its temporal and spatial characteristics (Doll, 2008; Rashed et al., 2005).

Very high-resolution imagery has been tested as a way to assess inner-city growth by detecting new buildings within neighborhoods using object-based approaches (Doxani et al., 2012; Tsai et al., 2011). Zhu et al. (2012) found that the classification accuracies of urban and peri-urban land improve when SAR data are integrated with optical satellite imagery. Leinenkugel et al. (2011) integrated SAR data with Quickbird and SPOT-5 imagery for impervious surface estimation and settlement detection and proposed a methodology for continuous monitoring of impervious cover in Can Tho Province, Vietnam. SAR and Landsat data have also been integrated to analyze the urbanization process in 27 mega-cities worldwide using object-oriented and pixel-based classification approaches and post-classification change detection to derive urban footprint products with accuracies over 80% (Taubenböck et al., 2012). Urban growth trends can be detected easily using medium spatial resolution satellite images, such as SPOT and Landsat TM, taking advantage of the very different spectral properties of urban manmade structures versus rural spaces. Land use patterns derived from remote sensing data provide additional means to describe urban growth types of peri-urban areas (Shi et al., 2012; Sun et al., 2013) and urban sprawl (Besussi et al., 2010; Soundranayagam et al., 2011). Night-time satellite imagery from the Defense Meteorological Satellite Programs Operational Linescan System (DMSP OLS) has also been used to assess urban sprawl in cities in the US (Sutton, 2003; Sutton et al., 2006) and Australia (Sutton et al., 2010). It is also remarkable that the whole Landsat archive is now freely available through the websites of the US Geological Survey and the Global Land Cover Facility at the University of Maryland, allowing researchers to perform urban growth studies for most cities world- wide from 1984 to the present using Landsat TM/ETM+ images.

Remotely sensed data and spatial modeling techniques have been applied to regional planning issues and specifically to estimations of future urban growth (Phinn et al., 2002; Ward et al., 2000). The need to estimate urban growth not only in terms of population growth, but also in terms of where most people are likely to be located, can be met using remote sensing data. Ward et al. (2000) and Phinn et al. (2002) developed a framework for urban growth modeling that was tested in Queensland, Australia. Information about the urban extent and urban land-cover was obtained from Landsat TM images from two different dates using a hierarchical classification approach; this information was used as inputs to a cellular automata

(CA) model that simulated local decision-making processes within an optimization framework that takes into account issues of sustainable urban development. One remarkable advantage of this approach is that it can be applied globally because it is supported on medium spatial resolution images that are now freely available.

Madhavan et al. (2001) also used Landsat TM images to measure and analyze the spatial growth of the Bangkok metropolitan area of Thailand. Their approach was based on an image classification scheme and the VIS model (Gluch and Ridd, 2010; Ridd, 1995). Urban land-use/land-cover maps for 1988 and 1994 were produced from the classification processes, and changes in each of the urban land classes were measured with traditional post-classification methods. The VIS model was used in a later stage to visualize the trends of those changes in terms of the changes in vegetation, impervious surface and soil fractions and their trajectories for selected sites between the two dates. An analysis of these outputs led the researchers to conclude that the observed change in land-use led to improper urban land development in some areas of the city. Weng and Lu (2006) applied the same concept to characterize urban landscape patterns and to quantify spatial and temporal changes in urban landscape compositions in Indianapolis, Indiana.

Major determinants of urban growth in Wuhan city, PR China, from 1993 to 2000 were modeled using exploratory data analysis and spatial logistic regression techniques with remotely sensed data from a 2000 SPOT image and GIS layers including planning schemes, a 1993 land-cover map and population and statistics from census data (Cheng and Masser, 2003). This study showed that the major determinants of urban growth were urban road infrastructure and developed area and that master planning did not play a role in that specific period. On a broader spatial scale, Schneider et al. (2005) evaluated the urban growth in the Chinese Western provinces in response to the Go West policy that started in the 1990s. They used Landsat imagery (MSS, TM and ETM+) to map the land cover changes in Chengdu, Sichuan province and to investigate the spatial distribution of the development zones from 1978 to 2002. Major spatial trends were identified, including spatial clustering, specialization of land uses and periurban development. They were related to a growing private sector and uncoordinated large-scale investment by public agencies. A major reexamination of planning and policy was proposed to improve resource use and management, provide urban services where needed, reduce environmental degradation and promote urban sustainability.

Time series of Landsat imagery have been used to map and analyze urban expansion (Tian et al., 2011; Wu and Zhang, 2012) and to analyze the spatio-temporal dynamics of green spaces within a city (Zhou and Wang, 2011). Michishita et al. (2012) examined two decades of urbanization in four cities located in the Poyang

Lake area, PR China, using a longitudinal dataset of Landsat imagery to derive a time series of urban land-cover fractions through Multiple Endmember Spectral Mixture Analysis (MESMA) and compared the fractions with socioeconomic statistics. They found that the four cities were urbanized through different mechanisms and obtained positive correlations between built-up fractions and urban population, gross GDP and GDPs in secondary and tertiary industries.

Bhatta et al. (2010b) highlighted the differences in definitions and approaches used to measure urban sprawl among the research community. They mentioned that the general consensus is that urban sprawl is characterized by an unplanned and uneven pattern of growth that leads to inefficient resource utilization.[6] The spatial and temporal characteristics of urban sprawl have been monitored and modeled with medium-resolution images from Landsat MSS, TM, ETM+ and IRS LISS-III for the city of Ajmer, India, from 1977 to 2002 (Jat et al., 2008). Urban sprawl was derived from classified satellite images and was analyzed using landscape metrics and multivariate statistical techniques to establish relationships with its causative factors. Their results showed that land development increased almost three times as much as the population did. The obtained relationships between the present trends of urban sprawl and their causative factors can be used by local authorities for land policy reexamination and improvement.

Sutton (2003) and Sutton et al. (2006) used night-time satellite imagery from DMSP OLS to characterize urban sprawl in the US using night-time lighting as a proxy of urban extent. Sutton (2003) measured urban sprawl adjusted to the total population of an urban area and derived aggregate measures of per capita land consumption that allowed comparisons between different metropolitan areas across the US. He found that western cities such as Los Angeles and San Francisco have lower levels of urban sprawl than do inland cities such as Atlanta, St. Louis and Minneapolis. Sutton et al. (2006) characterized the exurban areas in the US, defined as the expansive areas of low light surrounding all major metropolitan areas, and proposed a classification scheme of land cover into urban, exurban and rural areas based on DMSP OLS data. Well-lit areas were characterized as urban, low-lit areas as exurban and dark areas as rural. They stated that this scheme is better than an urban, suburban and rural classification scheme based on census data for characterizing the population and land use of industrialized countries because the distinction between urban and exurban is more pronounced than the distinction between urban and suburban. Sutton et al. (2010) used the same approach to quantify urban sprawl in several cities in Australia and to contrast the results with those of their previous

[6]For more information on remote sensing-derived measures of urban sprawl, see Bhatta et al. (2010b).

investigation in the US. They found that cities with an equivalent population in the US occupy larger areal extents than Australian cities, which suggests that sprawl is lower in Australian cities than in American cities.

There are many approaches and measures to evaluate urban growth and sprawl (Belal and Moghanm, 2011; Bhatta et al., 2010b,a; Jat et al., 2008; Lu et al., 2011; Sudhira et al., 2004; Tole, 2008), and satellite remote sensing has proven to be a useful tool to derive ground information to evaluate planning and policy outcomes on a large scale. The studies reviewed so far show the usefulness of medium spatial resolution satellite imagery for urban growth and land development assessment.

2.5 Population estimation

An understanding of the population size and spatial distribution in urban areas is essential for social, economic and environmental applications (Li and Weng, 2005; Liu et al., 2008). The use of remote sensing for population estimation started in the mid-1950s as a way to address the shortcomings of the decennial census, such as its high cost, low frequency and intense labor requirements (Liu et al., 2006). Regional-scale population estimates can be achieved through the quantification of urbanized built-up area from remote sensing data using statistical relationships with the total settlement population. In urban areas, population estimation by remote sensing often involve counting dwelling units, measuring land areas, and classifying land use. The last requirement results from the close correlation of land use with population density in urban areas. Ground data on the average number of persons per dwelling unit and for different land uses are always needed to calibrate models (Jensen and Cowen, 1999). The general strategy to achieve population estimations from remote sensing data is to establish statistical relationships between census or field survey data and remotely sensed data using regression analysis techniques (Harvey, 2002; Liu and Herold, 2006; Lo, 1995; Pozzi and Small, 2005; Yuan et al., 1997). Data from SPOT, Landsat TM and ETM+ sensors have been used along with census or survey data on population counts to estimate population size. Recently, building heights derived from LiDAR (Light Detection and Ranging) data and very high resolution imagery have also been used to estimate population counts at an intra-urban scale.

Lo (1995) developed methods to extract population and dwelling unit data from a SPOT image of the Kowloon metropolitan area in Hong Kong. Different regression models were applied to link the spectral radiance values of image pixels with population densities. Accurate estimates of population and dwelling units were obtained at the macro level for the whole study area, but intra-urban variations and micro-level estimates tended to be of low accuracy because of the difficulty of discriminating resi-

dential from non-residential use in multifunctional buildings in the satellite image. In a regional-scale study, Yuan et al. (1997) found high correlations between land cover classified from a Landsat TM image and population counts from census data. By applying regression and scaling techniques, they were able to obtain a population distribution map with much more detail than census data alone offered for four counties in central Arkansas.

Harvey (2002) used a Landsat TM image and census data for Ballarat and Geelong, Australia, to allocate population estimates to each pixel of the image and overcome the problem of spatial aggregation of census data. The satellite image was first classified into residential and non-residential classes, and initial reference populations were distributed uniformly for each census zone across its residential pixels only. Then, pixel populations were related to pixel spectral values and re-estimated pixel populations using an expectation–maximization regression algorithm, and the regression equation was tested in a second image. The relative error of global estimations was less than 1% in both images, but it rose to approximately 16% and 21% in the individual census zones. Wu and Murray (2005) followed a similar approach with a Landsat ETM+ image, but instead of using pixel spectral values, they used the fraction of impervious surfaces in residential areas. Li and Weng (2005) integrated a Landsat ETM+ image and census data for estimating intra-urban variations in population density in Indianapolis, Indiana, using several remote sensing-derived variables as predictive indicators, correlation analysis to explore their relationships with population data, and step-wise regression analysis to develop models. Liu et al. (2006) stated that these results were valuable for improving population estimations, but the spatial resolution of 30m limits their utility in urban applications. Liu and Herold (2006) highlighted the use of high spatial resolution satellite imagery to study intra-urban population characteristics.

Although very high spatial resolution imagery has been available for at least 10 years, since the launch of the Ikonos and Quickbird satellite programs, little research has explored the use of this type of imagery for population estimation purposes (Liu et al., 2008). Liu et al. (2006) and Galeon (2008) are two examples of this type of research. Liu et al. (2006) explored the correlation between census population density and image texture with a very high spatial resolution Ikonos image of Santa Barbara, California, using linear regression. The spatial unit used in this analysis was the census block with homogeneous land use. Different methods for describing image texture were tested, and the correlation was found to vary depending on the method used. The obtained correlation between image texture and population was not strong enough to predict or forecast residential population size, but the researchers concluded that image texture could be used to refine census-reported population distributions

using remote sensing and to support smart interpolation programs to estimate human population distributions in areas where detailed information is not available. Galeon (2008) used a Quickbird satellite image to estimate the population size in informal settlements using a field survey and regression analysis for the University of Philippines campus area. The researchers obtained accurate estimates of population size with first-order equations for the slum areas but not for the semi-formal housing areas present on the campus.

Although LiDAR techniques have been available since the 1960s, they have only become commonly used in the past few years (Lwin and Murayama, 2009). LiDAR data can now provide very accurate height information for land surface features such as buildings and trees. Digital volume models (DVMs) derived from LiDAR data are increasingly being used for population estimation at the urban scale and are being integrated with very high resolution imagery with good results (Lwin and Murayama, 2009, 2011; Qiu et al., 2010; Ramesh, 2009; Weng, 2012). SPOT and Landsat TM imagery has been popular among population estimation applications of satellite remote sensing because of its relative success in regional- and medium-scale studies. At the urban or intra-urban scales, further research is needed to establish the best methods and procedures for population estimation, taking advantage of the very high spatial resolution satellite imagery and LiDAR data that are now widely available. Ground truth, in the form of field surveys and census data, is always needed, but the goal is to achieve the highest possible accuracy while minimizing fieldwork to keep the research both cost effective and operationally practical.

2.6 Social vulnerability assessment

The concepts of hazard, disaster, risk and vulnerability are often used interchangeably and have different implications in different scientific disciplines (Rashed, 2005; Rashed et al., 2007). However, in the broad and multidisciplinary field of risk management, researchers agree on the definitions of the concepts of hazard, vulnerability and risk. Hazard is related to the occurrence of a physical phenomenon or event that constitutes a threat to society, whereas vulnerability is related to the ability of the society to cope with the impact of the hazard, and risk is understood as the combination or product of the probability of a hazard's occurrence and the degree of vulnerability of the society or community exposed to the hazard (Rashed, 2005; Taubenböck et al., 2008). The coping ability of the society involves several factors and relationships of a physical, economic, social and political nature. Social vulnerability refers to the social aspects of a community that are related to the conditions in which it can face a hazard Botero (2009).

Although it has been stated that remote sensing can provide a fast and cost-effective way to derive information about many of the factors that shape social vulnerability to natural hazards (Taubenböck et al., 2008), there are few published works on the assessment of social vulnerability to natural hazards in urban settings with remote sensing data. Social vulnerability to different types of natural hazards has been estimated using remotely sensed data from the Landsat TM, Quickbird, and Ikonos platforms and complementary data from censuses and surveys. (Taubenböck et al., 2008) listed the possible uses of remote sensing for measuring social status indicators of vulnerability to earthquakes and tested them using very high and medium spatial resolution satellite images of Istanbul, Turkey, from Ikonos and Landsat. An automatic object-oriented and fuzzy-based approach was used to classify the Ikonos image into land-cover classes, and the researchers used this information to derive indicators related to the physical and demographic components of vulnerability. A time series of Landsat data was used to map and assess building ages and urbanization rates. Built-up density, accessibility, population density, building age, and urbanization rate indicators were derived from satellite imagery. Their results showed the capability of remote sensing to assess several aspects of social vulnerability and their spatial distributions, but the researchers stated that the main limitation of this approach lies in the lack of knowledge of how to derive reliable values for the socioeconomic and political components of vulnerability.

Ebert et al. (2007, 2009) tested whether it was possible to evaluate social vulnerability to floods and landslides in Tegucigalpa, Honduras, using remote sensing data. They used very high resolution images from Quickbird and ResourceSat P-6 satellites as well as elevation models and risk maps of floods and landslides to estimate proxy variables to describe the physical aspects of the urban environment. The proportions of vegetation and built area, road conditions, roof and building materials, the position of the building relative to the slope, the building height, the number of evacuation routes and the distance between the buildings and evacuation routes, among other variables, were derived from satellite images. Although remotely sensed data by themselves cannot replace traditional methods such as census and surveys, the synergy of the use of remote-sensed data with field surveys, the application of a census and the combination of these data in a GIS helped to improve the efficiency, frequency, and coverage of the evaluation of social vulnerability at different spatial scales. Object-oriented analysis techniques proved useful for deriving proxy variables that describe non-physical indicators of social vulnerability. The importance of using these techniques that combine traditional approaches with the analysis of remotely sensed data lies in the possibility of transference to other regions and the applicability of these methods in a sustainable way over time (Ebert and Kerle, 2008).

Rashed (2005), Rashed et al. (2007) and Taubenböck et al. (2008) have analyzed the potential of remote sensing to derive useful information on the social and physical characteristics of urban settings that increase or decrease the capacity of a local community to cope with a natural hazard. Rashed (2005) emphasizes that many of the findings of urban change studies that have used remote sensing can help researchers to better understand the implications of these changes for urban vulnerability because urban changes are related to land-use practices, resource management and hazard mitigation policies in cities. Advances in using landscape metrics in urban morphology characterization for the city of Los Angeles, California, have been made to infer socioeconomic indicators that could be related to social vulnerability to earthquakes (Rashed et al., 2003, 2007). Research in different cities and environments is still needed to develop a robust framework in which remote sensing-derived measures can play a leading role in the assessment of social vulnerability to natural hazards.

2.7 Crime

Remote sensing imagery is very well suited for the quantitative characterization of environmental features such as surface temperature, vegetation cover, land uses and the spatial pattern of the urban layout. Environmental criminologists explored the relationships between crime in urban settings and physical variables such as altitude and slope (Breetzke, 2012), temperature and weather (Anderson and Anderson, 1984; Anderson et al., 2000; Butke and Sheridan, 2010; Carlsmith and Anderson, 1979; Cohn and Rotton, 2000; DeFronzo, 1984; Field, 1992; Salleh et al., 2012; Sorg and Taylor, 2011), land use patterns (Kurtz et al., 1998; Wilcox et al., 2004), night-time lighting (Weeks, 2003), vegetation coverage (Donovan and Prestemon, 2012; Kuo and Sullivan, 2001; Troy et al., 2012; Wolfe and Mennis, 2012), and the spatial pattern of the built environment (Browning et al., 2010; Foster et al., 2010; Matthews et al., 2010; Shu and Huang, 2003; Taylor and Harrell, 1996).

Kuo and Sullivan (2001) examined the relationship between vegetation and crime in Chicago, Illinois, and found that the greener a buildings surroundings are, the fewer crimes are reported (property and violent crimes). Trees and grass density were computed from 35 mm oblique aerial photographs. The relationship of vegetation to crime held after the number of apartments per building, building height, vacancy rate, and number of occupied units per building were accounted for. Donovan and Prestemon (2012) related tree presence derived from aerial photographs to crime aggregates (all crime, violent crime, and property crime), and individual crimes (burglary and vandalism) in Portland, Oregon. They reported that trees in the public right of way are associated with lower crime rates, while this relationship on private lots is mixed.

Smaller, view-obstructing trees are associated with increased crime, whereas larger trees are associated with reduced crime. Troy et al. (2012) investigated the urban tree cover influence in crime over Baltimore, Maryland, using color infrared imagery from the National Agricultural Imagery Program (NAIP) along with surface models generated from light detection and ranging (LiDAR) data. They reported a strong inverse relationship between tree canopy and an index of robbery, burglary, theft and shooting. Results indicate a negative relationship between crime and trees in the vast majority of block groups of the study area, but there are a few patches with mixed industrial and residential use where the opposite relationship is true. Wolfe and Mennis (2012) analyzed the association of vegetation abundance with crime (assault, robbery, burglary, and theft) using a normalized difference vegetation index (NDVI) calculated from a Landsat ETM+ image. The reported that vegetation abundance was significantly associated with lower rates of assault, robbery and burglary, but not with theft.

Weeks (2003) used night-time imagery to investigate the relationship between nighttime lighting and crime in San Diego, California. Their results did not support the idea that nighttime lighting has a deterrent effect, but rather that those places with the highest crime rates against persons are those places that are most well lighted at night. This suggests that light follows crime and may therefore discourage higher rates of crime, even though a direct deterrent effect could not be established. Sparks (2011) analyzed the contextual influence of the built environment on violent crimes using a land use diversity index derived from remote sensing data. His results indicate that the determinants of violent crime depend on the model specification, but are primarily related to the built environment and neighborhood socioeconomic conditions.

Several works inquire about the relationship between the spatial pattern of the built environment and crime (Browning et al., 2010; Foster et al., 2010; Matthews et al., 2010; Shu and Huang, 2003; Taylor and Harrell, 1996). Browning et al. (2010) explored the link between commercial and residential density and violent crime in Columbus, Ohio. They reported a curvilinear association between commercial and residential density and both homicide and aggravated assault. At low levels, increasing commercial and residential density is positively associated with homicide and aggravated assault. Beyond a threshold, increasing commercial and residential density serves to reduce the likelihood of both outcomes. Matthews et al. (2010) focused on the relationship between the spatial patterning and predictors of property crime in Seattle, Washington. They analyzed aggregate property crime data as well as specific types: residential burglary, nonresidential burglary, theft, auto theft, and arson; and found that built environment variables are significant predictors of property crime,

especially the presence of a highway on auto theft and burglary. Night-time lighting, vegetation measures and land use diversity are the main remote sensing variables used in crime studies. Indices in those works have been derived from remote sensing imagery, while the spatial pattern descriptors have been computed from expensive and time consuming field surveys and appraisals. Spatial pattern descriptors from VHR imagery could be used to analyze the relationship between the spatial pattern of the built environment and crime at a lower cost than field surveys.

2.8 Chapter conclusions

This chapter describes major research themes of regional science and remote sensing in urban environments: relationships between land cover and socioeconomic variables, slum and urban deprivation hot spots (urban poverty), urban quality of life, estimation of housing values, urban growth analysis, population estimation, urban social vulnerability assessment, and crime. Human health applications of remote sensing tend to operate on a regional scale and are oriented toward the detection of environmental conditions that influence vector disease propagation and early warning systems (de Sherbinin et al., 2002; Johnson, 2007). To date, urban public health issues have been examined through satellite remote sensing by mapping urban vegetation and impervious surfaces and relating them to the urban heat island phenomenon and the health conditions of local residents (Dadvand et al., 2012; Harlan et al., 2006; Liu et al., 2007; Lo and Quattrochi, 2003; Rhew et al., 2011; Weeks et al., 2010). Tables 2.3 and 2.4 summarize the data and methods most commonly used in regional science applications of remote sensing data in urban settings. Linking socioeconomic and demographic information from field surveys with satellite data from different sensors and relating empirical measurements, spatial theory, and modeling have been successful approaches (Herold, 2009).

Researchers have found consistent relationships between environmental factors derived from satellite remote sensing data and socioeconomic variables in several cities in the US (Jenerette et al., 2007; Jensen et al., 2004; Li and Weng, 2007; Lo, 1997; Lo and Faber, 1997; Mennis, 2006). Object-oriented image classification, image texture measures and spatial metrics have been successfully applied to identify deprivation hot spots or slums in cities in developing countries (Barros, 2008; Hofmann et al., 2008; Kit et al., 2012; Kux et al., 2010; Novack and Kux, 2010; Rhinane, 2011; Stow et al., 2007; Taubenböck et al., 2009b; Weeks et al., 2007). The assumption underlying these findings is supported by Tobler's first law of geography, Everything is related to everything else, but near things are more related that distant things Tobler (1970), and was explained by Taubenböck et al. (2009b) as a self-segregation of social

Table 2.3: Regional science applications in urban settings and remote sensing data.

Application	RS data	Complementary data	References
Slum detection	Ikonos	Census	Weeks et al. (2007)
	Quickbird	Field survey	Hofmann et al. (2008)
	SPOT 5		Novack and Kux (2010)
			Kux et al. (2010)
			Barros (2008)
			Taubenböck et al. (2009b)
			Rhinane (2011)
			Kit et al. (2012)
			Kohli et al. (2012)
Quality of Life index	Landsat MSS	Census	Forster (1983)
	Landsat TM		Weber and Hirsch (1992)
	Landsat ETM+		Lo (1997)
	SPOT 2		Lo and Faber (1997)
	ASTER		Jensen et al. (2004)
			Stathopoulou and Cartalis (2006)
			Li and Weng (2006)
			Nichol and Wong (2006)
			Li and Weng (2007)
House value estimation	Landsat MSS	Census	Forster (1983)
	Landsat ETM+	Field Survey	Jensen et al. (2004)
	ASTER		Yu and Wu (2006)
	Ikonos		Taubenböck et al. (2009b)
Urban growth	Landsat TM	Census	Ward et al. (2000)
	Landsat ETM+		Phinn et al. (2002)
	IRS-1C		Madhavan et al. (2001)
	SPOT 4		Cheng and Masser (2003)
	SAR		Schneider et al. (2005)
	DSML OLS		Jat et al. (2008)
			Bhatta et al. (2010a)
			Sutton et al. (2010)
			Tsai et al. (2011)
			Zhu et al. (2012)
			Taubenböck et al. (2012)
			Leinenkugel et al. (2011)
Population estimation	Landsat TM	Census	Lo (1995)
	Landsat ETM+	Field survey	Yuan et al. (2008)
	SPOT 3		Harvey (2002)
	Ikonos		Wu and Murray (2005)
	Quickbird		Liu et al. (2006)
	LiDAR		Galeon (2008)
			Ramesh (2009)
			Lwin and Murayama (2009)
			Qiu et al. (2010)
			Lwin and Murayama (2011)
			Weng (2012)
Social vulnerability assessment	Ikonos	Field survey	Taubenböck et al. (2008)
	Quickbird		Ebert et al. (2007, 2009)
			Rashed and Weeks (2003a)
			Rashed et al. (2007)
Crime	Aerial photographs	Field survey	Kuo and Sullivan (2001)
	Color infrared imagery	Police records	Donovan and Prestemon (2012)
	Night-time imagery		Troy et al. (2012)
	Landsat ETM+		Weeks (2003)
	LiDAR		Wolfe and Mennis (2012)

Table 2.4: Image processing methods and regional science applications of remote sensing data in urban settings.

Method	RS data	Application	References
Direct radiance measures and regression analysis	Landsat MSS SPOT 3	Quality of life House value estimation Population estimation	Forster (1983) Lo (1995)
Per-pixel classification	Landsat TM Landsat ETM+ SPOT 4	Land policy assessment Population estimation	Ward et al. (2000) Phinn et al. (2002) Cheng and Masser (2003) Schneider et al. (2005) Yuan et al. (2008) Harvey (2002)
Vegetation index (NDVI/SAVI)	Landsat TM Landsat ETM+ SPOT 2	Quality of life Crime	Weber and Hirsch (1992) Lo (1997) Lo and Faber (1997) Jensen et al. (2004) Stathopoulou and Cartalis (2006) Nichol and Wong (2006) Li and Weng (2007) Wolfe and Mennis (2012)
Vegetation index (LAI) Object-oriented classification	ASTER Quickbird SPOT 5 Ikonos LiDAR SAR color infrared	House value estimation Slum detection Social vulnerability assessment Population estimation Urban growth Crime	Jensen et al. (2004) Weeks et al. (2007) Hofmann et al. (2008) Novack and Kux (2010) Kux et al. (2010) Rhinane (2011) Taubenböck et al. (2008) Taubenböck et al. (2009a) Ebert et al. (2007, 2009) Ramesh (2009) Tsai et al. (2011) Doxani et al. (2012) Zhu et al. (2012) Troy et al. (2012)
VIS model	Landsat ETM+ Landsat TM	House value estimation Land policy assessment Population estimation Urban growth	Yu and Wu (2006) Madhavan et al. (2001) Wu and Murray (2005) Gluch and Ridd (2010)
Landscape metrics	Landsat MSS Landsat TM Landsat ETM+ IRS-1C	Land policy assessment Urban growth and sprawl	Jat et al. (2008) Besussi et al. (2010) Soundranayagam et al. (2011) Shi et al. (2012) Sun et al. (2013)
Texture measures Lacunarity Spatial metrics Co-ocurrence matrix Semi-variance	Ikonos Quickbird	Slum detection Population estimation Land policy assessment Social vulnerability assessment	Barros (2008) Kit et al. (2012) Liu et al. (2006) Rashed and Weeks (2003b) Rashed et al. (2007)

groups in urban environments according to their economic status, amongst other aspects. This assumption is also used in geodemographics, in which it is said that the consumer behavior of a person living in a certain neighborhood can be inferred from information about other people living in the same neighborhood, or even in the same type of neighborhood (Harris et al., 2005). However, these relationships always need to be investigated and tested in each city where they are to be applied because the phenomena are highly specific. As stated by de Sherbinin et al. (2002), findings for one city can be quite the opposite of those for another city, even in the same country; e.g., the amount of vegetation in Detroit, Michigan, is an indicator of urban decay, whereas vegetation is positively correlated with income, house value and educational attainment in Denver, Colorado, and Phoenix, Arizona, among other cities.

Urban growth dynamics and urban sprawl have been assessed using satellite remote sensing data. Even though the Landsat sensor's spatial resolution is coarser than the requirements stated by Jensen and Cowen (1999), its images have been used successfully in several studies using the VIS model approach (Ridd, 1995) and spectral mixture analysis techniques to assess longitudinal changes in land use-land cover and urban growth quantitatively (Madhavan et al., 2001; Phinn et al., 2002; Ward et al., 2000; Wu and Murray, 2005; Yang and Liu, 2005; Yuan et al., 2008). Policy makers and urban planners can use the results of this research to better understand the urban growth trends in their cities and to plan better for future growth.

Demographers and other regional scientists can also benefit from using satellite remote sensing data to estimate population counts in urban settlements where detailed information is not available (Liu et al., 2006). Economists could also benefit from integrating remote sensing data into urban housing market models to account for environmental factors that influence property values, and they can use urban morphology measures derived from remote sensing imagery to identify housing submarkets Yu and Wu (2006). Further research is needed to demonstrate the utility and transferability of this approach.

Satellite remote sensing provides a fast and reliable way to derive useful information for social vulnerability assessment Ebert et al. (2007, 2009); Rashed and Weeks (2003b); Rashed et al. (2007); Taubenböck et al. (2008). Particularity is one of the most important concepts in the assessment of social vulnerability, and satellite remote sensing can provide data for place-based analysis (Rashed et al., 2007). Further research in cities with different environments is needed to develop a robust framework in which measures derived from remote sensing can play a leading role in the assessment of social vulnerability to natural hazards.

This review shows that, with the exception of urban growth analysis, little research has used remote sensing to analyze space–time dynamics. For example, the space–time

variations in a quality of life index before and after a government investment in public spaces and other urban services could be used to assess whether government actions have produced the expected results. Policy makers and urban planners could benefit from similar analyses of other topics, such as housing markets, deprivation hot spots, and vulnerability to natural hazards in urban settings, to help them allocate resources where they are most needed. Satellite remote sensing data have been successfully used in all of the aforementioned research areas, but some regional scientists remain skeptical about using remote sensing to derive information for their work. Although many local and regional government agencies are now using remote sensing to map, monitor and model the composition and growth of cities, more local research is needed to show the real potential and utility of satellite remote sensing for regional science in urban environments.

Chapter 3

Data preprocessing for quantitative analysis

This chapter is an edited compilation of some sections of the papers Duque, J. C., Patino, J. E., Ruiz, L. A., Pardo-Pascual, J. E. (2015). Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. Landscape and Urban Planning, 135, 11-21. doi:10.1016/j.landurbplan.2014.11.009; Patino, J. E., Duque, J. C., Pardo-Pascual, J. E, Ruiz, L. A. (2014). Using remote sensing to assess the relationship between crime and the urban layout. Applied Geography, 55C, 48-60. doi:10.1016/j.apgeog.2014.08.016; and Duque, J. C., Patino, J. E., Ruiz, L. A., Pardo-Pascual, J. E. (2013). Quantifying slumness with remote sensing data. CIEF working paper, EAFIT University, No. 13-23 (No. 13-23). Medellin.

This chapter presents the general data flow for quantitative analysis, the study site description, and the input data used in the applications of intra-urban poverty index estimation (chapter 4) and the assessment of the relationships between crime and the urban layout (chapter 5).

3.1 Data integration at object level

The general data flow for integration at object level starts with the selection or definition of the spatial unit of analysis. The socioeconomic variables are then aggregated at object level, i.e., the spatial unit of analysis via selection or using a regionalization process (explained in section 3.4). The remote sensing data processing starts with the image classification into land cover classes and the extraction of other image features (i.e., swimming pool locations) that do not depend of the selection of the spatial unit

of analysis. Then, the boundaries of the spatial units of analysis and the input VHR image are used to extract structure and texture descriptors for each spatial unit, and to build the land cover variables aggregated at the same level. At this point, all variables, socioeconomic and remote sensing-derived, are integrated to a single database for econometric modeling. Figure 3.1 shows the flow diagram of this general process.

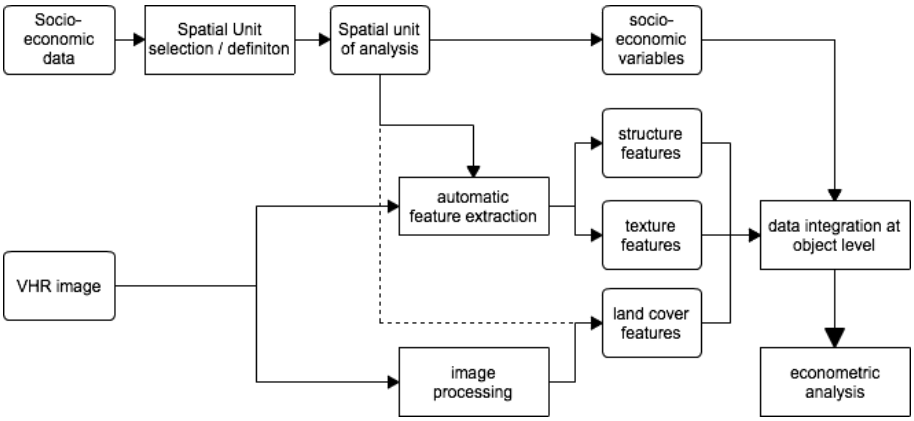

Figure 3.1: General data flow for remote sensing and socioeconomic data integration at object level.

3.2 Study site

Medellin is the second largest city in Colombia and has a population of 2.4 million (DANE, 2012), with 93.4% of its population considering themselves white descendants (which includes mixed Spanish and indigenous Amerindian descent); another 6.5% claim African descent, and 0.1% claim to be indigenous Amerindian (DANE, 2010). Medellin is located at -75.58º of Longitude and 6.22º 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 3.2). The average annual temperature is around 22 ºC, with minimal variations due to its proximity to the Equator. This 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, with the most western and eastern neighborhoods of the city built over slopes steeper than 20%.

Medellin is a useful location for conducting intra-urban variability studies because it has experienced high population growth rates since the 1950s owing to a combination of factors such as industrial development, which offered job opportunities, and then the political violence and the drug war in the rural areas of the country, which

forced many farmers and peasants to move from rural areas to the city. The rapid and unplanned urban growth exceeded the capacity of the local authorities to deliver affordable housing, public services and infrastructure in some parts of the city and resulted in the current high degree of spatial heterogeneity in both the socioeconomic and physical characteristics of its neighborhoods (Duque et al., 2013): the more affluent neighborhoods are located in the West and the South parts of the city, while the less affluent are in the North and towards the urban-rural fringe in the steepest slopes of the valley in the East and West.

3.3 Socioeconomic data for statistical analysis

We used two different sets of socioeconomic data in this work: one dataset for intra-urban poverty analysis and another for crime analysis. Socioeconomic data from the 2007 Quality of Life Survey of Medellin were used for intra-urban poverty index estimation using VHR imagery in chapter 4. This survey is developed annually since 2004 by the municipality of Medellin to measure standards of living throughout the city, and to have updated information at inter-census dates to inform policy making. Homicide counts for the year 2010 and 2011 from the *Sistema de Información de Seguridad y Convivencia, SISC* (Security and Conviviality Information System) of Medellin City Hall, and socioeconomic data from the 2005 National Census and from the 2009 Quality of Life Survey of Medellin were used for analyzing the relationship between homicide rates and urban layout features computed from VHR imagery in chapter 5.

3.3.1 Intra-urban poverty: the slum index

The slum index was computed using data from the 2007 Quality of Life Survey of Medellin. This survey includes 184 questions on the following nine 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., 2013). The 2007 Quality of Life Survey considers a wide list of questions related with poverty, from which Duque et al. (2013) 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 3.1 describes the variables used to calculate slum index and the proportion of housing units in the city under those conditions. The proportion of households not connected to piped water or to the sewer system is very low in Medellin because of

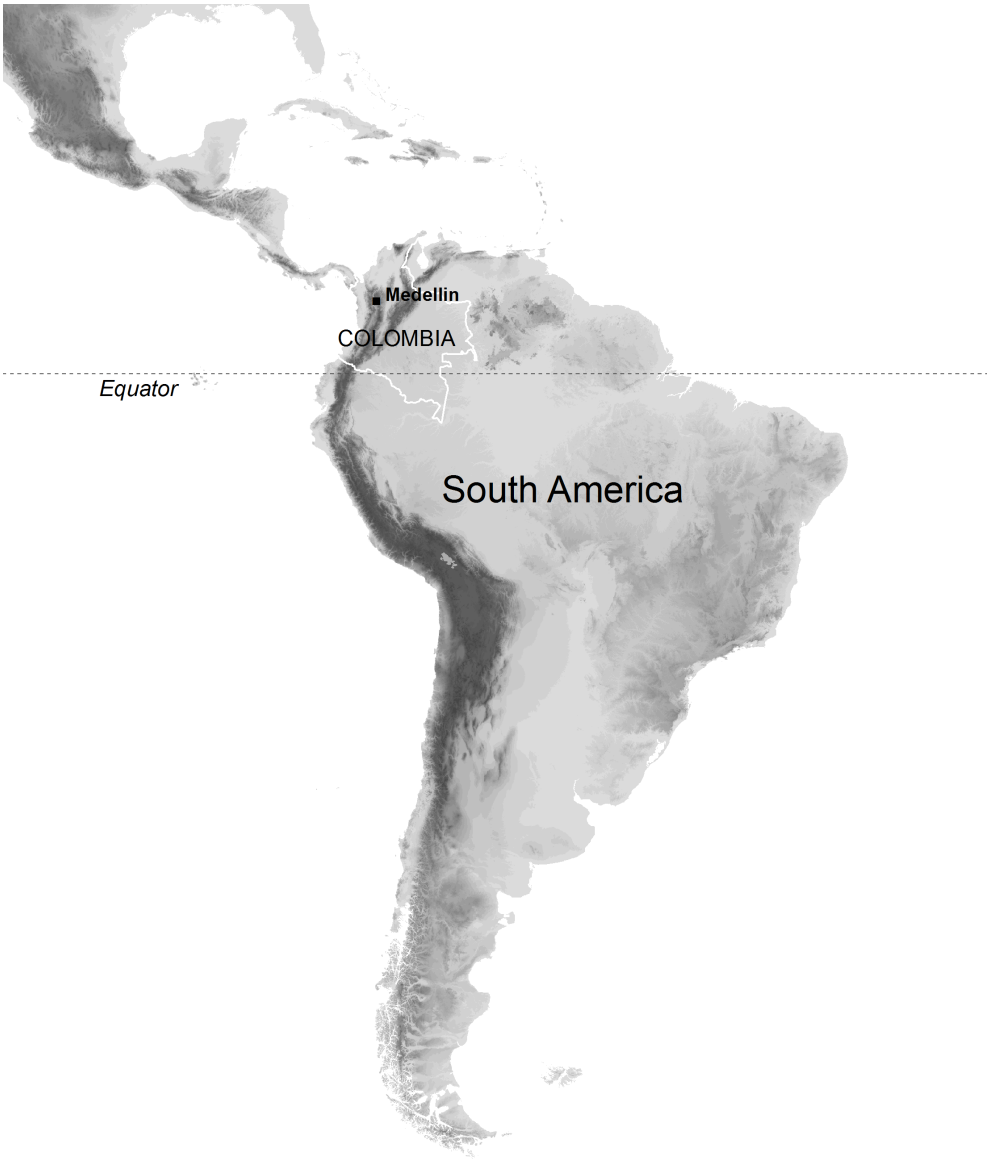

Figure 3.2: Location of Medellin City in Colombia and South America.

the city's high coverage of public services, which is atypical of Colombian cities. The proportion of housing units that are constructed from non-durable materials is also very low for this city. Figure 3.3 shows the Slum Index map at the analytical region level. Slum Index values are highly variable throughout the city, ranging from 0.17 to 0.86, and the map shows a directional spatial pattern with wealthier neighborhoods located in the south and higher values of the Slum Index in the north.

Table 3.1: Variables used to calculate slum index from the 2007 Quality of Life Survey of Medellin and proportion of housing units under those conditions. Modified after Duque et al. (2013).

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

3.3.2 Crime: background, intra-urban homicide counts and socioeconomic variables

During the last two decades, Colombia has sustained a reputation as one of the most violent countries in the world, and Medellin is considered to be its most violent city (Giraldo Ramírez, 2010). The average annual homicide rate in the country between 1998 and 2003 was 60 homicides per 100,000 inhabitants, whereas in Medellin, the rate reached 156; from 2004 to 2009, the rates decreased to 37 for Colombia and 50 for Medellin (SISC, 2010). Homicide rates in Medellin in the late 1980s correlate with the rise in narco-violence and the strategy of guerrillas and paramilitary groups of bringing the conflict to the cities, whereas the decreasing trends have their starting points in the peace agreements with guerrillas in 1990, the end of the Medellin drug cartel and the death of Pablo Escobar in December 1993, the joint Orion operation between the police and the army to retake control over the most western part of the city in 2002, and the demobilization of paramilitary groups in December 2003 (Giraldo Ramírez, 2010). Statistics for Medellin in recent years show a decreasing trend, from 380 homicides per 100,000 inhabitants in the early 1990s (Gaviria, 2000) to 34 in 2007, and this number peaked again at 94 in 2009 (Giraldo Ramírez, 2010) (Figure 3.4).

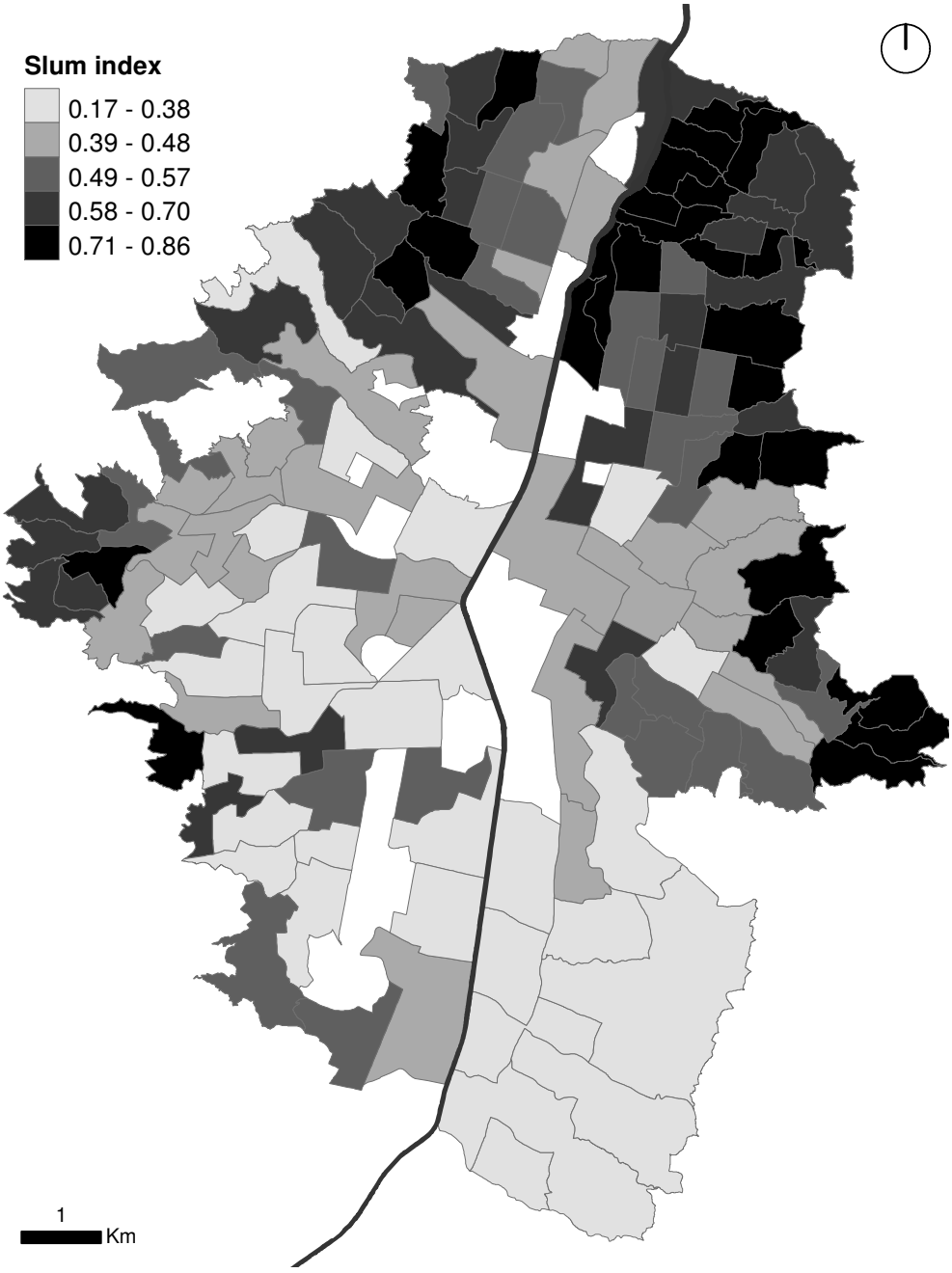

Figure 3.3: Medellin's Slum Index map at the analytical region level.

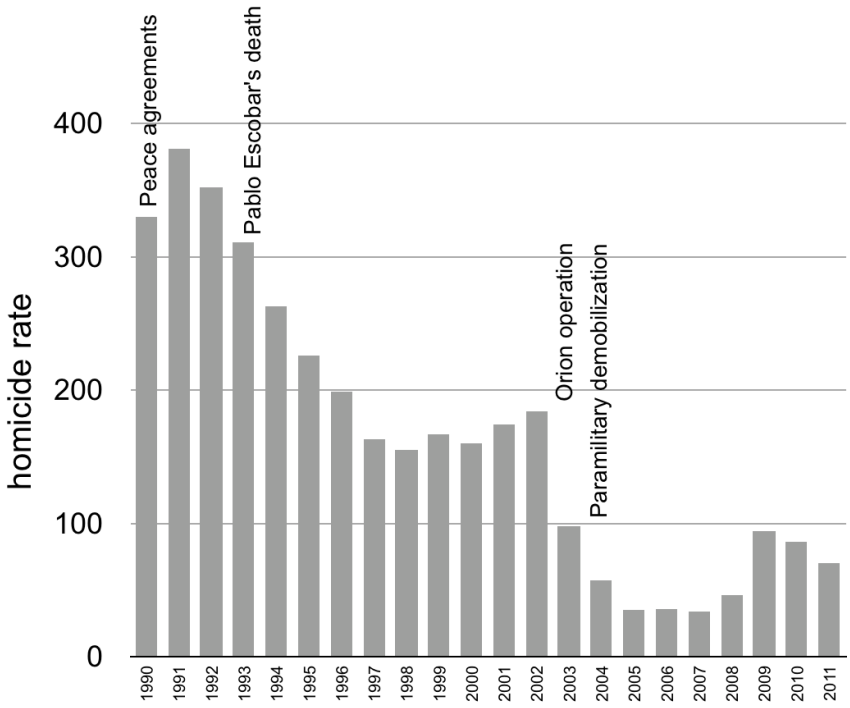

Figure 3.4: Medellin's annual homicide rates per 100,000 people from 1990 to 2011. Data source: Giraldo Ramírez (2010) and Medellín Cómo Vamos (2012).

Homicide counts by administrative neighborhood for Medellin for the years 2010 and 2011 were provided by the SISC of Medellin City Hall. Figure 3.5 shows the intra-urban variations in average homicide counts for the years 2010 and 2011 at the administrative neighborhood level in Medellin. The highest homicide counts are found in the downtown, as well as in some neighborhoods located in the North of the city and near the urban-rural fringe in the East, West, and Southwest. Blank areas in the map are institutional areas such as the local airport, public parks, and City Hall facilities.

Socioeconomic variables were computed from the 2005 National Census and the 2009 Quality of Life Survey of Medellin to study the relationship between homicide and the urban layout. The 2009 Quality of Life survey comprises 237 questions on 11 dimensions: housing, households, demographics, gender and family violence, education, social security, social participation, mobility, security and crime victimization, income and employment, and nutrition. The sample contains 20,782 households that represent 84,474 persons. These data were used to calculate socioeconomic variables to control for potential confounders when analyzing the influence of the urban layout characteristics on homicide.

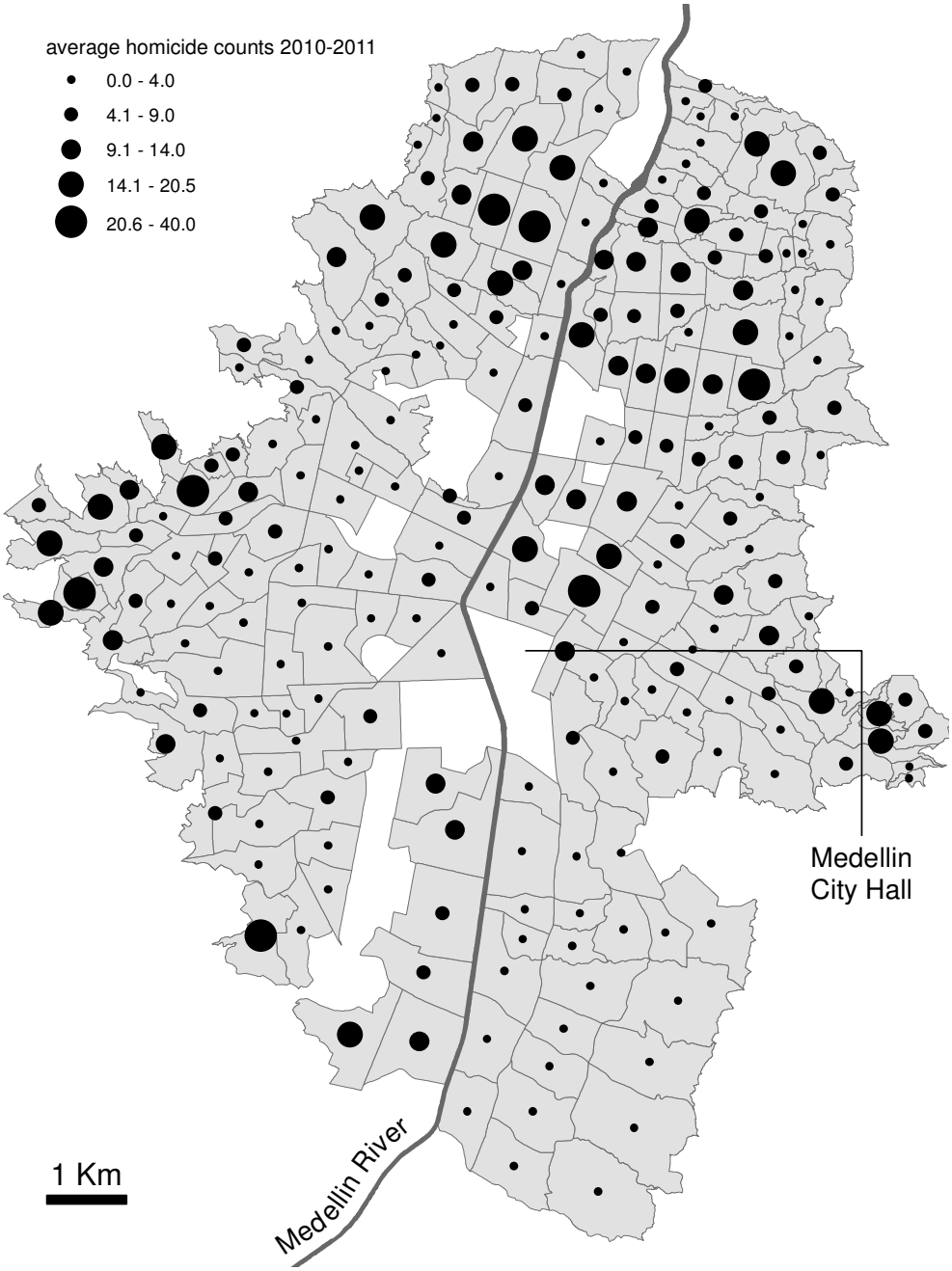

Figure 3.5: Average homicide counts for the years 2010 and 2011 at the administrative neighborhood level. Data source: SISC.

3.4 Spatial unit of analysis

Medellin has two levels of urban administrative units: communes (16) and neighborhoods (243). Socioeconomic variables and indicators from the 2007 and 2009 Quality of Life surveys were reported at the commune level. Two main drawbacks arise when using these units: they are too large to study the spatial patterns of intra-urban variations in socioeconomic conditions, and the statistical inference based on these units may be affected by aggregation problems such as the ecological fallacy (Robinson, 1950) and aggregation bias (Amrhein and Flowerdew, 1992; Fotheringham and Wong, 1991; Paelinck and Klaassen, 1979; Paelinck, 2000). The average number of surveyed households in 2007 was 84 ± 57, with values ranging from 3 to 296, and 15.64% of the neighborhoods had fewer than 30 surveyed households; while in 2009, the average number of surveyed households by neighborhood was 86 ± 54, with values ranging from 2 to 277, and 28 neighborhoods had fewer than 30 surveyed households; thus, the Medellin's administrative neighborhoods are unsuitable as spatial units of analysis because of the lack of statistical validity, the small numbers problem for rates calculation (Diehr, 1984), and the potential appearance of spurious autocorrelation (Weeks et al., 2007).

An alternative means of addressing the problems described above involves the use of analytical regions, which are spatial units that fulfill specific criteria (e.g. size, shape, attribute homogeneity, among others) that are relevant to the phenomena of study (Duque et al., 2006; Weeks et al., 2007). Regionalization can reduce the spurious spatial autocorrelation that is present when analyzing smaller administrative spatial units (Weeks et al., 2007). Analytical regions are delineated by grouping the administrative neighborhoods using the max-p-regions algorithm in ClusterPy software (Duque et al., 2011; **?**)[1]. This process has been used in previous research on the urban spatial structure and spatial dynamics of major U.S. metropolitan areas (Arribas-Bel and Schmidt, 2013; Rey et al., 2011), pediatric mortality in Brazil (Leyk et al., 2012), interregional inequality in Mexico (Rey and Sastré-Gutiérrez, 2010), intra-urban inequalities in Accra, Ghana (Stow et al., 2007; Weeks et al., 2006, 2007), and intra-urban poverty in Medellin, Colombia (Duque et al., 2013).

The max-p-regions seeks to aggregate n areas into the maximum amount of spatially contiguous regions, such that each region satisfies a predefined minimum threshold value for some spatially extensive regional attribute (e.g. the number of surveyed households per region). The max-p-regions also seeks to minimize intra-regional heterogeneity measured as:

[1] ClusterPy is an open source library of spatially constrained clustering algorithms that runs in the Python programming language (Duque et al., 2011)

$$H = \sum_{k=1}^{p} \sum_{ij \in R_k | i<j} d_{ij} \tag{3.1}$$

Where d_{ij} is the multidimensional squared Euclidean distance between areas i and j assigned to the same region k, R_{k}. Both objectives, the number of regions and the intra-regional heterogeneity, are merged into a single objective function as follows:

$$Minimize\, Z = \left(-\sum_{k=1}^{n} \sum_{i=1}^{n} x_i^{k0} \right) \times 10^h + \sum_i \sum_{j|j>i} d_{ij} t_{ij} \tag{3.2}$$

With two binary decision variables: $x_i^{k0} = 1$ if area i is assigned to region k in order 0 (each region has one and only one area assigned at this order), and $t_{ij} = 1$ if areas i and j are assigned to the same region. The first term represents the number of regions and the second term represents the intra-regional heterogeneity. Both terms are merged into a single value in such a way that an increment in the number of regions is preferred over any other solution involving a lower number of regions, and for a given number of regions, the solution with lower intra-regional heterogeneity is preferred. This hierarchy in the elements of the objective function are ensured by multiplying the first term by a scaling factor 10^h, with $h = 1 + \left[\log \sum_i \sum_{j|j>i} d_{ij}\right]$. In this work, each analytical region was required to meet the following criteria: socioeconomic homogeneity, measured using a wide set of socioeconomic variables, and a minimum of 100 surveyed households in the 2007 and the 2009 Quality of Life surveys as requirement for statistical validity.

In chapter 4, the analytical regions were delineated by applying the max-p-regions algorithm to the administrative neighborhoods with the 2007 Quality of Life Survey data, which grouped the 243 city neighborhoods to 139 analytical regions. The Slum Index was computed for these 139 spatial units. In chapter 5, the same procedure was applied with the 2005 Census and the 2009 Quality of Life Survey data, which resulted in 136 analytical regions. Homicide rates and socioeconomic variables for the years 2005 and 2009 were computed for these 136 analytical regions.

3.5 Remote sensing data and image processing

A very high spatial-resolution (VHR) image was used to build a database of variables at the analytical region level. The image was an RGB composition of a Quickbird scene with spatial resolution of 0.60 m captured in May 2008. VHR imagery allows an accurate classification of urban landscape features. Medellin City Hall Officers of the Sigame Group preprocessed the original Quickbird image using standard procedures.

They used the RCPs provided by the distributor and a local DEM to orthorectify the image. Then they did a fusion of the panchromatic band with the multispectral bands to produce the pan sharpened natural color RGB composition. For this work, we did not have access to the original image with near infrared (NIR) information, but to this natural color composition with spatial resolution of 0.6 m.

We applied principal component analysis to summarize the spectral information of the red, green and blue bands into one single band (the first principal component, PC1) and the band ratios red/blue and green/red to enhance the spectral information for a better differentiation of surfaces that reflected red and green. The image-derived variables were grouped into a set of land cover features, a set of structure features and a set of texture features.

3.5.1 Land cover features

Land cover features characterize the composition of an urban scene in terms of three basic land cover types, vegetation, soil, and impervious surfaces, and an additional land cover: clay roofs. Land cover features were obtained through pixel-based supervised classification using the maximum likelihood algorithm in ENVI with the composite image of PC1 and the red/blue and green/red band ratios (Figure 3.6). We collected on-screen circular regions of interest (ROIs) of 1-meter radius that contains approximately 9 pixels, as training and test samples for image classification. The classification accuracy was assessed using a traditional point- based technique with a reference dataset of randomly selected points (the sample of ground truth points was randomly divided into two datasets: 80% as a classification training set and the other 20% for accuracy assessment). The overall accuracy of this classification was 95%, with a kappa statistic value of 0.94. Table 3.2 shows the confusion matrix for the classification results. The producer's accuracy for vegetation, impervious surfaces, soil and the clay roof classes was over 93%; user accuracy was over 91% for the vegetation class; over 98% for the impervious surface and clay roof classes; and approximately 77% for the soil class.

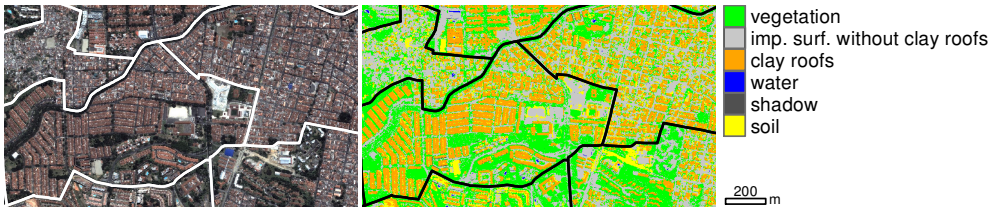

Figure 3.6: Example of the input VHR image and classified land cover types within analytical regions.

Table 3.2: Accuracy assessment of the classified image. The number of training and validation samples are expressed in pixels.

Ground truth	Vegetation	Imp. surfaces without clay roofs	Clay roofs	Water	Shadow	Soil	Total
Training samples	2,952	5,636	2,920	1,732	768	800	14,808
Validation samples	738	1,409	730	433	192	200	3,702
Classified as							
Vegetation	712	24	14	0	26	0	776
Impervious surfaces without clay roofs	10	1,356	0	6	1	9	1,382
Clay roofs	0	0	708	0	0	5	713
Water	0	0	0	400	0	0	400
Shadow	6	0	0	18	165	0	189
Soil	10	29	8	9	0	186	242
Total	738	1,409	730	433	192	200	
Producer accuracy (%)	96.48	96.24	96.99	92.38	85.94	93.00	
User accuracy (%)	91.75	98.12	99.29	100.00	87.30	76.86	
Overall classification accuracy (%)	95.27						
Kappa value	0.94						

We then calculated the percentage cover of vegetation, impervious surfaces, soil and clay roofs within analytical regions to build the following variables: percent of vegetation, percent of impervious surfaces, percent of soil, percent of clay roof surfaces, percent of impervious surfaces other than clay roofs and the fraction of clay roofs among the impervious surfaces. 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 also computed. Table 3.3 lists the land cover features derived from the input VHR image at the analytical region level.

Table 3.3: Land cover variables derived from VHR image.

Variable	Description
S.POOL.DENS	Mean swimming pool density
P.VEG	Percentage of vegetation cover
P.IMP.SURF	Percentage of impervious surface cover including clay roof cover
P.SOIL	Percentage of bare soil cover
P.CLAY.ROOFS	Percentage of clay roof cover
P.OTHER.IMP.S	Percentage of impervious surfaces without clay roof cover
F.CLAYR.IMPS	Fraction of clay roof cover within the impervious surface cover

3.5.2 Structure features

We used the PC1 and the analytical region boundaries to extract the structure and texture features using FETEX 2.0.[2] These features quantify different aspects of the spatial arrangement of the image's intensity values within the analytical regions, and we used them to quantify the differences in the spatial pattern of the urban layout.

Structure features inform about the spatial arrangement of elements within regions in terms of the randomness or regularity of the distribution of the elements (Balaguer et al., 2010; Balaguer-Beser et al., 2013; Ruiz et al., 2011); structure features are calculated in FETEX 2.0 using the experimental semivariogram approach. The semivariogram quantifies the spatial associations of the intensity values, measures the degree of spatial correlation between different pixels in an image and is a suitable tool for characterizing spatial patterns (Ruiz et al., 2011). 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°; then, each semivariogram curve is smoothed using a Gaussian filter to reduce experimental fluctuations (Ruiz et al., 2011). Structure features extracted from the semivariogram are based on a 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).[3] Table 3.4 lists the structure features derived from the input VHR image at the analytical region level. The section A.1 in Appendix A presents an explanation of the structure features used in this work.

[2]FETEX 2.0 is a free software tool for object-based image analysis that runs on ENVI image processing software and the IDL programming language (Ruiz et al., 2011).

[3]For a full description of these features see Appendix A, section A.1 and Balaguer et al. (2010); Balaguer-Beser et al. (2013); and Ruiz et al. (2011).

Table 3.4: Structure variables derived from VHR image.

Variable	Description
RVF	Ratio variance at first lag
RSF	Ratio between semivariance values at second and first lag
FDO	First derivative near the origin
SDT	Second derivative at third lag
MFM	Mean of the semivariogram values up to the first maximum
VFM	Variance of the semivariogram values up to the first maximum
DMF	Difference between the mean of the semivariogram values up to the first maximum and the semivariance at first lag
RMM	Ratio between the semivariance at first local maximum and the mean semivariogram values up to this maximum
SDF	Second order difference between first lag and first maximum
AFM	Area between the semivariogram value in the first lag and the semivariogram function until the first maximum

3.5.3 Texture features

Texture features quantify the spatial distribution of intensity values in the image and provide information about contrast, uniformity, roughness, etc. (Baraldi and Parmiggiani, 1995; Haralick et al., 1973). 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 GLCM tabulates the frequency of one gray tone appearing in a specified spatial relationship with another gray tone within the analyzed polygon (Baraldi and Parmiggiani, 1995). FETEX 2.0 uses the GLCM to calculate a set of variables that are widely used in image processing: uniformity, entropy, contrast, inverse difference moment (IDM), covariance, variance, and correlation (Haralick et al., 1973). The kurtosis and skewness features are based on the histogram of the pixel values inside the polygon. The edgeness factor is another feature that represents the density of edges present in a region (Sutton and Hall, 1972), and the mean and standard deviation of the edgeness factor (MEAN EDG, STDEV EDG) are also computed within this set of features (Ruiz et al., 2011). Table 3.5 lists the texture features derived from the input VHR image at the analytical region level. The section A.2 in Appendix A presents a full description of the texture features used in this work.

Table 3.5: Texture variables derived from VHR image.

Variable	Description
SKEWNESS	Skewness value of the histogram
KURTOSIS	Kurtosis value of the histogram
UNIFOR	GLCM uniformity
ENTROP	GLCM entropy
CONTRAS	GLCM contrast
IDM	GLCM inverse difference moment
COVAR	GLCM covariance
VARIAN	GLCM variance
CORRELAT	GLCM correlation
MEAN_EDG	Mean of the edgeness factor
STDEV_EDG	Standard deviation of the edgeness factor

Chapter 4

Measuring intra-urban poverty from space

This chapter is a compilation of some sections of the papers Duque, J. C., Patino, J. E., Ruiz, L. A., Pardo-Pascual, J. E. (2015). Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. Landscape and Urban Planning, 135, 11-21. doi:10.1016/j.landurbplan.2014.11.009; and Duque, J. C., Patino, J. E., Ruiz, L. A., Pardo-Pascual, J. E. (2013). Quantifying slumness with remote sensing data. CIEF working paper, EAFIT University, No. 13-23 (No. 13-23). Medellin.

The majority of the global population today is urban. The percentage of urban dwellers increased from 43% in 1990 to 52% in 2011, and it is expected to grow to 67% by 2050 (United Nations, 2007, 2008, 2012, 2014). All population growth from 2011 to 2050 is expected to be absorbed by urban areas, and most of this growth will occur in cities of less developed regions (United Nations, 2012). In developing countries, rapid urban growth normally exceeds the capacity for local governments to deliver services and infrastructure, which increases urban poverty and intra-urban inequalities (Duque et al., 2013).

The monitoring of poverty is a key issue for policy makers because it can help prevent poverty traps and crime nests and allocate public investments where they are needed most (Duque et al., 2013). Urban poverty is a multidimensional phenomenon; as such, there are many ways to measure it. These measures usually include information from at least one of the following dimensions: income/consumption, health/education, and housing (Carr-Hill and Chalmers-Dixon, 2005; Moser, 1998). They are computed from survey or census data, which are quite expensive, time con-

suming, less frequently produced, and often statistically significant for spatial units that are too large to capture the intra-urban variability of phenomena. This last feature creates inference problems such as the ecological fallacy (Baud et al., 2010; Robinson, 1950) or aggregation bias (Fotheringham and Wong, 1991; Paelinck and Klaassen, 1979).

This study works toward overcoming these problems by exploring the possibility of using remote sensing imagery to measure urban poverty. This proposal is based on the premise that the physical appearance of a human settlement is a reflection of the society in which it was created and on the assumption that people living in urban areas with similar physical housing conditions have similar social and demographic characteristics (Jain, 2008; Taubenböck et al., 2009b). The main advantage of using remote sensing imagery for urban poverty quantification is that this type of data can be obtained faster, at higher frequencies, and for a fraction of the cost required for field surveys and censuses. Poverty mapping usually follows two types of approaches: the expenditure-based econometric approach linked to a poverty line used by World Bank, and the value-focused approach used by United Nations Development Programme (UNDP) based on the Human Development Index (Baud et al., 2009). The Index of Multiple Deprivations (Baud et al., 2008), the Slum Index (Weeks et al., 2007), and the Slum Severity Index (Patel et al., 2014) all follow the value-focused approach that integrates several dimensions of deprivation in one single measure.

We chose the Slum Index to corroborate this possibility because this measure is based on the physical aspects of dwelling units. A slum household is defined as 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 at sufficient amounts and 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 (UN-HABITAT, 2006). We implement spatial econometric models using data from Medellin (one of the most unequal cities in the world) to assess whether the Slum Index can be estimated using image-derived measures.

4.1 Poverty and neighborhood appearance

Weeks et al. (2007) presented the calculation of the Slum Index from census and survey data as the sum of the fractions of households that lack one or more of the five conditions mentioned above. The value can range from 0, meaning that no slum-like households are present in an area, to 5, where all households in an area lack all five of the features defined by UN-Habitat. The proportion of slum dwellers in cities

is strongly correlated with the Human Development Index, which integrates three development indicators: per capita GDP, longevity and educational attainment (UN-Habitat, 2003). Thus, the presence of slums in a city is an indicator of poverty, and the Slum Index is a good proxy variable for urban poverty at the intra-urban level.

Weeks et al. (2007) and Stoler et al. (2012) used land cover descriptors and texture measures from medium to very high spatial resolution satellite imagery to develop spatial econometric models for predicting the Slum Index as a function of remote sensing-derived variables. This work builds on these previous studies by analyzing a wider set of remote sensing variables on land cover composition, image texture and urban layout spatial pattern descriptors to provide empirical evidence that either supports or refutes the hypothesis that remote sensing could be used to estimate the Slum Index at the intra-urban scale. As our intention was to lower the costs of this approach as much as possible, we use data drawn from an RGB composition of a Quickbird scene with a spatial resolution of 0.60 m captured in May of 2008. The imagery is similar in color and spatial resolution to Google Earth and Microsoft Bing images (Quickbird is a commercial Earth-observation satellite that collects very high spatial resolution -VHR- imagery). Although the conclusions of this exercise may not be valid worldwide, we seek to present new, innovative and low-cost means of measuring urban poverty.

4.2 Modeling the slum index from remote sensing variables

We developed a model for Slum Index estimation based solely on remote sensing data from a VHR image to assess the performance of the remote sensing-derived variables for this purpose and to examine the associations between urban fabric characteristics and urban poverty. The proposed model has the following general form:

$$Slum\,Index = F\,(land\,cover,\ structure,\ texture) \tag{4.1}$$

We used survey data to compute the Slum Index from socioeconomic variables following the method outlined by Weeks et al. (2007) and a VHR image to extract a wide set of remote sensing-derived measures. Although VHR imagery of higher spectral resolution exists, we deliberately chose to work with an image of a spatial and spectral resolution similar to many aerial color photographs and to Google Earth and Microsoft Bing imagery to make this approach repeatable for analyses of other urban areas around the world. Although the value of the variables that comprise the Slum Index cannot be measured directly from remote sensing data, 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 can provide a framework for the analysis of intra-urban variations of poverty so that urban planners and policy makers can identify which areas require the most attention (Stoler et al., 2012).

Stoler et al. (2012), Stow et al. (2007) and Weeks et al. (2007) developed spatial econometric models to estimate the Slum Index in Accra, Ghana as a function of within-neighborhood measures of vegetation, impervious surface and soil as well as a texture measure from remote sensing imagery. Owen and Wong (2013) reviewed indicators derived from remote sensing imagery and digital elevation data that have been reported to be useful for slum detection. Using information on Guatemala City, the authors found that the variables that best distinguished informal from formal settlements include the entropy of roads, vegetation patch size, vegetation compactness, profile convexity, road density and soil coverage. Following these previous studies, a model for Slum Index estimation using remote sensing imagery should include information on not only land cover materials but also on the spatial pattern of the urban fabric. Table 4.1 shows the complete list of remote sensing variables used in the statistical analysis and their summary statistics.

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).

The use of Ordinary Least Squares (OLS) regression with spatial data is problematic due to the potential presence of spatial autocorrelation within the model residuals, which violates the assumption on the independence of residuals. We evaluated the OLS residuals for evidence of spatial autocorrelation following the specification search strategy proposed by Florax et al. (2003) and using the Lagrange Multiplier tests and their robust forms to determine the appropriate type of spatial model to use, and we ran the spatially adjusted regression using OpenGeoda software (Anselin et al., 2006). Following the notation of Anselin (1988), the general form of the spatial model is:

$$Slum\ Index = X\beta + \rho W\ Slum\ Index + u \tag{4.2}$$

$$u = \lambda W u + \epsilon \tag{4.3}$$

where X is an n by k matrix of observations on explanatory variables (in our case, image-derived variables), β is a k by 1 vector of parameters associated with the independent variables, W is a spatial weight matrix that represents the spatial association between regions, ρ is a measure of the strength of that spatial association,

Table 4.1: Summary statistics of remote sensing derived variables used in the analysis.

Group	Variable	Mean	Standard deviation
Land cover	S.POOL.DENS	5.04	6.50
features	P.VEG	36.66	13.19
	P.IMP.SURF	54.09	12.75
	P.SOIL	8.04	2.60
	P.CLAY.ROOFS	13.22	9.24
	F.CLAYR.IMPS	0.25	0.15
Structure	RVF	12.23	2.80
features	RSF	1.80	0.02
	FDO	116.93	20.46
	SDT	15.15	3.18
	MFM	822.65	132.71
	VFM	1.44	0.22
	DMF	676.61	108.08
	RMM	1.41	0.04
	SDF	-338.46	73.52
	AFM	5,236.57	890.74
Texture	SKEWNESS	-0.75	0.26
features	KURTOSIS	0.33	0.69
	UNIFOR	0.01	0.002
	ENTROP	2.61	0.07
	CONTRAS	15.23	2.73
	IDM	0.43	0.04
	COVAR	97.27	14.35
	VARIAN	104.89	14.41
	CORRELAT	0.93	0.02
	MEAN.EDG	8.37	1.02
	STDEV.EDG	7.42	0.54

and u is the remaining error term with λ being the parameter measuring the strength of the spatial association in the error term. Finally, ϵ is a vector of independently distributed random terms. When neighboring values of the dependent variable have a direct effect on the value of the dependent variable itself, the λ parameter equals zero, and we obtain a spatial autoregressive or spatial lag model. When the spatial dependence enters through the errors rather than through the systematic component of the model, the ρ parameter equals 0, and we obtain a spatial error model.

We first parameterized the spatial weight matrix, which defines the areas that are considered neighbors of a given area, using the Rook and Queen specification of contiguity matrices. The Rook specification states that two polygons are neighbors if they share a border, while the Queen specification states that two polygons are neighbors if they share a border or a vertex. As these specifications created confusing results regarding the type of spatial dependence present in the models tested, we conducted the specification of a neighbor weight matrix based on a fixed distance band; thus, all neighbors within a specified distance of a given polygon are considered neighbors. To determine an appropriate distance band, we followed the strategy

of Troy et al. (2012) and built the semivariogram of the endogenous variable, the Slum Index, to determine the distance at which spatial autocorrelation decays. The semivariogram shows how variance between observations pairs varies as a function of distance, and in this case it showed a terrace-type pattern with a sill at approximately 2 km (Figure 4.1). This indicated that the autocorrelation of observation pairs leveled off at that distance. We performed a sensitivity analysis using distance bands ranging from 1.3 to 2 km and found that the models were very consistent in terms of the magnitude and sign of the coefficients. However, the 1.7 km distance band provided conclusive results on the type of spatial model to be used.

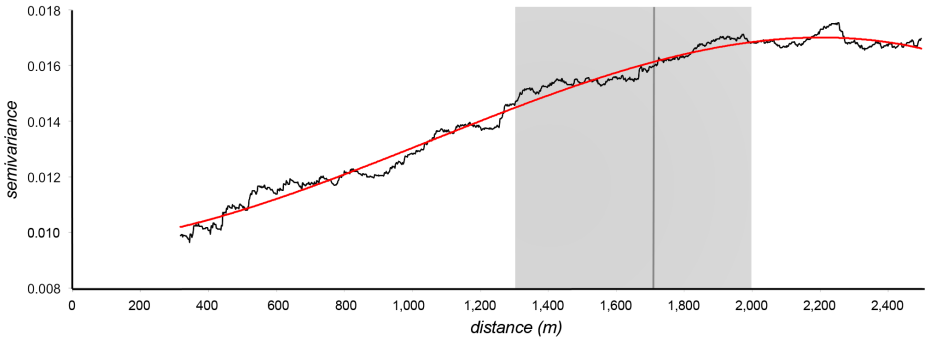

Figure 4.1: Empirical omnidirectional semivariogram of Slum Index. Smoothed trend line in red. Gray area shows the distance interval tested for the weight contiguity matrix specification.

4.2.1 Statistical correlations between slum index and remote sensing variables

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

The last row in Figure 4.2 shows bivariate scatter plots of remote sensing derived variables (plotted in the X-axis) vs. slum index (plotted in the Y-axis). The last

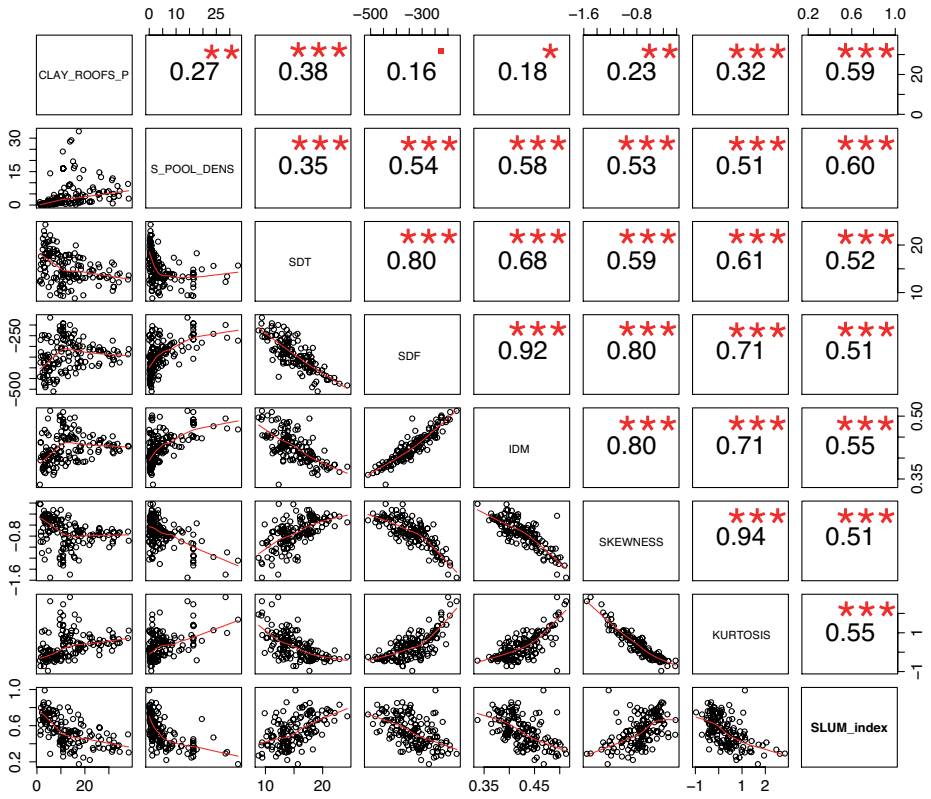

Figure 4.2: Graph of the correlation matrix of the selected remote sensing variables and Slum index. Upper right part shows the absolute value of the Pearson's correlation coefficient and significance levels as symbols in red: 0.1%: ***, 1%: **, 5%: *, 10%: · . Lower left part shows bivariate scatter plots with a fitted line in red.

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 4.3 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.

4.2.2 Models using VHR Image derived variables

A number of statistical models were run to test the relationship between slum index and the selected remote sensing variables. Figure 4.4 outlines the image processing methodology, neighborhood delineation and statistical analysis used to develop these econometric models. The section of the diagram with a gray background portrays the basic steps applied by Duque et al. (2013) that were used to delineate the analytical regions.

Ordinary Least Squares (OLS) regressions were used with slum index as a dependent variable and the remote sensing variables as independent variables. Table 4.2 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 (Model 1) and the other five models with one added variable at a time from the other groups (texture and structure, models 2 to 6). We did not include different structure and texture variables in the same model in order to avoid multicollinearity, because it is clear from figure 4.2 that they are highly correlated among them. The first OLS model (Model 1) shows that swimming pool density and clay roofs percentage cover together account for 55% of the variability in the slum index.

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 4.2, Model 2 and Model 3). 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 4.2, Models 4 to 6). 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 (Model 1) to 0.60 in the case of IDM (Model 4), and to 0.58 for skewness (Model 5),

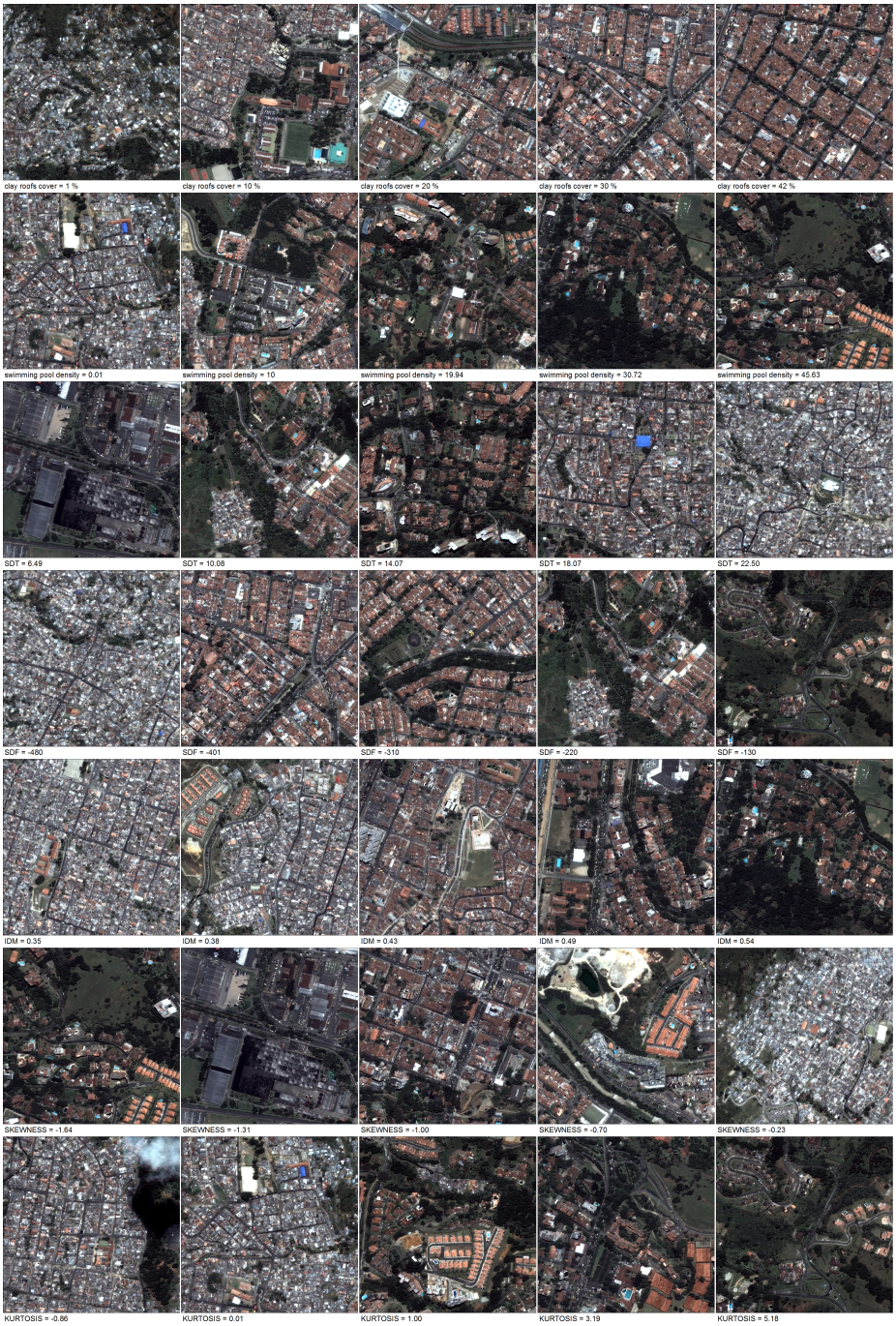

Figure 4.3: Square image tiles showing urban areas (500 x 500 meters) that have different values of the remote sensing variables selected in model specification. From top to bottom: clay roofs percentage cover, swimming pool density, SDT, SDF, IDM, skewness and kurtosis. Tiles organized according to increasing values of each variable from left to right.

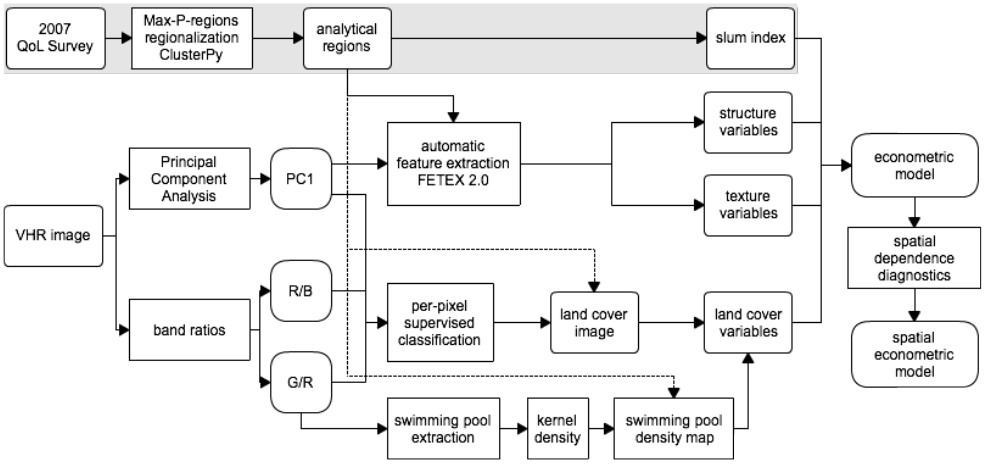

Figure 4.4: Steps for model specification of slum index as a function of remote sensing-derived variables from a VHR image of the urban scene.

Table 4.2: Multivariate OLS 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	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	0.7061***	0.5039***	0.4904***	1.2442***	0.7899***	0.6997***
P.CLAY.ROOFS	-0.0081***	-0.0069***	-0.0081***	-0.0080***	-0.0078***	-0.0073***
S.POOL.DENS	-0.0121***	-0.0105***	-0.0086***	-0.0079***	-0.0093***	-0.0094***
SDT		0.0118***				
SDF			-0.0006***			
IDM				-1.3149***		
SKEWNESS					0.1362***	
KURTOSIS						-0.0537***
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
Diagnostics for spatial dependence:						
Lagrange Multiplier - lag	12.45***	4.05**	5.39**	5.19**	8.77**	8.36***
Robust LM - lag	10.93***	1.95	2.62*	3.49*	4.46**	4.70**
Lagrange Multiplier - error	2.51	2.12	2.84*	1.72	4.36**	3.66**
Robust LM - error	0.99	0.02	0.08	0.03	0.05	0.00

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

and kurtosis (Model 6). 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 4.2 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. We performed spatial dependence diagnostics using the Lagrange Multiplier with a contiguity matrix based on euclidean distance with threshold at 1,700 m. 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 4.3 shows the spatially adjusted regressions of Slum index as a function of remote sensing derived variables for all previous model specifications except for Model 4, 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. The 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 4.3: Spatial Lag models (SML) 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	SLM 1	SML 2	SLM 3	SLM 5	SLM 6
Constant	0.4836***	0.4093***	0.3731***	0.5882***	0.5174***
P.CLAY.ROOFS	-0.0064***	-0.0061***	-0.0069***	-0.0064***	-0.0060***
S.POOL.DENS	-0.0089***	-0.0088***	-0.0070***	-0.0071***	-0.0073***
SDT		0.0089***			
SDF			-0.0005***		
SKEWNESS				0.1133***	
KURTOSIS					-0.0447***
W Slum Index	0.3446***	0.2225**	0.2407**	0.2906***	0.2839***
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% levels as indicated by ***, **, and * respectively.

Table 4.4: Univariate OLS 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	Model 7	Model 8	Model 9	Model 10	Model 11
Constant	0.1306**	0.1492***	1.6155***	0.7821***	0.5804***
SDT	0.0269***				
SDF		-0.0011***			
IDM			-2.5257***		
SKEWNESS				0.3250***	
KURTOSIS					-0.1310***
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
Diagnostics for spatial dependence:					
Lagrange Multiplier - lag	53.20***	66.79***	58.64***	76.16***	66.03***
Robust LM - lag	11.74***	15.01***	22.23***	5.24**	7.25***
Lagrange Multiplier - error	42.16***	52.46***	36.44***	79.30***	65.09***
Robust LM - error	0.69	0.69	0.03	8.38***	6.32**

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

We then 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 4.4). 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.

Again, the results of the Breusch-Pagan test indicated that there isn't evidence of heteroskedasticity in these regressions (Models 7 to 11), but the Lagrange Multiplier tests pointed to the presence of spatial autocorrelation that should be addressed through spatially adjusted regressions. Table 4.5 shows the respective spatial lag models for each structure and texture variables (SLM 7 to SLM 11). 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-R^2 values indicate that there is a significant improvement of the explanatory power of these models when spatial effects are taken into account.

Table 4.5: Spatial Lag models (SML) 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	SLM 7	SML 8	SLM 9	SLM 10	SLM 11
Constant	0.0034	-0.0139	0.8055***	0.3233***	0.2265***
SDT	0.0127***				
SDF		-0.0006***			
IDM			-1.4079***		
SKEWNESS				0.1811***	
KURTOSIS					-0.0758***
W Slum Index	0.6417***	0.6507***	0.6238***	0.6567***	0.6292***
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% levels as indicated by ***, **, and * respectively.

4.2.3 Models using Factor Analysis to summarize structure and texture features

Due to collinearity among the structural and texture variables, we conducted a factor analysis using the principal components method to consolidate the variables into their main uncorrelated factors. For easier interpretation of factors, we applied Orthogonal Varimax rotation to maximize the loading of each variable on one factor while minimizing it in others. Figure 4.5 shows the methodology of this approach. Then, we used stepwise regression (backward, forward and both) based in the Akaike Information Criterion (AIC) as the variable selection criteria for model specification using R software (R Core Team, 2013) and ran an OLS regression, the diagnostics for spatial dependence and the corresponding spatially adjusted regression. We excluded one analytical region at the city's border (ID = 94) because it included large portions of unoccupied green spaces that were not part of the urban fabric. The variable S.POOL.DENS was also excluded from this analysis due to its high correlations with structure and texture variables (see figure 4.2, second row).

The principal components factor analysis of image structure and texture variables resulted in four retained factors with eigenvalues above 1. These four factors account for 95% of the variance for the original 21 variables. Table 4.6 shows the rotated factor analysis results. Factor 1 accounts for 57% of the variance among these variables. It was labeled *overall complexity (OC)* as it captures most of the structural variables with the exception of RSF and VFM as well as most of the texture variables with the exception of uniformity, variance and covariance. Factor 2 accounts for 17% of the variance and was labeled *variance* as it captured the texture variables VARIAN and

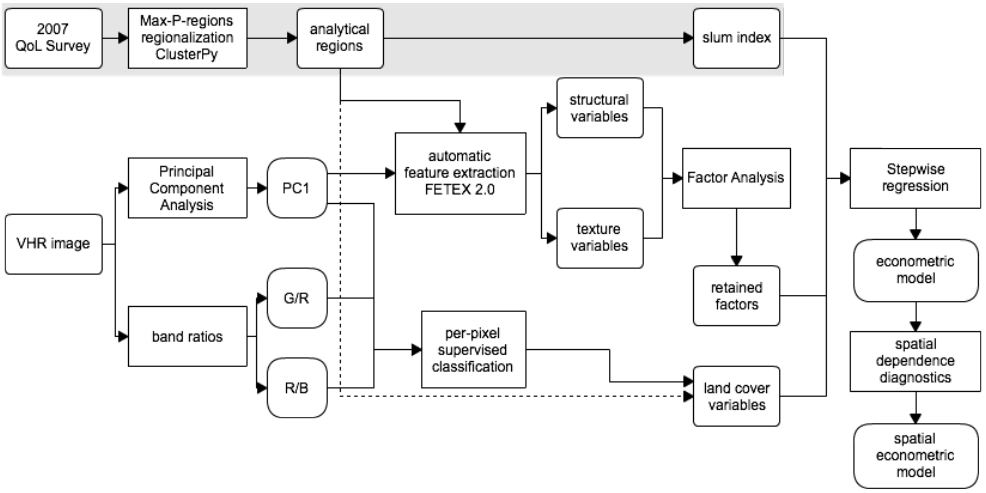

Figure 4.5: Steps for model specification of slum index as a function of land cover variables and summarized structure and texture variables from a VHR image of the urban scene.

COVAR. The structural variables RSF and VFM were captured by factor 3, which accounts for 15% of the variance. Because RSF pertains to the variability of intensity values at short distances and VFM is related to the variation in changes of the image intensity values as a function of distance (Balaguer et al., 2010), we labeled factor 3 *variation of heterogeneity as a function of the distance (VHD)*. In this context, factor 3 represents the diversity of landscape elements, and we interpret this factor as a local measure, which complements the global aspects already captured in the OC factor. Factor 4 accounts for 6% of the variance and captured only one texture variable, *uniformity*.

The stepwise regression procedure involves selecting the variables that minimize the model's AIC value through progressive steps, wherein improvements are made to the fitness of the model as each variable is included (forward option) or excluded (backward option) from the initial model. We run stepwise regressions through forward selection, backward elimination and bidirectional elimination using R software and obtained the same results. The variables that were included in the model are the percentage of impervious surfaces (P.IMP.SURF), the fraction of clay roof-cover over impervious surfaces (F.CLAYR.IMPS), the overall complexity factor (OC), and the variation of heterogeneity as a function of the distance factor (VHD). This linear model yields a multiple R^2 of 0.59 and an adjusted R^2 of 0.58, which indicates that these four variables together can explain up to 58% of the variability of the Slum Index in Medellin. Table 4.7 shows a summary of the model's parameters estimated using

Table 4.6: Rotated factor loadings (orthogonal Varimax rotation) of image structure and texture variables.

Variable	Factor 1: *Overall Complexity (OC)*	Factor 2: *Variance*	Factor 3: *Variation of heterogeneity as a function of distance (VHD)*	Factor 4: *Uniformity*
RVF	**-0.6776**	0.5939	0.3585	0.1426
RSF	-0.0613	-0.0453	**0.9596**	0.0973
FDO	**0.9671**	-0.0581	-0.2174	-0.0261
SDT	**0.8796**	0.1118	0.4252	0.0448
MFM	**0.9876**	0.0837	-0.0453	-0.0570
VFM	-0.2237	0.3901	**0.8681**	-0.0340
DMF	**0.9843**	0.1156	0.1156	-0.0617
RMM	**-0.6349**	0.4211	0.4938	-0.0544
SDF	**-0.9806**	0.0894	0.1219	0.0143
AFM	**0.9950**	0.0412	-0.0453	-0.0410
SKEWNESS	**0.7958**	0.3833	-0.3372	-0.1677
KURTOSIS	**-0.7338**	-0.5810	0.2399	0.1552
UNIFOR	-0.1785	0.3127	0.0117	**0.9072**
ENTROP	**0.8269**	0.1594	-0.2220	-0.4743
CONTRAS	**0.9324**	-0.0576	-0.3315	-0.0327
IDM	**-0.8896**	0.0873	0.1759	0.3158
COVAR	0.1014	**0.9733**	0.0968	0.1397
VARIAN	0.1880	**0.9632**	0.0654	0.1359
CORRELAT	**-0.6808**	0.6199	0.3517	0.0299
MEAN EDG	**0.9294**	-0.0378	-0.2767	-0.1728
STDEV EDG	**0.7039**	0.0208	-0.4758	0.0775

Note: Numbers in bold represent the highest loading of each variable on one factor.

OLS and the results of several tests for the normality of errors, heteroskedasticity and the specification robust test as well as the Lagrange Multiplier tests.

The non-significant values of the Breusch-Pagan, Koenker-Bassett and White tests indicate that there was no evidence of heteroskedasticity in the model. The robust version Lagrange Multiplier tests indicate the presence of spatial autocorrelation in the form of a spatially lagged dependent variable. We adjusted this using a spatial lag model that included the lagged dependent variable (W Slum Index). The highly significant value of the spatial lag variable coefficient indicates that the spatial autocorrelation was properly addressed in the spatial lag model, and a careful examination of the spatial lag model residuals did not show signs of remaining spatial patterns. The Likelihood Ratio test, the AIC and the Schwarz criterion all indicate a better

Table 4.7: Multivariate OLS model of Slum Index as a function of summarized remote sensing variables. Spatial unit: analytical regions. N = 138. Contiguity matrix: 1,700 m distance band, row-standardized weights.

Exogenous variables	Coefficients
Constant	0.9005***
P.IMP.SURF	-0.0042***
F.CLAYR.IMPS	-0.5492***
OC	0.0888***
VHD	-0.0307***
R^2	0.59
Adjusted R^2	0.58
Akaike information criterion	-228.949
Schwarz criterion	-214.312
Multicollinearity Condition Number	17
Test on normality of errors:	
Jarque-Bera test	1.25
Diagnostics for heteroskedasticity	
Breusch-Pagan test	1.29
Koenker-Bassett test	1.68
Specification robust test:	
White	13.18
Diagnostics for spatial dependence:	
Lagrange Multiplier - lag	11.07***
Robust LM - lag	4.78**
Lagrange Multiplier - error	6.63***
Robust LM - error	0.35

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

fit of the spatially adjusted regression over its non-spatial counterpart, and the non-significant value of the Breusch-Pagan test again indicates that heteroskedasticity is not present in the spatial model. Table 4.8 shows the spatial lag model coefficients and the Akaike information criterion, the Schwarz criterion and the results of the Breusch-Pagan test for heteroskedasticity and the Likelihood Ratio test.

Table 4.8: Spatial Lag model of Slum Index as a function of summarized remote sensing variables. Spatial unit: analytical regions. N = 138. Contiguity matrix: 1,700 m distance band, row-standardized weights.

Exogenous variables	Coefficients
Constant	0.6523***
P.IMP.SURF	-0.0031***
F.CLAYR.IMPS	-0.4608***
OC	0.0650***
VHD	-0.0286***
W Slum Index	0.3126***
Pseudo R^2	0.62
Akaike information criterion	-236.098
Schwarz criterion	-218.534
Diagnostics for heteroskedasticity	
Breusch-Pagan test	1.22
Likelihood Ratio test	9.15***

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

4.3 Discussion

We found spillover effects regarding the Slum Index in Medellin, which is in agreement with previous works that have found the same effect for different poverty measures (Duncan et al., 2012; Holt, 2007; Okwi et al., 2007; Orford, 2004; Sowunmi et al., 2012; Voss et al., 2006).

The results of the statistical analysis using VHR image-derived variables directly indicate that the most important predictors of slum index at the analytic region level in Medellin 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 4.2, Model 2 and Model 3, and table 4.3, SLM 2 and SLM 3), or from the texture variable set (skewness or kurtosis, table 4.2, Model 5 and Model 6, and table 4.3, SLM 5 and SLM 6), 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, 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.

Since ancient times, roofing material and style choice has been viewed as a reflection of the owner's wealth and power (Fiumi, 2012). The land cover variables such as the percentage of impervious surfaces and fraction of clay roof cover over impervious surfaces are highly significant in all models, which means that the association between roofing materials and urban poverty is present in Medellin. Lower values of clay roof cover are related to more deprived neighborhoods in Medellin. This is the case because clay roofs are more expensive to install and maintain than other roofing materials such as industrial roof tiles made from zinc or asbestos and because clay roofs are rarely present in the most deprived areas of the city. However, this does not mean that a house with a different type of roof is more deprived than a house constructed of clay roof tiles. The association that we found holds at the analytical region level, which can include hundreds of houses or buildings. Moreover, the presence and abundance of orange and red clay tile roofs in a city is closely related to its cultural heritage as well as the environmental conditions of each city. This is the case because the characteristics of tile roofs constructed depends on the availability and mineral composition of clays and sands used as the primary source of building materials that are located within a city's proximity. One can expect that this relationship could be similar in other Latin American cities that share a similar cultural heritage and that also have close access to orange and red clays.

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 socioeconomic 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 structure and texture descriptors that were used to characterize the spatial pattern of the urban layout proved to be useful for Slum Index

estimation, and these descriptors are easily computed from VHR imagery and polygon boundary data using FETEX 2.0. The use of image texture measures for Slum Index estimation has been reported for Accra, Ghana (Weeks et al., 2007) and their use for slum detection and mapping has been reported in other different cities around the world such as Campinas and Rio de Janeiro in Brazil (Barros and Sobreira, 2005; Hofmann et al., 2008); Cape Town, South Africa (Hofmann, 2001); Casablanca, Morocco (Rhinane, 2011); and Ahmedabad, Delhi, Pune and Hyderabad in India (Kit et al., 2012; Kohli et al., 2013; Niebergall et al., 2007; Shekhar, 2012).

To assess the performance of this approach using VHR image-derived variables directly, we took the best performing spatial regression (SLM 3, from table 4.3) and we built a box map of the residuals in order to examine the location of the analytical regions with the greatest errors (figure 4.6). The box map shows that there are no spatial pattern in this residuals, as indicated by the Anselin-Kelejian test reported in table 4.3, 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.

The results of the statistical analysis using Factor Analysis to summarize structure and texture variables corroborate the results obtained using the remote sensing variables directly. However, the resulting models using Factor Analysis and Stepwise Regression techniques show some improvements when compared to the models using the structure and texture variables directly, such as minimizing multicollinear-

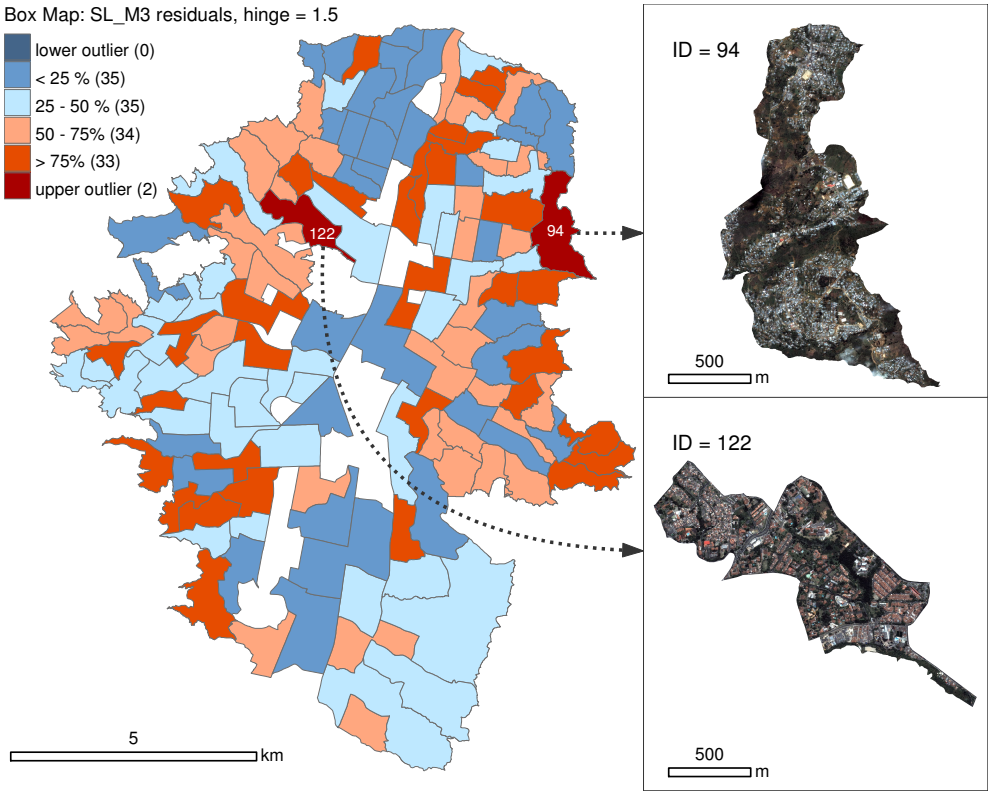

Figure 4.6: Box map of SML 3 residuals and image details of analytical regions ID = 94 and ID ID = 122.

ity, ensuring normality of errors, minimizing spatial heterogeneity, and easiness of interpretation. When using the retained factors, the most important remote sensing predictors of the Slum Index at the analytical region level for Medellin include the percentage of impervious surfaces, the fraction of clay roofs over impervious surfaces, the overall complexity factor and the variation of heterogeneity as a function of the distance factor. These variables describe aspects of the land cover composition within the analytical region as well as the spatial pattern of the urban layout in terms of texture and structure.

The model's coefficients indicate a negative association between the Slum Index and the percentage of impervious surfaces, the fraction of clay roof cover over impervious surfaces and the VHD factor. A positive association was found between the Slum Index and the OC factor at the analytical region level. These associations were expected. The Slum Index is higher in the analytical regions with a lower percentage of impervious surface. However, this is only the case when the impervious surface is mostly comprised of surfaces other than clay roof and when the urban layout has

both higher overall complexity and lower diversity of landscape elements within the analytical region.

Higher values of the fraction of clay roof over impervious surfaces are related to wealthier neighborhoods in Medellin. Whereas the most important remote sensing predictor of the Slum Index in Accra is the amount of vegetation (Stoler et al., 2012; Weeks et al., 2007), this variable does not perform in the same way for Medellin. In fact, this variable is barely related to the Slum Index at the analytical region level for this city, with a Pearson's correlation coefficient of 0.14. This may be explained by differences in climate conditions, which influences vegetation abundance in a given location, as well as by cultural differences between the cities' inhabitants.

The structure and texture descriptors summarized in the factors indicate that the Slum Index is higher in the analytical regions that registered simultaneously higher overall complexity (OC) and lower variation of heterogeneity as a function of distance (VHD). Figure 4 shows square image tiles of 500 meters in different areas of the city with low and high values of OC and VHD factors. The image tiles are organized according to increasing values of each factor from left to right to show changes in the urban scene as the value of the factors increases. From this figure, the reader can intuitively corroborate the associations of the Slum Index with OC and VHD values found in the statistical models.

High OC values in the image denote that different surface covers exist within close proximity to one another as well as high edge density and high local heterogeneity, which are typical characteristics of the city's most deprived neighborhoods. In these neighborhoods, building density is high, but the dwelling units are rather small with varying roofing materials and located in close proximity to each other. This increases the local heterogeneity and overall complexity of the image. Low overall complexity in the urban layout is associated with analytical regions that have large, homogenous surfaces such as wide roads and avenues, open green spaces, industrial areas that have large buildings with homogeneous roofs and large detached houses surrounded by vegetation. The lowest Slum Index values in this city are located in analytical regions where the existence of large houses surrounded by vegetation is most prevalent. High VHD values are associated with high landscape fragmentation due to a higher diversity of landscape elements. Common examples include residential areas with bigger houses or buildings and the presence of green areas and wide roads surrounded by vegetation, parks, pools, and other urban elements. Low VHD is associated with more crowded and organic urban layouts where small and diverse surfaces exist in close proximity, registering high variability over short distances but lower variance at a higher scale (see Figure 4.7, VHD tiles).

In Medellin, organic urban layouts are primarily, but not solely, the result of rural-

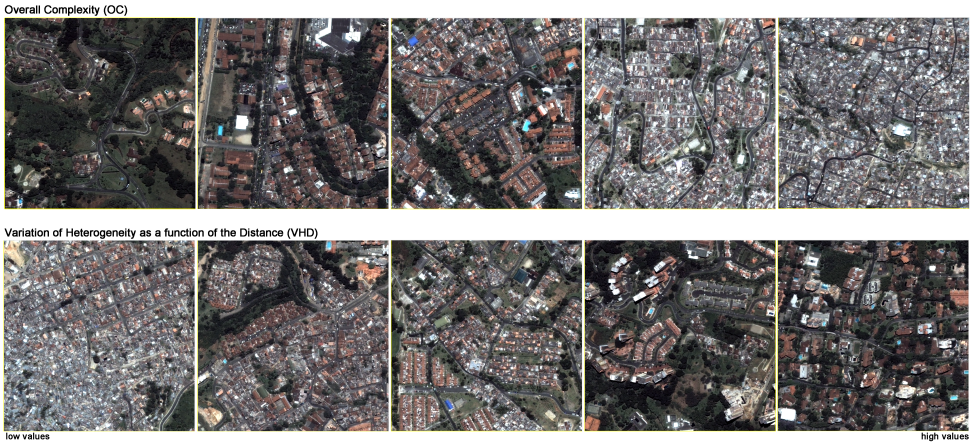

Figure 4.7: Square image tiles showing urban areas (500 x 500 meters) that have different values of overall complexity (OC) and variation of heterogeneity as a function of distance (VHD) factors. The tiles are organized from left to right according to increasing values of each factor.

to-urban migration to unplanned urban spaces (Duque et al., 2013). Political violence, an ongoing internal armed conflict with guerrillas and paramilitary groups, and the drug war in the rural areas of the country have caused important rural-to-urban migrations in Colombia (Ibáñez and Vélez, 2008). These migrations have impacted urban population growth rates since the mid-20th century in Medellin. Concentrated in the northern periphery of the city, this demographic growth has exceeded the capacity of the local government to deliver services and infrastructure and caused the appearance of slums (Duque et al., 2013). The natural population increase combined with structural barriers to socioeconomic advancement, such as poor access to health, education, and work, has also contributed to the perpetuation of slum-like neighborhoods. As these neighborhoods were born from self-build processes without any planning, they are characterized by organic layouts and almost no open spaces. After these areas were consolidated, local authorities provided access to electric energy, piped water, and the sewer system to replace illegal and unsafe electric and water connections while also seeking to decrease losses and increase collection of these services.

We followed the approach of (Orford, 2004) to assess the usefulness of this model for policy making. We used the local Moran I coefficient, a local indicator of spatial association (Anselin, 1995), to identify areas of concentrated poverty and affluence within the city and to see if the estimated Slum Index from remote sensing can identify the same areas than the survey-based Slum Index. Figure 4.8 shows a comparison of local Moran's I maps of the survey-based Slum Index versus the estimated Slum Index

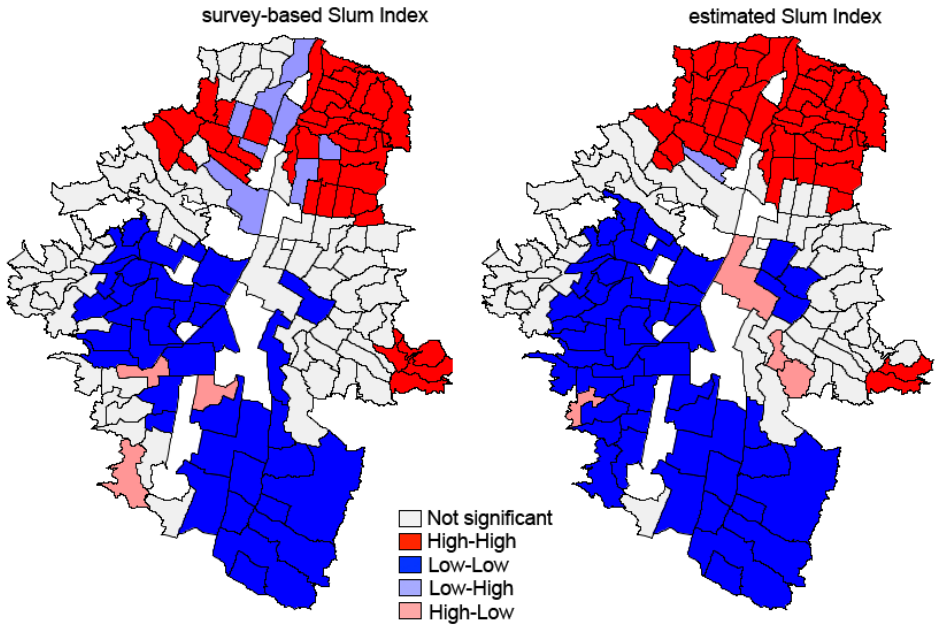

Figure 4.8: LISA maps of Survey-based vs. estimated Slum Index with remote sensing-derived variables. Spatial clusters of high values, showed in red, indicate areas of concentrated poverty; spatial clusters of low values, showed in blue, indicate areas of concentrated affluence.

with the remote sensing-derived variables; both created using OpenGeoda. Although there is not an exact match, the maps show a general good agreement between the High-High and Low-Low areas, which means that the Spatial Lag model is picking the spatial concentration of the poorest and wealthier areas of the city in a rather good way. The Spatial Lag Model residuals are normally distributed, with a mean of 0 and a standard deviation of 0.098. This means that rather than skewing toward low or high values, the model fitted quite well throughout the city.

4.4 Chapter conclusions

This chapter seeks to estimate the Slum Index for one Latin American city solely using remote sensing data. The usefulness of remote sensing data for estimating the Slum Index at the intra-urban level was previously tested in the case of Accra, Ghana. The findings of the present work corroborate these earlier results in a city with a different geographical and economic setting. In this work, we tested a wider set of remote sensing variables related not only to land cover composition but also to the spatial

pattern characteristics of the urban layout.

The presence of spatial autocorrelation with regard to the Slum Index in Medellin indicates that poverty in this city could be tackled through a policy of strategically located investments and programs designed to boost economic development in some of the city's neighborhoods as their outcomes can be expected to affect not only the neighborhoods where the investment is located but surrounding neighborhoods as well. From this point of view, municipal public investments made over the last decade including public library parks and high quality schools seem to be an appropriate strategy for reducing poverty in this city and should be encouraged and improved.

We highlight that the input remote sensing data used in this paper are rather simple. For these data, the most important requirement is a very high spatial resolution to capture the differences in urban layout complexity and heterogeneity. We did not use near-infrared spectral information or other special spectral features from the satellite imagery. The structure and texture variables were calculated from the first principal component band after processing the red, green and blue bands of the color image. Any aerial color image with a very high spatial resolution could serve this purpose. This creates the possibility for rapid Slum Index estimation and intra-urban poverty pattern analysis from remote sensing data given that similar imagery could easily be obtained from Internet services such as Google Earth, Yahoo, or Microsoft Bing Imagery.

The model developed so far for Medellin is based on the city's unique characteristics. Future research should address two main issues: to test whether the same variables relate to intra-urban poverty measures in a similar way in other cities around the world and to verify that this approach provides consistent results over time. When these aspects are resolved, a similar approach could be used for the following purposes in urban settings with sparse socioeconomic data: to lower the cost of socioeconomic surveys by developing an econometric model from a sample and applying that model to the rest of the city and to perform intercensal or intersurveys estimates of intra-urban Slum Index maps.

Chapter 5

Crime and the urban layout

This chapter is a compilation of some sections of the paper Patino, J. E., Duque, J. C., Pardo-Pascual, J. E, Ruiz, L. A. (2014). Using remote sensing to assess the relationship between crime and the urban layout. Applied Geography, 55C, 48-60. doi:10.1016/j.apgeog.2014.08.016.

Previous studies that addressed the relationship between place and crime using remote sensing focused in the quantification of night-time lighting Weeks (2003); some vegetation measures such as vegetation coverage, NDVI and LAI indices (Donovan and Prestemon, 2012; Kuo and Sullivan, 2001; Troy et al., 2012; Wolfe and Mennis, 2012); and land use diversity indicesSparks (2011). Previous research that analyzed the influence of the spatial pattern of the built environment on crime crime have used spatial pattern descriptors computed from expensive and time consuming field surveys and appraisals. This chapter shows an example of the use of land cover and spatial pattern descriptors derived from VHR imagery to analyze the relationship between the spatial pattern of the built environment and crime at a lower cost than field surveys and appraisals.

5.1 Crime and the urban environment

Macro-level or social ecology theories of crime are focused in how the characteristics of geographical areas influence crime rates, i.e., how crime rates vary by ecological units rather than how the characteristics of individuals relate to their involvement in criminal behavior (Cullen and Agnew, 2011). Since the 19th-century, early French social ecologists Guerry and Quetelet were interested in explaining differences in community crime levels in terms of the varying social conditions of the resident population (Anselin et al., 2000). Social disorganization theorists linked high crime rates to neigh-

borhood characteristics such as low income, high unemployment, ethnic heterogeneity, and residential mobility, and the impact these characteristics have on the community ability to exert social control for the prevention of crime (Shaw and McKay, 1942). The link between place and crime is also central to the perspective of Routine Activities theory, which states that criminal acts require convergence in space and time of a motivated offender, a suitable target, and the absence of capable guardians against crime, and relates crime patterns to the everyday patterns of social interaction (Cohen and Felson, 1979).

Urban neighborhoods differ by the types of households they contain and by the types of physical environments they provide (Roncek, 1981), and these physical environments can influence the likelihood of crime occurrence (Taylor and Harrell, 1996). Environmental criminology studies that have addressed the spatial distribution of crime demonstrate that certain land uses, i.e., the concentration of commercial and residential areas, are associated with crime hot spots (Anselin et al., 2000). Land use, sociodemographics, open spaces, and residential density affect the types and intensity of crimes (Salleh et al., 2012). The broken windows theory states that visible cues of physical and social disorder in a neighborhood can lead to an increase in crime (Wilson and Kelling, 1982). Physical disorder refers to the presence of abandoned and deteriorated houses and vehicles, graffiti-painted walls and litter, and this phenomenon is also related to social disorder and the occurrence of minor offenses within a neighborhood. Similar ideas support planning approaches that seek to deter criminal behavior by designing ordered, highly visible and easily defensible urban spaces, such as crime prevention through environmental design (CPTED), and New Urbanism (Cozens, 2008). DeFrances and Titus (1993) found that whether a burglary was completed or aborted was associated with neighborhood disorganization and home occupancy, and Keizer et al. (2008) tested the broken windows theory in six field experiments and reported that when people observe that others have violated certain social norms or legitimate rules, they are more likely to violate other norms and rules, causing disorder to spread and escalate.

Previous research addressed the relationships between crime in urban settings and physical and environmental variables such as altitude and slope (Breetzke, 2012), temperature and weather conditions (Anderson and Anderson, 1984; Anderson et al., 2000; Butke and Sheridan, 2010; Carlsmith and Anderson, 1979; Cohn and Rotton, 2000; DeFronzo, 1984; Field, 1992; Salleh et al., 2012; Sorg and Taylor, 2011), vegetation (Donovan and Prestemon, 2012; Kuo and Sullivan, 2001; Troy et al., 2012; Wolfe and Mennis, 2012), land use (Kurtz et al., 1998; Wilcox et al., 2004), night-time lighting (Weeks, 2003), and the spatial pattern of the built environment (Browning et al., 2010; Foster et al., 2010; Matthews et al., 2010; Shu and Huang, 2003; Taylor and

Harrell, 1996). Cerdá et al. (2012) found reduced violence outcomes in some neighborhoods of Medellin, Colombia, that were the focus of an urban development plan that included a transportation system and public space improvements such as additional lighting, new pedestrian bridges and street paths, a library park, and buildings for schools and recreational centers.

Spatial effects affect crime as a phenomenon -spatial dependence and spatial heterogeneity- and these effects should be considered when investigating the relationships between crime and place (Anselin, 1988; Anselin et al., 2000). However, few studies have considered spatial effects in the literature on crime and the physical features of urban places. Another undertaken issue is the potential effects of selecting improper spatial units of analysis, such as spurious autocorrelation (Anselin, 1988) and the "ecological fallacy" (Openshaw, 1984; Robinson, 1950). Based on the hypothesis that the surface appearance of a settlement is the result of the human population's social and cultural behavior and interactions with the environment, and that people living in urban areas with similar physical housing conditions have similar social and demographic characteristics (Jain, 2008; Taubenböck et al., 2009b), the work presented here uses remote sensing data to investigate the relationship between homicide and the urban environment while controlling for potential socioeconomic confounders in Medellin, Colombia. This city is well suited for such a study because it is affected by high crime rates and because of the high variability in the urbanism and socioeconomic conditions of its neighborhoods; these factors make Medellin a useful laboratory for studying urban issues from the spatial analysis perspective.

5.2 Modeling approach

We built an econometric model of homicide rates as a function of socioeconomic factors to assess if the remote sensing variables added to the model were significant and had the expected sign. The dependent variable was the average homicide rate per 100,000 people for the years 2010 and 2011 at the analytical region level, and the predictors were socioeconomic variables from the years 2005 and 2009 related to the homicide rate at the intra-urban level according to the literature. We used this time lag of one year between the independent variables and the dependent variable to avoid endogeneity effects. Figure 5.1 outlines the methodology.

5.2.1 A classic spatial model of homicide

Previous research related the concentration of interpersonal violence to neighborhoods characterized by poverty, the segregation of minority groups, and the presence of

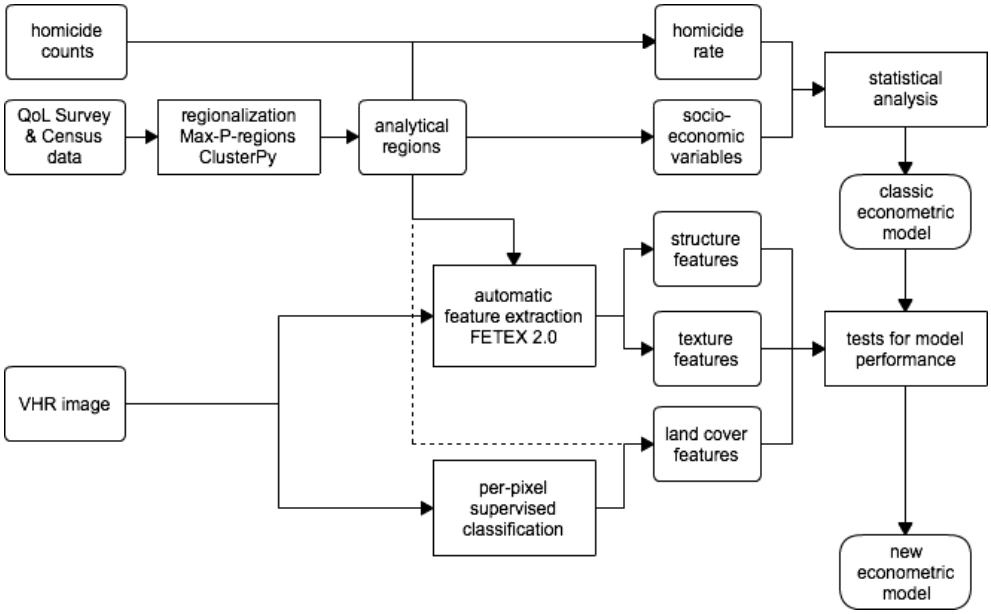

Figure 5.1: Main steps of modeling the homicide rate as a function of socioeconomic and remote sensing variables at the analytical region level.

single-parent families (Morenoff et al., 2001). The presence of young males, income inequality, unemployment, concentrated disadvantage, population density, residential instability, concentrated immigration, family disruption, educational attainment and ethnicity have been used in homicide econometric models in different cities around the world (Browning et al., 2010; DeFronzo, 1984; Kawachi et al., 1999; Kennedy et al., 1998; Menezes et al., 2013; Morenoff et al., 2001; Wang and Arnold, 2008; Ye and Wu, 2011). However, it is possible that some of the variables that explain intra-urban variations in the homicide rate in one city do not perform in the same way in different cities.

Ethnicity and immigration are of little importance in Medellin because the ethnic composition is very homogeneous throughout the city, with more than 90% of its inhabitants considering themselves white descendants, and these variables show little spatial variation across the city. In Colombia, the homicide rates at the city level have been related to the variables discussed above as well as to the justice system's inefficacy, the war on drugs, and the presence of illegal armed groups and gangs (Giraldo Ramírez, 2010; Levitt and Rubio, 2000; Medina and Martínez, 2003; Medina et al., 2011; Sánchez Torres and Núñez Méndez, 2001). The justice system's inefficacy and the war on drugs cannot be taken into account because they are constants throughout the city. Medina et al. (2011) reported that 75% of murder victims in Medellin in the

first six months of 2009 were assassinated in their own neighborhoods, which is related to the territorial control struggle between gangs. The presence of gangs does show intra-urban variations and was considered in the econometric model of the homicide rate at the intra-urban level in Medellin. The estimated model has the following form:

$$Y = c + X\beta + \epsilon \tag{5.1}$$

Where Y is an n by 1 vector of the dependent variable: the average homicide rate per 100,000 people for the years 2010 and 2011; X is an n by k matrix of observations on explanatory variables from the years 2005 and 2009, with an associated k by 1 vector of regression coefficients (β); and ϵ is an n by 1 vector of error terms. We built the classic model according to the following apriori hypothesis. All variables were computed at the analytical region level from the 2009 Quality of Life Survey, with the exception of population density, which was computed from the 2005 Census:

- Downtown (DOWNTOWN): dummy variable equal to 1 for the downtown area of the city. The homicide rates in the downtown area were expected to be unusually high because of the high floating population that is not recorded as living downtown in the Census.
- Population density (POP.DENS): expressed as people per hectare. A positive association was expected between population density and the homicide rate.
- Gang presence (GANGS): measured as the fraction of households that had been victims of criminal gangs. High gang presence in an area was expected to be related to high homicide rates.
- Residential instability (NO.OWNERS): measured as the fraction of households that did not own the house where they lived. Areas with high residential instability will have less social cohesion and more anonymity, which was expected to be associated with high homicide rates.
- Income inequality (GINI.INC): measured using the Gini coefficient of income, which was calculated from household income data within analytical regions. A positive association was expected.
- Young males (YOUNG.MALES): measured as the fraction of males between 20 and 24 years old as a proxy for the presence of young males in the neighborhood. The presence of young males was expected to have a positive association with the homicide rate.
- Family disruption (DIVORCED): measured as the fraction of households with divorced parents. A positive relationship with the homicide rate was expected.

- Unemployment (UNEMPL): computed as the fraction of unemployed people in the neighborhood. High unemployment was expected to be related to high homicide rates.

Table 5.1 presents the summary statistics of these variables and Figure 5.2 shows their corresponding maps at the analytical region level. The modeling process was implemented in OpenGeoda software (Anselin et al., 2006). We ran an ordinary least squares (OLS) regression and the diagnostics for spatial dependence using a Rook type contiguity matrix with row-standardized weights. We followed the hybrid specification search strategy based on robust Lagrange multiplier tests because the standard versions were both significant (Anselin, 2005; Florax et al., 2003). The robust Lagrange multiplier tests indicated the presence of residual spatial autocorrelation that had to be adjusted using a spatial error model. The results of the log likelihood ratio test, as well as the Akaike information criterion and Schwarz criterion values, showed the superiority of the spatially adjusted regression over its OLS counterpart, and according to Anselin (2005), in finite samples, the Wald test (W), the likelihood ratio test (LR) and the Lagrange multiplier test (LM) based on OLS residuals should follow the order: $W > LR > LM$, as they did in this model. Table 5.2 shows the information from both models.

Table 5.1: Summary statistics of the exogenous socioeconomic variables.

Variable	Mean	Standard deviation
POP.DENS	301.46	152.16
GANGS	0.11	0.13
NO.OWNERS	0.40	0.09
GINI.INC	0.43	0.12
YOUNG.MALES	0.10	0.02
DIVORCED	0.11	0.03
UNEMPL	0.07	0.03

5.2.2 Integrating urban fabric descriptors from VHR imagery to the classic model

We ran spatially adjusted regressions with the socioeconomic variables plus one remote sensing variable at a time to assess which included variables were significant without compromising model performance. Table 5.3 shows the summary statistics of the analyzed remote sensing variables. We considered that model performance was not compromised if the new specification fulfilled these conditions: the included variable had to be significant in the model specification, the inclusion of the variable could not increase the models multicollinearity (otherwise, the information it contributed to the model could have already been accounted for with the socioeconomic

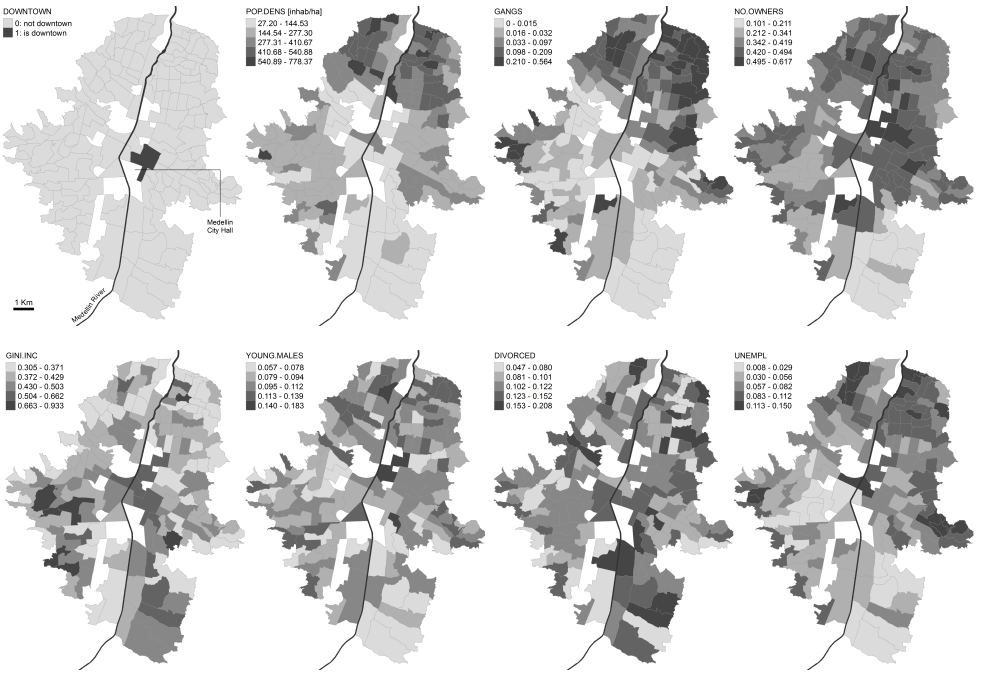

Figure 5.2: Choropleth maps of the socioeconomic variables included in the model specification of the average homicide rate per 100,000 people (2010 - 2011) in Medellin at the analytical region level. All variables classified into intervals using the Natural Breaks Method.

variables), and the information criteria (Akaike or Schwarz) and the likelihood ratio test had to indicate the improved fit of the new model over the classic specification. Table 5.4 summarizes the results of this exercise: an asterisk shows that the included variable was statistically significant at the 5% level in the spatial error model, that multicollinearity was not a serious concern, and that the log likelihood ratio test, the Akaike information criterion or the Schwarz criterion indicated improved fit over the classic model.

The remote sensing variables whose inclusion fulfilled all conditions were P.OTHER.IMP.S and F.CLAYR.IMPS from the land cover features group; SDT and SDF from the structure group; and UNIFOR from the texture group. These variables informed about the land cover composition, the spatial arrangement of the urban layouts elements (structure group), and the uniformity of the urban fabric within the analytical regions (texture group). Figure 5.3 shows the analytical region maps of the remote sensing variables included in the models. Table 5.5 shows the corresponding spatial error models.

Figure 5.3: Choropleth maps of the best performing remote sensing variables added to the classic model specification of the average homicide rate per 100,000 people (2010 - 2011) in Medellin at the analytical region level. All variables classified into intervals using the Natural Breaks method.

Table 5.2: Classic model specification of the average homicide rate per 100,000 people (2010 - 2011) as a function of the socioeconomic variables. Coefficients and significance levels for the OLS and spatial error models. Spatial unit: analytical regions. N = 136.

Exogenous variables	OLS model	Spatial Error model
Constant	-18.4223	-1.8289
DOWNTOWN	410.4631***	423.0517***
POP.DENS	0.1419***	0.1330***
GANGS	95.7671**	51.8867*
NO.OWNERS	249.5371***	245.0647***
GINI.INC	-50.7756	-51.8295**
YOUNG.MALES	201.2657	158.1371
DIVORCED	102.2078	112.8527
UNEMPL	228.0302	137.8414
Lambda		0.6907***
R^2	0.51	0.71
Adjusted R^2	0.48	
Multicollinearity Condition Number	24.88	
Koenker-Bassett test	7.24	
Diagnostics for spatial dependence:		
Lagrange Multiplier - lag	36.4435***	
Robust LM - lag	0.4038	
Lagrange Multiplier - error	47.8951***	
Robust LM - error	11.8554***	
Wald test		107.37
W > LR > LM		*yes*
Akaike information criterion	1,453.29	1,402.92
Schwarz criterion	1,479.51	1,429.13
Log Likelihood	-717.65	-692.46

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

We then tested the inclusion of combinations of two remote sensing variables from the previous selection in the classic model to assess whether they captured complementary information about the urban fabric or informed about similar aspects. The three combinations of variables that showed an improvement over the classic model are presented in Table 5.6. A discussion of these results is presented in the next section.

Table 5.3: Summary statistics of remote sensing derived variables used in the analysis.

Group	Variable	Mean	Standard deviation
Land cover	P.VEG	36.01	13.59
features	P.IMP.SURF	54.73	13.30
	P.SOIL	0.13	2.66
	P.CLAY.ROOFS	13.30	9.23
	P.OTHER.IMP.S	41.43	14.39
	F.CLAYR.IMPS	0.25	0.16
Structure	RVF	12.26	2.94
features	RSF	1.80	0.02
	FDO	117.68	20.09
	SDT	15.33	3.18
	MFM	828.19	129.96
	VFM	1.45	0.22
	DMF	681.39	106.08
	RMM	1.41	0.04
	SDF	-341.09	72.74
	AFM	5,274.48	875.59
Texture	SKEWNESS	-0.75	0.24
features	KURTOSIS	0.30	0.63
	UNIFOR	0.005	0.002
	ENTROP	2.61	0.07
	CONTRAS	15.30	2.66
	IDM	0.43	0.04
	COVAR	97.93	14.38
	VARIAN	105.58	14.25
	CORRELAT	0.93	0.02
	MEAN.EDG	8.39	0.99
	STDEV.EDG	7.45	0.52

Table 5.4: Summary of the results of including the remote sensing derived variables in the classic spatial error model. An asterisk indicates that including the variable fulfilled the criteria.

Group	Variable	Significant	Multicollinearity	LL Ratio test	Akaike inf. c.	Schwarz c.
Land cover	P.VEG					
features	P.IMP.SURF		*			
	P.SOIL		*			
	P.CLAY.ROOFS		*			
	P.OTHER.IMP.S	*	*	*	*	*
	F.CLAYR.IMPS	*	*	*	*	*
Structure	RVF	*		*	*	
features	RSF					
	FDO	*		*	*	
	SDT	*	*	*	*	
	MFM	*		*	*	*
	VFM					
	DMF	*		*	*	*
	RMM					
	SDF	*	*	*	*	
	AFM	*		*	*	*
Texture	SKEWNESS					
features	KURTOSIS		*			
	UNIFOR	*	*	*	*	*
	ENTROP	*		*	*	
	CONTRAS	*			*	
	IDM	*		*	*	
	COVAR					
	VARIAN					
	CORRELAT				*	
	MEAN.EDG	*		*	*	
	STDEV.EDG				*	

Table 5.5: Spatial error models of the average homicide rate per 100,000 people (2010 - 2011) as a function of the socioeconomic variables and one remote sensing variable. Spatial unit: analytical regions. N = 136. Models 1 and 2 include socioeconomic variables plus one land cover variable at a time. Models 3 and 4 include socioeconomic variables plus one structure variable at a time. Model 5 includes socioeconomic variables plus one texture variable.

Exogenous variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-8.3806	33.2173	-26.5559	-25.0011	26.5149
DOWNTOWN	402.0680***	412.7394***	419.9667***	411.4869***	419.3412***
POP.DENS	-0.1700***	-0.1508***	-0.1433***	-0.1702***	-0.1487***
GANGS	59.0273**	32.3126	48.9399	57.0203**	51.3281*
NO.OWNERS	223.6354***	259.6264***	241.1003***	217.4753***	211.6481***
GINI.INC	-49.5049**	-42.1245	-59.3078**	-55.7996**	-54.4743**
YOUNG.MALES	91.4376	97.9944	157.3962	140.8751	152.4899
DIVORCED	96.1949	82.9640	88.6220	112.6263	166.3085*
UNEMPL	105.9465	-3.1900	90.3745	149.1204	237.9309*
P.OTHER.IMP.S	0.8646***				
F.CLAYR.IMPS		-72.9792**			
SDT			2.5733**		
SDF				-0.1400**	
UNIFOR					-4,271.1080**
Lambda	0.7021***	0.7005***	0.7091***	0.7008***	0.6972***
R^2	0.72	0.72	0.72	0.72	0.72
W > LR > LM	*yes*	*yes*	*yes*	*yes*	*yes*
Akaike inf. criterion	1,398.72	1,399.48	1,400.77	1,400.46	1,399.97
Schwarz criterion	1,427.85	1,428.61	1,429.89	1,429.59	1,429.10
Log Likelihood	-689.36	-689.74	-690.38	-690.23	-689.99

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

Table 5.6: Spatial error models of the average homicide rate per 100,000 people (2010 - 2011) as a function of the socioeconomic variables and two remote sensing variables. Spatial unit: analytical regions. N = 136. Models 6 and 7 include socioeconomic variables plus one land cover variable and one texture variable. Model 8 includes socioeconomic variables plus one land cover variable and one structure variable.

Exogenous variables	Model 6	Model 7	Model 8
Constant	17.7203	69.8542**	10.2641
DOWNTOWN	400.1827***	407.1254***	400.0821***
POP.DENS	-0.1841***	-0.1709***	-0.1909***
GANGS	58.1735*	29.3208	37.0499
NO.OWNERS	194.8035***	223.3692***	231.1741***
GINI.INC	-52.0982**	-43.9087*	-45.9163*
YOUNG.MALES	91.1029	83.8527	76.9909
DIVORCED	145.9673	140.0600	81.4866
UNEMPL	199.3559	93.3247	1.9449
P.OTHER.IMP.S	0.8022**		
F.CLAYR.IMPS		-82.2869***	-76.4102***
SDF			-0.1481**
UNIFOR	-3,873.1150**	-4,859.0850***	
Lambda	0.7056***	0.7084***	0.7114***
R^2	0.73	0.73	0.73
W > LR > LM	*yes*	*yes*	*yes*
Akaike inf. criterion	1,396.48	1,394.81	1,396.24
Schwarz criterion	1,428.52	1,426.85	1,428.28
Log Likelihood	-687.24	-686.40	-687.12

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

5.3 Discussion

The implemented spatial error models (Table 5.2, Table 5.5, and Table 5.6) correspond to a spatial autoregressive (SAR) process (Anselin, 2005), and the highly significant spatial error coefficient (Lambda) values indicate that the spatial effects were properly accounted for in these spatial error models. The modeled spatial interaction implies that each error term was correlated with each other error term in the system, but the magnitude of the correlation decayed with distance. The error terms in the model were considered "ignored" or "unmeasurable" effects, and for the SAR specification, this means that if there were an unmeasurable neighborhood effect, the change in this effect in one location would affect all of the locations in the system following a distance decay effect (Anselin et al., 2000).

The coefficients of the classic model (Table 5.2) indicate that the homicide rates in Medellin were positively related to the presence of gangs in a neighborhood (measured as the fraction of households that had been victims of gangs), the fraction of non-owner households (proxy for residential instability), the fraction of young males, the fraction of households with divorced parents and the fraction of unemployed people. These findings are as expected and are supported by social disorganization theory (Cullen and Agnew, 2011; Shaw and McKay, 1942; Stark, 1987). The rates were negatively related to population density and to the Gini index of income within the analytical region, a result that is contrary to those from a previously reported model for Recife, Brazil (Menezes et al., 2013). In Medellin, the Gini index of income can be taken as a proxy of social mix within analytical regions because income is the most important factor for separating social groups in this city. This result indicates that the presence of different income groups in the same neighborhood could act as a deterrent for this type of criminal behavior in this city. Although empirical evidence about social mix and crime outcomes is complex and ambiguous (Atkinson, 2005), this finding is in agreement with reports of reduced crime rates in urban areas that shifted from public housing to mixed-income neighborhoods in the US (Holin et al., 2003; Popkin et al., 2004).

The negative association between the average homicide rate and the population density in Medellin also contradicted the conventional belief that the relationship should be positive. The same finding was reported for Chicago by Morenoff et al. (2001) and was explained by the depopulation of the most devastated and poorest areas of the city, which also showed high homicide rates. Cole and Marroquín Gramajo (2009) also reported a negative association between homicide rates and population density in a cross-sectional study at the country level. They found that the effect of population density on homicide rates was ambiguous at best, and they highlighted

the fact that although many theorists expect high density to be associated with high crime rates, low crime rates are found in both high- and low-density societies. In Medellin, the highest homicide rates occurred in neighborhoods with medium population density.

The associations with the socioeconomic factors remained the same when the remote sensing variables were added to the classic model, with the exception of unemployment, which changed when F.CLAYR.IMPS was included (Table 5.5, Model 2; and Table 5.6, Model 7 and Model 8). Roofing materials and styles have been seen as a reflection of the wealth and power of dwelling owners since ancient times (Fiumi, 2012), and clay roofs are more expensive to install and maintain than are other common roofing materials such as zinc or asbestos. In Medellin, the most deprived neighborhoods, where clay roofs are uncommon, showed the highest unemployment rates. This variable could be a proxy for urban poverty, which is in agreement with previously reported associations between poverty and homicide at the intra-urban level (Browning et al., 2010; Sparks, 2011; Ye and Wu, 2011) and is also supported by social disorganization theory concepts.

Previous works have reported significant relationships between measures of vegetation abundance and crime rates in a number of US cities: Chicago, Illinois (Kuo and Sullivan, 2001); Portland, Oregon(Donovan and Prestemon, 2012); Baltimore, Maryland (Troy et al., 2012); and Philadelphia, Pennsylvania (Wolfe and Mennis, 2012). The percentage of vegetation cover was not significant when it was added to the classic model, which indicates that there was no significant relationship with the homicide rates in Medellin. The best-performing variable from the land cover group was P.OTHER.IMP.S (Table 5.5, Model 1; and Table 5.6, Model 6), with a positive association with homicide rates: the average homicide rates tended to be higher in areas with higher percentages of impervious surfaces other than clay roof cover.

The best-performing structure and texture variables were SDT and SDF from the structure group (Table 5.5, Models 3 and Model 4; and Table 5.6, Model 8) and UNIFOR from the texture group (Table 5.5, Model 5; and Table 5.6, Model 6 and Model 7). The structure variable SDT provides information on the concavity level of the semivariogram at short distances, corresponding to the heterogeneity of the image objects (Balaguer et al., 2010). The SDT value increased as the image objects became more heterogeneous at short distances, i.e., where the presence of small dwellings with different roofing materials was common and the absence of homogeneous open spaces and wide roads was the rule. High SDT values were present in unplanned neighborhoods with few or no open spaces, narrow roads and irregular street networks. The association with the homicide rate was positive, which is in agreement with the CPTED concept in the sense that crowded and cluttered urban layouts are associated

with higher crime outcomes.

SDF informs about the semivariogram curvature in the interval between its first lag and its first maximum, and it represents the low-frequency values in the image (Balaguer et al., 2010). SDF informs about the spatial arrangement of the elements at short distances and provides another means of quantifying the homogeneity of the urban layout's spatial pattern. Higher SDF values were associated with regions that had open spaces, wide avenues or large industrial facilities. SDF showed a negative association with homicide rates.

The texture variable UNIFOR (uniformity) is also a measure of homogeneity in the image (Haralick et al., 1973), and high uniformity values occurred when the intensity values within analytical regions were mostly similar. The association of image uniformity with homicide rates was negative, indicating that more uniform and homogeneous urban layouts were associated with low average homicide rates.

The fraction of clay roof cover within the impervious surface cover and uniformity was the best-performing combination of remote sensing variables (Table 5.6, Model 7), as indicated by the Akaike information criterion, the Schwarz criterion and the log likelihood ratio test. When these two variables were included in the model at the same time, the associations of the other socioeconomic variables with the homicide rates remained the same as in the classic model specification, which means that this combination of variables added to the model by accounting for complementary information about the urban layouts composition and spatial pattern that also affected the homicide rates. Figure 5.4 shows square image tiles of 500 meters with the corresponding values for the remote sensing variables included in the models to illustrate the differences in urban landscapes corresponding to different values of these variables.

5.4 Chapter conclusions

The negative association between the Gini index of income and the average homicide rate at the analytical region level could indicate that the presence of different income groups in the same neighborhood could deter this type of criminal behavior. This association must be tested in other cities because it could support public housing policies that promote socially mixed neighborhoods to counter social exclusion and reduce crime.

The areas with high homicide rates tended to be more heterogeneous in general and at short distances: they had higher local variation and less general homogeneity. The urban fabric was more crowded and cluttered, with small buildings or housing units located in close proximity to one another and with different roofing materials,

Figure 5.4: Square image tiles showing urban areas (500 x 500 meters) that have different values of the best performing remote sensing variables in model specification. From top to bottom: P.OTHER.IMP.S, F.CLAYR.IMPS, SDT, SDF, and UNIFOR. The tiles are organized from left to right according to increasing values of each variable.

and these regions lacked other homogeneous surfaces such as open green spaces, wide roads, water surfaces or large industrial or commercial buildings. These results are in agreement with the ideas of urban planning approaches such as CPTED and New Urbanism, but at the neighborhood scale when seen from the space: more heterogeneous and disordered urban layouts were associated with higher homicide rates.

The statistically significant correlations between the remote sensing variables and the homicide rates in Medellin do not necessarily imply causational effects. But these associations do inform about the physical aspects that are common in crime prone areas around this city and the results of this empirical analysis are in agreement with the theoretical expected effects. A heterogeneous, organic, cluttered urban layout can provide easy concealment and escape routes for committing homicide. Wide avenues, open spaces and regular street networks will increase the urban layout homogeneity at short distances. This phenomenon is in agreement with the ideas that were implemented in the "Second Empire" reforms by Baron Georges-Eugène Haussmann in Paris in the 19th century that shaped the modern city with its wide boulevards, open spaces and a network of large avenues, which facilitated the control of the city because the troops and the army could maneuver more quickly in these wider spaces than they could in the previous medieval urban layout (Clout, 1977).

The emergence of automated methods and tools for image feature extraction at object level such as FETEX opens up new possibilities of regional science analysis. This chapter provides empirical evidence about the usefulness of remote sensing as a complementary information source for socioeconomic studies in urban settings. The remote sensing variables proved to be useful to quantify different aspects of the physical environment that relate to homicide rates in a city. These variables can add to classic crime explanatory models by quantifying complementary information on the urban composition and the spatial pattern of the urban layout. Spatial analysis of urban crime could benefit from this approach by including information on the physical environment using VHR imagery. Future work should focus in the development and assessment of new tools and techniques of image feature extraction for socioeconomic urban analysis. Regional scientists play an important role in directing that work to the most relevant topics.

Chapter 6

Conclusions

This work seeks to bring closer the remote sensing world to the regional scientists by providing new evidence about the usefulness of very high-resolution images as an alternative and complementary data source for the study of socioeconomic phenomena in urban settings. It illustrated the use of image-derived features in two different quantitative applications of regional science. The literature review, presented in chapter 2, described recent developments related to this field. It showed what remote sensing can do for regional science and how, and listed the main types of data, methods and approaches found in the literature. A work flow for data integration and analysis was presented in chapter 3. It described the main features of the study site and the used data sources in two specific applications. Chapters 4 and 5 presented applications that contribute to the growing literature in this field: chapter 4 showed the potential of remote sensing as an alternative data source for intra-urban poverty mapping, and chapter 5 showed an example of the use of image-derived metrics as a complementary data source for the analysis of urban crime. This chapter presents the conclusions of this work.

The literature review showed that regional scientists can benefit from using satellite remote sensing data for several applications: the estimation of population counts in urban settlements where detailed information is not available; by integrating remote sensing data into urban housing market models to account for environmental factors that affect property values, and to identify housing submarkets, among other topics. Satellite remote sensing also can provide consistent data for place-based analysis, which is very important for the assessment of vulnerability to natural hazards. It is clear that, with the exception of urban growth analysis, little regional science research has used remote sensing to analyze space-time dynamics of other urban issues. The use of remote sensing data to monitor or assess policy outcomes in urban

settings could be of great value. The space-time variations in a quality of life index, or in an intra-urban poverty index, estimated from remote sensing data before and after a government investment in infrastructure and social programmes, could be used to assess whether government actions have produced the expected results or to redirect the investment to new areas most in need. Policy makers and urban planners could benefit from similar analyses of other topics, such as crime, deprivation hot spots, and vulnerability to natural hazards in urban settings, to help them allocate resources in a smart way.

Consistent relationships between environmental factors derived from satellite remote sensing data and socioeconomic variables have been reported in several cities in the US. But these relationships always need to be investigated and tested in each city where they are to be applied because the phenomena are highly specific. Findings for one city can be quite the opposite for another city, even in the same country. However, the use of these place-specific descriptors is still important because it adds to the literature in the field, it helps to identify what image-derived features relate to socioeconomic variables in each place, and to identify patterns that could help to develop new methods of image feature extraction that take into account the cultural and environmental differences.

The most successful approaches in the field of remote sensing for regional science start by linking socioeconomic and demographic information from field surveys and censuses with remote sensing data from different sensors, and relating empirical measurements, spatial theory, and modeling. Image texture measures, spatial metrics and spatial pattern descriptors have been successfully applied for the detection of urban deprivation hot spots in different developing countries with different cultures and environmental conditions. These techniques exploit aspects of the urban layout, such as spatial heterogeneity, regularity and diversity of landscape elements, that are associated in the same direction almost world wide with poverty. However, those characteristics are also common in historic downtowns in many cities of the world and may not be related to poverty. Future research should address this problem by taking into account ancillary data to identify those areas and give them a special treatment in the analysis. LIDAR data has been used to properly identify historic downtowns and could provide a means to tackle this problem.

Although many local and regional government agencies are now using remote sensing to map, monitor and model the composition and growth of cities, more local research linking socioeconomic and remote sensing data is still needed to show the real potential and utility of satellite remote sensing for regional science in urban environments. This work adds to the growing literature in the field by contributing with new empirical evidence about the use of remote sensing data in two different

quantitative applications in urban settings.

The socioeconomic data used for the analysis of intra-urban differences is often associated with administrative spatial units such as administrative neighborhoods or census tracts. But the use of these units can be problematic due to a variety of issues exposed in chapter 3. The use of regionalization methods such as the Max-P-Regions algorithm provides a good way to delineate spatial units that are statistically robust and overcome those issues. Given that the socioeconomic data is always associated with a spatial unit of analysis with polygon geometry, the use of automated tools for image feature extraction techniques at object level, such as FETEX, is very well suited for the extraction of remote sensing metrics and for the integration with socioeconomic data for spatial econometric modeling.

Chapter 4 addressed the usefulness of remote sensing data as an alternative information source for estimating the Slum Index at intra-urban level. A similar exercise was previously tested in a very different city than Medellin: the case of Accra, Ghana. The findings of the present work corroborate these earlier results in the sense that remote sensing can provide reliable proxies for intra-urban poverty mapping. In this work, we tested a wider set of remote sensing variables related not only to land cover composition but also to the spatial pattern characteristics of the urban layout. The results also corroborated the statement that the findings for one city can be quite the opposite of those for another city: the amount of vegetation was the most important predictor of Slum Index in Accra, while in Medellin it wasn't statistically significant. On the other hand, the use of texture measures proved to be useful in both places, because they take advantage of the most general characteristics of the urban layout that relate to urban poverty almost world wide.

The spatial modeling involved in the analysis of intra-urban poverty is very important. Socioeconomic phenomena are often affected by spatial effects, and the understanding of these effects can give insights to policy makers to tackle the problems. The presence of spatial autocorrelation with regard to the Slum Index in Medellin indicates that poverty in this city could be addressed through a policy of strategically located investments and programs designed to boost economic development in some of the city's neighborhoods, because their outcomes can be expected to affect not only the neighborhoods where the investment is located but surrounding neighborhoods as well.

The spatial models for Slum Index estimation developed so far for Medellin are based on the city's unique characteristics. They take into account some land cover features, such as the presence of clay roofs, that are very specific of the culture and environment of this city, as well as other, more general descriptors of the spatial arrangement of the urban layout and of the diversity of landscape features. Future

research in this regard should address two main issues: to test whether the same variables relate to intra-urban poverty measures in a similar way in other cities around the world and to verify that this approach provides consistent results over time. When these aspects are resolved, a similar approach could be used for the following purposes in urban settings with sparse socioeconomic data: to lower the cost of socioeconomic surveys by developing an econometric model from a sample and applying that model to the rest of the city and to perform intercensal or intersurveys estimates of intra-urban Slum Index maps.

The usefulness of remote sensing-derived variables as a complementary data source was demonstrated in chapter 5. In relation to crime modeling, the classic model based solely in socioeconomic factors showed a negative association between the Gini index of income and the homicide rates. This result indicates that the presence of different income groups in the same neighborhood could deter this type of crime and adds support to the idea that social mix in neighborhoods can help reduce crime.

The statistically significant correlations between the remote sensing variables and the homicide rates in Medellin do not necessarily imply causational effects. But these associations do inform about the physical aspects that are common in crime prone areas around this city and the results of the empirical analysis are in agreement with the theoretical expected effects. The areas with high homicide rates in Medellin tended to be more heterogeneous in general and at short distances: they registered higher local variation and less general homogeneity in the image. The urban fabric was more crowded and cluttered, with small buildings or housing units located in close proximity to one another and with different roofing materials, and these regions lacked other homogeneous surfaces such as open green spaces, wide roads, water surfaces or large industrial or commercial buildings. These results corroborate the ideas of urban planning approaches such as CPTED and New Urbanism, but at the neighborhood scale when seen from the space: more heterogeneous and disordered urban layouts were associated with higher homicide rates.

We highlight that the input remote sensing data used in both applications, intra-urban poverty and crime, are rather simple. For these data, the most important requirement is the very high spatial resolution to capture differences in urban layout complexity and heterogeneity. We did not use near-infrared spectral information or other special spectral features from the satellite imagery. The structure and texture variables were calculated from the first principal component band after processing the red, green and blue bands of the color image. Any aerial color image with a very high spatial resolution could serve this purpose. This opens up the possibility of intra-urban poverty index estimation from remote sensing data given that similar imagery could easily be obtained from Internet services such as Google Earth, Yahoo, Microsoft

Bing Imagery, and using aerial imagery obtained with new low cost technologies such as platforms mounted in helium balloons or remote controlled drones that are easy to launch and retract on a highly flexible operating schedule.

Urban planners and policy makers from cities of less developed countries who have sparse or no access to survey data can benefit from this rapid approach for mapping intra-urban poverty variations and informed resource allocation. Firms and companies that use geodemographics and business intelligence could benefit from a similar approach to characterize new developed urban areas or cities with sparse socioeconomic information of the less developed countries. They could be more interested in other socioeconomic descriptors, such as household size and income, but these socioeconomic variables could also be modeled with the help of remote sensing imagery, using data from a well known area to estimate those descriptors in other areas lacking socioeconomic information, and for a fraction of the cost of a socioeconomic survey.

Finally, this research shows that the emergence of automated methods and tools for image feature extraction at object level such as FETEX opens up new possibilities of regional science analysis. This research provides empirical evidence about the usefulness of remote sensing as an alternative and as a complementary information source for socioeconomic studies in urban settings. The remote sensing variables proved to be useful to quantify different aspects of the physical environment that relate to socioeconomic factors within a city. These variables can add to classic explanatory models by quantifying complementary information on the urban composition and the spatial pattern of the urban layout, or can be used to estimate some socioeconomic features when the socioeconomic data are sparse.

Future work should focus in the development and assessment of new tools and techniques of image feature extraction for socioeconomic urban analysis. It should also address the development of space-time analysis techniques aimed at the understanding of the spatial dynamic of the socioeconomic phenomena: the availability of remote sensing imagery for different dates could allow the quantification of spatial patterns each date and thus to objectively identify whether the phenomenon under study is contracting or expanding, or if it is being redistributed across the city.

Chapter 7

Related publications

The following references have been produced during this research.

Published papers in top journals

- Duque, J. C., Patino, J. E., Ruiz, L. A., Pardo-Pascual, J. E. (2015). Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. Landscape and Urban Planning, 135, 11-21. doi:10.1016/j.landurbplan.2014.11.009.[1]

- Patino, J. E., Duque, J. C., Pardo-Pascual, J. E, Ruiz, L. A. (2014). Using remote sensing to assess the relationship between crime and the urban layout. Applied Geography, 55C, 48-60. doi:10.1016/j.apgeog.2014.08.016.

- Patino, J. E., & 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. doi:10.1016/j.compenvurbsys.2012.06.003.[2]

[1]This paper has been selected as the Editor's Choice for Volume 135. Editor's Choice is a pro bono program sponsored by the publisher of the journal that aims to disseminate quality scholarship among a broader community of international scholars and practitioners. Articles of Editor's Choice are available for free download on the journal's website.

[2]This paper has had a good welcome in the research community. The journal's website listed it as the most downloaded article in the past 90 days just four months after first published in electronic format. More than one year and a half later (by December 2014) it was still listed as the second most downloaded article of the journal in the past 90 days. Google scholar has registered 43 citations to this article by March 2015 in other peer-reviewed papers related to the fields of urban studies, environmental studies, regional science, urban planning, geography, GIS, and remote sensing.

Working papers

- Duque, J. C., Patino, J. E., Ruiz, L. A., & Pardo-Pascual, J. E. (2013). Quantifying slumness with remote sensing data. CIEF working paper, EAFIT University, No. 13-23 (No. 13-23). Medellin.

Conference presentations

- Patino, J. E., Duque, J. C., Ruiz Fernández, L. A. & Pardo Pascual, J.E. (2014) Medellín, homicidios y forma urbana: una evaluación con imágenes de satélite. Seminario Territorios y Sociabilidades violentas, EAFIT University, March 13th to 14th, 2014, Medellin, Colombia.

- Patino, J. E., Duque, J. C., Ruiz Fernández, L. A. & Pardo Pascual, J.E. (2013) Improving crime spatial models via remote sensing. El Norte de México y Colombia, III Coloquio Binacional sobre Economía y Empresa, EAFIT University, May 23th to 25th, 2013, Medellin, Colombia.

- Duque, J. C., Patino, J. E., Ruiz Fernández, L. A, & Pardo Pascual, J. E. (2013) Quantifying slumness with remote sensing data. North American Regional Science Conference (NARSC) Meeting 2013, November 13th to 16th, 2013, Atlanta, Georgia.

Bibliography

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.

Anderson, C. A. and Anderson, D. C. (1984). Ambient temperature and violent crime: Tests of the linear and curvilinear hypotheses. *Journal of Personality and Social Psychology*, 46(1):91–97.

Anderson, C. A., Anderson, K. B., Dorr, N., DeNeve, K. M., and Flanagan, M. (2000). Temperature and aggression. *Advances in Experimental Social Psychology*, 32:63–133.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic Publishers, Dordrecht.

Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis*, 27(2):93–115.

Anselin, L. (2005). *Exploring spatial data with GeoDa: A Workbook*. Center for Spatially Integrated Social Science, Urbana, IL, revised edition.

Anselin, L., Cohen, J., Cook, D., Gorr, W., and Tita, G. (2000). Spatial analyses of crime. *Criminal Justice*, 4:213–262.

Anselin, L., Syabri, I., and Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, 38(1):5–22.

Arribas-Bel, D. and Schmidt, C. R. (2013). Self-organizing maps and the US urban spatial structure. *Environment and Planning B: Planning and Design*, 40(2):362–371.

Atkinson, R. (2005). Occasional paper 1: Neighbourhoods and the impacts of social mix: crime, tenure diversification and assisted mobility. Technical report, Housing and Community Research Unit, ESRC Centre for Neighbourhood Research.

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 & Geosciences*, 50:115–127.

Baraldi, A. and Parmiggiani, F. (1995). An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2):293–304.

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.

Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R., and Karuppannan, S. (2010). Understanding heterogeneity in metropolitan India: The added value of remote sensing data for analyzing sub-standard residential areas. *International Journal of Applied Earth Observation and Geoinformation*, 12(5):359–374.

Baud, I., Sridharan, N., and Pfeffer, K. (2008). Mapping Urban Poverty for Local Governance in an Indian Mega-City: The Case of Delhi. *Urban Studies*, 45(7):1385–1412.

Baud, I. S., Pfeffer, K., Sridharan, N., and Nainan, N. (2009). Matching deprivation mapping to urban governance in three Indian mega-cities. *Habitat International*, 33(4):365–377.

Belal, A. A. and Moghanm, F. S. (2011). Detecting urban growth using remote sensing and GIS techniques in Al Gharbiya governorate, Egypt. *The Egyptian Journal of Remote Sensing and Space Science*, 14(2):73–79.

Besussi, E., Chin, N., Batty, M., and Longley, P. (2010). The structure and form of urban settlements. In Rashed, T. and Jürgens, C., editors, *Remote sensing of urban and suburban areas*, Remote Sensing and Digital Image Processing, chapter 2, pages 13–31. Springer Netherlands, Dordrecht.

Bhatta, B., Saraswati, S., and Bandyopadhyay, D. (2010a). Quantifying the degree-of-freedom, degree-of-sprawl, and degree-of-goodness of urban growth from remote sensing data. *Applied Geography*, 30(1):96–111.

Bhatta, B., Saraswati, S., and Bandyopadhyay, D. (2010b). Urban sprawl measurement from remote sensing data. *Applied Geography*, 30(4):731–740.

Botero, V. (2009). *Geo-information for measuring vulnerability to earthquakes: a fitness for use approach.* Doctoral dissertation, University of Utrech, The Netherlands.

Breetzke, G. D. (2012). The effect of altitude and slope on the spatial patterning of burglary. *Applied Geography*, 34:66–75.

Browning, C. R., Byron, R. A., Calder, C. A., Krivo, L. J., Kwan, M.-P., Lee, J.-Y., and Peterson, R. D. (2010). Commercial density, residential concentration, and crime: Land use patterns and violence in neighborhood context. *Journal of Research in Crime and Delinquency*, 47(3):329–357.

Butke, P. and Sheridan, S. C. (2010). An analysis of the relationship between weather and aggressive crime in Cleveland, Ohio. *Weather, Climate, and Society*, 2(2):127–139.

Carlsmith, J. M. and Anderson, C. A. (1979). Ambient temperature and the occurrence of collective violence: A new analysis. *Journal of Personality and Social Psychology*, 37(3):337–344.

Carr-Hill, R. and Chalmers-Dixon, P. (2005). *The Public Health Observatory Handbook of Health Inequalities Measurement*. The South East Public Health Observatory, Oxford.

Cerdá, M., Morenoff, J. D., Hansen, B. B., Tessari Hicks, K. J., Duque, L. F., Restrepo, A., and Diez-Roux, A. V. (2012). Reducing violence by transforming neighborhoods: a natural experiment in Medellín, Colombia. *American Journal of Epidemiology*, 175(10):1045–53.

Cheng, J. and Masser, I. (2003). Urban growth pattern modeling: a case study of Wuhan city, PR China. *Landscape and Urban Planning*, 62(4):199–217.

Cliff, A. D., Haggett, P., and Ord, J. K. (1975). *Elements of spatial structure: a quantitative approch*. Cambridge Geographical Studies;no. 6. Cambridge University Press, London.

Clout, H. D. (1977). Urban Growth, 1500 - 1900. In Clout, H. D., editor, *Themes in the Historical Geography of France*, chapter 13, pages 483–540. Academic Press Inc. Ltd., London.

Cohen, L. E. and Felson, M. (1979). Social change and crime rate trends: A routine acivity approach. *American Sociological Review*, 44(4):588–608.

Cohn, E. G. and Rotton, J. (2000). Weather, seasonal trends and property crimes in Minneapolis, 1987-1988. A moderator-variable time-series analysis of routine activities. *Journal of Environmental Psychology*, 20(3):257–272.

Cole, J. H. and Marroquín Gramajo, A. (2009). Homicide rates in a cross-section of countries: evidence and interpretations. *Population and Development Review*, 35(October):749–776.

Cozens, P. M. (2008). New Urbanism, crime and the suburbs: A review of the evidence. *Urban Policy and Research*, 26(4):429–444.

Craglia, M., Leontidou, L., Nuvolati, G., and Schweikart, J. (2004). Towards the development of quality of life indicators in the "digital" city. *Environment and Planning B: Planning and Design*, 31(1):51–64.

Cullen, F. T. and Agnew, R. (2011). The Chicago School: The city, social disorganization, and crime. In Cullen, F. T. and Agnew, R., editors, *Criminological theory: past to present. Essential readings*, chapter Part III, pages 89–117. Oxford University Press, New York, Oxford, 4th ed. edition.

Dadvand, P., de Nazelle, A., Figueras, F., Basagaña, X., Su, J., Amoly, E., Jerrett, M., Vrijheid, M., Sunyer, J., and Nieuwenhuijsen, M. J. (2012). Green space, health inequality and pregnancy. *Environment International*, 40:110–5.

DANE (2010). Boletín Censo general 2005. Perfil Medellin, Antioquia. Technical report, DANE, Bogota, Colombia.

DANE (2012). Estimaciones de población 1985 - 2005 y proyecciones de población 2005 - 2020, total municipal por área. Technical report, DANE, Bogota, Colombia.

de Sherbinin, A., Balk, D., Yager, K., Jaiteh, M., Pozzi, F., Giri, C., and Wannebo, A. (2002). *A CIESIN thematic guide to social science applications of remote sensing*. Center for International Earth Science Information Network (CIESIN), Columbia University, Palisades, NY.

DeFrances, C. J. and Titus, R. M. (1993). Urban planning and residential burglary outcomes. *Landscape and Urban Planning*, 26:179–191.

DeFronzo, J. (1984). Climate and crime: Tests of an FBI assumption. *Environment and Behavior*, 16(2):185–210.

Diehr, P. (1984). Small area statistics: large statistical problems. *American Journal of Public Health*, 74(4):313–4.

Doll, C. N. H. (2008). *CIESIN thematic guide to night-time light remote sensing and its applications*. Center for International Earth Science Information Network (CIESIN), Columbia University, Palisades, NY.

Doll, C. N. H., Muller, J.-P., and Elvidge, C. D. (2000). Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. *AMBIO: A Journal of the Human Environment*, 29(3):157–162.

Donnay, J.-P., Barnsley, M. J., and Longley, P. A., editors (2001). *Remote Sensing and Urban Analysis*. Taylor & Francis, London.

Donovan, G. H. and Prestemon, J. P. (2012). The effect of trees on crime in Portland, Oregon. *Environment and Behavior*, 44(1):3–30.

Doxani, G., Karantzalos, K., and Strati, M. T. (2012). Monitoring urban changes based on scale-space filtering and object-oriented classification. *International Journal of Applied Earth Observation and Geoinformation*, 15:38–48.

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. (2012). 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. RiSE-group (Research in Spatial Economics), EAFIT University (http://code.google.com/p/clusterpy/).

Duque, J. C., Royuela, V., and Noreña, M. (2013). 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 12, pages 237–254. Springer-Verlag, Berlin Heidelberg.

Ebert, A. and Kerle, N. (2008). Urban social vulnerability assessment using object-oriented analysis of remote sensing and GIS data. A case study for Tegucigalpa, Honduras. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(Part B7):1307–1312.

Ebert, A., Kerle, N., and Stein, A. (2007). Remote sensing based assessment of social vulnerability. In *Proceedings of the 5th international workshop on remote sensing applications to natural hazards*, pages 1–7, Washington, D.C. George Washington University: The Space Policy Institute.

Ebert, A., Kerle, N., and Stein, A. (2009). Urban social vulnerability assessment with physical proxies and spatial metrics derived from air- and spaceborne imagery and GIS data. *Natural Hazards*, 48(2):275–294.

Elvidge, C., Sutton, P., Ghosh, T., and Tuttle, B. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, 35:1652–1660.

Elvidge, C. D., Imhoff, M. L., Baugh, K. E., Ruth, V., Nelson, I., Safran, J., Dietz, J. B., and Tuttle, B. T. (2001). Night-time lights of the world : 1994-1995. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56:81–99.

Emmanuel, R. (1997). Urban vegetational change as an indicator of demographic trends in cities: the case of Detroit. *Environment and Planning B: Planning and Design*, 24(3):415–426.

Field, S. (1992). The effect of temperature on crime. *British Journal of Criminology*, 32(3):340–351.

Fiumi, L. (2012). Surveying the roofs of Rome. *Journal of Cultural Heritage*, 13:304–313.

Florax, R. J. G. M., Folmer, H., and Rey, S. J. (2003). Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Regional Science and Urban Economics*, 33:557–579.

Forster, B. (1983). Some urban measurements from Landsat data. *Photogrammetric Engineering and Remote Sensing*, 49(12):1693–1707.

Foster, S., Giles-Corti, B., and Knuiman, M. (2010). Neighbourhood design and fear of crime: a social-ecological examination of the correlates of residents' fear in new suburban housing developments. *Health & place*, 16(6):1156–1165.

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.

Fugate, D., Tarnavsky, E., and Stow, D. (2010). A survey of the evolution of remote sensing imaging systems and urban remote sensing applications. In Rashed, T. and Jürgens, C., editors, *Remote sensing of urban and suburban areas*, Remote Sensing and Digital Image Processing, chapter 7, pages 119–139. Springer Netherlands, Dordrecht.

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.

Gaviria, A. (2000). Increasing returns and the evolution of violent crime: the case of Colombia. *Journal of Development Economics*, 61(1):1–25.

Giraldo Ramírez, J. (2010). Cambios en la interpretación, el comportamiento y las políticas públicas respecto a la violencia homicida en Medellín. In Hermelin, M., Giraldo Ramírez, J., and Echeverry, A., editors, *Medellín, Medio Ambiente, Urbanismo y Sociedad*, chapter 3, pages 294–318. EAFIT University, Medellin.

Glover, F. (1977). Heuristics for Integer Programming Using Surrogate Constraints. *Decision Sciences*, 8(1):156–166.

Gluch, R. M. and Ridd, M. K. (2010). The V-I-S model: Quantifying the urban environment. In Rashed, T. and Jürgens, C., editors, *Remote sensing of urban and suburban areas*, Remote Sensing and Digital Image Processing, chapter 6, pages 85–116. Springer Netherlands, Dordrecht.

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.

Griffiths, P., Hostert, P., Gruebner, O., and der Linden, S. V. (2010). Mapping megacity growth with multi-sensor data. *Remote Sensing of Environment*, 114(2):426–439.

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.

Harlan, S. L., Brazel, A. J., Prashad, L., Stefanov, W. L., and Larsen, L. (2006). Neighborhood microclimates and vulnerability to heat stress. *Social Science & Medicine*, 63(11):2847–63.

Harris, R., Sleight, P., and Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting*. John Wiley & Sons, Ltd, Chichester, 1st edition.

Harvey, J. T. (2002). Population estimation models based on individual TM pixels. *Photogrammetric Engineering and Remote Sensing*, 68(11):1181–1192.

Hermosilla, T. (2011). *Detección automática de eficios y clasificación de usos del suelo en entornos urbanos con imágenes de alta resolución y datos LiDAR*. Doctoral dissertation, Universitat Politecnica de Valencia.

Herold, M. (2009). Some recommendations for global efforts in urban monitoring and assessments from remote sensing. In Gamba, P. and Herold, M., editors, *Global mapping of human settlement. Experiences, datasets and prospects*, chapter 2, pages 11–23. CRC Press - Taylor & Francis Group, Boca Raton.

Hofmann, P. (2001). Detecting informal settlements from IKONOS image data using methods of object oriented image analysis-an example from Cape Town (South Africa). *Jürgens, C.(Ed.): Remote Sensing of Urban Areas/ ...*, pages 107–118.

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.

Holin, M. J., Buron, L., Locke, G., and Cortes, A. (2003). Interim assessment of the HOPE VI Program cross-site report. Technical report, U.S. Department of Housing and Urban Development, Bethesda, MD.

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.

Ibáñez, A. M. and Vélez, C. E. (2008). Civil conflict and forced migration: The micro determinants and welfare losses of displacement in Colombia. *World Development*, 36(4):659–676.

Isard, W. (1975). *An introduction to regional science*. Prentice Hall, Englewood Cliffs, N.J.

Isserman, A. M. (2004). Intellectual leaders of regional science: A half-century citation study. *Papers in Regional Science*, 83(1):91–126.

Jain, S. (2008). Remote sensing application for property tax evaluation. *International Journal of Applied Earth Observation and Geoinformation*, 10(1):109–121.

Jat, K. M., Garg, P. K., and Khare, D. (2008). Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. *International Journal of Applied Earth Observation and Geoinformation*, 10:26–43.

Jenerette, G. D., Harlan, S. L., Brazel, A., Jones, N., Larsen, L., and Stefanov, W. L. (2007). Regional relationships between surface temperature, vegetation, and human settlement in a rapidly urbanizing ecosystem. *Landscape Ecology*, 22(3):353–365.

Jensen, J. R. and Cowen, D. C. (1999). Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, 65(5):611–622.

Jensen, R., Gatrell, J., Boulton, J., and Harper, B. (2004). Using remote sensing and geographic information systems to study urban quality of life and urban forest amenities. *Ecology and Society*, 9(5):5–15.

Johnson, D. P. (2007). Public participation geographic information systems as surveillance tools in urban health. In Jensen, R., Gatrell, J., and McLean, D., editors, *Geospatial technologies in urban environments. Policy, practice and pixels.*, chapter 6, pages 109–120. Springer-Verlag, Terre-Haute, IN, second edition.

Kawachi, I., Kennedy, B. P., and Wilkinson, R. G. (1999). Crime: social disorganization and relative deprivation. *Social science & medicine (1982)*, 48(6):719–31.

Keane, M. (1975). The size of the region-building problem. *Environment and Planning A*, 7(5):575–577.

Keizer, K., Lindenberg, S., and Steg, L. (2008). The spreading of disorder. *Science*, 322(5908):1681–1685.

Kennedy, B. P., Kawachi, I., Prothrow-Stith, D., Lochner, K., and Gupta, V. (1998). Social capital, income inequality, and firearm violent crime. *Social science & medicine (1982)*, 47(1):7–17.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science (New York, N.Y.)*, 220(4598):671–680.

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.

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.

Kohli, D., Warwadekar, P., Kerle, N., Sliuzas, R., and Stein, A. (2013). Transferability of Object-Oriented Image Analysis Methods for Slum Identification. *Remote Sensing*, 5(9):4209–4228.

Kuo, F. E. and Sullivan, W. C. (2001). Environment and crime in the inner city. Does vegetation reduce crime ? *Environment and Behavior*, 33(3):343–367.

Kurtz, E. M., Koons, B. A., and Taylor, R. B. (1998). Land use, physical deterioration, resident-based control, and calls for service on urban streetblocks. *Justice Quarterly*, 15(1):121–150.

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).

Leinenkugel, P., Esch, T., and Kuenzer, C. (2011). Settlement detection and impervious surface estimation in the Mekong Delta using optical and SAR remote sensing data. *Remote Sensing of Environment*, 115(12):3007–3019.

Levin, N. and Duke, Y. (2012). High spatial resolution night-time light images for demographic and socio-economic studies. *Remote Sensing of Environment*, 119:1–10.

Levitt, S. and Rubio, M. (2000). Understanding crime in Colombia and what can be done about it. FEDESARROLLO working paper No. 20.

Leyk, S., Norlund, P. U., and Nuckols, J. R. (2012). Robust assessment of spatial non-stationarity in model associations related to pediatric mortality due to diarrheal disease in Brazil. *Spatial and Spatio-temporal Epidemiology*, 3(2):95–105.

Li, G. and Weng, Q. (2005). Using Landsat ETM+ imagery to measure population density in Indianapolis , Indiana, USA. *Photogrammetric Engineering & Remote Sensing*, 71(8):947–958.

Li, G. and Weng, Q. (2007). Measuring the quality of life in city of Indianapolis by integration of remote sensing and census data. *International Journal of Remote Sensing*, 28(2):249–267.

Li, G. M. and Weng, Q. (2006). Integration of remote sensing and census data for assessing urban quality of life: Model development and validation. In Weng, Q. and Quattrochi, D. A., editors, *Urban remote sensing*, chapter 15, pages 311–336. CRC Press, Boca Raton.

Liu, G. C., Taylor C, J., Wilson, J. S., Yamada, I., and Hoch, S. C. (2007). Examining urban environment correlates of chilhood physical activity and walkability perception with GIS and remote sensing. In Jensen, R., Gatrell, J., and McLean, D., editors, *Geospatial technologies in urban environments. Policy, practice and pixels.*, pages 121–139. Springer-Verlag, Terre-Haute, IN, second edition.

Liu, X., Clarke, K. C., and Herold, M. (2006). Population density and image texture: A comparison study. *Photogrammetric Engineering and Remote Sensing*, 72(2):187–196.

Liu, X. and Herold, M. (2006). Population estimation and interpolation using remote sensing. In Weng, Q. and Quattrochi, D. A., editors, *Urban remote sensing*, chapter 13, pages 269–290. CRC Press, Boca Raton.

Liu, X. H., Kyriakidis, P. C., and Goodchild, M. F. (2008). Population-density estimation using regression and area-to-point residual kriging. *International Journal of Geographical Information Science*, 22(4):431–447.

Lo, C. P. (1995). Automated population and dwelling unit estimation from high-resolution satellite images: a GIS approach. *International Journal of Remote Sensing*, 16(1):17–34.

Lo, C. P. (1997). Application of LandSat TM data for quality of life assessment in an urban environment. *Computers, Environment and Urban Systems*, 21(3-4):259–276.

Lo, C. P. and Faber, B. J. (1997). Integration of Landsat Thematic Mapper and census data for quality of life assessment. *Remote Sensing of Environment*, 62(2):143–157.

Lo, C. P. and Quattrochi, D. A. (2003). Land-use and land-cover change, urban heat island phenomenon, and health implications: A remote sensing approach. *Photogrammetric Engineering and Remote Sensing*, 69(9):1053–1063.

Lu, D., Moran, E., and Hetrick, S. (2011). Detection of impervious surface change with multitemporal Landsat images in an urban-rural frontier. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):298–306.

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.

Lwin, K. K. and Murayama, Y. (2009). A GIS approach to estimation building population for micro-spatial analysis. *Transactions in GIS*, 13:401–414.

Lwin, K. K. and Murayama, Y. (2011). Estimation of building population from LIDAR derived digital volume model. In Muruyama, Y. and Thapa, R. B., editors, *Spatial analysis and modeling in geographical transformation process*, volume 100, chapter 6, pages 87–98. Springer Netherlands, Dordrecht.

Madhavan, B. B., Kubo, S., Kurisaki, N., and Sivakumar, T. V. L. N. (2001). Appraising the anatomy and spatial growth of the Bangkok Metropolitan area using a vegetation-impervious-soil model through remote sensing. *International Journal of Remote Sensing*, 22(5):789–806.

Materka, A. and Strzelecki, M. (1998). Texture Analysis Methods. A Review. Technical report, Technical University of Lodz, Institute of Electronics, Brussels.

Matthews, S. A., Yang, T.-C., Hayslett, K. L., and Ruback, R. B. (2010). Built environment and property crime in Seattle, 1998 - 2000: a Bayesian analysis. *Environment and Planning A*, 42(6):1403–1420.

Medellín Cómo Vamos (2012). Análisis de la evolución de la calidad de vida en Medellín, 2008-2011. Seguridad ciudadana. Technical report, Medellín Cómo Vamos, Medellín.

Medina, C. and Martínez, H. (2003). Violence and drug prohibition in Colombia. CEDE working paper, Crisis States Programme. Los Andes University, No. 32.

Medina, C., Posso, C., and Tamayo, J. A. (2011). Costos de la violencia urbana y políticas públicas: algunas lecciones de Medellín. *Borradores de Economía*, (674):1–42.

Menezes, T., Silveira-Neto, R., Monteiro, C., and Ratton, J. L. (2013). Spatial correlation between homicide rates and inequality: Evidence from urban neighborhoods. *Economics Letters*, 120(1):97–99.

Mennis, J. (2006). Socioeconomic-vegetation relationships in urban, residential land: the case of Denver, Colorado. *Photogrammetric Engineering and Remote Sensing*, 72(8):911–921.

Mennis, J. and Liu, J. W. (2005). Mining association rules in spatio-temporal data: An analysis of urban socioeconomic and land cover change. *Transactions in GIS*, 9(1):5–17.

Metivier, E. D. and McCoy, R. M. (1971). Mapping urban poverty housing from aerial photographs. In *Proceedings of the Seventh International Symposium of Remote Sensing of Environment*, pages 1563–1569, Ann Harbor, MI. University of Michigan.

Michishita, R., Jiang, Z., and Xu, B. (2012). Monitoring two decades of urbanization in the Poyang Lake area, China through spectral unmixing. *Remote Sensing of Environment*, 117:3–18.

Miller, P. R. and Winer, A. M. (1984). Composition and dominance in Los Angeles basin urban vegetation. *Urban Ecology*, 8:29–54.

Miller, R. B. and Small, C. (2003). Cities from space: potential applications of remote sensing in urban environmental research and policy. *Environmental Science & Policy*, 6(2):129–137.

Monier, R. B. and Green, N. E. (1957). Aerial photographic interpretation and the human geography of the city. *The Professional Geographer*, 9(5):2–5.

Montgomery, M. R. and Hewett, P. C. (2005). Urban poverty and health in developing countries : Household and neighbourhood effects. *Demography*, 43(3):397–425.

Morenoff, J. D., Sampson, R. J., and Raudenbush, S. W. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology*, 39(3):517–560.

Moser, C. O. (1998). The Asset Vulnerability Framework : Reassessing Urban Poverty Reduction Strategies. *World Development*, 26(1):1–19.

Mullens, R. H. J. and Senger, L. W. (1969). Analysis of urban residential environments using color infrared aerial photography: An examination of socioeconomic variable and physical characteristics of selected areas in Los Angeles basin. Interagency report from the U.S. Geological Survey to NASA. Technical report, U.S. Geological Survey, Reston, Virginia.

Mumbower, L. and Donoghue, J. (1967). Urban poverty study. *Photogrammetric Engineering*, 33(6):610–618.

Netzband, M. and Jürgens, C. (2010). Urban and suburban areas as a research topic for remote sensing. In Rashed, T. and Jürgens, C., editors, *Remote sensing of urban and suburban areas*, Remote Sensing and Digital Image Processing, chapter 1, pages 1–9. Springer Netherlands, Dordrecht.

Nichol, J. E. and Wong, M. S. (2006). Assessing urban environmental quality with multiple parameters. In Weng, Q. and Quattrochi, D. A., editors, *Urban remote sensing*, chapter 12, pages 253–268. CRC Press - Taylor & Francis Group, Boca Raton.

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. (1984). *The modifiable areal unit problem, CATMOG 38*. Geo Books, Norwich, vol. 38. edition.

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.

Pacione, M. (2003). Urban environmental quality and human wellbeing-a social geographical perspective. *Landscape and Urban Planning*, 65(1-2):19–30.

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.

Patel, A., Koizumi, N., and Crooks, A. (2014). Measuring slum severity in Mumbai and Kolkata: A household-based approach. *Habitat International*, 41:300–306.

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 Ramírez, J., editors, *Medellín, Medio Ambiente, Urbanismo y Sociedad*, chapter 1, pages 112–127. Fondo Editorial EAFIT, Medellin, 1st edition.

Pham, H. M., Yamaguchi, Y., and Bui, T. Q. (2011). A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. *Landscape and Urban Planning*, 100(3):223–230.

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.

Popkin, S. J., Katz, B., Cunningham, M. K., Brown, K. D., Gustafson, J., and Turner, M. A. (2004). A decade of HOPE VI: Research findings and policy challenges. Technical report, The Urban Institute, Washington, D.C.

Pozzi, F. and Small, C. (2005). Analysis of urban land cover and population density in the United States. *Photogrammetric Engineering and Remote Sensing*, 71(6):719–726.

Qiu, F., Sridharan, H., and Chun, Y. (2010). Spatial autoregressive model for population estimation at the census block level using LIDAR-derived building volume information. *Cartography and Geographic Information Science*, 37(3):239–257.

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/.

Rajasekar, U. and Weng, Q. (2009). Application of association rule mining for exploring the relationship between urban land surface temperature and biophysical/social parameters. *Photogrammetric Engineering and Remote Sensing*, 75(3):385–396.

Ramesh, S. (2009). *High resolution satellite images and LiDAR data for small-area building extraction and population estimation*. Master of science (applied geography), University of North Texas.

Rashed, T. (2005). Sustainable hazards mitigation. In Campagna, M., editor, *GIS for sustainable development: Bringing geographic information science into practice towards sustainability*, chapter 17, pages 287–309. Taylor & Francis (CRC Press), New York.

Rashed, T., Weeks, J., Couclelis, H., and Herold, M. (2007). An integrative GIS and remote sensing model for place-based urban vulnerability analysis. In Mesev, V., editor, *Integration of GIS and remote sensing*, chapter 9, pages 199–233. John Wiley & Sons, Ltd.

Rashed, T. and Weeks, J. R. (2003a). Assessing vulnerability to earthquake hazards through spatial multicriteria analysis of urban areas. *International Journal of Geographical Information Science*, 17(6):547–576.

Rashed, T. and Weeks, J. R. (2003b). Exploring the spatial association betwen measures from satellite imagery and patterns of urban vulnerability to earthquake hazards. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIV(7/W9):144–152.

Rashed, T., Weeks, J. R., Gadalla, M. S., and Hill, A. G. (2001). Revealing the anatomy of cities through spectral mixture analysis of multispectral satellite imagery: A case study of the Greater Cairo Region, Egypt. *Geocarto*, 16(4):5–15.

Rashed, T., Weeks, J. R., Roberts, D. A., Rogan, J., and Powell, R. (2003). Measuring the physical composition of urban morphology using multiple endmember spectral mixture models. *Photogrammetric Engineering and Remote Sensing*, 69(9):1011–1020.

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.

Rey, S. J., Anselin, L., Folch, D. C., Arribas-Bel, D., Sastre Gutierrez, M. L., and Interlante, L. (2011). Measuring spatial dynamics in metropolitan areas. *Economic Development Quarterly*, 25(1):54–64.

Rey, S. J. and Sastré-Gutiérrez, M. L. (2010). Interregional inequality dynamics in Mexico. *Spatial Economic Analysis*, 5(3):277–298.

Rhew, I. C., Vander Stoep, A., Kearney, A., Smith, N. L., and Dunbar, M. D. (2011). Validation of the normalized difference vegetation index as a measure of neighborhood greenness. *Annals of epidemiology*, 21(12):946–52.

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.

Rindfuss, R. R. and Stern, P. C. (1998). Linking remote sensing and social science: The need and the challenges. In Liverman, D., Moran, E. F., Rindfuss, R. R., and Stern, P. C., editors, *People and pixels: Linking remote sensing and social science*, chapter 1, pages 1–27. National Academy Press, Washington, D.C.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3):351–357.

Roncek, D. W. (1981). Dangerous places: Crime and residential environment. *Social Forces*, 60(1):74–96.

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.

Ruiz Fdez, L. A., Fdez-Sarría, A., and Recio Recio, J. A. (2004). Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study. *20th ISPRS Congress*, (1).

Sabet, M., Ibrahim, A. L., and Kanaroglou, P. (2011). Three decades of urban growth in the city of Shiraz, Iran: A remote sensing and geographic information systems application. *Cities*, 28(4):320–329.

Salleh, S. A., Mansor, N. S., Yusoff, Z., and Nasir, R. A. (2012). The crime ecology: Ambient temperature vs. spatial setting of crime burglary. *Procedia - Social and Behavioral Sciences*, 42:212–222.

Sánchez Torres, F. and Núñez Méndez, J. (2001). Determinantes del crimen violento en un país altamente violento: el caso de Colombia. CEDE working paper, Los Andes University, No. 2001-02.

Schneider, A., Seto, K. C., and Webster, D. R. (2005). Urban growth in Chengdu, Western China: application of remote sensing to assess planning and policy outcomes. *Environment and Planning B: Planning and Design*, 32(3):323–345.

Sembler, C. (2006). Estratificación social y clases sociales. Una revisión analítica de los sectores medios. Technical report, Naciones Unidas - CEPAL, Santiago de Chile.

Setiawan, H., Mathieu, R., and Thompson-Fawcett, M. (2006). Assessing the applicability of the V-I-S model to map urban land use in the developing world: Case study of Yogyakarta, Indonesia. *Computers, Environment and Urban Systems*, 30(4):503–522.

Shaw, C. R. and McKay, H. D. (1942). *Juvenile Delinquency and Urban Areas*. University of Chicago Press, Chicago.

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.

Shi, Y., Sun, X., Zhu, X., Li, Y., and Mei, L. (2012). Characterizing growth types and analyzing growth density distribution in response to urban growth patterns in peri-urban areas of Lianyungang City. *Landscape and Urban Planning*, 105(4):425–433.

Short, N. M. (2010). The Remote Sensing Tutorial.

Shu, S. C. F. and Huang, J. N. H. (2003). Spatial configuration and vulnerability of residential burglary: A case study of a city in Taiwan. In *4th International Space Syntax Symposium*, pages 46.1–46.14, London.

SISC (2010). Dinámica del homicidio en 2009. Technical report, Sistema de Información de Seguridad y Convivencia, Medellin.

Sliuzas, R., Kuffer, M., and Masser, I. (2010). The spatial and temporal nature of urban objects. In Rashed, T. and Jürgens, C., editors, *Remote sensing of urban and suburban areas*, Remote Sensing and Digital Image Processing, chapter 5, pages 67 – 84. Springer Netherlands, Dordrecht.

Sorg, E. T. and Taylor, R. B. (2011). Community-level impacts of temperature on urban street robbery. *Journal of Criminal Justice*, 39(6):463–470.

Soundranayagam, J. P., Sivasubramanian, P., Chandrasekar, N., and Durairaj, K. S. P. (2011). An analysis of land use pattern in the industrial development city using high resolution satellite imagery. *Journal of Geographical Sciences*, 21(1):79–88.

Sowunmi, F. A., Akinyosoye, 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.

Sparks, C. S. (2011). Violent crime in San Antonio, Texas: An application of spatial epidemiological methods. *Spatial and Spatio-temporal Epidemiology*, 2(4):301–309.

Stark, R. (1987). Deviant places: A theory of the ecology of crime. *Criminology*, (25):893–909.

Stathopoulou, M. and Cartalis, C. (2006). Mapping quality of life in Metropolitan Athens using satellite and census data. In *1st EARSeL Workshop of the SIG Urban Remote Sensing*, Berlin. Humboldt-Universität zu Berlin.

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.

Sudhira, H. S., Ramachandra, T. V., and Jagadish, K. S. (2004). Urban sprawl: metrics, dynamics and modelling using GIS. *International Journal of Applied Earth Observation and Geoinformation*, 5(1):29–39.

Sun, C., Wu, Z.-f., Lv, Z.-q., Yao, N., and Wei, J.-b. (2013). Quantifying different types of urban growth and the change dynamic in Guangzhou using multi-temporal remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 21:409–417.

Sutton, P. C. (2003). A scale-adjusted measure of "urban sprawl" using nighttime satellite imagery. *Remote Sensing of Environment*, 86(3):353–369.

Sutton, P. C., Cova, T. J., and Elvidge, C. D. (2006). Mapping "Exurbia" in the conterminous United States using nighttime satellite imagery. *Geocarto International*, 21(2):39–45.

Sutton, P. C., Goetz, A. R., Fildes, S., Forster, C., and Ghosh, T. (2010). Darkness on the edge of town: Mapping urban and peri-urban Australia using nighttime satellite imagery. *The Professional Geographer*, 62(1):119–133.

Sutton, R. and Hall, E. (1972). Texture Measures for Automatic Classification of Pulmonary Disease. *IEEE Transactions on Computers*, C-21(7):667–676.

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., Esch, T., Felbier, A., Wiesner, M., Roth, A., and Dech, S. (2012). Monitoring urbanization in mega cities from space. *Remote Sensing of Environment*, 117:162–176.

Taubenböck, H., Post, J., Roth, A., Zosseder, K., Strunz, G., and Dech, S. (2008). A conceptual vulnerability and risk framework as outline to identify capabilities of remote sensing. *Natural Hazards and Earth System Science*, 8(3):409–420.

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.

Taylor, R. B. and Harrell, A. V. (1996). Physical Environment and Crime. Technical report, U.S. Department of Justice, Rockville.

Tian, G., Jiang, J., Yang, Z., and Zhang, Y. (2011). The urban growth, size distribution and spatio-temporal dynamic pattern of the Yangtze River Delta megalopolitan region, China. *Ecological Modelling*, 222(3):865–878.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240.

Tole, L. (2008). Changes in the built vs. non-built environment in a rapidly urbanizing region: A case study of the Greater Toronto Area. *Computers, Environment and Urban Systems*, 32(5):355–364.

Troy, A., Grove, J. M., and O'Neil-Dunne, J. (2012). The relationship between tree canopy and crime rates across an urban rural gradient in the greater Baltimore region. *Landscape and Urban Planning*, 106:262–270.

Tsai, Y. H., Stow, D., and Weeks, J. (2011). Comparison of object-based image analysis approaches to mapping new buildings in Accra, Ghana using multi-temporal QuickBird satellite imagery. *Remote Sensing*, 3(12):2707–2726.

UN-Habitat (2003). Slums of the World: the face of urban poverty in the new millennium?

UN-HABITAT (2006). *State of the World's Cities 2006/7*. UN-HABITAT.

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.

United Nations (2014). *World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352)*. United Nations, Department of Economic and Social Affaris, Populartion Division, New York.

Van de Voorde, T., Jacquet, W., and Canters, F. (2011). Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data. *Landscape and Urban Planning*, 102(3):143–155.

Van Kamp, I., Leidelmeijer, K., Marsman, G., and de Hollander, A. (2003). Urban environmental quality and human well-being. Towards a conceptual framework and demarcation of concepts; a literature study. *Landscape and Urban Planning*, 65(1-2):5–18.

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.

Wang, F. and Arnold, M. T. (2008). Localized income inequality, concentrated disadvantage and homicide. *Applied Geography*, 28(4):259–270.

Ward, D., Phinn, S. R., and Murray, A. T. (2000). Monitoring growth in rapidly urbanizing areas using remotely sensed data. *The Professional Geographer*, 52(3):371–386.

Weber, C. and Hirsch, J. (1992). Some urban measurements from SPOT data: urban life quality indices. *International Journal of Remote Sensing*, 13(17):3251–3261.

Weeks, J. R. (2003). Does nighttime lighting deter crime? An analysis of remotely-sensed imagery and crime data. In Mesev, V., editor, *Remotely-sensed cities*, chapter 16, pages 355–372. Taylor & Francis, London.

Weeks, J. R., Getis, A., Hill, A. G., Agyei-Mensah, S., and Rain, D. (2010). Neighborhoods and fertility in Accra, Ghana: An AMOEBA-based approach. *Annals of the Association of American Geographers*, 100(3):558–578.

Weeks, J. R., Getis, A., Hill, A. G., Gadalla, M. S., and Rashed, T. (2004). The fertility transition in Egypt: Intra-urban patterns in Cairo. *Annals of the Association of American Geographers*, 94(1):74–93.

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.

Weeks, J. R., Hill, A. G., Stow, D. A., Getis, A., and Fugate, D. (2006). The Impact of Neighborhood Structure on Health Inequalities in Accra, Ghana. In *Annual Meeting of the Population Association of America*, pages 1–34, Los Angeles. San Diego State University.

Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, 117:34–49.

Weng, Q. and Lu, D. (2006). Subpixel analysis of urban landscapes. In Weng, Q. and Quattrochi, D. A., editors, *Urban Remote Sensing*, chapter 4, pages 71–90. CRC Press - Taylor & Francis Group, Boca Raton.

Weng, Q. and Quattrochi, D. A. (2006). *Urban remote sensing*. CRC Press - Taylor & Francis Group, Boca Raton.

Wilcox, P., Quisenberry, N., Cabrera, D. T., and Jones, S. (2004). Busy places and broken windows? Toward defining the role of physical structure and process in community crime models. *The Sociological Quarterly*, 45(2):185–207.

Wilson, J. Q. and Kelling, G. L. (1982). Broken windows: The police and neighborhood safety. *Atlantic Monthly*, 249(3):29–38.

Wolfe, M. K. and Mennis, J. (2012). Does vegetation encourage or suppress urban crime? Evidence from Philadelphia, PA. *Landscape and Urban Planning*, 108(2-4):112–122.

Wu, C. (2004). Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. *Remote Sensing of Environment*, 93(4):480–492.

Wu, C. and Murray, A. T. (2005). A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, 29(5):558–579.

Wu, K.-y. and Zhang, H. (2012). Land use dynamics, built-up land expansion patterns, and driving forces analysis of the fast-growing Hangzhou metropolitan area, eastern China (1978-2008). *Applied Geography*, 34:137–145.

Yang, X. and Liu, Z. (2005). Use of satellite-derived landscape imperviousness index to characterize urban spatial growth. *Computers, Environment and Urban Systems*, 29(5):524–540.

Ye, X. and Wu, L. (2011). Analyzing the dynamics of homicide patterns in Chicago: ESDA and spatial panel approaches. *Applied Geography*, 31(2):800–807.

Yin, Z.-Y., Stewart, D. J., Bullard, S., and MacLachlan, J. T. (2005). Changes in urban built-up surface and population distribution patterns during 1986 - 1999: A case study of Cairo, Egypt. *Computers, Environment and Urban Systems*, 29(5):595–616.

Yu, D. and Wu, C. (2006). Incorporating remote sensing information in modeling house values : A regression tree approach. *Photogrammetric Engineering and Remote Sensing*, 72(2):129–138.

Yuan, F., Wu, C., and Bauer, M. (2008). Comparison of spectral analysis techniques for impervious surface estimation using Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, 74(8):1045–1055.

Yuan, Y., Smith, R. M., and Limp, W. F. (1997). Remodeling census population with spatial information from Landsat TM imagery. *Computers, Environment and Urban Systems*, 21(3-4):245–258.

Zhang, C. and Zhang, S. (2002). *Association rule mining. Models and algorithms*. Springer-Verlag, Berlin.

Zhang, Q. and Seto, K. C. (2011). Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sensing of Environment*, 115(9):2320–2329.

Zhou, X. and Wang, Y.-C. (2011). Spatial-temporal dynamics of urban green space in response to rapid urbanization and greening policies. *Landscape and Urban Planning*, 100(3):268–277.

Zhu, Z., Woodcock, C. E., Rogan, J., and Kellndorfer, J. (2012). Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. *Remote Sensing of Environment*, 117:72–82.

Appendix A

FETEX variables

This appendix describes the image-derived variables, extracted with FETEX 2.0, that were used to quantify the differences in the spatial pattern of the urban layout. Two different sets of variables were computed using FETEX: a set of structure features and a set of texture features.

A.1 Structure features

According to Ruiz et al. (2011), structure features provide information of the spatial arrangement of different elements in the object (the analytical region in this work), in terms of randomness or regularity of the distribution of the elements, which is the case of alignments or regular patterns that are present in different man-made landscapes. These features quantify different aspects of the spatial arrangement of the image's intensity values within the analytical regions, and were used to quantify the differences in the spatial pattern of the urban layout.

Structure features are calculated in FETEX 2.0 using the experimental semivariogram approach. The semivariogram quantifies the spatial associations of the intensity values, measures the degree of spatial correlation between different pixels in an image and is a suitable tool for characterizing spatial patterns (Ruiz et al., 2011).

According to Balaguer et al. (2010), the experimental semivariogram is computed as

$$\gamma(\overrightarrow{h}) = \frac{1}{2N(\overrightarrow{h})} \sum_{i=1}^{N(\overrightarrow{h})} \left[z(x_i) - z(x_i + \overrightarrow{h}) \right]^2 \tag{A.1}$$

where $z(x_i)$ represents the value of the variable at the location x_i, $\overrightarrow{h}$ the separation

between elements in a given direction, and $N(\overrightarrow{h})$ the number of data pairs occurring at locations x_i and $x_i + \overrightarrow{h}$. It is considered that x_i represents the central position of pixel i, $z(x_i + \overrightarrow{h})$ is equal to $z(x_k)$, and x_k is located at the central position of the nearest pixel to the location $x_i + \overrightarrow{h}$.

As each object or region is characterized with one object-specific multidirectional experimental semivariogram, only the pixels inside a region are used for the computation of the semivariogram. 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°; then, each semivariogram curve is smoothed using a Gaussian filter to smooth its shape and to reduce experimental fluctuations (Balaguer et al., 2010; Ruiz et al., 2011). The omnidirectional experimental semivariogram is computed as

$$\gamma(h_d) = \frac{1}{6}\sum_{i=0}^{5}\gamma^{i30}(h_d) \tag{A.2}$$

with $\gamma(h_d)$ being the semivariogram computed in A.1, within a distance of module h_d and direction $(i \cdot 30)$.

After computing the experimental semivariogram (A.2), a set of n points is available. Assuming that $\gamma_i = \gamma(h_i)$ and the value on the lags are equally spaced, the points that belong to the experimental semivariogram can be expressed as

$$\{(h, \gamma_1), (2h, \gamma_2), (3h, \gamma_3), (4h, \gamma_4), \ldots, (nh, \gamma_n)\} \tag{A.3}$$

Structure features extracted from the semivariogram are based on a zonal analysis defined by a set of singular points on the semivariogram, such as the first maximum, the first minimum, and the second maximum. Local maxima values are extracted considering linear neighbourhoods of 5 positions to reduce the detection of false maxima due to slight local fluctuations in the experimental semivariogram values as follows:

$$\gamma_{max\_j} = \{\gamma_i, \in \{3, 4, \ldots, n-2\} : \gamma_{i-1} > \gamma_i < \gamma_{i+1}\, and\, \gamma_{i-2} > \gamma_i < \gamma_{i+2}\}\, j = 1, \ldots, M \tag{A.4}$$

Assuming that these local maxima are ordered, $h_{max\_1} < h_{max\_2} < \cdots < h_{max\_M}$, the first local minimum is defined as

$$\gamma_{min\_1} = \{\gamma_i \in \{(max\_1) + 2, \ldots, (Peak\_2) - 2\} : \gamma_{i-1} > \gamma_i < \gamma_{i-1}\, and\, \gamma_{i-2} > \gamma_i < \gamma_{i+2}\} \tag{A.5}$$

where $Peak\_2$ is equal to max_2 if this exists, or equal to n when there is only one local maximum (M=1 in equation A.4). The parameters that characterize the semivariogram behaviour are defined using the set of points given in equation A.3, and

are grouped according to the position of the lags used in their definition as (i) near the origin, (ii) up to the first maximum, and (iii) between first and second maxima. The 10 parameters used in this research are those belonging to the (i) and (ii) groups. A description of the structure variables used in this work is given next according to Balaguer et al. (2010).

A.1.1 Parameters near the origin

A.1.1.1 RVF: Ratio variance - first lag

RVF is the ratio between the values of the total variance and the semivariance at first lag:

$$RVF = \frac{Variance}{\gamma_1} \tag{A.6}$$

where $Variance$ is the value of the total variance of the group of pixels belonging to the object (i.e., the analytical region). This parameter is an indicator of the relationship between the spatial correlation at long and short distances. High variability at long distances and low variability at short distances are related to increases in RVF values.

A.1.1.2 RSF: Ratio between semivariance values at second and first lag

$$RSF = \frac{\gamma_2}{\gamma_1} \tag{A.7}$$

This feature represents the proportion of the semivariogram value at second lag in relation to the first lag. It informs about changes in the variability of data at short distances.

A.1.1.3 FDO: First derivative near the origin

$$FDO = \frac{\gamma_2 - \gamma_1}{h} \tag{A.8}$$

This parameter represents the slope of the semivariogram at the first two lags, approximating the first derivative near the origin. It shows the variability changes of the pixel values at short distances.

A.1.1.4 SDT: Second derivative at third lag

$$SDT = \frac{\gamma_4 - 2\gamma_3 + \gamma_2}{h^2} \tag{A.9}$$

This feature approximates the value of the second derivative of the semivariogram at the third lag. It quantifies the concavity (or convexity) level of the semivariogram at short distances (between 2 and 4 pixels for $h = 1$), corresponding with the heterogeneity objects of the image. Positive values of this feature indicate that the curve is concave, which means that the image is homogeneous at short distances; while negative values indicate that the semivariogram is convex and the image is heterogeneous at short distances.

A.1.2 Parameters up to the first maximum

A.1.2.1 MFM: Mean of the semivariogram values up to the first maximum

$$MFM = \gamma_{max\_1}^{mean} = \frac{1}{max\_1} \sum_{i=1}^{max\_1} \gamma_i \tag{A.10}$$

MFM is an indicator of the average of the semivariogram values between the first lag and the first maximum. This feature informs about the changes in the variability of the data, and is related to the concavity or convexity of the semivariogram in that interval.

A.1.2.2 VFM: Variance of the semivariogram up to the first maximum

$$VFM = \frac{1}{max\_1} \sum_{i=1}^{max\_1} \left(\gamma_i - \gamma_{max\_1}^{mean}\right)^2 \tag{A.11}$$

This feature computes the variance of the semivariogram values between the first value and the first maximum. It complements the information provided by MFM and is directly related to the homogeneity of the values in the image (i.e., the object or analytical region in this research).

A.1.2.3 DMF: Difference mean of semivariogram and first lag semivariance

DMF is the difference between the mean of the semivariogram values up to the first maximum (MFM) and the semivariance at first lag:

$$DMF = \gamma_{max\_1}^{mean} - \gamma_1 \tag{A.12}$$

It shows the decreasing rate of the spatial correlation in the image up to the lags where the semivariogram theoretically tends to be stabilized. This feature complements the information provided by RVF.

A.1.2.4 RMM: Ratio between the semivariance at first local maximum and the mean semivariogram values up to this maximum

$$RMM = \frac{\gamma_{max\_1}}{\gamma_{max\_1}^{mean}} \tag{A.13}$$

This feature complements the information provided by RVF, which takes into account the influence of the total variability of the analyzed data.

A.1.2.5 SDF: Second-order difference betweeen first lag and first maximum

$$SDF = \gamma_{max\_1} - 2\gamma_{\frac{max\_1}{2}} + \gamma_1 \tag{A.14}$$

In the computation of this feature, if max_1 is an odd number, the integer part of $\frac{max\_1}{2}$ is taken to compute $\gamma_{\frac{max\_1}{2}}$. This parameter informs about the semivariogram curvature in the first lag - first maximum interval and represents the low frequency values in the image.

A.1.2.6 AFM: Area between the semivariogram value in the first lag and the semivariogram function until first maximum

$$AFM = \frac{h}{2}\left(\gamma_1 + 2\left(\sum_{i=2}^{max\_1-1}\gamma_i\right) + \gamma_{max\_1}\right) - \left(\gamma_1\left(h_{max\_1} - h_1\right)\right) \tag{A.15}$$

This feature informs about the semivariogram curvature and is also related to the variability of the data (i.e., the pixel values within the object or analytical region).

A.2 Texture features

Texture features quantify the spatial distribution of the intensity values in the image and provide information about contrast, uniformity, roughness, etc. (Haralick et al., 1973; Baraldi and Parmiggiani, 1995). 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 GLCM tabulates the frequency of one gray tone appearing in a specified spatial relationship with another gray tone within the analyzed polygon (Baraldi and Parmiggiani, 1995). FETEX 2.0 uses the GLCM, averaged from four principal orientations (0^{o}, 45^{o}, 90^{o}, and 135^{o}) to calculate a set of variables that are widely used in image processing: uniformity, entropy, contrast, inverse difference moment (IDM), covariance, variance, and correlation (Haralick et al., 1973; Ruiz et al., 2011). The kurtosis and skewness features are based on the histogram

of the pixel values inside the polygon. The edgeness factor is another feature that represents the density of edges present in a region (Sutton and Hall, 1972), and the mean and standard deviation of the edgeness factor (MEAN EDG, STDEV EDG) are also computed within this set of features (Ruiz et al., 2011). A brief description of this variables follows.

A.2.1 Parameters based on the histogram

The first-order statistics are descriptive features of the image texture directly related to the form of the histogram of image pixel values or intensity levels (Hermosilla, 2011). The histogram contains the frequency of occurrence of each single pixel value in the image. The approximate probability density of occurrence of intensity levels or pixel values is obtained when the pixel values are divided by the total number of pixels in the image (Materka and Strzelecki, 1998). Skewness and kurtosis are the two parameters computed from the histogram that were used in this work. A brief description and their formulas is given next.

A.2.1.1 Skewness

This parameter indicates if the distribution of pixel values is symmetric around the value of maximum frequency. Positive values indicate that the distribution is displaced to the rigth with a larger tail towars the maximum's rigth, i.e., with brigther values than the average, while negative values indicate that the distribution is displaced to the left, with a larger tail towards this side, i.e., with darker pixel values than the average. This feature is computed as follows:

$$Skewness = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(x_i - \bar{x})}{\sigma}\right)^3 \tag{A.16}$$

where x_i represents the pixel value, $\bar{x}$ the mean of all pixel values in the image, and σ the standard deviation of pixel values.

A.2.1.2 Kurtosis

The kurtosis of a histogram is defined as the degree in which a statistical frequency distribution is peaked (sharp) or flattened, and indicates how uniform is the distribution of pixel values in the image. This parameter is computed using the expression:

$$Kurtosis = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(x_i - \bar{x})}{\sigma}\right)^4 - 3 \tag{A.17}$$

A.2.2 Parameters based on the gray level co-ocurrence matrix

The gray level coocurrence matrix is used in a statistical approach to quantify texture features, which represent the texture of the image indirectly by the non-deterministic properties that govern the distributions and relationships between the gray levels or intensity values of the image (Materka and Strzelecki, 1998). The texture features obtained form the GLCM are called second-order statistics, because they are based in statistics given by pairs of pixels. The elements of the gray level coocurrence matrix, *p(i,j)*, represent the relative frequency by wich two pixels with gray levels "i" and "j", that are located at a distance "d" in a given direction, are in the image or neighborhood (Baraldi and Parmiggiani, 1995; Ruiz Fdez et al., 2004). The GLCM is a symmetrical matrix of dimension equal to the number of intensity levels in the image and its elements are expressed by the following equations (Haralick et al., 1973):

$$p(i,j) = \frac{P(i,j)}{\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} P(i,j)} \tag{A.18}$$

$$\mu = \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} i \cdot p(i,j) \tag{A.19}$$

$$\sigma^2 = \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} (i-\mu)^2 \cdot p(i,j) \tag{A.20}$$

where *Ng* is the total number of gray levels, μ is the mean of the GLCM, and σ^2 is the standard deviation. Haralick et al. (1973) proposed several statistical features that represent texture properties like *contrast, uniformity, mean, variance, inertia moments,* etc. A brief description of the second-order texture features used in this work is given next, as interpreted by Haralick et al. (1973), and Baraldi and Parmiggiani (1995).

A.2.2.1 UNIFOR: Uniformity

Also known as "angular second moment" or "energy". This parameter measures textural uniformity, i.e., pixel pairs repetition; high values occur when the gray level distribution has eather a constant or a periodic form. This feature is computed using the following expression:

$$UNIFOR = \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} [p(i,j)]^2 \tag{A.21}$$

A.2.2.2 ENTROP: Entropy

This parameter measures the disorder of the image. High values are associated with not uniform textures, and Entropy is strongly and inversely correlated to Uniformity. This feature is computed as follows:

$$ENTROP = -\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} p(i,j)\cdot\log\left(p(i,j)\right) \tag{A.22}$$

A.2.2.3 CONTRAS: Contrast

Contrast measures the difference between the highest and the lowest values of a contiguous set of pixels. It is associated with the average value difference between neighbor pixels, and high contrast values imply a high contrast texture. This feature is computed usign the expression:

$$CONTRAS = \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} (i-j)^2\cdot p(i,j) \tag{A.23}$$

A.2.2.4 IDM: Inverse Difference Moment

This parameter is also known as "homogeneity", and it measures image homogeneity as it takes larger values for smaller gray tone differences in pixel pairs. This feature is computed as follows:

$$IDM = \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} \left[1/(1+(i-j)^2\right] p(i,j) \tag{A.24}$$

A.2.2.5 VARIAN: Variance

This feature is a measure of heterogeneity and is strongly correlated to the first order statistical variable standard deviation, and the value of this parameter increases when the pixel values differ from their mean. It is computed using the expression:

$$VARIAN = \sigma^2 = \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} (i-\mu)^2\cdot p(i,j) \tag{A.25}$$

A.2.2.6 COVAR: Covariance

This parameter is a measure of local homogeneity, and high values of this feature imply a high probability of occurrence of similar pixel values together in the image or region

under analysis, while low values are associated with images where neighboring pixels have very different values. This feature is computed using the following expression:

$$COVAR = -\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1}(i-\mu)\cdot(j-\mu)\cdot p(i,j) \tag{A.26}$$

A.2.2.7 CORRELAT: Correlation

This feature is expressed by the correlation coefficient between two random variables i and j, where i stands for the possible outcomes in the measurement of the pixel value for the first element of the displacement vector, while j is associated with pixel values of the second element of the displacement vector. Correlation is a measure of pixel value linear-dependencies in the image, and high correlation values imply a linear relationship between the values of pixel pairs (a completely homogeneous region will have a correlation value of 1). It is computed as follows:

$$CORRELAT = \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1}(i-\mu)\cdot(j-\mu)\cdot p(i,j)/\sigma^2 \tag{A.27}$$

A.2.3 Parameters based on the edgeness factor

According to Sutton and Hall (1972), the edgeness factor represents the density of edges present in a neighbourhood of an image. The change between pixel values or the gradient of an image I is computed as a function of the distance *"d"* between neighbour pixels using the following expression:

$$\begin{aligned} g(i,j,d) \quad = \quad & \sum_{(i,j)\in N}\{\mid I(i,j)-I(i+d,j)\mid+\mid I(i,j)-I(i-d,j)\mid+ \\ & \mid I(i,j)-I(i,j+d)\mid+\mid I(i,j)-I(i,j-d)\mid\} \end{aligned}$$

where *g(i,d,j)* is the edgeness per unit area surrounding a generic pixel *(i,j)*. FETEX 2.0 calculates the mean and the standard deviation of the edgeness factor (named MEAN EDG and STDEV EDG) from the values of the edgeness per unit area of all the pixels inside an objetc, i.e., a polygon, parcel or analytic region (Ruiz et al., 2011).

Appendix B

Regionalization with max-p-regions

The max-p-regions, proposed by Duque et al. (2012), is a Mixed Integer Programming model for spatial clustering that seeks to aggregate n areas (e.g. administrative neighborhoods) into the maximum number of regions (analytical regions) such that each region satisfies the following conditions: (1) the regions must be as homogeneous as possible in terms of a given set of areal attributes, e.g. socioeconomic attributes, and (2) each region must satisfy a minimum threshold value imposed on a predefined spatially extensive attribute, e.g. number of households, area, population, etc. The first condition seeks to minimize the loss of information generated in the aggregation process (aggregation bias), and the second condition imposes that each region must contain a minimum amount of observation units (households) to guaranty a proper statistical inference (i.e., it seeks to solve the small numbers problem). The max-p-regions model is of great utility in helping practitioners to design spatial units that fit the minimum requirements for performing spatial econometric analysis.

As the majority of the clustering algorithms, the max-p-regions model maximizes the level of intraregional homogeneity by minimizing a measure of intraregional heterogeneity (H) defined as follows:

$$H = \sum_{k=1}^{p} \sum_{ij \in R_k | i<j} d_{ij} \tag{B.1}$$

Where d_{ij} is the multidimensional squared Euclidean distance between all the areas assigned to the same region k, R_{k}. For the sake of completeness, the original formulation and explanation is as follows (after Duque et al., 2012):

Parameters:

i, I, Index and set of areas, $I = \{1, \ldots, n\}$;

k = index of potential regions, $k = \{1, \ldots, n\}$;

c = index of contiguity order, $c = \{0, \ldots, q\}$, with $q = (n - 1)$;

$$\omega_{ij} = \begin{cases} 1 & , \text{ if areas } i \text{ and } j \text{ share a border, with } i, j \in I \text{ and } i \neq j \\ 0 & , \text{ otherwise;} \end{cases}$$

$N_i = \{j \mid \omega_{ij} = 1\}$, the set of areas that are adjacent to area i;

d_{ij} =dissimilarity relationships between areas i and j, with $i, j \in I$ and $i < j$;

$h = 1 + \left[\log \sum_i \sum_{j|j>i} d_{ij}\right]$, which is the number of digits of the floor function of $\sum_i \sum_{j|j>i} d_{ij}$, with $i, j \in I$;

l_i = spatially extensive attribute value of area i, with $i \in I$;

threshold = minimum value for attribute l at regional scale.

Decision variables:

$$t_{ij} = \begin{cases} 1, & \text{if areas } i \text{ and } j \text{ belong to the same region } k, \text{ with } i < j \\ 0, & \text{otherwise;} \end{cases}$$

$$x_i^{kc} = \begin{cases} 1, & \text{if area } i \text{ is asigned to region } k \text{ in order } c \\ 0, & \text{otherwise.} \end{cases}$$

Minimize:

$$(1)\ Z = \left(-\sum_{k=1}^{n} \sum_{i=1}^{n} x_i^{k0}\right) * 10^h + \sum_i \sum_{j|j>i} d_{ij} t_{ij}$$

Subject to:

$$(2)\ \sum_{i=1}^{n} x_i^{k0} \leq 1 \qquad \forall k = 1, \ldots, n;$$

$$(3)\ \sum_{k=1}^{n} \sum_{c=0}^{q} x_i^{kc} = 1 \qquad \forall i = 1, \ldots, n;$$

(4) $x_i^{kc} \leq \sum_{j \in N_i} x_j^{k(c-1)} \qquad \forall i = 1, \ldots, n;\ \forall k = 1, \ldots, n;\ \forall c = 1, \ldots, q;$

(5) $\sum_{i=1}^{n} \sum_{c=0}^{q} x_i^{kc} l_i \geq \text{threshold} * \sum_{i=1}^{n} x_i^{k0} \qquad \forall k = 1, \ldots, n;$

(6) $t_{ij} \geq \sum_{c=0}^{q} x_i^{kc} + \sum_{c=0}^{q} x_j^{kc} - 1 \qquad \forall i, j = 1, \ldots, n \mid i < j;\ \forall k = 1, \ldots, n;$

(7) $x_i^{kc} \in \{0,\ 1\} \qquad \forall i = 1, \ldots, n;\ \forall k = 1, \ldots, n;\ \forall c = 0, \ldots, q;$

(8) $t_{ij} \in \{0,\ 1\} \qquad \forall i, j = 1, \ldots, n \mid i < j.$

The objective function (1) is composed of two terms, the first term captures the number of resulting regions by adding those areas i assigned to a region in order 0 (each region has one and only one area assigned at this order). The second term captures the level of intraregional heterogeneity by adding the d_{ij} distances between those areas assigned to the same region (i.e., $t_{ij} = 1$). Both terms are merged into a single value in such a way that an increment in the number of regions is preferred over any other solution involving a lower number of regions, and for a given number of regions, the solution with lower intra-regional heterogeneity is preferred. This hierarchy in the elements of the objective function is obtained by multiplying the first term by a scaling factor 10^h, with $h = 1 + \left[\log \sum_i \sum_{j|j>i} d_{ij}\right]$.

Constraints set (2) to (4) guarantee the spatial contiguity of the regions by using a strategy based on contiguity order. Following this strategy Constraints set (2) imposes that each region must contain one and only one area assigned to a region k in order 0. These areas are known as root areas. Constraints set (3) dictates that every area must be assigned to a region k and an order c, and Constraints set (4) imposes that an area i can be assigned to region k in order c if and only if there exist an area j that belongs to the neighborhood of i (N_i) that is assigned to the same region k in order $c - 1$ (this condition does not apply to the root area). Constraints set (5) guarantees that each region satisfies the minimum threshold value imposed over a spatially extensive attribute (l_i). Constraints set (6) activates the pairwise links between the areas assigned to the same region ($t_{ij} = 1$). Constraints set (7) and (8) guarantee variable integrity.

This model is proven to be Non-deterministic Polinomial-time hard (NP-hard) (Cliff et al., 1975; Keane, 1975), which implies that the necessary time to solve an instance grows exponentially with the instance size. According to computational experiments reported by Duque et al. (2012), the largest instance that achieved an optimum solution was for a 16 areas size. The solution of large instances requires the

use of heuristic methods, which are iterative methods that start from an initial feasible solution and then make modifications by moving areas from one region to another region seeking to improve the objective function. These methods have systems that allow them to escape from "local optima," by allowing a temporal detriment of the objective function to escape from local optima and then maybe find better solutions than the one they had already reached.

The heuristic in Duque et al. (2012) is composed of two large phases: one that builds the initial feasible solution, which is based on the generation of regions by seeds. It generates one region at a time, by selecting one random area that grows by the aggregation of neighbor areas until the threshold is reached. When a region reaches the threshold, it stops growing and a new seed is selected. This process is repeated until no more new feasible regions can be generated. At the end of this process, some areas are left without being assigned to any region because they can not be converted to a feasible region. These areas are known as "enclaves". Then these enclaves are assigned to a region that already exists. When all the enclaves are assigned, an initial feasible solution is obtained. This process is repeated a determined number of times to obtain a set of different initial feasible solutions. The best initial feasible solution is extracted from that set.

Then, the best initial feasible solution passes to a process of local search. Duque et al. (2012) present three different methods of local search: one uses the Simulated Anealing algorithm (SA), a local search method proposed by Kirkpatrick et al. (1983), another based in the Tabu Search method proposed by Glover (1977), and the Greedy Algorithm method which is a very basic local search method that only makes movements when they improve the objective function.

The computational experiments performed by Duque et al. (2012) showed that the local search estrategy applied by the Tabu Search method is the most effective. As the movement of areas is done by moving only one area at a time, each time an area i is moved from a donnor region to a receiving region, that movement enters to a "Tabu list", which is a list of forbidden movements for a determined number of iterations; for example, when an area i is moved from region A to region B, the reverse movement of area i to region A is forbidden for the next 10 iterations (a determined number previously defined). When the movement of an area worsens the objective function, the algorithm would try to reverse that movement; but, as that specific movement is forbidden, the algorithm will try to search for a better solution while maintaining that area in the receiving region. In this way the algorithm is forced to escape local optima so it can find a better solution. The algorithm stops when it does not find a substantial improvement after a given number of movements without improvement.

In this work, we used the number of surveyed households as the spatially exten-

sive attribute with a threshold of 100 surveyed households per region. In chapter 4, the analytical regions were delineated by applying the max-p-regions algorithm to the administrative neighborhoods with the 2007 Quality of Life Survey data, which grouped the 243 city neighborhoods to 139 analytical regions. In chapter 5, the same procedure was applied with the 2005 Census and the 2009 Quality of Life Survey data, which resulted in 136 analytical regions.