GEOSPATIAL AND REMOTE SENSING-BASED INDICATORS OF SETTLEMENT TYPE – DIFFERENTIATING INFORMAL AND FORMAL SETTLEMENTS IN GUATEMALA CITY

by

Karen K. Owen
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
the Requirements for the Degree
of
Doctor of Philosophy
Earth Systems and Geoinformation Science

Committee:

| | |
|---|---|
| ________________ | Dr. David Wong, Dissertation Director |
| P. Agouris | Dr. Peggy Agouris, Committee Member |
| G. E. Taylor | Dr. George Taylor, Committee Member |
| Kathryn H. Jacobsen | Dr. Kathryn Jacobsen, Committee Member |
| P. Agouris | Dr. Peggy Agouris, Department Chairperson |
| ________________ | Dr. Timothy L. Born, Associate Dean for Student and Academic Affairs, College of Science |
| ________________ | Dr. Vikas Chandhoke, Dean, College of Science |
| Date: April 13, 2012 | Spring Semester 2012<br>George Mason University<br>Fairfax, VA |

# Geospatial and Remote Sensing-based Indicators of Settlement Type – Differentiating Informal and Formal Settlements in Guatemala City

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Karen K. Owen
Master of Science
George Mason University, 2006

Director: David Wong, Professor
Department of Earth Systems and Geoinformation Science

Spring Semester 2012
George Mason University
Fairfax, VA

## DEDICATION

This effort is dedicated foremost to my husband, Dr. Mark Owen, whose boundless encouragement was a major factor toward completion, and whose love of teaching and science continuously supported my learning. Thanks for not letting me quit. I would also like to dedicate this to Rebecca, Timothy, Lucas, and William, who accepted without complaint my evening and weekend work as a way of life for more years than I care to admit. May you continue to learn and be inspired. As you go out into the world, remember to share your knowledge and do good works.

## ACKNOWLEDGEMENTS

I would like to thank Dr. David Wong, whose standards for excellence served as the impetus for my final product, and who encouraged me to seek a topic that interested me personally so that I would be inspired until the end. His guidance and dedication toward this research were tremendous. It was his instruction in spatial aspects of population and quantitative methods that urged me in the direction of applied human geography. I would not have been able to complete this without the help of Greg Terrie, Exelis Inc., and Dr. Devin White. They propelled me forward with the ENVI software. Ramón López provided invaluable assistance, enabled by Ángel Santiago, during my field visits to Guatemala. Rick MacGregor was instrumental in the beginning by helping to locate and identify study sites in Guatemala. Partner for Surgery, and especially Frank and Linda Peterson, opened my eyes to the beauty and dignity of Guatemala and its indigenous population. Dr. Kathryn Jacobsen expertly guided me on the path to my first publication, also creating an adaptive, flexible radiometer base throughout the rural areas of Alta Verapáz. Dr. Peggy Agouris afforded me a spectacular experience as a co-instructor in the Department of Geography and Geoinformation Science, and Dr. Taylor's assistance with scholarly review kept the dissertation well-organized. I must surely thank my parents, George and Clare Kettell, for the positive environment they provided growing up and their willingness to support my early college career. I would like to thank Dr. Carrie Root for sincerely answering pressing questions about balancing family and research demands, and her unwavering confidence in my ability to finish the job. And finally, it was God's grace that served as my foundation. He continues to guide and uplift me through all of life's challenges.

## TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

# ABSTRACT

GEOSPATIAL AND REMOTE SENSING-BASED INDICATORS OF SETTLEMENT TYPE – DIFFERENTIATING INFORMAL AND FORMAL SETTLEMENTS IN GUATEMALA CITY

Karen K. Owen, Ph.D.

George Mason University, 2012

Dissertation Director: Dr. David Wong

This research addresses the need for reliable, repeatable, quantitative measures to differentiate informal (slum) from formal (planned) settlements using commercial very high resolution imagery and elevation data. Measuring the physical, spatial and spectral qualities of informal settlements is an important precursor for evaluating success toward improving the lives of 100 million slum dwellers worldwide, as pledged by the United Nations Millennium Development Goal Target 7D.

A variety of measures were tested based on surface material spectral properties, texture, built-up structure, road network accessibility, and geomorphology from twelve communities in Guatemala City to reveal statistically significant differences between informal and formal settlements that could be applied to other parts of the world without the need for costly or dangerous field surveys.

When information from satellite imagery is constrained to roads and residential boundaries, a more precise understanding of human habitation is produced. A classification and regression tree (CART) approach and linear discriminant function analysis enabled a variable dimensionality reduction from the original 23 to 6 variables that are sufficient to differentiate a settlement as informal or formal.

The results demonstrate that the entropy texture of roads, the degree of asphalt road surface, the vegetation patch compactness and patch size, the percent of bare soil land cover, the geomorphic profile convexity of the terrain, and the road density distinguish informal from formal settlements with 87-92% accuracy when results are cross-validated.

The variables with highest contribution to model outcome that are common to both approaches are entropy texture of roads, vegetation patch size, and vegetation compactness suggesting that road surfaces and vegetation provide the necessary characteristics to distinguish the level of informality of a settlement. The results will assist urban planners and settlement analysts who must process vast amounts of imagery worldwide, enabling them to report annually on slum conditions. An added benefit is the ability to use the measures in data-poor regions of the world without field surveys.

# INTRODUCTION

The intent of the United Nations Millennium Development Goal 7, Target 7D is "to have achieved a significant improvement in the lives of at least 100 million slum dwellers" by 2020 (UN-HABITAT, 2003). In 2001, 924 million people, or 31.6% of the world's urban population, lived in slums and accounted for 43% of the urban population in developing countries (UN-HABITAT, 2001). In 2007, the number of slum dwellers worldwide crossed the one billion mark, with 1 of every 3 city-dwellers worldwide lacking basic services such as clean drinking water or adequate living space (UN-HABITAT, 2006; Martínez et al., 2008). In Guatemala, where this research is focused, 61.8% of the urban population was estimated to reside in slums (UN-HABITAT, 2001). In 2010 the UN HABITAT reported significant progress toward meeting this goal, illustrated in Figure 1, but acknowledged it is insufficient to counter the continued growth of informal settlements in the developing world (UN-HABITAT, 2010). The number of slum dwellers in absolute terms is still expected to grow to 889 million by 2020, and the MDG target 7D is likely to be revised to address percent improvement rather than absolute numbers of inhabitants (United Nations, 2010, p. 64). The explosive growth of these populations has created living conditions that contribute to poor health, high mortality, excessive crime and economic degradation with inadequate response mechanisms during times of natural disaster (National Research Council, 2007b; Owen, Obregón, & Jacobsen, 2010).

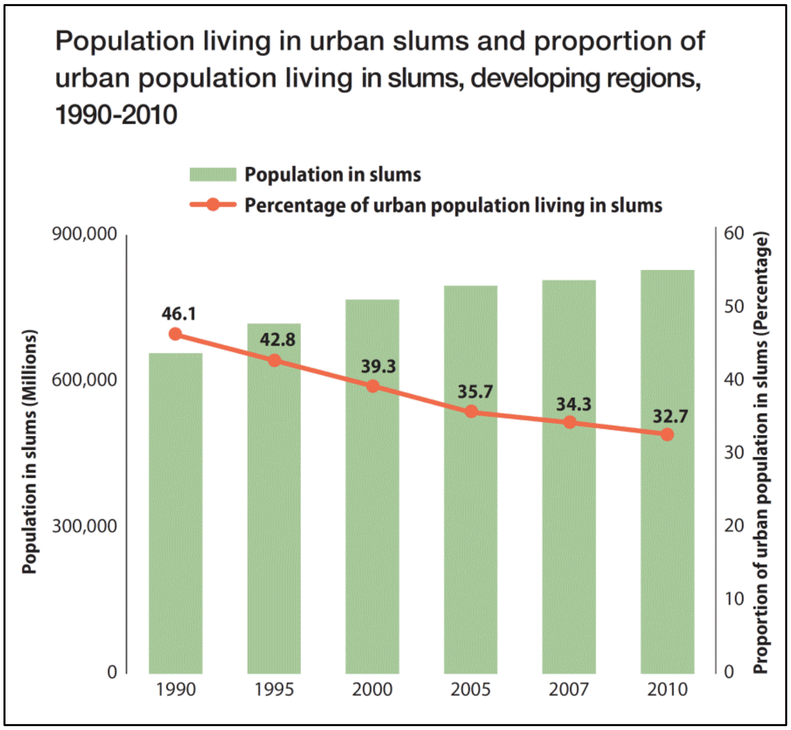



**Figure 1 Change in Urban Slum Population as Percent of urban population (United Nations, 2010)**

Using GDP per capita, Rice (2008) evaluated the under-five mortality rates of the 15 best-performing and 15 worst-performing developing countries. Rice was able to demonstrate a positive correlation between the worst-performing countries' urban slum population and child mortality (28.4 percent vs. 5.9 percent in the best-performing countries), and a negative correlation with socio-economic level (Rice, 2008). Martínez et al. (2008) revealed a positive correlation between child mortality and poor environmental conditions such as overcrowding, a lack of safe water and inadequate sanitation. A strong positive relationship therefore exists between under-five mortality and the proportion of a city's population living in slums.

Slums, known as 'informal settlements', must be characterized in order to assess their vulnerability to natural and man-made hazards, adequacy of urban infrastructure, estimate traditionally marginalized populations, and inform sustainable development in an era of increasingly limited natural resources.

The demographic shift revealing that more of the world's population lives in urban instead of rural areas occurred in late 2007 (UN-HABITAT, 2008). As migration trends continue from rural areas to cities, urban planners must accurately monitor a large number of urban areas to determine where conditions are worsening or improving. Reliable spatial indicators of informal settlements are needed to focus international aid and to help design improved infrastructure and community services. The UN-HABITAT is the only organization that monitors Millennium Development Goal performance worldwide, using methodology "structured on collaborative data collection between national, local and metropolitan governments in each country" and it has become an enormous undertaking (Hoornweg et al., 2007; European Commission JRC, Int'l Centre for Remote Sensing of Environment, & ISPRS, 2009). These urban monitoring projects have also relied heavily on expensive field-based surveys that are difficult to undertake because of the need for direct observation often in dangerous areas (Galeon, 2008; Rice, 2008; Yeh & Li, 2001; UN-HABITAT, 2010).

Efforts to quantify characteristics of slums or informal communities in comparison to their formal counterparts are continuing. In addition to each country's national datasets that are collected without common standards, the following statistical sources are currently used for measuring slums: the UN's World Urbanization Prospects; WHO/UNICEF Joint Monitoring Programme on Water Supply and Sanitation (JMP); demographic and health surveys (DHS); Multiple Indicator Cluster Survey (MICS); the UN-HABITAT's Urban Inequities Surveys (UIS) and census data when available (Sliuzas, Mboup, & de Sherbinin, 2008).

## Research Objectives

The three main objectives of this research are:

1. To test the hypothesis that foundational measures derived from roads, vegetation, soil, image texture, and geomorphology evaluated with multivariate methods can distinguish settlement structure of residential areas.

2. To determine if imagery and elevation data alone are sufficient to distinguish informal from formal settlements.

3. To develop a small set of statistically significant indicators that distinguish settlement type to be used by organizations such as the UN HABITAT and urban planners without the need for field work or surveys.

The UN-HABITAT urban slum indicators are aggregated to the national and continental level, and methods unfortunately vary by each country's own national data-gathering strategy. This research advances the ability to distinguish informal (slum) from formal areas by analyzing shape (form), texture, road networks and built-up areas using geographic information systems (GIS) and remote sensing image analysis.

In 2008, the combination of semi-automated information extraction from very high resolution (VHR) satellite imagery, expert knowledge applied to GIS analysis, and fusion of sensor and ancillary data were recommended as the most effective approaches for modeling and measuring informal settlements (Sliuzas et al., 2008). Myint et al. (2006) mention the limitations of traditional spectral approaches that only identify homogeneous features without concern for shape, asserting that "textural and spatial algorithms [should] measure both the variance within and the geometric configuration of whole urban objects, respectively". This dissertation research treats the residential area and its component roads and intervening spaces as major contextual pieces of the physical urban neighborhood. It is accomplished without using ancillary information (e.g. spatial data infrastructure in the form of urban planning data and regional GIS databases), which is often not available in developing countries.

To lay out the basis for this research the problem will first be described in detail, followed by a definition of informal settlements and a characterization of the settlement process and pattern in Latin America. Then the methods of informal settlement modeling in the scientific literature will be critiqued and a connection will be made to other informal settlement measures previously proposed, as well as to sprawl measures. The most relevant measures found in the literature will be described, followed by an evaluation of methodological shortcomings. Next is a description of the datasets used, research methods applied, and an overview of the proposed indicators. A results section reveals indicator performance between settlement types and by sampling methods, then conclusions are summarized, followed by limitations of this research, and recommendations for future research.

## Informal Settlements Defined

The study of informal settlements, also known as slums, squatter settlements, congested communities, shanty towns, self-help housing (Ward & Peters, 2007) and spontaneous settlements (Rapoport, 1988) requires a formal definition followed by a discussion of the process of their formation and pattern. The conventional definition approved by the United Nations defines a slum household as:

> "…a group of individuals living under the same roof lacking one or more of the following conditions:
>
> - Access to improved water: This indicator concerns the access to sufficient water for family use, at an affordable price, available to household members without being subject to extreme effort.
> - Access to improved sanitation: This indicator concerns access to an excreta disposal system, either in the form of a private toilet or a public toilet shared with a reasonable number of people.
> - Sufficient living space: A dwelling unit is considered to provide a sufficient living area for the household members if there are fewer than three people per habitable room.
> - Durability of housing: This indicator shows the percentage of households living in a housing unit considered as 'durable', i.e. those houses built on a non-hazardous location and with a permanent structure that is adequate to protect its inhabitants from the extremes of local climatic conditions such as rain, heat, cold, humidity.
> - Security of tenure: Evidence of documentation to prove secure tenure status or de facto or perceived protection from evictions."
>   (Martínez et al., 2008)

In Brazil, where much research on informal settlement process and pattern has been conducted on *favelas* (the Brazilian name for slums), such settlements are defined as "areas of unorganized expansion… where no urban design can be observed" (Junior & Filho, 2005).

Terms used to define informal settlements in Guatemala where this research is focused include: *barrio marginal* (marginal neighborhood), *palomar* (tenement), *toma* (land grab), *invasores* (invasion of lands with or without tacit permission), and *champa* (precarious housing, self-built with waste material) (Valladares Cerezo, 2003; UN-HABITAT, 2003). The primary reasons for slum formation in Guatemala are listed in Table 1.

**Table 1 Slum Formation in Guatemala**

| **Historical Reasons for Slum Formation in Guatemala** |
|---|
| Settlement on private lands not authorized by the municipality because of contamination or hazardous location |
| Invasions of state or private lands |
| Low cost government housing designated for the poor where green space is occupied by *invasores* (invaders) |
| Rural villages absorbed by city growth, but lacking city infrastructure services |
| Dwelling units leased from private owners but frequently lacking services (water, sanitation, electricity) |
| Permitted occupation of state-owned land |

(UN-HABITAT, 2003)

## History of Informal Settlement Process and Pattern in Latin America

The formation and expansion of informal settlements in urban areas generally follow a predictable sequence where the defined process of expansion creates a predictable pattern or structure (Bhatta, Saraswati, & Yopadhayay, 2010). Slums are initiated when population growth exceeds available shelter and land, and newer migrants settle peripherally (Ooi & Phua, 2007). In Latin America, squatter settlements usually start through illegal subdivision of land which is marketed to the urban poor and then subdivided into plots for which legal title is ultimately not provided while zoning laws are ignored and public infrastructure services remain undeveloped (Dickenson, 1996). Such settlements are first created on the periphery of the city where unclaimed land may still be available, often in undesirable locations near natural or other hazards.

The classic pedagogical model of the Latin American city (Griffin & Ford, 1980) is depicted in Figure 2 and directly addresses the integration of process with pattern. The Griffin and Ford model is comprised of a central business district (CBD), a dominant "elite" residential sector built around a commercial *spine*. The elite residential sector consisted of those residents who were socially worthy indicating that social and economic status decreased with distance from the CBD (Griffin & Ford, 1980). A series of concentric zones with residential quality proportional to city center proximity then radiate outward from the urban core (Griffin & Ford, 1980). In this model, the spine and elite residential sector, termed the *spine/sector*, "constitutes a morphological response to the limited ability to extend

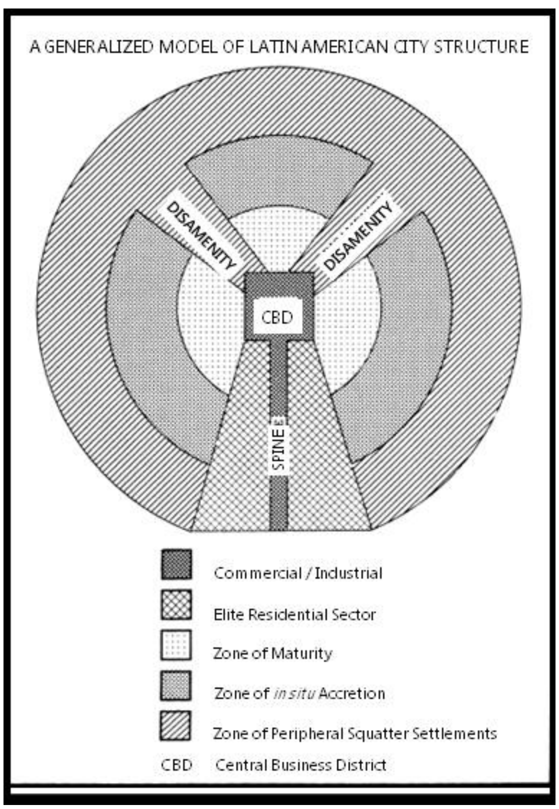



**Figure 2 Model of Latin American City Structure (Griffin and Ford, 1980)**

urban services” and “to the inability of Latin American cities to accommodate massive growth” (Griffin & Ford, 1980) . The spine of the model contains the elite housing sector supported by modern infrastructure services. In Latin American cities the three residential zones or rings are 1) a zone of maturity, 2) a zone of accretion, and 3) a zone of informal or squatter settlements. The size of the zones is a function of the in-migration rate (primarily rural to urban) and the residents’ ability to improve the quality of their dwellings, combined with the city’s ability to expand urban infrastructure and services to support them. The zone of maturity is fully served by infrastructure services and contains improved and upgraded housing generally constructed of large, solid, concrete or brick houses (Griffin & Ford, 1980). The next zone of ‘in situ accretion’ is in some process of maturing while in a constant state of upgrade and construction, with main streets normally paved but side streets unpaved. Development in the in-situ accretion zone appears chaotic due to the variation in the degree of improvement and the beginnings of public infrastructure construction.

In The Mega-City in Latin America, Gilbert acknowledges this in-situ improvement, stating that “what begins as a shanty soon becomes a consolidated house” (Gilbert, 1996). In the Griffin and Ford model, the peripheral squatter settlements exist outside the zone of accretion, unlike the Anglo-American concept of the “suburbs” with manicured lawns, parks, paved streets and public services. Griffin & Ford (1980) describe these squatter areas as devoid of vegetation, packed with small houses, unpaved streets, non-existent public infrastructure services, and located far from job sites at urban centers. Field visits to Guatemala City by the author of this research confirm this model is still relevant today.

Government-constructed low-income settlements built specifically to house the poor either during massive relocation efforts or in response to natural disasters such as Hurricane Mitch in 1998 or the Guatemalan earthquake of 1976 are unique situations of forced relocation. In those situations, the building materials and dwelling footprint are initially homogenous, with varying levels of disrepair as time passes or when government investment subsequently ceases.

## Spanish Colonial Influence on Guatemalan City Structure

Spanish Colonial-influenced Guatemalan cities follow the traditional vernacular architecture of the regular grid pattern with a cathedral, government buildings and the town square in the center of the town with more affluent residences in close proximity. Instead of suffering urban decay the city center in Latin American cities is often a desirable location for the wealthy to reside. Traditionally, the poor settled on the periphery in agricultural areas providing needed labor. This historical configuration can be seen in declassified 1941 images of Guatemala City where mansions were located in the city center near the Presidential Palace, Cathedral and Spanish Colonial Military fortress while much smaller dwellings are embedded in the surrounding agricultural landscape (Declassified NIMA imagery of British Honduras and Guatemala, 1941) In contrast to this model, urban blight and deterioration in the US and Europe forced the housing elite to create settlements on the outskirts of the city center, creating *sprawl* communities, though specific targeted cases of urban renewal are the exception.  In post-modern times, affluent areas in Latin American cities are still found in the central business district as the prior models describe.

These differences are in stark contrast to the North American or First World phenomenon of suburban sprawl[1]. Suburban sprawl is a condition of developed countries heavily reliant on automobile transportation and a mature interstate highway system. As Barros states, “while the problem of urban growth in Europe and North America has been formulated in terms of sprawl, in the Third World and more specifically in Latin America, the main focus has been the rapidity of growth of cities as well as the social inequalities in urban space produced by this process” (J. Barros, 2004). The lack of a sophisticated transportation infrastructure and the cost of vehicle transportation prevents the vast majority of the working poor from living very far from the city center, and the sprawl concepts of *leapfrog development* and *highway strip development* along secondary highways are rarely applicable (Hasse, 2002). Heuristic processes that contribute to the evolution of informal settlements will be described next.

## Heuristics of Growth

Informal settlement growth has been documented by numerous researchers and a variety of heuristics have been used to describe it. *Clustering*, *densification* and *peripherisation* are examples of such descriptors. Clustering is the initial formation of self-built dwellings near one-another, densification is in-fill or in-situ expansion or accretion. Peripherisation is a particular type of growth often found in Latin American cities (Barros, 2004). In peripherisation, settlements (usually low-income) form on the

[1] First World refers to developed countries as opposed to Third World, developing countries.

periphery of the city, are then incorporated into the city through longer term expansion of city boundaries, eventually a few of these settlements become 'recontextualized' within the urban fabric and may be occupied by groups with higher economic status, so that newer low-income settlements continue to appear on the periphery (Barros & Sobreira, 2008). These processes explain why informal settlement enclaves are sometimes found in the same general vicinity as higher-rent areas which have grown around them.

There is a need for improved understanding of the process, and a need to analyze space-filling and growth patterns over time. Barros and Sobreira (2008) contend that spontaneous settlements "can be classified according to locational and morphological characteristics in 'inner city' and 'peripheral settlements'". But they also acknowledge there is no generally accepted theory of spontaneous settlement location. Factors influencing migration into these settlements are availability of land, proximity to high-density mixed land use and proximity to job opportunities. A survey conducted of informal settlements in the Ivory Coast concluded that incoming residents are attracted by (1) the availability of land (31%), (2) proximity to employment (22%) and (3) strong social ties (20%) (UN High Commission on Refugees, 1996). Similar results were reported for the informal settlements on the urban fringe of Tehran, where respondents indicated the greatest reason for relocating to slums was either access to affordable housing (in Saleh Abad) or proximity to place of work (in Khatoon Abad) (Zebardast, 2006). Eventually the peripheral areas that first became settled are then consumed by urban expansion, remaining as pockets of poverty that continue to pack more densely (Sietchiping, 2000) .

The gradual creation of informal settlements is also described as *accretion* (J. Barros, 2004). Figure 3, derived from (Sliuzas, 2008), notionally depicts this process of informal settlement formation over time, capturing the concepts of *in-situ* accretion and haphazard densification.

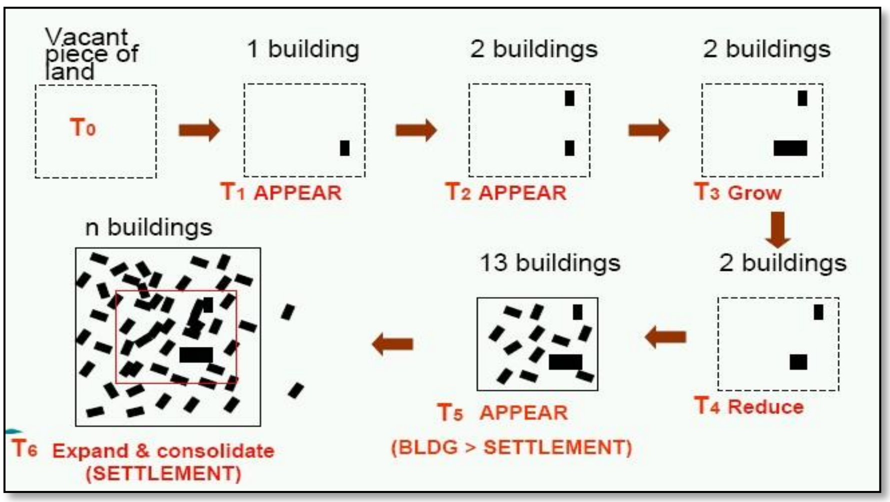



**Figure 3 Notional process of Informal Settlement Formation (Sliuzas, 2008)**

Informal settlement growth or process can be decomposed into four phases in mostly sequential fashion:

1. Land Invasion
2. Social Formation
3. Physical Consolidation
4. Urban Maturity

(Alsayyad, 1993)

As an example, land invasion in Cobán in north-central Guatemala by *invasores* has occurred gradually on the outskirts of town adjacent to the garbage dump on municipal land where the local government has chosen not to intervene (Cable, 2009). Social formation occurs when family groups move together and relatives or friends from the same original location follow. Physical consolidation occurs when external pressures or topographic constraints such as rivers or super highways prevent settlement expansion, causing continued infill. Urban maturity is the final phase often characterized by extreme lack of housing durability when the settlement infrastructure is not upgraded.

The response of the state to the development of squatter communities can take a variety of forms. The state may ignore, demolish, relocate, legalize or upgrade informal settlements (Alsayyad, 1993). Infill can be an organized invasion or gradual. Alsayyad (1993) concludes that “In Latin America, the most effective approach for squatters is to connect with the political system either through the official or oppositional parties and use electoral competitions to advance their interests”. Success in obtaining state-sponsored assistance is predicated on increases in scale of participation of the residents in the political process. The more involved they are, the more likely they will gain recognition (Alsayyad, 1993; Guisti de Pérez & Pérez, 2008).

The morphology of informal settlements is a result of clearly-defined processes common to Latin American cities that were initially constructed by Spaniards under the “Laws of the Indies”. The affluent areas are still found near the city center and the impoverished areas exist on the periphery and are gradually being consumed by urban growth and densification. This process is quite different than many developed cities where affluent communities are found on the outskirts of town (Thomas, Tannier, & Frankhauser, 2008; Pumain & Tannier, 2005; Thomas, Frankhauser, & Biernacki, 2008; De Keersmaecker, Frankhauser, & Thomas, 2003). The ability to distinguish formal from informal settlements in the same region provides the ability to measure quantitatively how the urban form has changed over time, and whether in fact such models as Griffin & Ford (1980) and Barros & Sobreira (2008) are applicable in the 21$^{st}$ century.

## The Need for Quantitative Spatial Measures of Informal Settlements

There is no doubt that slum areas are expanding in major urban centers of developing countries worldwide, but significant challenges remain in evaluating them and measuring expansion of built-up areas with scientific methods. Although shape-based measures (fractal dimension, lacunarity, mathematical morphology) and texture measures (landscape fragmentation analysis, grey-level co-occurrence measures) have been used to identify individual slum communities in the past two decades, minimal progress has been made on development of geospatial measures that can distinguish differences between informal and planned settlements using readily-available very high resolution (VHR) multispectral imagery. There is also little use of topographic-related measures to explore differences within settlements. Development of methods that fully exploit image processing and require minimal field work can be more efficient due to the extensive geographic coverage of VHR satellite imagery and can preserve the safety of researchers whose visits to these often dangerous areas can be minimized.

It should be acknowledged that some amount of field-based cultural and spatial observation is still required. In fact, this research was enriched by a number of field visits to Guatemala by the author (Jacobsen, Nelson, & Owen, 2011). However, physical entry into some settled regions can still be dangerous, and remote evaluation is often necessary. While the body of literature contributing to the measurement of formal settlements manifested as suburban sprawl continues to grow (see subsequent section 'Relationship to Sprawl Research'), conspicuously absent are summarized quantitative measures that may be applied to informal settlements.

The study areas in this research have been selected from urban sites in Guatemala, a Latin American country whose slum population is projected to increase by nearly 20 million in the years from 2010 to 2020 (UN Habitat, 2009). Despite its middle ranking in per capita GDP among poor countries defined by the United Nations, the proportion of slums in Guatemala is severely elevated, where 62 percent of urban dwellers were living in slums in a nation that also suffers from a high GINI inequality coefficient of 0.60 (UN Habitat, 2003)[2]. The slums of Guatemala City suffer an under-5 mortality rate of 56% and an acute respiratory infection rate of 20.2% in children under 5 (Martínez et al., 2008).

Limited access to GIS software and training (Bishop et al., 2002; Baltsavias & Mason, 1997) and a lack of spatial data (Busgeeth, Brits, & Whisken, 2008; Mason, Baltsavias, & Bishop, 1997; Hasan, 2006; Barros Filho & Sobreira, 2007) in developing countries creates challenges for planners and decision-makers. These groups must measure settlement growth, patterns and intensity as a foundational step toward improving these areas and the lives of their inhabitants using less labor-intensive and less costly means. Methods and indicators to characterize this type of urban growth could be used by planners in disaster management, slum monitoring and human-environment

[2] The GINI coefficient of 0 indicates perfect income equality among a population, whereas 1 indicates perfect *in*equality.

interactions as a way to monitor impacts on nearby natural resources (Owen, 2012). The United Nations Global Urban Observatory project has acknowledged there is a lack of data and an immaturity of applicable methodology to measure durability of housing because settlement variables are not examined collectively (Sliuzas et al., 2008). Serving as a peer review of the UN-HABITAT methodology on slum estimation, a Global Urban Observatory report also emphasized the need to "be able to identify and define slums spatially in a consistent manner to be able to use geographical targeting for slum intervention programmes" (Sliuzas et al., 2008). A comprehensive World Bank report summarizing the status of urban indicators worldwide also emphasized that new indicators must be "measurable and replicable, easily quantifiable, and systematically observable" and should "begin incorporating GIS techniques in future data collection to permit a better understanding of geographical changes over time" (Hoornweg et al., 2007, p. 13). Besussi et al. (2010) assert that "improvements in the resolution of satellite images have not been matched by commensurate improvement in the detail of socioeconomic data on urban distributions". This makes our understanding of the built form disjoint from the ability to measure details of intra-urban socioeconomic conditions (Besussi, Chin, Batty, & Longley, 2010).

Human settlements are comprised of three fundamental physical elements clearly visible in remote sensing imagery: buildings, roads and open spaces (Pesaresi & Ehrlich, 2009). This observation, and the need to exploit physical context, structure and spectral information together lead to the main hypothesis of this research, summarized here:

*Empirical measures of a settlement's physical structure can positively distinguish urban neighborhoods as formal or informal*.

Indeed, other research has also emphasized the need to evaluate structure and spectral information in tandem (Jin & Davis, 2005). Inhabitants' poverty level - a proxy for many other variables related to wellbeing - could later be derived from distinctions between settlement types, and conditional measurement scales could be developed for specific regions or urban areas. Proving the hypothesis of this research will create robust indicators to differentiate among settlement types without the reliance on expensive or dangerous field work that is difficult to undertake. This is confirmed in prior research where slum-dwellers are hesitant to provide a response to field-based surveys for "fear that anything they say might be used against them" (Galeon, 2008). A final complication is that a larger than normal sample size in informal settlements has been needed to justify statistical significance of survey results (National Research Council, 2007a). This "oversampling imperative" also underscores the need for alternative methodologies.

# LITERATURE REVIEW

This literature review section is divided into two focus areas: (1) informal settlement literature review, and (2) formal settlement literature review exemplified by suburban sprawl. The main objectives of the literature review are to determine whether prior research revealed geospatial and remote-sensing based indicators that could be used to discriminate settlement type (informal or formal) in the same image scene without the need for extensive field work or ancillary datasets. This section will review the locations of prior GIS and remote-sensing based informal settlement research and group those efforts by their approach. The relationship to prior measures of suburban sprawl will also be explored, followed by a summary of the limitations of their applicability to the current research problem. A listing of prior indicators that may be relevant is provided followed by suggestions for new indicators.

GIS became popular in the 1990's and was soon integrated with remote sensing for research into mapping and modeling informal settlements (Yaakup & Healey, RG, 1994; Bharathi & Lakshmi, 2005; Hasan, 2006; Vicente, Villarin, Galgana, Guzman, & de Mesa, 2006; Martínez et al., 2008; Dare & Fraser, 2001). Research was applied to problems of estimating the size of the settlement (geographic extent in the landscape), estimating the need for government services, estimating population through dwelling feature extraction supplemented by survey data, and included modeling to predict growth and understand the process of infill. The majority of this research requires prior ancillary GIS datasets, administrative boundaries, and field surveys. Despite the fact that many settlements can be detected through visual imagery interpretation, the human eye cannot quantitatively measure informal settlement properties. No prior research effort has performed settlement typology discrimination using geospatial metrics to differentiate formal and informal areas in the same region as a basis to model and estimate the intensity of slums. This research fills that gap.

## Location of Prior Informal Settlement Modeling Efforts

Informal settlement modeling and analysis with GIS is largely focused on developing countries where slums are prevalent, with much research focused in Brazil. Table 2 summarizes the locations of these studies.

**Table 2 Location of GIS-based Informal Settlement Studies**

| **Country** | **City** | **Author & Year** |
|---|---|---|
| Tunisia | Metropolitan Area Tunis | Weber and Pouissant, 2003 |
| China | Pearl River Delta | Yeh and Li, 2001 |
| Thailand | Bangkok | Thomson and Hardin, 2000 |
| Malaysia | | Yaakup and Healey, 1994 |
| Philippines | Manila | Vicente et al., 2001; Galeon, 2008 |
| Pakistan | Karachi | Hasan, 2006 |
| Morocco | Marrakech | Baudot, 1993 |
| South Africa | Cape Town | Baltsavias and Mason, 1997; Li et al., 2005; Hofmann, 2001 |
| Kenya | Voi | Hurskainen and Pellikka, 2004 |
| | Nairobi | Pesaresi and Ehrlich, 2009 |
| Ghana | Accra | Stow et al., 2007 |
| Tanzania | Dar es Salam | Mayunga, Coleman, & Zhang, 2007; Augustijn-Beckers, Flacke, & Retsios, 2011 |
| Sudan | Northern Darfur | Stasolla and Gamba, 2008; Sulik & Edwards, 2010 |
| Argentina | Rosario | Martínez, 2009; Martínez, 2004; Hall et al., 2001 |

| | | |
|---|---|---|
| Brazil | Mato Grosso | Zeilhofer and Topanetti, 2008 |
| | Rio de Janeiro | Junior and Filho, 2005; Hofmann et al., 2008 |
| | Recife | Filho and Sobreira, 2008 |
| | Belo Horizonte | Kux and Araújo, 2008 |
| Greece | | Ioannidis et al., 2009 |
| Turkey | Istanbul | Dubovyk et al., 2011 |
| India | Bengal | Bhatta et al., 2010 |
| | Delhi | Niebergall et al., 2007 |
| México | Tijuana | Monkkonen, 2008 |
| | Morelia City | López et al., 2001 |

No studies have incorporated VHR satellite imagery into informal settlement modeling in Central American except for México (Monkkonen, 2008; López, Bocco, Mendoza, & Duhau, 2001), which is still considered North America.  Figure 4 depicts the locations of remote sensing analysis of slum areas in Latin America identified with green boxes.

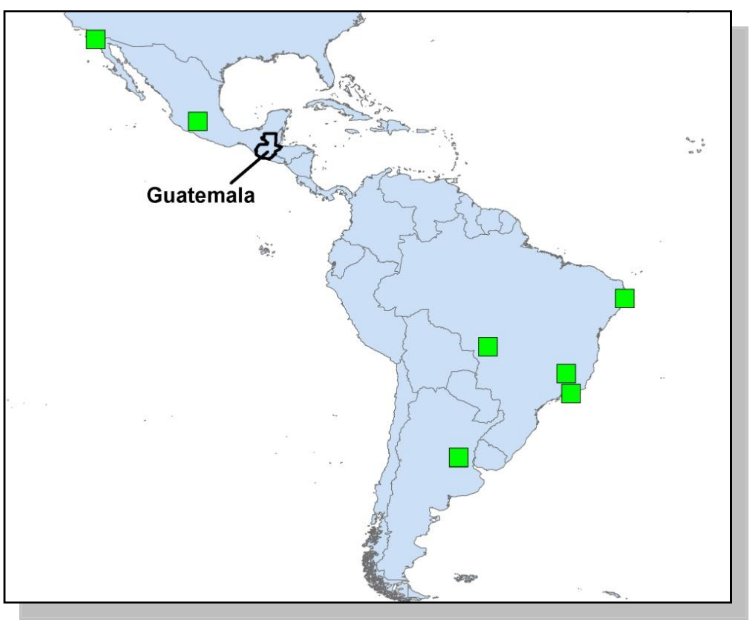



**Figure 4 Locations of GIS-based Informal Settlement Research - Latin America**

In Monkkonen (2008) the feasibility of using Google Earth imagery as a tool to confirm the spatial extent and boundary delineation of informal settlements was explored. At the time of publication in 2008, he determined that Google Earth imagery as a tool was not yet sufficiently mature for reliable boundary delineation due to the absence of image date information and the fact that Google did not yet publish details on imagery sources. López et al. (2001) focused on measuring the relationship between growth of urban land cover and growth of population reported in the Méxican census. These studies were limited to evaluating the shape of built-up areas using land use/ land cover classification without examining individual neighborhood components or structural interrelationships of their component parts, revealing the absence of remote sensing-based informal settlement analysis in Central America.

## Informal Settlement Literature Review

Earlier research in informal settlement modeling focused on boundary delineation and land use/land cover classification and was limited by the spectral and spatial characteristics of older satellite sensors at the expense of structural characterization. As sensor spatial resolution has increased, research methods have naturally improved. In general, methods used in the literature for spatial modeling of informal settlements are grouped as follows:

**Table 3 Informal Settlement Modeling Methods**

| Grouping | Methods Used |
|---|---|
| Direct Mapping and Demographic Studies | • Cadastral Micro-scale maps<br>• Cartographic Products<br>• Field Surveys |
| Dynamic Growth Models | • Cellular Automata<br>• Complexity Models<br>• Predictive and Temporal Models<br>• Logistic Regression |
| Multi-scale Approaches | • Fractal Analysis<br>• Lacunarity |
| Object-based Image Analysis and Building Feature Classification | • Digital surface models, building heights, settlement structure height (LIDAR and RADAR-based)<br>• Dwelling object-based approaches<br>• Spectral properties of roofs to extract dwellings |
| Image Texture and Mathematical Morphology Measures | • Grey-level co-occurrence measures<br>• Impervious Surface detection<br>• Mathematical Morphology of objects |
| Socioeconomic Measures and Landscape Analysis | • Entropy<br>• Patch Dynamics |

This section explores these literature groupings, pointing out their contributions and limitations.

### Direct Informal Settlement Mapping and Demographic Studies

GIS is described as a necessary tool for slum upgrading (Vicente et al., 2006; Hall, Malcolm, & Piwowar, 2001; Melesse, 2005; Lemma, Sliuzas, & Kuffer, 2006) and has been used to support cadastral efforts and to measure areal extent of homogeneous slum neighborhoods (Barry & Rüther, 2001). Many studies aimed at mapping the slum areas delineate a rectangular boundary in imagery and simply measure change over time. Some of these studies predict where growth might occur while others attempt to gain an understanding of a slum area by cadastral mapping and then measuring the lack of services through field surveys, which can be costly and labor intensive. A related branch of research also focuses on urban vulnerability and environmental degradation in slum areas (Pauchard, Aguayo, Peña, & Urrutia, 2006; Zeilhofer & Topanotti, 2008) caused by earthquakes, air pollution, and the [lack of] elimination of solid waste (Parker, 1995). The delineation of property boundaries and the accurate mapping of informal settlements are relevant to land tenure studies and the creation of city maps (Hasan, 2006). Martínez, et al. (2008) developed indicators requiring the additional step of analyzing demographic and health surveys of living conditions. The goal of some research that focused on counting dwellings was part of local planning efforts to estimate population, as in Galeon (2008) where random household sampling was conducted for the creation of an accurate basemap, and Qadeer (2000) whose focus was the evaluation of population density in developing countries to evaluate the phenomenon of a ‘ruralopolis’. There are many examples of advanced techniques to estimate population in informal settlements using GIS and remote sensing, but the need for field surveys, land records, or census data to supplement the analysis can be costly, may require local government permission, and may need compliance to strict institutional review board rules when humans are subjects of research.

### Dynamic Growth Models

Cellular automata have also been used to model growth trajectories of informal settlement areas in an attempt to understand the dynamics of change and visualize future urban forms. Cellular automata and complexity modeling support "the idea of a structure emerging from a bottom-up process where local actions and interactions produce the global pattern" (UN High Commission on Refugees, 1996). A cellular automata approach was tested by several authors to model growth of informal settlements (Sietchiping, 2000; Sietchiping, 2004; Barros, 2004).

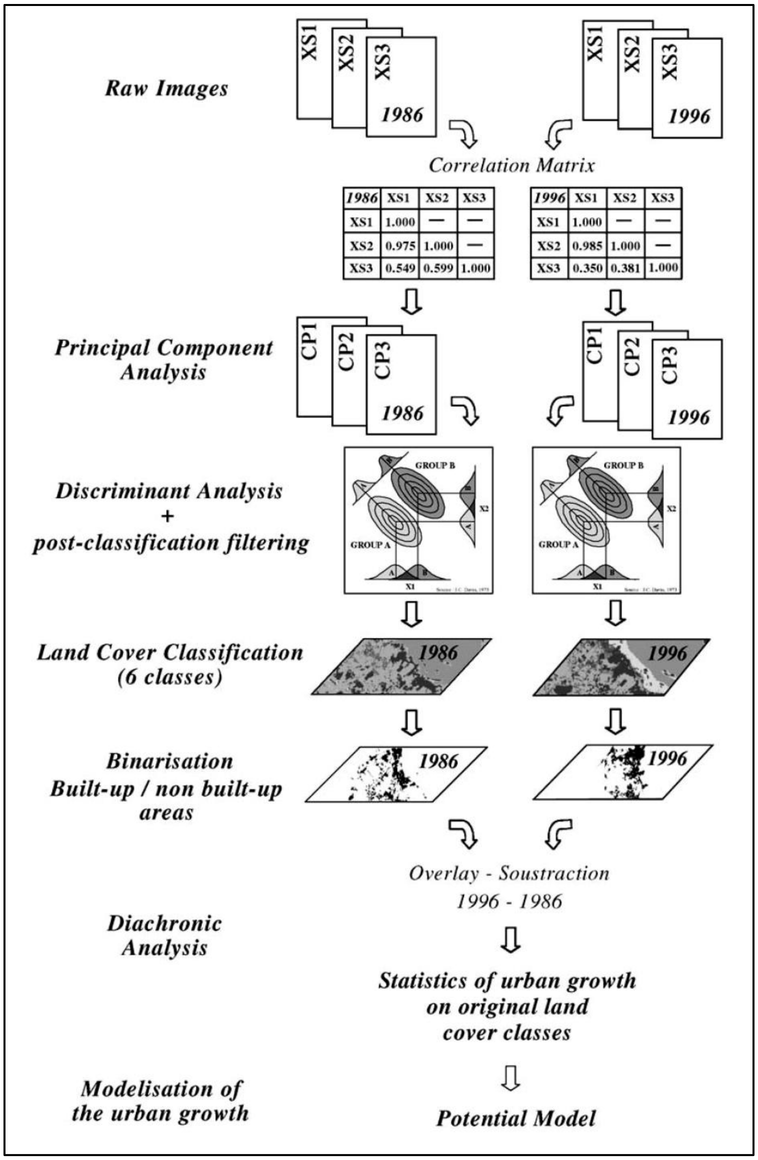
Agent-based models have also been used to simulate the dynamical process of infill and extension based on settlers' preference to limit distance to roads or footpaths, or maximize distance to flood zones and regions subject to inundation (Augustijn-Beckers, Flacke, & Retsios, 2011). Expansion is most evident in the early stages of settlement formation, while infill occurs as time progresses, though both processes are at play over the temporal dimension of many years (Augustijn-Beckers et al., 2011). Other simulations use agent-based models to predict how the land cover changes as slums proliferate in a city (Diuana et al., 2006; Mayunga, 2007). Although dynamic models are useful to simulate process and for predictive purposes, their contribution differs from the present research which addresses the need to first measure and quantify the physical elements before the change process can be modeled.

Sensors such as SPOT 4 & 5 multispectral and Landsat Thematic Mapper have frequently been used to evaluate land cover change (area and shape) at the city level but cannot be used discriminate dwelling structures due to their spatial resolution (10, 20 and 28.4m, respectively). Land cover change and city-level expansion in the informal areas of the Tunis metropolitan area were measured in one study, but with moderate resolution data, they were unable to evaluate the internal structure of these settlements (Weber &

Puissant, 2003). The authors generalized that informal settlement patterns can be recognized by the following:

- Lack of obvious planned structure
- Lack of vegetation in housing areas
- High building density

They tested a dynamic model that spatially weighted the interaction between man-made and natural growth potential by land cover class (high-density built-up, low-density built-up, spontaneous vegetation / forested, salt lagoon, and wetlands). The interaction potential between two time periods was based strictly on land cover class, with the ultimate goal of predicting where growth would occur in the succeeding ten years following their analysis. They applied a form of discriminant analysis and created binary images of "built-up" and "not built-up" to compute change, as depicted in Figure 5.




| 1986 | XS1 | XS2 | XS3 |
|---|---|---|---|
| XS1 | 1.000 | — | — |
| XS2 | 0.975 | 1.000 | — |
| XS3 | 0.549 | 0.599 | 1.000 |

| 1996 | XS1 | XS2 | XS3 |
|---|---|---|---|
| XS1 | 1.000 | — | — |
| XS2 | 0.985 | 1.000 | — |
| XS3 | 0.350 | 0.381 | 1.000 |

**Figure 5  Informal Settlement Modeling Methods (Weber & Puissant, 2003)**

Characteristics of residential areas were limited to high density and low density without evaluating roads, greenspace or relevant anthropogenic features.  The authors' use of principal components as variables to create land cover classes served as synthetic and indirect indicators for measuring these foundational parts of settlements.

The impact of slum upgrading plans on the lives of slum-dwellers was also evaluated (Martínez, Boerboom, & Sliuzas, 2007). The Martínez et al. (2007) work expanded the use of modeling tools to predict the impact of slum improvements in sanitation but they did not integrate GIS or imagery in their work. More recently, researchers built a logistic regression model to predict future locations of informal settlements in Istanbul, Turkey, aided in part by the inclusion of detailed urban planning, census, and demographic data for the study area, which is not readily available in many developing countries (Dubovyk, Sliuzas, & Flacke, 2011). Exploring growth patterns and urban extent are valuable in monitoring change, but the intrinsic value of first knowing what structural characteristics to measure must be foundational to such efforts.

### Multiscale Approaches

Several approaches have attempted to estimate the appropriate spatial scale to model informal settlements, with the view that techniques appropriate at multiple scales could be more effective and more broadly applied. Table 4 lists the spatial scales that are recommended based on the target of analysis and suggested spatial resolution.

**Table 4 Recommended Spatial Scale based on target of analysis (Stein, 2008)**

| Target of Analysis | Spatial Scale | Pixel Spatial Resolution |
|---|---|---|
| City | 1:15,000 | 2.4 – 5.8m Imagery |
| Ward | 1:10,000 | 2.5 – 5.8m Imagery |
| Enumeration Area (Census,Survey Unit) | 1:5,000 | 2.4 – 4m Imagery |
| Slum Neighborhood | 1:2,000 | ≤ 2.5m Imagery |
| Individual Household | 1:500 | Survey Instrument |

The spatial scales recommended for neighborhood characterization push the limits of readily available commercial remote sensing products. Based in part on Table 4, the scales of analysis useful to model characteristics of informal settlements are regional, neighborhood, feature object, and individual household. Regional scale studies focus on boundary delineation and growth, depicting settlements within the context of an urban area, city or ward. In the literature, land use and land cover segmentation was normally performed at a regional scale. Neighborhood scale analysis focused on identifying housing clusters. Here it is important to note that housing clusters are embedded within roads and natural land qualities such as slope or convexity, embedded vegetation, soil and building materials. Neighborhood scale is appropriate for vulnerability analysis and understanding dynamics of neighborhoods. Feature Object scale refers to the study of individual dwellings, instead of clusters of dwellings, in relation to each other and to their surroundings. Individual household scale is needed to evaluate the United Nations slum characteristics of security of tenure, access to sanitation, and access to clean water by tying a household and their actions to a place. Sample surveys of households and cadastral mapping through more intense field work are needed at this scale (Lemma et al., 2006). This dissertation research is conducted at the neighborhood scale.

#### ***Fractal Dimension of Settlement Structure.***

Evaluating the fractal dimension of settlement areas, both formal and informal is considered a scale-invariant technique (Thomas, Frankhauser, et al., 2008; Tannier & Pumain, 2005; Cooper, 2005). In one study, the fractal dimension of built up areas at varying scales showed a decrease in value as the grid size, or unit of analysis, increased, and the authors concluded the interpretive value of fractal dimension was retained at all grid sizes (Wallace, Morris, & Howarth, 2004). Fractal analysis generally applies fractal theory at the regional scale, but has not been used in the published literature to test differences in settlement structure between formal and informal. Tannier & Pumain (2005) state "the main advantage of fractal geometry is to provide a model of reference which seems more adapted than Euclidean geometry to the description of spatial forms created by societies: features of heterogeneity, self-similarity and hierarchy are included from the very beginning in fractal structures".

In both Tannier and Pumain (2005) and Thomas et al. (2008), fractal analysis of settlement structure was computed after the morphological transformation of *dilation* was performed on the image, which does not preserve the initial shape of the built-up areas, but did allow the authors to measure scale variations using the ratio of fractality of the settlement's edge to its built-up surface. The $D^{Surf\text{-}dil0}$ represented undilated fractal dimension of the built-up surface areas of a settlement. It was found that values of $D^{Surf\text{-}dil0}$ close to 2.0 represented homogeneity while $D^{Surf\text{-}dil0}$ close to 1.0 corresponded to accentuated street/ribbon village patterns (more dendritic in nature) (Thomas, Frankhauser, et al., 2008). These authors focused on edge patterns of settlements and on comparing urbanization in the central parts of settlements with their peri-urban counterparts on the urban fringe. The authors used ancillary socio-economic information to include land rent and median income. High rent and high income were proxies for higher socio-economic status in census units called *communes* in Belgium. However, their analysis was applied to a first world country and was not tested in developing cities.

One additional fractal-related study of urban morphology found higher fractal dimension of the urban surface near the city center, suggesting that the city center is more "homogeneous and densely built" (De Keersmaecker, Frankhauser, & Thomas, 2003). A measure of diversity proposed that "housing in the periphery is not planned and its structure is spatially more heterogeneous" with lower fractality (De Keersmaecker et al., 2003). Their work was also applied to a first world country where the pattern in the city center resulted from a greater degree of urban planning and regularized spatial arrangement, and could therefore be described as more homogeneous. It is interesting to note that in Brussels, higher rents are associated with peri-urban areas (De Keersmaecker et al., 2003). This is the reverse of the Latin American city model mentioned previously. Fractal analysis of urban areas has thus been applied at the city level and found to be higher in urban first world regions, but there is no mention of its use in developing cities.

Another scale-related measure found in the literature is lacunarity. Where fractal shapes are a measure of self-similarity at all scales, lacunarity is expected to change by scale, and is roughly described as a measure of image 'gappiness'. Lacunarity has been applied specifically to model the pattern of informal settlements.

#### ***Lacunarity of Settlement Structure.***

Lacunarity, represented as $\Lambda$, is derived by comparing built-up areas of different types of settlements. Simply stated, lacunarity measures gaps that vary by scale, and is a measure of translational or rotational invariance or heterogeneity in an image. Lacunarity supplements fractal dimension to characterize settlement patterns and has been used in the literature to evaluate informal settlement housing, to compare neighborhood racial segregation, and to determine whether it can help improve classification accuracy (Filho & Sobreira, 2005; Filho & Sobreira, 2007; Filho & Sobreira, 2008; Junior & Filho, 2005; Wu & Sui, 2001; Myint et al., 2006). After converting an image to binary (black and white) so that built-up areas or housing become foreground pixels (white), lacunarity is used to identify the degree of heterogeneity or homogeneity found in an image (Owen, 2011). Additionally, lacunarity has been applied to grayscale images. This requires greater computing power but may not necessarily provide better results. In a general sense, lacunarity represents the variation in foreground pixel density over various box sizes moving over the image window (Karperien, 2007).

The most-studied lacunarity algorithms include the Gliding Box algorithm and the Differential Box-Counting (DBC) algorithm. Originally, the gliding box algorithm was recommended over the box-counting method in areas with limited data samples when evaluating geochemical data as multifractals (Cheng, 1999). In contrast, the DBC algorithm was found to correctly classify 90% of image sub-scenes in the region of Recife, Brazil as informal settlements, versus only 80% correctly identified using the Gliding Box algorithm (Filho & Sobreira, 2008). These image sub-scenes are rectangular in shape, and do not follow the natural irregular contours of settlement boundaries. In other work, an “inhabitability index” of settlements in Brazil was developed that concluded the DBC algorithm applied to binarized Quickbird images derived from greyscale was best able to discriminate texture in urban areas of differing inhabitability conditions (Filho & Sobreira, 2007). In that case, slum areas exhibited lower lacunarity values resulting from stated "lower permeability" and lower separation of structures, or less "gappiness". The authors assigned an inhabitability index based on socioeconomic values of a census region's central pixel, aggregated to the geographic coordinates of its regional capital, and then developed a kriged surface. This artificially assumed that economic status as a measure of inhabitability can be derived from census data that have been aggregated to a reporting point. It also assumed that inhabitability is inversely proportional (to a scaling factor) to distance from the central pixel, but that inhabitability is also related by proximity to a nearby capital's mean socioeconomic level. Small samples were then selected randomly and assigned high or low inhabitability based on the kriged values they overlaid. The samples were histogram-equalized and converted to binary. The authors did not classify the built and non-built environments, which could potentially have an impact on ability to delineate urban features. Although this novel method is interesting in its use of lacunarity to evaluate urban texture, the kriged inhabitability index may not serve as a true proxy for socio-economic status (SES) because it is unclear whether SES is related spatially to the centroid of the census unit.

Lacunarity has also been shown to vary by scale and by settlement type using extracted building shapes in rectangular-shaped image subsets (Junior & Filho, 1997). Higher lacunarity values were reported in the more regularized settlements found in the city center and not the squatter/informal settlements (Junior & Filho, 1997). Myint et al. (2006) reported a different result, finding higher spatial heterogeneity, as expected in informal communities, resulted in higher lacunarity values. The conflicting results found in the informal settlement lacunarity literature underscore the need for further study.

### Object-based Image Analysis and Building Feature Classification

The study of the urban form, or urban morphology theory, developed historically using the lower spatial resolution imagery of Landsat TM (30m) and SPOT-HRV (20m) sensors during a period when pixel-based analysis was the focus and multiple small informal dwelling structures fit into a single pixel. Imagery available from these sensors contributed to the study of built-up areas or urban agglomerations and their growth (Aplin, 2003; Wilson, Hurd, Civco, Prisloe, & Arnold, 2003; Hofmann, 2001) and were measured as clusters of development (Weber & Puissant, 2003) and not as individual feature objects. The pixel-based approach for the study of informal settlement built-up areas has now given way to the object-based approach (Hurskainen, 2004). The availability of very high resolution (VHR) satellite imagery such as DigitalGlobe Quickbird (0.6m panchromatic), GeoEye-1 (0.4m panchromatic), and IKONOS (4m) has enabled greater focus on object-based image analysis, or OBIA (Hay & Castilla, 2006; Blaschke & Lang, 2006) that identifies discrete features such as buildings, roads (Lopez-Ornclas & Flouzat, 2008) and vehicles at sub-meter spatial scales (Marangoz, Oruc, & Buyuksalih, 2004). The ability to extract objects and measure their relationships can offer valuable information for the integrated study of informal settlements, but limitations mainly deal with excessive clutter and heterogeneity of such settlements, as illustrated in Figure 6.

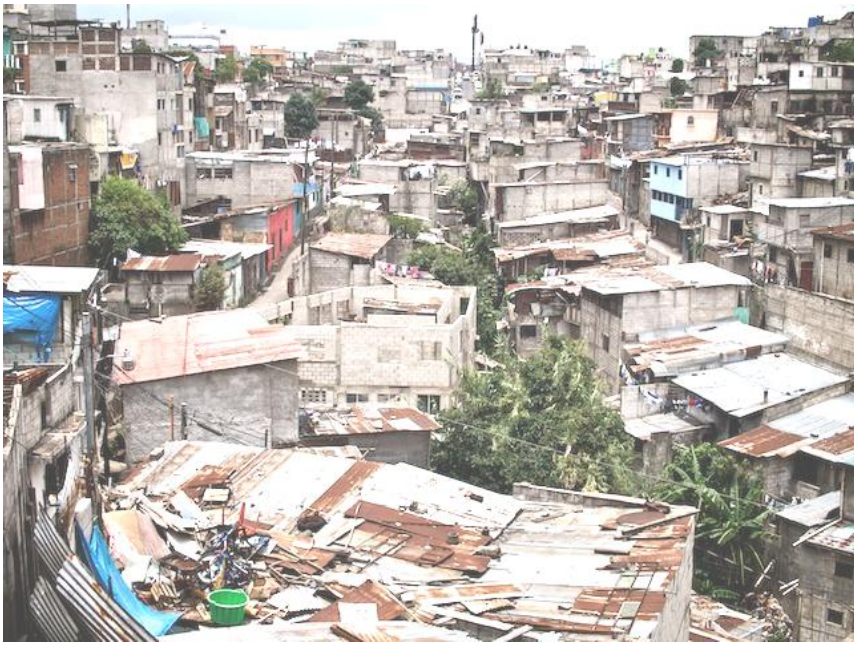


**Figure 6 Informal Settlement Clutter, La Limonada slum, Guatemala City (Evertsz, 2009)**

Buildings are the minimum mapping unit of an informal settlement, and their extraction has been modeled through energy functions called 'snakes' (Rüther, Martine, & Mtalo, 2002; H.-Y. Li, Wang, & Ding, 2006; Mayunga, Coleman, & Zhang, 2010) or shadows (Baltsavias & Mason, 1997) and with digital surface models in the determination of building heights to estimate and characterize slum populations (Galeon, 2008; Qadeer, 2000). A foundational work by Baltsavias and Mason (1997) spurred research on informal settlement measurement and modeling by using aerial half-meter film-based photographs of informal settlements in South Africa that were georectified by their roof corner geometries (Baltsavias & Mason, 1997). The authors used cues in object space (elevation contours and shadows) to extract the outlines of informal settlement shacks and then hypothesized their height. The authors' list of common properties of informal settlement dwellings in Cape Town, South Africa included the following:

- Single storied structures with primarily flat roofs (very few hip roofs)
- Simple geometries (85% of dwellings were 4-sided and only 15% 6-sided or more)
- Roof corner geometries deviating from orthogonality by 30° or more
- Small size (approximately 4 x 4m) with smaller-sized outhouses
- High diversity and texture of construction materials (e.g. plastic, iron sheeting, timber) with variable textures and colors for individual dwellings and for the settlement area in general
- High building density with only 2-3m separation
- General lack of vegetation

(Baltsavias & Mason, 1997)

To date, no work has attempted to quantify the effectiveness of these or similar attributes collectively in a manner that correctly identifies settlement type. The authors' stated goal was to map the settlement areas and assess the utility of a particular film-based camera product for photogrammetric mapping of the informal dwellings within a neighborhood. This effort was one of the few early ones applying remote sensing to the study of informal settlements using aerial photography – considered a very costly means of measuring the settlements compared to satellite-derived products (see also: Holz & Huff, 1973). Interestingly, Hofmann compared the performance of IKONOS and Quickbird to evaluate differences between the informal settlement areas and the rest of an image scene in Cape Town, South Africa (Hofmann, 2001). His study is one of the earliest object-based image analyses of informal settlements. The data were further analyzed by Hofmann et al. (2008) using newly-available object-based measures

(Hofmann, Strobl, Blaschke, & Kux, 2008). Table 5 summarizes object characteristics of the Quickbird data that were segmented with image feature extraction software.

**Table 5 Object Measures for Informal Areas vs. Whole-scene (Hofmann et al., 2008)**

| Mean Metric Type | Informal Settlement Value | Whole Scene Value |
|---|---|---|
| Object Size | 15m$^2$ | 25m$^2$ |
| Length (m) | 6.47 | 7.73 |
| Width (m) | 3.78 | 4.61 |
| Border Length | 21.56 | 25.57 |
| Asymmetry | 0.56 | 0.55 |
| Mean NIR | 506.99 | 553.94 |
| StDev Red Band | 67.46 | 51.7 |

From a purely object-based approach, the informal settlement dwelling objects were smaller in all dimensions. The mean near infra-red (NIR), a good indicator of vegetation, was lower in the informal areas, and the standard deviation of the red band was also 30% higher in the informal parts of the scene compared to the whole scene. If reliable techniques of building footprint extraction (e.g., polygon boundaries) existed at the time, the characteristics in Table 5 could have been enhanced with more precise measures of edges, house orientation to road segment, and heights of dwelling structures that would likely vary considerably between informal and formal areas. Nevertheless, a major limitation in Hofmann is the authors used fuzzy logic to negate ‘informal’ in order to classify ‘formal’ (Hofmann et al., 2008). This approach did not compare pre-defined informal and formal settlement areas within the same image scene, and was therefore not suitable for the purposes of settlement differentiation. Like all other research reviewed here, they also measured rectangular settlement areas and did not follow the contours of actual residential settlement boundaries that may have produced different results.

A snake contour on smoothed (filtered) Quickbird imagery was used to extract building outlines in informal settlements in Tanzania (Mayunga et al., 2007). The semi-automated approach was focused on comparison of building extraction models using an energy minimization function to estimate outlines for each building whose approximate centroid was manually seeded before the algorithm was run. The authors demonstrated a 32% improvement over time spent performing manual building digitization and still about 15% of buildings extracted from aerial orthoimagery of the study area could not be extracted from Quickbird imagery. Most building extraction research is only successful in very small image sub-scenes with homogeneous characteristics, or requires pre-seeding where an operator selects the estimated centroid of a building, followed by a region-growing technique that detects the outer edges. The Mayunga (2007) study extracted only 78 buildings. Most informal settlements contain hundreds or thousands of dwellings and a great deal of clutter and continuous rooflines where multiple discarded materials comprise a single roof. Such image characteristics tend to confuse automated feature extraction tools and classifiers.

In a subsequent study the prior energy function was modified to use a radial casting algorithm to define building edges in informal settlements of Tanzania, but each individual building still required manual seeding of its approximate centroid, and the authors acknowledge that research is still needed to clearly delineate building corners (Mayunga et al., 2010). The major drawbacks to the building extraction approach:

1) complex post-processing required to split improperly-merged buildings into the appropriate separate structures
2) smoothing of pixilated shapes after the extraction process to produce the expected orthogonal or semi-orthogonal corners, and
3) inability to determine correct exterior wall locations beneath continuous rooflines
4) locally over-trained classifiers not appropriate for other images

The majority of prior research regarding dwelling extraction has been applied in small areas of 100-200m$^2$ and not entire residential areas, making this method less useful for complex urban environments of diverse settlement typologies.

Informal settlements have also been classified according to the spectral properties of roofs, including plastic sheeting and reinforced concrete in Dehradun, India (Jain, 2007) and zinc metal sheeting in Bangkok (Thomson & Hardin, 2000). The studies identifying roofing materials common to informal housing did not also compare roofing materials common to formal housing, and could therefore not be used to distinguish two settlement types at opposite ends of the scale. A further complication is that roofing materials are part of the vernacular architecture of a neighborhood or city and vary regionally depending upon the self-help materials available during home construction (Kellett & Napier, 1995). This was evident in a 1991 field survey of informal settlements in Guatemala that reported wall materials varied by age of settlement. In the first year of settlement, perishable waste materials and plastic predominate, while wood, bark and corrugated iron occur most often in years 1 through 10 of inhabitance (Valladares Cerezo, 2003). In the Guatemala informal settlement region surveyed by Valladares Cerezo (2003), concrete block, brick and adobe is found most often in settlements greater than ten years old. Corrugated metal was found in 86% of homes, with the remaining roofs comprised of waste materials, clay tiles or concrete in the oldest homes. Building construction and roof materials are a key component in understanding informal settlements from an imagery standpoint, but their analysis may require accurate building footprint extraction as a precursor. A separate method that has been used to improve the extraction of discrete objects such as dwellings is mathematical morphology.

### Mathematical Morphology and Image Texture Operators

Mathematical morphology and image texture operators are two methods that transform the original image. Mathematical morphology is applied to improve extraction of image elements, while measuring texture evaluates the change in spectral intensity of regions within the scene. These methods use either a kernel or matrix that moves through the image - a structuring element for mathematical morphology, and a moving window for co-occurrence texture operators. The field of mathematical morphology is well-summarized by Sulik and Edwards (2010):

> "Morphological operators are based on set theory and are similar to smoothing filters. However, unlike […] filters that act on spectral properties, morphological filters modify the spatial properties of foreground pixels (set *A*) relative to background pixels (set *B*). Morphological transformations are applied through a structural element that dictates the connectivity (topology) of pixel groups from set *A* that are allowed to pass through the filter. The connectivity within these structural elements defines what information is retained from the original image" (Sulik & Edwards, 2010, p. 2527)

Mathematical morphology pre-processes the image to create a new binary image from the original. It has been used to improve object segmentation (Liu, Wang, & Luan, 2007; Pesaresi & Ehrlich, 2009; Jin & Davis, 2005), and has also been applied specifically to informal settlements (Sulik & Edwards, 2010; J. Li, Li, Chapman, & Rüther, 2005;

Lefèvre, Weber, & Sheeren, 2007). Morphological properties of settlements that have been studied include area, perimeter and compactness. Research into the use of morphological operators to examine or extract structural relations using some combination of adjacency, containment, distance, or direction has also been attempted by (Z. Li, Yan, Ai, & Chen, 2004; Pesaresi & Ehrlich, 2009; Jin & Davis, 2005; Lefèvre, Weber, & Sheeren, 2007; Pesaresi & Benediktsson, 2001; Klaric, Matt, Scott, Grant, Shyu, Chi-Ren, & Davis, Curt, 2005). Morphological operators of opening, closing, dilation and erosion are applied to binary images using a structuring element to improve extraction of features.

One effort utilized an iterative histogram clustering to identify a cluster's mode, selecting the highest local maxima compared to neighboring values, then proceeding to the next cluster until all pixels in a greyscale image were assigned to a specific cluster, creating a binary image (Lefèvre, Weber, & Sheeren, 2007). Sets of binary images were thus produced for further filtering, but heterogeneous roofs (prevalent in informal settlements) complicated the clustering mechanism. Subsequently, the *opening* morphological operator was applied using a variable size and shape structuring element to remove possible non-building objects, followed by the building extraction process on the binary images. Their work was performed on regular-sized and shaped buildings from Strassbourg, France and not an informal settlement. The authors acknowledged that buildings located close together may have confused the clustering step by creating improper aggregation into a single building object. Morphological operators have been used to perform image segmentation to classify and then extract housing (Lopez-Ornclas

& Flouzat, 2008). Hofmann et al. (2008) performed an iterative rule-based semi-automated image segmentation and classification of housing objects and confirmed that the need for expertise in remote sensing, image analysis, and image processing make the object extraction process a non-trivial affair (Hofmann et al., 2008). Mathematical morphology does not consider the heterogeneity that exists by feature class, such as housing, roads, and soil, or the inter-class variation. The role of mathematical morphology is limited to improving the segmentation of image structural components, and is commonly applied to a single feature class, such as housing.

A related thrust of the literature focuses on variation of local texture in slum areas, with image texture defined as repetition of image elements (usually spectral intensity) in a local spatial domain. Texture analysis has been performed using grey level co-occurrence matrices (GLCM) to determine if texture variation helps identify informal settlement regions in an image scene (Haralick, Shanmugam, & Dinstein, 1973; Stasolla, Mattia & Gamba, Paolo, 2007; Pesaresi & Benediktsson, 2001). Stasolla et al. (2007) reported preliminary results that the GLCM homogeneity texture measure of SPOT imagery in an arid environment in Sudan is effective for the purpose of extracting built-up areas considered to be informal or slum (Stasolla, Mattia & Gamba, Paolo, 2007). Figure 7 displays the two settlements, showing the predominance of soil and the relative lack of vegetation, but also a clear difference in structural pattern between the two insets.

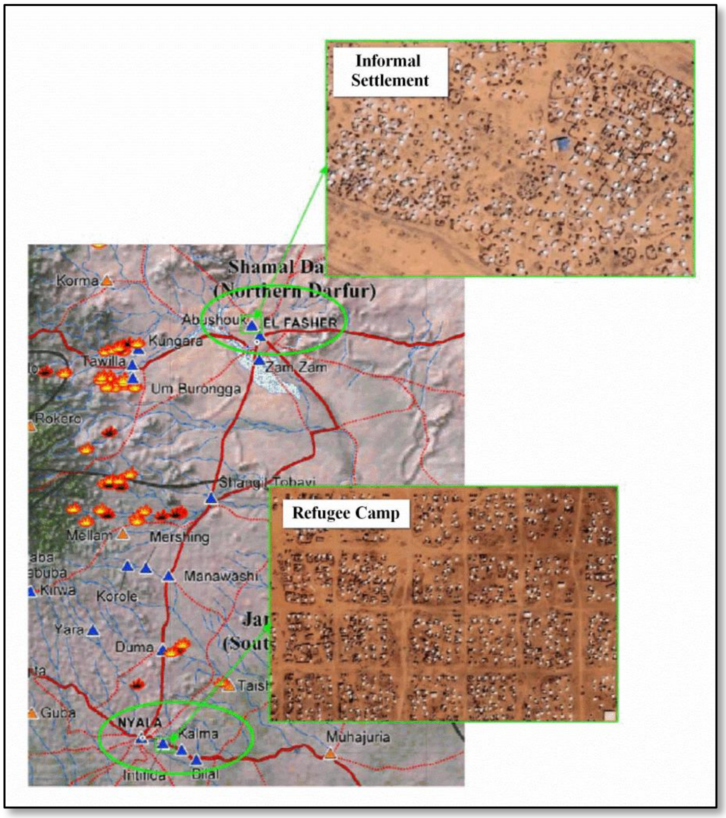



**Figure 7  Informal settlements and refugee camps using GLCM texture measure (Stasolla & Gamba, 2007).**

The authors found the GLCM texture metrics performed better in the formal area than the informal, but problems resulted in distinguishing between housing and rocks/shrubs using the SPOT-5 dataset where the pixel size was too large to extract individual dwellings. The value of the Stasolla & Gamba (2007) research is that it underscores the need for very high spatial resolution imagery, and demonstrates the usefulness of the GLCM texture measures for informal settlement analysis. It also shows semi-automated and knowledge-based approaches are needed to better accommodate the great variety of informal settlement typologies, from those in arid desert-like environments such as Sudan where their research was conducted, to more complex highly urbanized areas such as in Rio de Janeiro, Brazil, or Guatemala City. Despite the fact that the Stasolla & Gamba (2007) research was mainly focused on extracting informal areas by exploiting differences in GLCM homogeneity texture, the work was one of the few that acknowledged the need to evaluate differences in settlement patterns between regularized, gridded settlements, and more informal random settlements. Their research in an arid environment also confirmed that soil and vegetation were important to the classification.

The primary intent in Stasolla and Gamba (2008) was to apply indices of spatial autocorrelation (Morans $I_i$, Getis-Ord $G_i$, and Geary's $c_i$) as an approach to classifying built up areas of formal and informal settlements using radar data. The authors compared a formal settlement (Pavia in Northern Italy) to a starkly different informal settlement area (Al Fashir, Northern Darfur, Sudan) on another continent using fine beam single polarization 6.25m RADARSAT and ALOS PALSAR radar images respectively. The Sudan site contained refugee camps and rocky areas on the outskirts of the city which produced signatures that yielded confusing results and made the settlements hard to detect. RADAR images rely on the scattering properties of varying object heights for detection, but due to the low spatial resolution, this imagery type was not useful to distinguish individual dwellings, only built-up areas. The authors again tested the GLCM homogeneity texture operator which did not satisfactorily identify built up areas from the radar data (Stasolla & Gamba, 2008).

It is clear from more recent research using higher resolution imagery that structure, soil, and vegetation, in addition to housing are foundational to the measurement of informal settlements.

### Socioeconomic Measures and Landscape Analysis

In one study a complex processing framework was developed to extract informal settlement areas from Delhi, India using Quickbird VHR imagery with the objective of rating the quality of population density indicators in order to estimate water consumption and waste water disposal needs (Niebergall, Loew, & Mauser, 2007). The house mean size (area) and the degree of imperviousness were used to feed into a decision framework. To accurately measure these indicators, intense field research was conducted to sample and identify housing and settlement types, water-related structures and distinct features *in situ* that could be identified in the Quickbird scenes. Family size and water consumption survey responses were gathered as well. The authors sought to integrate objects from the segmentation into a GIS to evaluate vulnerability of informal settlements within mega cities, and to determine if their methods could derive socio-economic indicator values indirectly via image analysis. The authors' test sites were selected for their high socio-economic gradient - settlement structures from upper and middle class residential areas were adjacent to informal settlements.

Quickbird imagery was pan-sharpened followed by semi-automated supervised classification of residential structures (imperviousness), roads (soil), vegetation (using the Normalized Difference Vegetation Index, or NDVI), and shadows. The authors found that texture was an essential parameter in the detection of informal settlements, and that classification at varying scales enabled them to extract small features such as houses and larger features such as streets. Using some amount of trial and error they determined the appropriate classification level for each feature, calibrated the rule set on the training site, and integrated the survey results with the imagery analysis to estimate urban vulnerability to unmet water resource needs.

This research is significant because it sought to determine if socio-economic indicators can be measured from imagery. Household and door-to-door questionnaires increased the time and effort to identify living quarters of vulnerable populations. The use of measures such as house size in $m^2$, imperviousness, and NDVI underscores the importance of these remote sensing indicators to measuring informal settlements. However, the research did not identify specific differences between settlement types, despite their juxtaposition in the same image scene.

The potential for informal settlements to degrade the surrounding environment was investigated using variables quantified through prior research on secondary data, field work, air photo interpretation and GIS techniques (Zeilhofer & Topanotti, 2008). The use of detailed municipal planning data enabled the development of very specific indicators. Indicators related to slum housing included distance to street, amount of greenspace or public squares, and sidewalks. This work is useful if extremely detailed urban planning datasets are available, but also emphasized the importance of measuring green space in relation to street networks and housing topologies (Zeilhofer & Topanotti, 2008).

Slum areas, as well as formal settlements have been modeled using algorithms developed for landscape analysis and borrowed from the study of patch dynamics: area, density, form, edge, core area, proximity, subdivision and diversity (Lang, Walz, Klug, Blaschke, & Syrbe, 2009; Sudhira, Ramachandra, & Jagadish, 2004; Yeh & Li, 2001). Entropy of housing patches was studied with entropy given as $H_n = -\Sigma P_i \log_e(P_i)$ where; $P_i$ is the proportion of the variable in the $i$th zone and $n$ is the total number of zones (Yeh & Li, 2001). The result ranges from $0$ to $\log(n)$ where a compact distribution of patches is represented by values closer to 0. Entropy, patchiness, and density/growth of built-up areas were measured in other research (Sudhira et al., 2004). Although Sudhira et al. (2004) applied previously developed metrics for landscape analysis to characterize the sprawl growth in Mangalore, India, they did not focus specifically on informally-settled or slum areas.

In their research, similar buildings were segmented, classified and clumped into a homogenous landscape patch. Then patch types were classified using maximum-likelihood metric into the five classes of built-up, vegetation, water, agricultural land and open land. A measure they called ‘patchiness’ clumped all housing together into an aggregated patch and then compared patches to each other (Sudhira et al., 2004). The major limitation in the Sudhira et al. (2004) study was the lack of evaluating the relationship between the housing classes and the remaining classes that would have provided the needed context to the interrelationships between urban feature areas. Other research has emphasized multivariate methods that include vegetation patches, anthropogenic features, roads, and water bodies provide useful predictive power and contribute to model strength (Owen, 2009).

Table 6 summarizes all of the indicators available in the literature that are useful for informal settlement modeling, numbered for reference, including the originating author or study. The first 6 indicator classes shaded in gray are directly relevant to the current research. Indicator classes 7-14 require extraction of accurate building outlines, or ancillary data on locations of natural and man-made hazards or locations of social services (hospitals, schools, government services, markets), and will not be used for this dissertation due to the limitations previously listed: 1) complex post-processing required to split improperly-merged buildings, 2) smoothing of pixilated shapes needed after extraction, and 3) inability to determine correct exterior wall locations beneath continuous rooflines.

**Table 6 Summary of Indicators Used in Literature to Evaluate Informal Settlements**

| # | Indicator | Description – Expected Informal Values | Author/Study |
|---|---|---|---|
| 1 | Lacunarity of Housing Structures | Heterogeneity or 'gappiness' of empty spaces (lacunae) between built-up structures; degree of rotational or translational invariance. | Junior & Filho 2005; Weber & Puissant 2003; Filho & Sobreira 2005; Filho & Sobreira 2008; Owen, 2011 |
| 2 | Vegetation | Lack of vegetation per measured housing area. In Guisti de Perez et al. (2008), squatter settlements typically have 5-10% public spaces whereas planned settlements have 30%. | Weber & Puissant 2003; Guisti de Perez & Perez 2008; Niebergall 2007; Baltsavias & Mason, 1997; Griffin & Ford, 1980; Owen, 2011 |
| 3 | Road Segment Type and Materials | Elongation of roads, more regular road segments (although Hofmann, Strobl et al. (2008) do not define meaning of 'regular'). | Hofmann et al., 2008; Kux & Araújo 2008; Sliuzas et al., 2008; Strobl et al., 2008 |
| 4 | Road Accessibility Measures | Road width too narrow for vehicular traffic in informal areas. | Erickson & Lloyd-Jones, 1997; UN Habitat Expert Working Group on Slums – Breakout Group A – Spatial Criteria and Indicators. |

| | | | |
|---|---|---|---|
| 5 | Slope of Terrain | Settlements built on gullies & ravines, unstable soils. | Valladares Cerezo, 2003; |
| 6 | Texture Measures (Entropy, Homogeneity) | Entropy - where dwelling areas merged into agglomerated housing patches. | Yeh & Li, 2001; Stasolla & Gamba, 2007 |
| 7 | Proximity to hazards | Hazards include floodzones, hydrologic setbacks, landslide/earthquake, garbage-mountains, high industrial pollution, proximity to airports, energy transmission lines, major transportation corridors; areas susceptible to debris flows, rock and block falls, mass movements | Guisti de Perez & Perez 2008; Sliuzas et al., 2008; Duvadi, et al., ACRS 2002; Dubovyk et al., 2011 |
| 8 | Consistency of housing orientation | Borrowing from computer vision, the angles and lengths of line segments characterize informal settlements | UN Habitat Expert Working Group on Slums, Breakout Group A – Spatial Criteria and Indicators; Cheriyadat, Vatsavi & Bright, 2010. |
| 9 | Proximity to City Center and Services | Network analysis of distance to city services, market area or city center – derived from research on access to healthcare. Requires prior point locations of schools, community centers, health facilities, market areas. | UN Habitat Expert Working Group on Slums, Breakout Group A – Spatial Criteria and Indicators; Griffin & Ford, 1980; Owen et al., 2010 |
| 10 | Dwelling Size | Mean, $R^2$ of extracted dwelling sizes; dwellings between $16m^2$ and $40m^2$ classified as slum ($15m^2$ mean building size in slums of Cape Town) | Junior & Filho 2005; Ioannidis et al 2009; Martínez et al., 2008; Kemper & Pesaresi, 2008a; Baltsavias & Mason, 1997; Griffin and Ford, 1980; Hofmann et al., 2008 |
| 11 | Dwelling Shape | Height of dwellings measured by image shadows, or radar/LIDAR; simplicity of shape (4-sidedness). | Baltsavias and Mason, 1997; Hofmann et al., 2008; |
| 12 | Dwelling consistency of orientation; Dwelling road setback | Precarious house placement. | UN Habitat Expert Working Group on Slums, Breakout Group A – Spatial Criteria & Indicators; Zeilhofer and Topanotti 2008 |
| 13 | Building Density – Dwelling separation | Dwelling features extracted as polygons –nearest neighbor distance using centroid of dwelling polygons. | Weber & Puissant, 2003; Baltsavias and Mason, 1997 |
| 14 | Roofing Materials | Spectral properties of informal roofing materials, roof texture has more heterogeneity. | Weber a&Puissant, 2003; Thomson, 2000; Hofmann ,2008; Baltsavias & Mason, 1997 |

## Relationship to Formal Settlement Research

To understand how metrics previously developed for formal communities embodied by suburban sprawl might inform this research, an evaluation of related literature was conducted. Extensive research has been directed at characterizing sprawl as an undesirable settlement type in formal settlements of developed countries (Sutton, 2003; Ewing, Pendall, & Chen, 2002; Hasse, 2002; Transportation Research Board, National Research Council, 1998; Ewing, Schmid, Killingsworth, Zlot, & Raudenbush, 2008; Glaeser & Kahn, 2003), and much work has been applied to the development of techniques to map and measure sprawl aided by GIS (Hasse, 2002; Hasse & Kornbluh, 2004; Forsyth, Schmitz, Oakes, Zimmerman, & Koepp, 2006; Galster et al., 2001; Wilson et al., 2003; Epstein, Payne, & Kramer, 2002; Ewing et al., 2002; Fang, Liu, Hong, & Qing, 2007; Zeng, Sui, & Li, 2005). Quantitative indicators of sprawl in formal settlements have already been developed from remote sensing datasets. Similar to informal settlements, sprawl is also undesirable and negatively impacts residents' wellbeing.

A major premise of this dissertation research is that indicators for informal settlements can be adapted from sprawl metrics. Some results will differ between informal and formal (sprawl) settlements due to the variation of structural compactness in informal settlements and dispersion in formal settlements. At the end of this section, the settlement measures found in the formal settlement literature as they relate to sprawl are summarized and metrics for informal settlements are recommended.

## Formal Settlement Literature Review

The landscape of sprawl has four main characteristics (Ewing et al., 2002):

- Population is widely dispersed in low-density development
- Housing areas are rigidly separated from shops and work facilities or locations
- The road network is characterized by poor accessibility and expansive block sizes
- There exists a lack of defined and thriving activity centers

The result of these conditions is a lack of efficient transportation, lack of variety in housing options, and difficulty in walking from housing areas to activity centers. In effect, sprawl causes isolation and inefficient use of transportation and housing space. It is argued here that informal housing causes the same isolation and lack of access. But with informal housing the reasons are instead due to absence of land tenure, absence of housing durability, insufficient living area and lack of sanitation and clean water. Similar to much research on informal settlements, measuring sprawl requires the evaluation of *built structures, roads* and *land use / land cover* that are best extracted using remote sensing combined with GIS techniques (Weng, 2002; Hasse & Lathrop, 2003; Wilson et al., 2003; Epstein et al., 2002). Five major contributions to the development of standard measures of sprawl aided by GIS and remote sensing are described next (Hasse, 2002; Hasse & Lathrop, 2003; Forsyth et al., 2006; Ewing et al., 2002; Fang et al., 2007; Wilson et al., 2003).

Hasse (2002) developed specific spatial metrics of sprawl in the US state of New Jersey. Of the thirteen measures developed, five could be applied or adapted to informal settlement morphology to assess growth, pattern and severity. The *urban density index* measures the space between houses. Contrasting to sprawl, the space should be extremely small in slum areas. The *community node inaccessibility index* measures mean distance of housing communities to schools, health care, post offices and supermarkets. In the context of slum areas, this index could be adapted to instead measure the shortest path that also accounts for road surface (paved, unpaved and trails) to medical facilities, markets, and schools, or the city-center as a surrogate. The *loss of important land resources* indicator used to measure loss of agricultural, forested, or sensitive habitat areas can be adapted to measure proximity to environmental hazards, creating greater risk for the population. This indicator can be combined with one that measures *encroachment on sensitive open space* to reveal how much vegetation cover has been lost to informal housing. *Per-unit impervious surface* is another indicator that can be adapted by measuring ratio of paved to unpaved roads and the ratio of housing to roads, and assumes limited road accessibility per unit of housing. The *per-unit impervious surface* indicator can be adapted to measure other road characteristics such as width, efficiency, relative asymmetry, etc. The Hasse (2002) indicators could thus provide a good start for measuring informal settlements, but would require extraction of discrete house outlines and prior knowledge of spatial location of city centers, schools, and health facilities. A focus on measuring impervious or built-up features, roads, and the areas in-between would therefore be more appropriate for the current research.

Forsyth et al. (2006) sought to develop measures of sprawl for the purpose of assessing how the built environment affects health. Forsyth's research is applicable because of the development of indicators that measure walking accessibility using the number of *4-way intersections* and the number of *access points*. Walking is critical in slum areas due to severely reduced automobile ownership. In the Forsyth study, four-way intersections were measured as a representation of a gridded street pattern, and it was hypothesized that accessibility improves as node valence increases. Node valence is the sum of the number of road segments converging to create a node. In other words, a 1-valent intersection is a dead-end or cul-de-sac, also known as a *dangle* in the literature (Cao, Song, Wang, & Wang, 2010; Tresidder, 2005). Access points are points of intersection between paved (non-highway) roads and housing site boundaries. The number of access points measures the amount of connectedness of an area to its surroundings. Although adapting such measures to informal settlements is possible, the formal planning process in developed countries that prevents housing along major highways may be non-existent in many third world or developing countries, forcing a qualitative, site-specific identification of access points.

The Ewing (2002) measures of sprawl benefited from a rich source of statistical data available in the United States from the Census Bureau and other sources. Ewing did not rely on remote sensing per se, but did focus on the level of isolation resulting from sprawl. In effect, this isolation of the residential areas from important community nodes could be modeled a number of ways in informal areas. The value of the Ewing (2002) research is in the formal definition of sprawl supported by empirical research. Three of Ewing's four definitions of sprawl are applicable to informal settlements:

1) Housing areas rigidly separated from shops and work facilities or locations,
2) Road network with poor accessibility (informal areas have unpaved dirt roads that become

impassible during certain times of year, are limited in length compared to number of structures supported, often have inadequate width to allow automobile passage), and

3) Lack of defined and thriving activity centers co-located with housing areas (isolation from infrastructure services, parkland, and recreation).

Fang et al. (2007) developed geospatial measures of sprawl in rapidly-growing Beijing, considered to be transitioning from a third world to a first world city (Fang et al., 2007). The authors developed an integrated urban sprawl index using several GIS-based data sources. Specifically, the authors sought to examine three major areas:

1) fragmentation and irregularity of landscape (discontinuous development)
2) low efficiency of development
3) negative effects on agriculture, environment, and urban living.

We would expect to find fragmentation and irregularity of informal settlements in slums, low access to infrastructure (as a measure of settlement efficiency), and negative effects on surrounding environment caused by high housing density and lack of proper infrastructure that creates a lower quality of urban life. Their measures were grouped by Spatial Configuration, Growth Efficiency, and External Impacts as portrayed in Figure 8 (Fang et al., 2007).

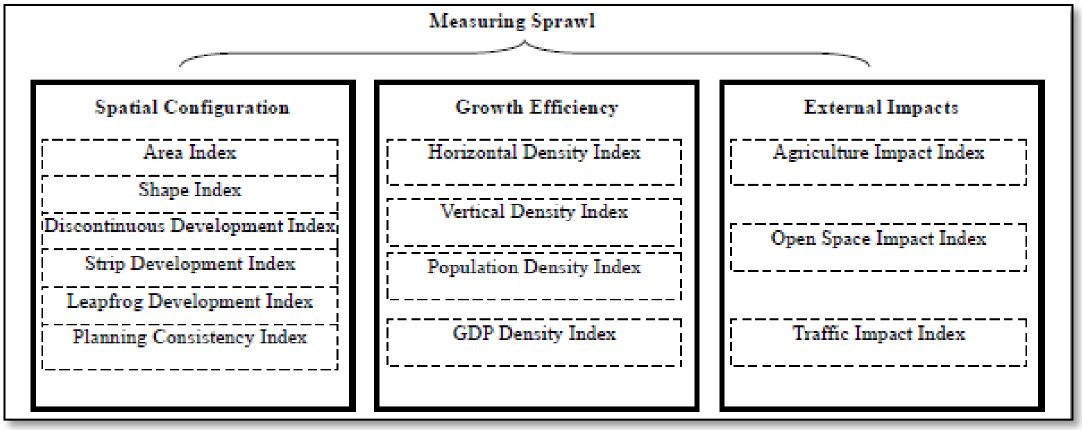



**Figure 8 Sprawl Measures (Fang et al., 2007)**

The spatial configuration indices of *Area, Shape, and Discontinuous Development*, the Growth Efficiency index of *Horizontal Density*[3] and the External Impacts index of *Open Space Impact* could be adapted for the assessment of informal settlements. The remaining indices would be extremely difficult to measure given the lack of ancillary urban planning data and the sparse availability of pre-populated GIS datasets in many developing countries. The *Open Space Impact* index could be measured as a ratio of housing to NDVI, similar to the Haase (2002) indicators that measure encroachment onto sensitive open space. We expect to find the open space encroached upon by informal settlements to instead represent the less-desirable and more hazardous areas. In this way, the external impacts are such that the effects of environmental hazards are magnified when adjacent to dense human settlements.

Wilson et al. (2003) developed an urban growth model to quantify the manner in which urban sprawl expands. The authors identified three categories of urban growth:

[3] The Fang et al. (2007) index of vertical density could only be measured with adequate planning data and either LIDAR or other datasets able to measure building heights and knowledge of residential vs. commercial areas.

*infill*, *expansion*, and *outlying*. Outlying urban growth was further divided into *isolated*, *linear branch*, and *clustered branch* to study the shape and direction of growth in a part of Connecticut. Again borrowing from landscape patch dynamics and adapting a forest fragmentation model originally developed by Riitters et al. (2000), Wilson et al.(2003) developed a roving window model to analyze each Landsat pixel according to its neighboring pixels and its surrounding landscape that improved upon previous per-pixel change classification methods by incorporating a framework of neighborhood pattern (Riitters at al., 2000). The authors used coarse but widely available Landsat data with 30m pixel size, and acknowledged that "varying the window size influences the scale and type of spatial patterns that are subsequently detected and classified" (Wilson et al., 2003, p. 280).

The classification of growth patterns is still an extremely useful means of examining the process by which slum areas fill over time. According to Wilson, image segmentation and object-based classification can result in under-classification of isolated growth areas (Wilson et al., 2003). They also cautioned that extensive filtering or other methods that decrease heterogeneity of the urban areas may result in over-classification of linear branch and under-classification of the clustered-branch class because interior-to-patch change pixels necessary for class clustering may be under-represented. To apply these methods to informal settlements, one would first locate *outlying* areas containing new land invasions, and then measure *infill* to understand how the settlement, once established, grows organically.

By relying on multi-temporal datasets, Wilson's methodology is useful to evaluate the settlement process together with pattern, but the use of moderate resolution imagery (Landsat) obscures smaller patches and road features, and is therefore insufficient.

Remote sensing methods have also been used to inform quality of life (Owen, 2011). Table 7 lists related remote-sensing based indicators summarized by the Cowen & Jensen indicators (2001), with a column added to suggest adaptation for informal settlement modeling:

**Table 7 Jensen & Cowen Quality of Life Indicators with Informal Settlement Modeling**

| **Quality of Life Indicator** | **Measurement** | **Can be Used for Informal Settlements?** |
|---|---|---|
| **Building size** | $m^2$ | Yes – need building object shape |
| **Lot Size** | acres or hectares | Would require cadastral data. |
| **Presence of Swimming Pool** | $m^2$ | Depends on climate. |
| **Vacant lots per city block** | Integer | Would measure vacant buildings; vacant lots rare in informal settlements |
| **Frontage** | $m^2$ | Would require cadastral data |
| **House distance to street** | Linear distance | Yes – need building object shape |
| **Building density** | Nearest Neighbor Distance | Yes – need building object shape |
| **Presence of Driveway** | % of all dwellings | Would require cadastral data |
| **Presence of Garage** | % of all dwellings | Would require cadastral data |
| **# Autos visible per house** | Ratio | Would require excluding street-based vehicular traffic. |
| **Unpaved Roads** | % of all roads | Yes – spectral differences in road surface materials needed |
| **Road Width** | Mean width per segment using station points | Yes – need road classified as object |
| **Vegetation health** | NDVI or Vegetation Ratio Index | Yes |
| **Proximity to manufacturing and retail activity** | Linear distance | Yes, Industrial facility proximity, requires feature point data |

A close relationship exists between quality of life indicators and possible indicators of informal settlements. Table 8 summarizes the indicators and their literature source, and suggests adaptations for distinguishing informal from formal settlement types.

**Table 8 Summary of Sprawl Indicators adaptable to Informal Settlement Measurement**

| Sprawl Indicator | How to Measure in Informal Settlements | Source |
|---|---|---|
| (a) Urban Density Index<br>(b) Horizontal Density | Distance between houses (nearest neighbor distance between houses on same street must be adapted to nearest neighbor for slum areas where house placement is irregular and not necessarily anchored to street frontage) | Hasse (2002);<br>Feng et al., (2007) |
| (c) Community Node Inaccessibility<br>(d) Rigid separation between housing and shops | Distance to community centers (schools, health, shopping, market, etc.). Distance using transportation network expected to be greater in informal settlements. | Hasse (2002);<br>Ewing (2002) |
| (e)Loss of Important Land resources<br>(f)Encroachment on sensitive open space<br>(g) Open Space Impact | Loss of green space, vegetation indices would be low in Informal compared to formal; are measured as percent vegetation per settlement. | Hasse (2002);<br>Feng et al., (2007) |
| (h)Per-Unit Impervious Surface (IS) (includes houses & driveways) | Informal settlements expected to have low per-dwelling impervious surface, formal would have higher per-dwelling impervious surface. Since informal settlements rarely have driveways, per-unit-IS should be adapted to measure [roads:housing] IS. If roads are all paved, this value should be high. If not, this value will be low. Calculate impervious surfaces from self-help roofing materials. | Hasse (2002) |
| (i) number of 4-way Intersections<br>(j) number of of Access Points<br>(k) Road Network Accessibility | Node valence + road distance to market centers from each household;<br>Adapt for impassibility during rainy seasons, calculate NSDI (normalized soil difference index) and compare to Impervious surface – if soil high then roads unpaved – method may not work in arid, compacted environments where roads ALL unpaved. | Forsyth et al., (2006)<br>Ewing (2002) |
| (l)Area<br>(m)Shape<br>(n)Discontinuous development | Measure size (extent), shape of open spaces, and irregularity / heterogeneity of house placement. | Feng et al., (2007) |
| (o)Infill | Temporal measure: shows how open space fills with housing. A locally random pattern or greater | Wilson et al., (2003) |

| | | |
|---|---|---|
| | heterogeneity is expected to describe pattern of informal settlement space-filling. | |
| (p)Expansion | Temporal measure: measure expansion of settlement or community outer boundary using rate over time, and ratio of perimeter to area. | Wilson et al., (2003) |
| (q)Outlying | Temporal measure: morphology & directionality of areas in the urban fringe. | Wilson et al., (2003) |

## Summary of Limitations from Literature Review

Recall the objectives of this dissertation research were to determine which foundational measures derived from roads, vegetation, soil, image texture and geomorphology can explain settlement structure of residential areas to determine which ones are informal and which ones are formal. The need for ancillary datasets limits this measurement in data-poor or developing regions worldwide, so the goal was to determine what indicators are statistically significant without requiring field survey instruments.

Given these objectives, the related limitations of prior research are:

1) Household surveys are costly, not practical, and the data are not readily available or replicable in informal settlements of the developing world.
2) Research has not focused on the differences in settlement structure – informal vs. formal/planned, but instead has focused on characteristics of slums in isolation.
3) Most research ignores useful foundational settlement characteristics, such as the transportation network, the open spaces between dwellings, and features such as soil and vegetation, failing to evaluate them collectively.
4) No applicable remote sensing-based research on informal settlements in Central America has been conducted

The reliance on household surveys greatly increases cost and in many cases, risk, for those performing the surveys. In Martínez (2009) GIS was used to analyze intra-urban inequalities in slum areas of Argentina, but was supplemented with detailed field surveys (Martínez, 2009). Although the research helped quantify the inequalities and analyzed sentiment, it is another example of the cost of administering survey instruments to inform the GIS analysis, and highlights the increased labor needed to operationalize research variables at the household scale.

Only two remote sensing and GIS studies could be found that compared performance of models in informal vs. formal areas in the same study area (Filho & Sobreira, 2008; Filho & Sobreira, 2007). However, they were focused on scale-based characteristics only. A third effort required household survey data and it did not develop indicator variables (Niebergall et al., 2007). The remaining relevant literature is devoted to measuring the characteristics of slums in isolation and does not provide useful information on settlement differences.

Another limitation of past research is that housing is often consolidated into a single class, or at best, into high, medium and low density. By merging all housing together into a class without understanding the shape of built-up areas, their relationship with vegetation and the interconnected network of roads ignores the interrelationships of basic settlement primitives as potentially powerful discriminators. Instead, informal settlements should be viewed collectively in context with their built-up areas, roads, soils, vegetation, and topography from which spectral and spatial properties can be derived. Prior research also ignored the irregularity of settlement boundaries. The irregular shape of residential settlements provides a focused and more nuanced understanding of inhabitants in terms of residential accessibility, open space, and feature compactness.

Finally, there is also a lack of remote sensing and GIS-based understanding of informal settlements in Central America. By adding Guatemala City to the current body of literature, the results could be applied and tested elsewhere in Latin America. Such comparisons may help understand how indicators could vary by terrain, building materials, culture, suitability of building plots, and human-environment interactions.

# RESEARCH METHODS

Research methods for this study were devised to answer the research objective of developing indicators from VHR satellite imagery that differentiate informal from formal communities without the need for survey or census data. A multi-stage approach was performed to accomplish this objective. Sampling units were first comprised of the entire settlement, defined a-priori as informal (slum) or formal (planned). Second, random sub-sampling was conducted. Indicator categories were developed from an exhaustive literature evaluation of informal settlement measurement as well as from sprawl indicators.

This section begins with a description and history of the study area, followed by a review of the datasets used, and the settlement area delineation methods. Following that is a detailed discussion of the sampling design, indicator categories, and modeling methods.

## Description and History of the study area

The study area was chosen due to the high percentage of urban slums in Guatemala and also due to the availability of high resolution imagery over this specific footprint that contained several formal and informal, slum areas. Neither SPOT nor IKONOS imagery was available in this specific area, so Quickbird imagery was chosen.

Figure 9 shows the outline of Guatemala's international borders and the placement of the capital in the southern part of the country.


**Figure 9 Guatemala City, adapted from Wikipedia.**

Figure 10 displays the *departamentos*, or first level administrative boundaries of Guatemala in relation to the image scene, which is positioned in *Zona* 11 located in the western part of the capital *departamento* of Guatemala, in green.

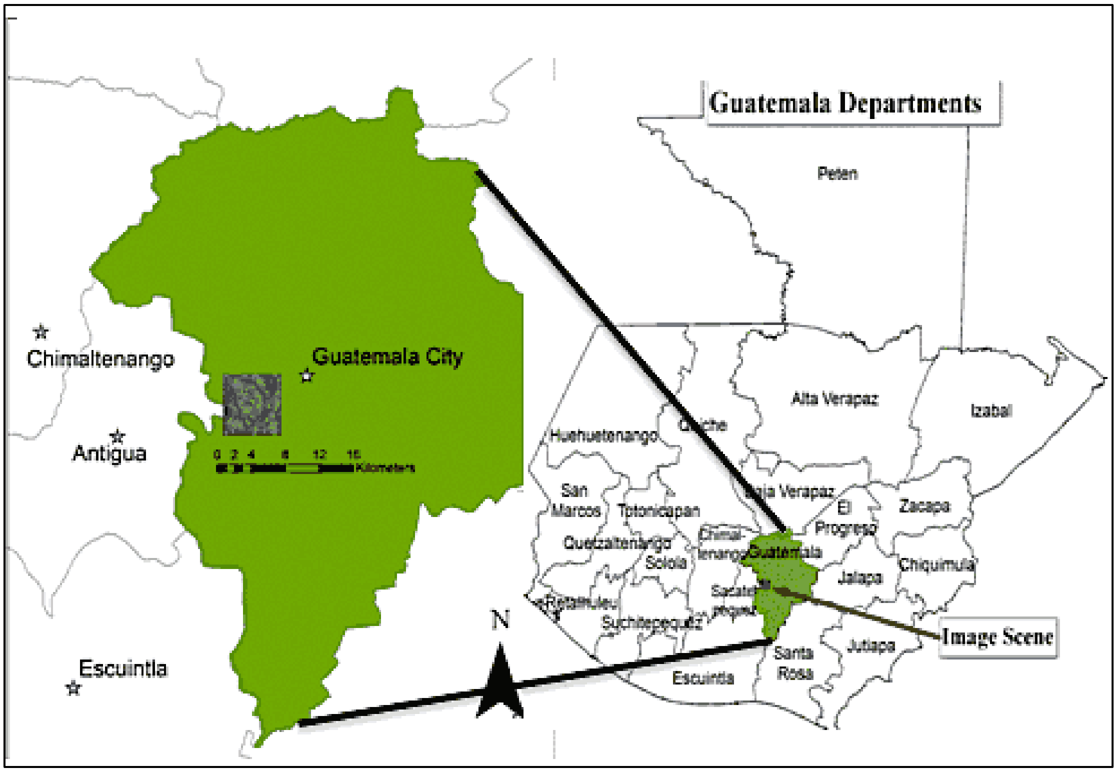



**Figure 10 Guatemalan *Departamentos* with Quickbird image Scene (Owen, K.)**

Settlement morphology must be understood not only by shape but also by historical context.  Guatemala City was originally settled in accordance with the Law of the Indies by the Spaniards in 1776 after the old capital in Antigua was partially

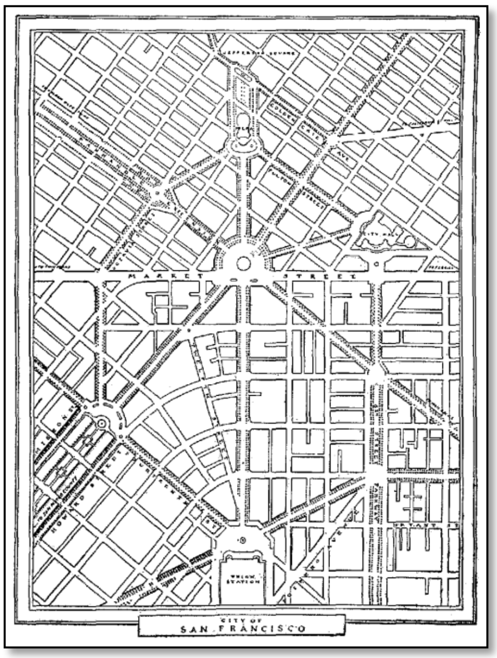



**Figure 11 Example of Gridiron City Pattern**
**(http://www.library.cornell.edu/Reps/DOCS/schermer.htm)**

destroyed by the Santa Martha earthquakes. According to the law, the city was laid out in a gridiron pattern similar to what is shown in Figure 11 with a central square and three smaller squares (Valladares Cerezo, 2003).   The Mudejar-influenced houses were characterized by a wide doorway with vestibule, a central open patio surrounded by a veranda leading to the bedrooms, and a kitchen located at the end of the house. The

Mayan Indians, whose culture and heritage differed from the Spaniards, traditionally lived in villages of small huts in the surrounding countryside. In 1917 another earthquake destroyed the city, temporarily restricting development, but its basic gridiron pattern remained. In 1954, President Arbenz was overthrown, ending a period of agrarian reform, also known as “Decree 900” that had been favorable to peasants (Gleijeses, 1992). A sharp increase in rural to urban migration resulted. The loss of opportunity for small agricultural landholders caused the city to double its population in 14 years (Valladares Cerezo, 2003). Some migrants converted vacant land to informal settlements, while others began to occupy large homes vacated by the wealthy, turning them into tenements (*palomares*) and adding improvised enlargements with an entire family often living in a single room.

Four factors contributed to the creation of the largest city in the Central American isthmus with the largest slum population: (1) a subsequent 1976 earthquake; (2) social upheaval caused by a twenty year civil war; (3) government plans gone awry to relocate slum dwellers; and (4) invasion by ‘squatters’ onto steep slopes of gullies and ravines. This large slum population and current conformance to Griffin & Ford’s (Griffin & Ford, 1980) Latin American City Model with wealthy elite neighborhoods located nearest the city center make Guatemala City an ideal test case for the proposed research objectives.

## Description of Imagery Data Used

### Quickbird

The author was provided a series of panchromatic and multispectral Quickbird image scenes (Digital Globe, 2009) through the US Geospatial Intelligence Foundation. The scenes cover settlement areas in Guatemala City ($46km^2$) defined a-priori with the help of a remote sensing analyst who resided there, and augmented by discussions with a non-government volunteer organization (NGO) employee who had traveled through the region for several years. The area was selected because of the known existence of established slums and more formal communities in the same vicinity, and the selected area was the only urban region in Guatemala where the Quickbird VHR imagery was captured from multiple years, providing future opportunities for temporal analysis. Table 9 lists the sensor features during data acquisition for the entire selected area in Guatemala City.

**Table 9 Quickbird Image Characteristics**

| Urban Area | Area | Acquisition Date & Overpass Time | Center Lat/ Long | Avg off-nadir Angle | Total Max off-Nadir Angle | Avg. Target Azimuth | Min Sun Elev | Day of Year=n, solar hour angle |
|---|---|---|---|---|---|---|---|---|
| Guate-mala City | 46 $km^2$ | 03/23/09 T16:53:00 | -90.591° W 14.607° N | 12 | 11.9 | 104 | 67.13 | 82 |

The scene from 23 March 2009 at 16:53 Zulu (11:53AM local time) was selected when cloud cover and shadows were minimized during the time of capture. Small areas of sun glint (white circles in Figure 12) can be found in a few spots in the scene, but were sufficiently limited in scope to not materially alter the results of the classification and image processing, so no attempt was made to correct them.  The average off-nadir angle was 12% which minimized shadow effects and shape distortion.

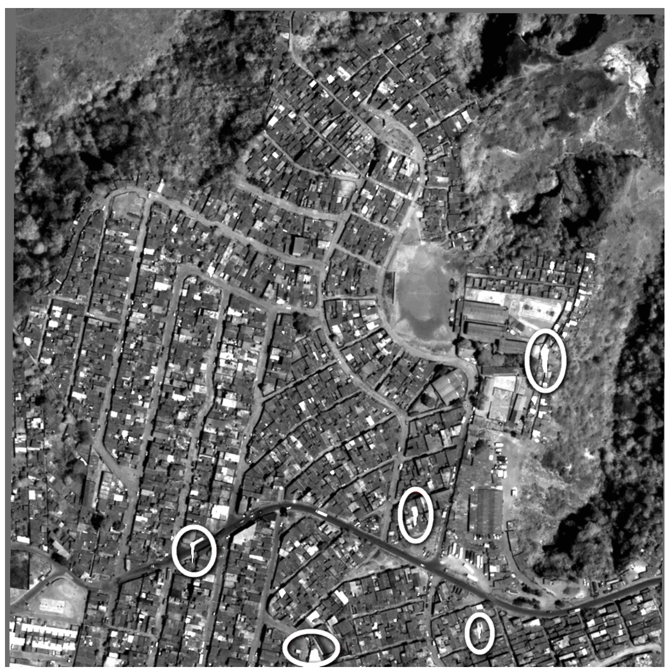


**Figure 12 Sun Glint Areas Circled in white, Peronia2 Settlement**

### Orthorectification and Elevation Data

Using the commercial ENVI™ software (*ENVI*, 2009) multi-spectral images were pan-sharpened using their corresponding 0.6m spatial resolution panchromatic band using the rational polynomial coefficients (RPC's) delivered with the image product. The image was orthorectified from an ASTER digital elevation model (DEM) in GeoTIFF format with 30m spatial resolution using a GEOID offset of 2.54m calculated from coordinates of a verified central pixel in Guatemala City. The DEM was referenced to WGS84 ellipsoid and delivered in 1° tiles, identified as GTM_N14W91 for Guatemala City[4]. Vertical accuracy of the ASTER DEM is reported as 19.1m with 95% confidence. The QA file in Figure 13 displays the histogram of the number of scene-based DEMS used to compute the final pixel elevation value for all pixels in the 1° tile that includes the Guatemala City scene.

[4] The ASTER GDEM was contributed by METI and NASA to the Global Earth Observation System of Systems (GEOSS) and is available at no charge to users via electronic download from the Earth Remote Sensing Data Analysis Center (ERSDAC) of Japan and NASA's Land Processes Distributed Active Archive Center (LP DAAC at http://gdex.cr.usgs.gov/gdex/).

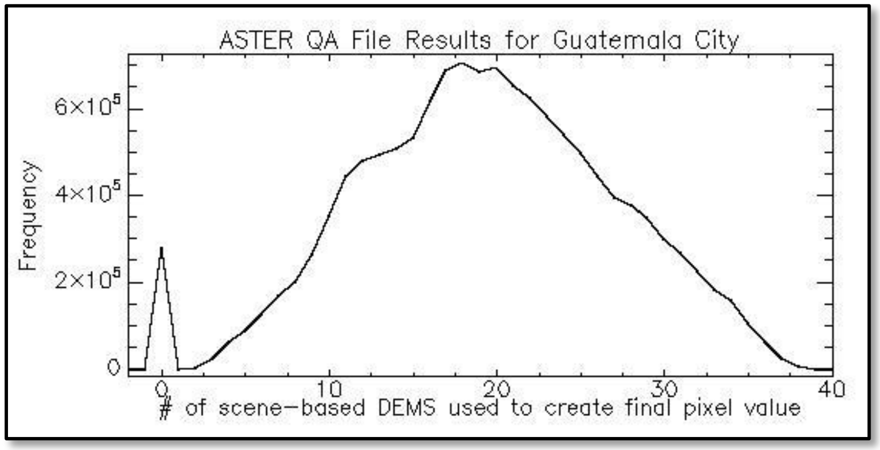



**Figure 13 QA Aster Histogram**

No other publicly available 30m DEM data were available, and the Shuttle Radar Topography Mission (SRTM) DEM at 90m grid spacing was too coarse. The area of the smallest settlement, at $0.05km^2$, would have been represented by only 9 SRTM pixels, so the decision was made to use the ASTER DEM, despite its reduced vertical accuracy, and because only 4 of the measured indicators relied upon elevation (degree slope, slope on roads, plan convexity and profile convexity).

Orthophotos from 2006 (0.5-meter) of the focus region in Guatemala City were provided through an informal GIS and cartographic training exchange with Mercy Corps Guatemala during the summer of 2009.  This imagery was used to visually confirm and validate road surfaces, built-up areas, and vegetation in the Quickbird scene.

## Settlement Delineation Methods.

Six formal and six informal settlement areas were identified in Guatemala City through field work during April 21st-30th, 2010 in consultation with professionals from the Guatemalan Instituto Geográfico Nacional. These settlement areas were defined and verified by driving through many of the locations, or viewing from a distance, in order to conduct an initial assessment of differences in indicator results at the settlement level. Indicators that showed promise in discriminating between settlement types could then be evaluated using a random sub-sampling design.

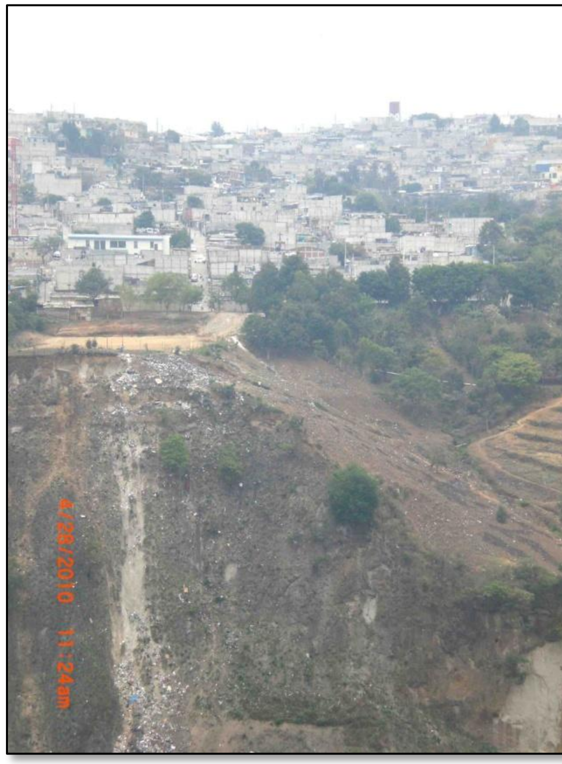



**Figure 14 Image of Ciudad Satélite, garbage dump in foreground**

The terms ‘informal’ and ‘formal’ were qualitatively determined. ‘Informal’ refers to poor, unplanned neighborhoods of low economic value with little, if any, zoning enforcement for dwellings. By contrast, ‘formal’ areas exhibited characteristics of higher economic value with more expensive building materials, adequate parking facilities, safety features such as guard shacks, and better road quality. Settlement names were determined from maps depicting official *lotificación* neighborhood names, also known as '*subdivisions*' in North American cadastral and zoning classification systems.

During ground verification, the entrances to half of the formal communities were guarded by security guards with automatic weapons who barred our access. This was not unexpected as gated communities often employ armed security in many urban parts of Guatemala. The driver refused to enter three of the informal communities by car (*Peronia1*, *Peronia2*, and *Ciudad Satélite*) for safety reasons, and because it was felt that narcotics traffickers controlled some of those areas (See Figure 14 of *Ciudad Satélite*). Photos of those settlements were taken from a distance, and their neighborhood boundaries were estimated through a combination of Google Earth visual interpretation aided by 0.5m orthoimage comparison and discussions with local experts. Neighborhoods were limited to residential areas and purposefully excluded surrounding greenspace, industrial or warehouse complexes, and major highways. This research differs from related efforts because it focuses on dwelling-areas to gain deeper understanding of the characteristics where residents choose to locate their homes. Therefore, irregular shaped polygons were drawn around settlements according to the following qualitative guidelines aptly acknowledging Weeks et al. (2007) contention that "Slum neighborhoods are defined largely by their physical and infrastructural environment, rather than by the demographic characteristics of residents":

- Exclude larger buildings adjacent to major roads – many are commercial areas, not dwellings
- Exclude surrounding highways – generally not considered part of the neighborhood
- Exclude wooded or vegetated areas adjacent to or just outside the boundaries of residential areas

Despite knowledge that external features influence small populations, this effort aims to evaluate differences inside dwelling area boundaries, which requires an important but somewhat tedious masking step for each measured settlement.  Figure 15 displays the 12 settlements, six informal and six formal.  Figure 15 is the main study area discussed in subsequent sections.

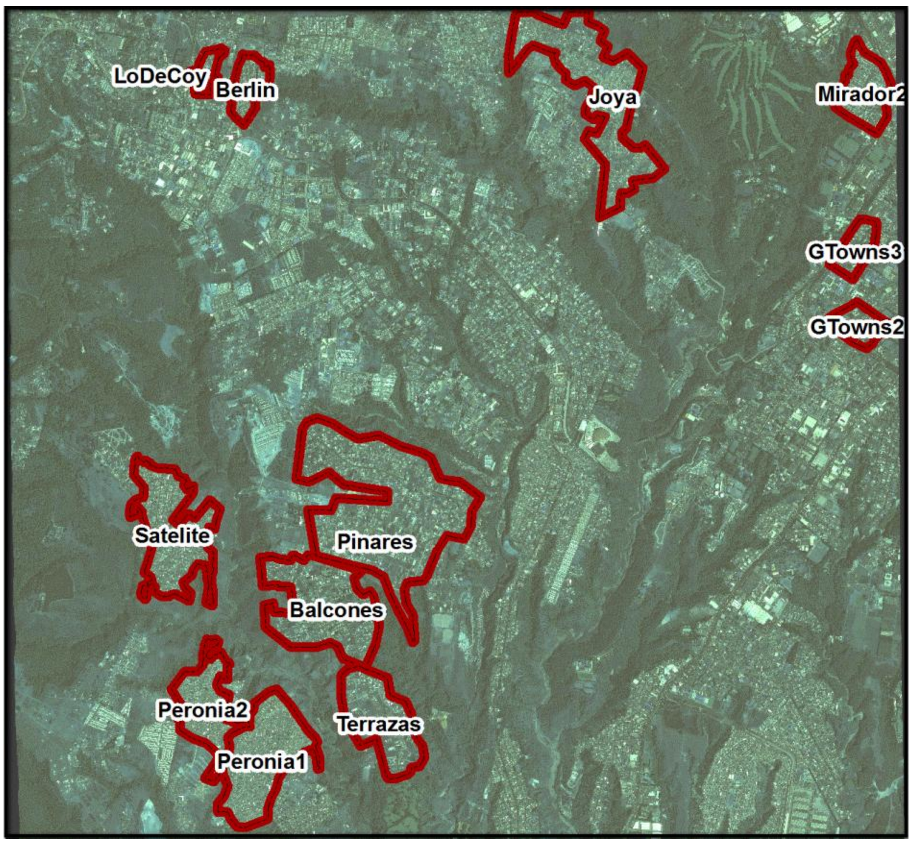



**Figure 15 Settlements delineated in Quickbird imagery**

Table 10 lists the settlement subset names by settlement type (formal or informal), the area, number of pan-sharpened 0.6m pixels, and perimeter.

**Table 10 Settlement Names and Characteristics**

| Settlement Name | Formal / Informal | Area ($km^2$) | Total Pixels | Perimeter (m) |
|---|---|---|---|---|
| **Mirador2** | Formal | 0.18 | 501724 | 1841 |
| **GTowns2** | Formal | 0.09 | 258689 | 1235 |
| **GTowns3** | Formal | 0.09 | 245976 | 1248 |
| **Balcones** | Formal | 0.51 | 1411246 | 3865 |
| **Pinares** | Formal | 1.08 | 5656576 | 6468 |
| **Terrazas** | Formal | 0.31 | 877677 | 2488 |
| **Peronia1** | Informal | 0.43 | 1189581 | 3267 |
| **Berlin** | Informal | 0.09 | 275525 | 1510 |
| **LoDeCoy** | Informal | 0.05 | 144067 | 1165 |
| **Peronia2** | Informal | 0.26 | 717203 | 3417 |
| **Joya** | Informal | 0.61 | 1720115 | 6007 |
| **Satelite** | Informal | 0.41 | 1153932 | 5071 |

### Masking Methods

Masking plays a key role in this research where values outside of a polygon boundary are removed from the computation of a metric. It is similar to the 'clipping' function in a GIS, whereby feature values are only preserved when they fall inside a designated polygon boundary. The masking transformation is accomplished in image analysis by applying a polygon boundary to an image in which all values inside the polygon result in preserving the image pixel values, and all values outside the polygon are assigned a non-value NaN (Not a Number). Masking becomes especially important when statistics are needed inside a settled area. Numerous other measures related to settlements and their inhabitants, including segregation measures, acknowledge the importance of using neighborhood boundaries (Wong, 2008). In the case of some image processing software, masking or clipping will apply either a zero-value (0) to all outside pixels or in the case of ArcGIS, a highly negative number, −9999. If a statistical computation for an image subset includes a large region of 0's or a large region of −9999's, errors will occur. The following example from the panchromatic image illustrates why correct masking matters. *Lo de Coy* is a small informal *lotificación* (subdivision) in the north-western edge of the original image scene, and shown in Figure 16. Masking is accomplished using the process also described in the figure.

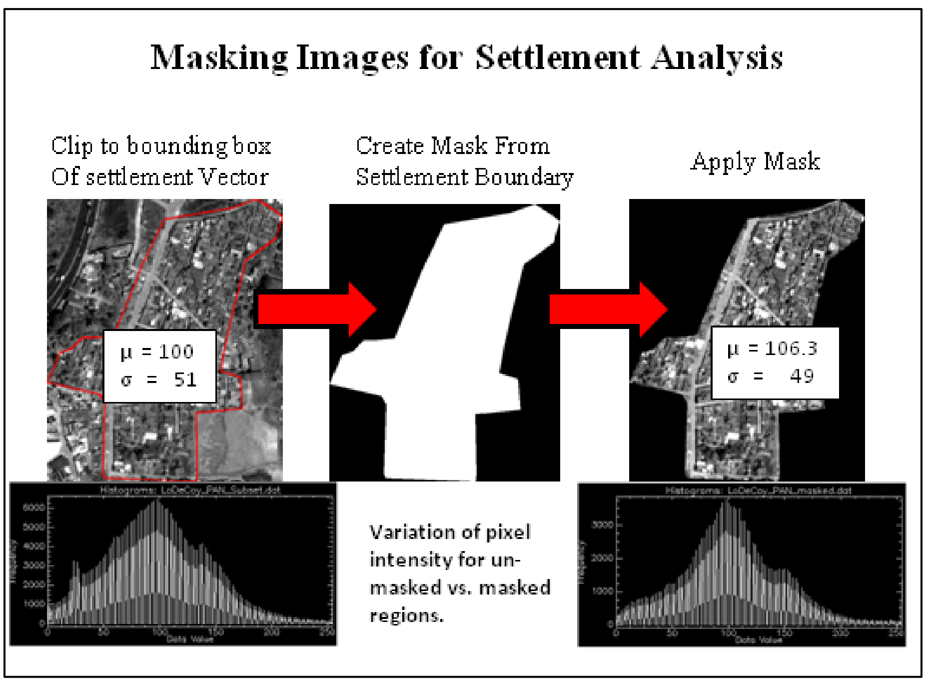



**Figure 16 Masking Procedure for Image Analysis (source: Owen, K.)**

In Figure 16 the distribution of pixel intensity from the panchromatic orthorectified image subset has a small peak at values of approximately 20-30 roughly corresponding to the highway pavement in the northwestern part of the unmasked image on the left of the figure. This peak does not occur in the masked image on the right. The unmasked image also has a lower mean panchromatic intensity value, demonstrating that masking plays an important role in improving the accuracy of the measures by providing more specificity of results.

## Sampling Design

In order to evaluate statistical significance, a breakdown of the six informal and six formal settlements into smaller, randomly selected sampling units was needed. A grid size was selected so that a sufficient number of sample grid cells intersect each settlement once the stratified random sampling is performed. This requires calculating the needed sample size. The sample size assumes a 5-6% margin of error is acceptable. Despite the fact that varying degrees of formality and informality exist in any settlement, for purposes of this analysis a two-category scale (Informal / Formal) was adopted for binary classification of categorical data. It was assumed that 3 standard deviations (+/−) from the mean would capture 98% of all possible values in the samples when estimating the necessary sample size. Sample size was computed from Cochran's Sample Size formula for categorical data (Bartlett, II, Kotrlik, & Higgins, 2001). Without knowing the variance in advance, it is estimated from: $= \frac{n}{\sigma}$. Given:

$n$ = 2 (number of categories: Informal or Formal)
$\sigma$ = 6 (3 to each side of the mean), so
$\mathcal{S}$ = $(0.333)^2$ = 0.111 = estimated variance
$t$ = 1.96 for the selected α of 0.025 in each tail of the distribution

So the minimum required sample size was calculated as:

$$n_0 = \frac{t^2 * (\mathcal{S})}{d^2} \Rightarrow \frac{1.96^2 * (.111)}{.05^2} = 170.5$$ (Bartlett, II et al., 2001).

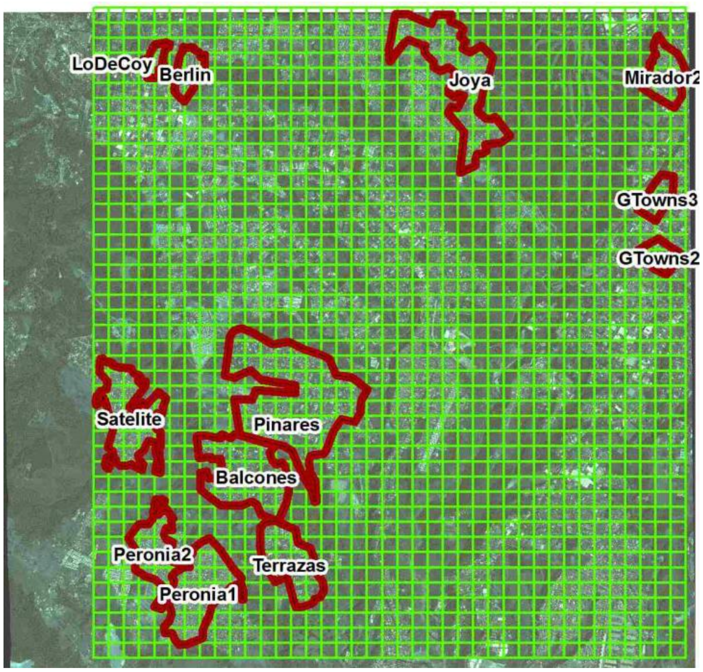



**Figure 17 150m2 grid overlaying the entire study boundary**

Where $(\mathcal{S}) = 0.111$, and d, the acceptable margin of error is 5%. If the margin of error is increased to 6%, a sample size of 118 is acceptable. Therefore, 170.5 > sample_size > 118 is sufficient to accept a margin of error between 5 to 6%. A 150m$^2$ sample grid contains 22,500 m of area and 62,500 0.6m pixels. A 150m$^2$ grid pattern overlaid upon the study area yields 1,677 samples, including many that do not intersect the delineated settlements of interest, as shown in Figure 17. This grid size is necessary to ensure sufficient inclusion of the ASTER 30m pixels and the road network indicators. Each grid cell then becomes a sampling unit with up to 62,500 pixels at a 0.6m spatial resolution. However, samples

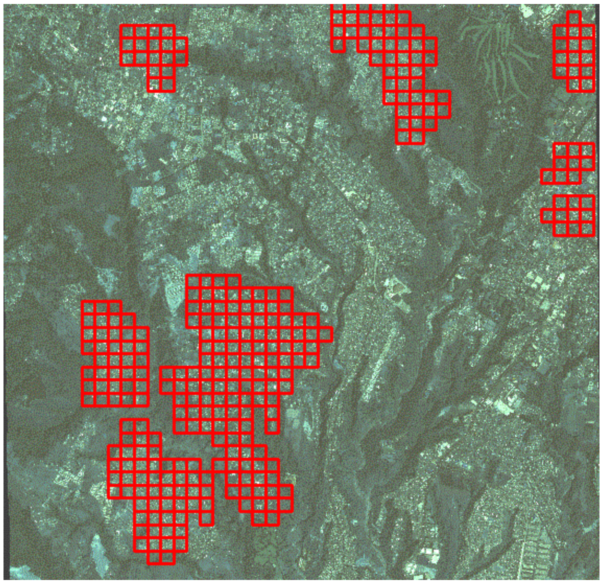


**Figure 18 Sample cells intersecting settlement boundaries**

must also intersect the settlement. The rectangular matrix of grids was sub-sampled to only include samples intersecting the settlement boundaries, yielding 316 samples, shown in red (Figure 18). Grid cells intersecting a given settlement were then assigned the same '*settle_id*' value for that settlement. In order to support random sampling with the sufficient number of samples required for robust statistical analysis, 50% of samples were selected randomly from each settlement (by *settle_id*) using the "sampling by subset" method from Hawth's Tools which produced 161 samples, yielding a sample size sufficient to explain results with < 6% margin of error (Beyer, 2004; Bartlett, II et al., 2001). A further reduction of sample grid cells was required to ensure adequate intersection area with the settlement. To explain this, a zoomed-in view of several selected grids overlaid on the image scene shows image and road features. In the example in Figure 19, the grids are labeled "Yes" if they were randomly selected and "No" if they were not.

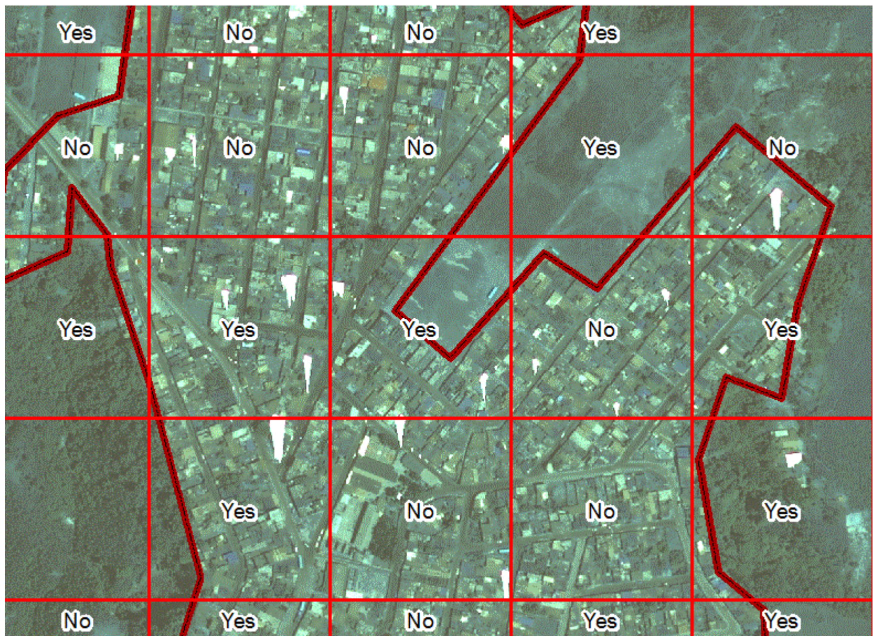



**Figure 19  Sample Grids Randomly selected = "Yes"**

Some "Yes" samples that were selected contain sizable area outside the settlement boundary.  If sample grids with < 25% intersection with a settlement are eliminated, the sample size is reduced to 127.  Based on Cochran's formula discussed previously, the acceptance of a 5% margin of error requires a sample size of 170.5, while a 6% margin of error requires a sample size of 118.  Therefore, the sample size of 127 is sufficient for two evaluation classes (informal / formal) with < 6% margin of error.

Figure 20 displays the final cells selected as samples highlighted in blue. Using this strategy, each settlement has at least 2 samples. Of all 127 samples used in this sub-

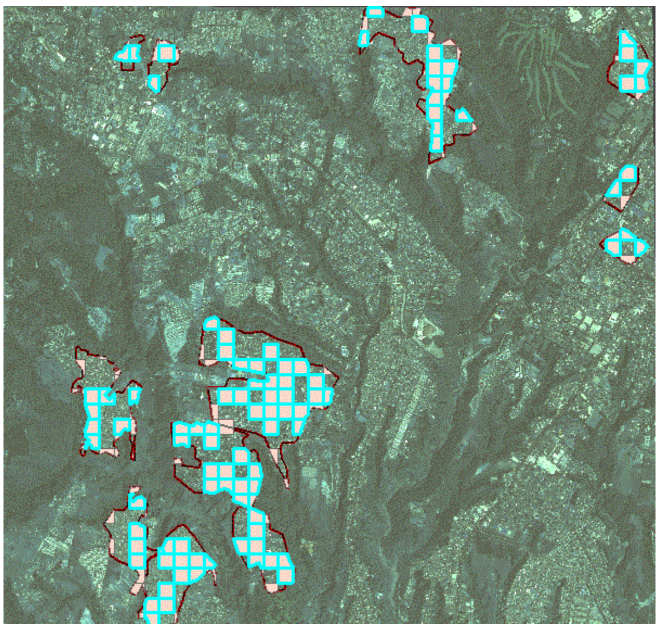


**Figure 20 Sample Grids with > 25% Coverage Intersecting Settlement Boundaries**

sampling method, the mean sample grid cell overlap with the actual settlement was 73%. Only 9 sample grids had < 30% overlap but > 25% overlap. To summarize, proportional stratified random sampling was performed using 50% of grid cells randomly selected from each settlement to achieve a sample size large enough to test statistical significance of the model (Beyer, 2004). Of the 50% selected, only grid cells with at least 25% overlapping coverage of the settlement boundary were included in the sample.

In the early stages of this research, a number of indicators based on the shape and size of individual building footprints in the settlement were considered. These indicators included area in $m^2$, roof corner angles, house distance to road, nearest neighbor distance using house centroid, among others found in Table 6. Theoretically, reliable techniques to extract such building footprints from large areas and multiple neighborhood morphologies are possible. However, initial efforts of semi-automated

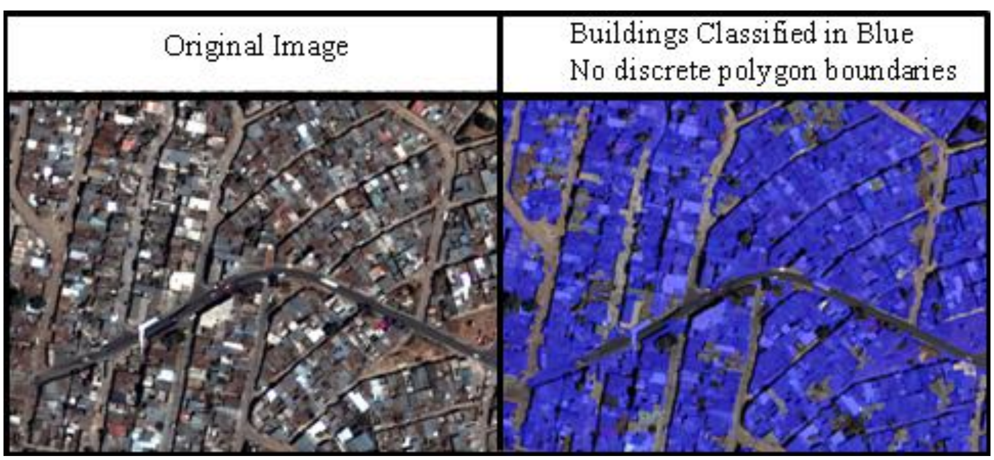



**Figure 21 Building Feature Extraction Performance**

extraction of dwellings produced unsatisfactory results. Buildings were incorrectly merged which inflated their sizes, and building edges became pixilated so that additional smoothing would have artificially altered their original shape. Figure 21 is an example of building extraction results from the Quickbird data using commercial image processing software.  Diverse construction materials, numerous rooftop objects, continuous rooflines, and building separation less than the minimum mapping unit of 2.4m in the Quickbird multispectral image rendered the building footprint extraction for the entire study area unworkable. The decision was therefore made to focus on measures of image spectra, roads, texture, scale, and topography, which do not require extraction of building footprints.

Roads were digitized in a semi-automated fashion because automated approaches did not perform well, creating incorrect vertices and discontinuities where road surfaces were spectrally mixed and roads were occluded by building and vegetation shadows. Topologic correctness was enforced at intersection vertices in order to compute some of the road network accessibility measures. The resulting road vector files were also visually compared to 0.5m orthoimagery for accuracy.

## Settlement Indicator Categories

These indicators were developed using a knowledge-based approach that merges prior research on sprawl (see Table 8 Summary of Sprawl Indicators adaptable to Informal Settlement Measurement) with research on informal settlements (see Table 6 Summary of Indicators Used in Literature to Evaluate Informal Settlements ). This approach takes full advantage of the integration of remote sensing and GIS techniques.  A total of 23 indicators grouped into 5 categories were measured and tested (Table 11). These categories were developed according to the following principles:

1. Spectral information from imagery is especially useful in understanding and classifying surface materials that represent patterns of human habitation. These include vegetation, roads, and built-up areas.

2. Roads are a direct representation of the movement of people across a landscape, and their composition, shape, and topology helps understand residents' interaction with their environment.

3. The built-up areas comprised of man-made structures can yield important patterns related to scale and dimensionality that could expose differences in underlying socioeconomic status of residents.

4. Texture of the high resolution panchromatic band in an image provides an enhanced view of intensity variation and arrangement that may be used to evaluate the continuous patterns of human settlement.

5. Underlying geomorphology is generally a fixed aspect of the landscape, and the topography upon which structures are built impacts living conditions as they relate to weather, climate, and natural or man-made hazards and is an indicator of the economic value of the land.

The following table summarizes the categories derived from the above principles, and provides a description of the individual indicators under each category type.

**Table 11 Settlement Type Discriminator Categories**

| Category Type | Description |
|---|---|
| MEASURES DERIVED FROM SPECTRAL ANALYSIS | • Vegetation percent<br>• Mean vegetation patch size<br>• Mean Vegetation patch Compactness Ratio<br>• Soil percent<br>• Soil on Roads<br>• Asphalt on Roads |
| ROAD MEASURES | • Connected node ratio<br>• Mean Node Valence<br>• Percent 4-Way Intersections<br>• Dangle ratio<br>• Road density per area<br>• Unpaved-to-paved road ratio |
| TEXTURE MEASURES | • GLCM Mean<br>• GLCM Correlation<br>• Entropy |

| | |
|---|---|
| | • Contrast on Roads<br>• Entropy on Roads |
| SCALE-RELATED MEASURES | • Lacunarity of Pseudo Built-up areas<br>• Fractal dimension of Pseudo Built-up areas |
| TOPOGRAPHIC MEASURES | • Mean degree slope<br>• Slope of roads<br>• Plan Convexity<br>• Profile Convexity |

Some indicators overlap categories. For example, degree-slope is considered a topographic measure, but slope on roads relates to both topography and roads. Entropy is a texture measure, but is calculated both as scene entropy and entropy of road surfaces. As discussed in the sampling design section, the measures originally created for the entire settlement boundary had to be clipped to support sub-sampling according the following methods listed in Table 12.

**Table 12 Sample Cell Clipping Methods by Indicator Type**

| Indicator Category | Indicator Name | Measurement Unit | Clipping Method |
|---|---|---|---|
| Measures Derived from Spectral Analysis | Vegetation percent | Percent area | Clip vegetation ROI to grid. Percent of vegetated pixels. Raster-derived. |
| | Mean Vegetation Patch Size | $m^2$ | Use patch size of all patches that intersect grid. Do not clip patches. |
| | Vegetation patch Compactness ratio | Ratio | Use Compactness ratio of all patches that intersect grid. Do not clip patches. |
| | Soil percent | Percent area | Clip soil ROI to grid, then recompute total percent of soil pixels. (Not vectorized). |
| | Asphalt road content | Percent | Proportion of asphalt pixels intersecting the road pixels, clipped to grid. |
| | Dirt road content | Percent | Proportion of soil pixels intersecting the road pixels and clipped to grid |
| Road Measures | Connected Node Ratio | Ratio | Ratio of all connected nodes to unconnected nodes intersecting the sample grid cell. |
| | Mean Node Valence | Mean | Mean of number of roads meeting at an intersection that intersect the sample grid cell. |

| | | | |
|---|---|---|---|
| Road Measures | Ratio of 4-way intersections | Ratio | Ratio of 4-way to all intersections that intersect the sample grid cell. |
| | Dangle Ratio | Ratio | Ratio of dangle nodes to all nodes that intersect the sample grid cell. |
| | Road Density per area | Ratio, length:area | Clip and re-node the road network to each sample grid cell, recompute length, sum length of all road segments clipped to grid, calculate ratio of length to area of sample grid. |
| | Unpaved-to-paved road ratio | Ratio | Ratio of dirt roads to asphalt roads – ratio per sample grid cell based on dirt and asphalt pixels intersecting the roads pixels. |
| Texture Measures | Contrast | Mean, unitless | Compute GLCM Contrast for entire image (avoid edge effects). Mask contrast pixels intersecting the sample grid cell, calculate mean. |
| | Entropy | Mean, Unitless, always positive and < 3 | Compute Entropy for entire image (avoid edge effects). Mask entropy pixels intersecting the grid cell, calculate mean |
| | Contrast on Roads | Mean, Unitless | Mean of GLCM Contrast pixels intersecting roads pixels clipped to the sample grid cell. |
| | Entropy on Roads | Mean, Unitless, always positive and < 3 | Mean of Entropy pixels intersecting roads pixels clipped to the sample grid cell. |
| Scale-Related Measures | Fractal Dimension | Fractal(D) is $-\lim[\log N_{\varepsilon}/\log_{\varepsilon}]$ | Clip binary image to sample grid cell, then compute Fractal (D) and Lacunarity(Λ) for box size 25, 50 and 75, then evaluate. A box with <45% foreground pixels is excluded from the calculation. |
| | Lacunarity | Lacunarity is $\Lambda\varepsilon = (\sigma/\mu)^2$ | (same as for Fractal D) |
| Topographic Measures | Degree Slope | Degrees (0 – 90) | First compute Degree slope from ASTER masked to original image boundary (avoid edge effects). Select Degree Slope pixels intersecting sample grid cell. |
| | Slope on Roads | Degrees (0 – 90) | First compute Degree slope from ASTER masked to original image boundary (avoid edge effects). Select Degree Slope pixels intersecting roads in sample grid cell. |
| | Plan Convexity | Unitless (+ if convex, - if concave) | First compute Plan Convexity from ASTER masked to original image boundary. Calculate mean of pixels that intersect sample grid. |
| | Profile Convexity | Unitless (+ if convex, - if concave) | First compute Plan Convexity from ASTER masked to original image boundary. Calculate mean of pixels that intersect sample grid. |

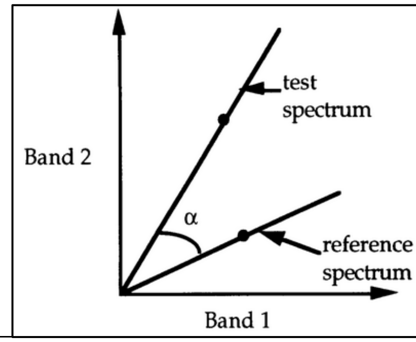
### Settlement Indicator Expected Values and Measurement Methods.

In this section, the indicators for each category are described by indicator name, expected informal settlement value as compared to formal, and methods or formulas for computation. Each table is followed by a narrative summary of methods for each indicator.

#### *Measures Derived from Spectral Analysis*

**Table 13 Measures Derived from Spectral Analysis – expected Informal Values and Methods**

| Indicator Name | Expected Informal Values | Methods and Tools |
|---|---|---|
| MEASURES DERIVED FROM SPECTRAL ANALYSIS | | |
| Vegetation Percent | Expect 5-10% vegetation, less than formal | Normalized Difference Vegetation Index (NDVI) computed from Red (0.63 – 0.69$\mu m$) and NIR (0.76 – 0.9 $\mu m$) bands in Quickbird $$\frac{(NIR - Red)}{(NIR + Red)}$$ Used a threshold of 0.06. $-1 \geq$ NDVI $\leq 1$ |
| Mean Vegetation Patch Size | Smaller than formal | Vectorization of NDVI result, with ( < 10m holes filled) to reduce pixilation |
| Vegetation Patch Compactness Ratio | More compact than formal | $$C_i = \sqrt{\frac{A_i}{B_i}}$$ where $A_i$ = Area of Patch, $B_i$ = Area of Circle with same circumference as patch. (*compactness* = $4\pi(area/perimeter^2)$) |
| Soil Percent | Higher soil than formal | ENVI Spectral Angle Mapper classifies soil areas from training (reference) polygons of known soil areas. (Kruse et al., 1993; Puetz & Olsen, 2006) $\alpha = cos^{-1}\left(\frac{\vec{t}\cdot\vec{r}}{\lVert\vec{t}\rVert\cdot\lVert\vec{r}\rVert}\right)$ |

<table>
<tr><td></td><td></td><td>Where t is the test spectrum and r is the reference spectrum. The smaller the $\alpha$ in radians, the higher the similarity between the test and reference spectra. Max angle of 0.07 rad was used.


</td></tr>
<tr><td>Asphalt Road Content</td><td>Less asphalt on roads than formal</td><td>Rasterize 1m buffered road layer, measure intersection with SAM-classified asphalt layer for entire scene. The 1m buffer ensures sufficient road material from the centerline is captured in the measure.</td></tr>
<tr><td>Dirt Road Content</td><td>Greater dirt roads than formal</td><td>Rasterize 1m buffered road layer, measure intersection with SAM - classified soil layer for entire scene. The 1m buffer ensures sufficient road material from the centerline is captured in the measure.</td></tr>
</table>

Spectral analysis provides the ability to classify surface materials of roads and of the spaces in-between buildings which are comprised of mostly soil (dirt) or vegetation. It was hypothesized that road surfaces would consist of more dirt or soil in the informal areas due to limited infrastructure development. In many parts of the world, living conditions with extreme poverty create overcrowded, narrow roads that are difficult to navigate, and would therefore be difficult to pave, especially when dwelling setbacks are non-existent or not enforced. Two methods were used for surface material detection. Vegetation was detected using the reliable NDVI vegetation index from the red (0.630 – 0.690 µm) and near infra-red (NIR) (0.760 – 0.900 µm) bands. Dark object subtraction was applied (ENVI, 2009). This is an image enhancement technique that searches each band for the darkest pixel value. Assuming that dark objects reflect no light, the darkest pixel value (> 0) must result from atmospheric scattering. The scattering is removed by subtracting this value from every pixel in the band. This simple technique is effectively used for haze correction in multispectral images (Kruse et al., 1993; Exelis, 2009). The entire image scene was classified once in ENVI™, using a threshold value of 0.06 which appeared to correctly delineate most vegetated areas and resulted in acceptable accuracy. After classification, the raster .tif was polygonized. A selected informal (*Peronia2*) and formal (*Balcones*) vegetation classification result are displayed to provide a visual understanding of vegetation polygon differences between formal and informal settlements as measured for shape-based vegetation metrics.

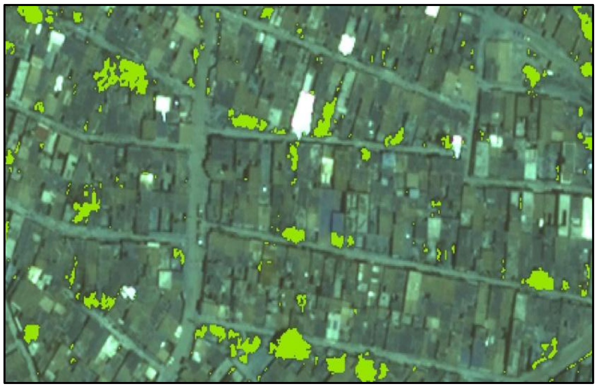


**Figure 23  Informal Settlement vegetation example**

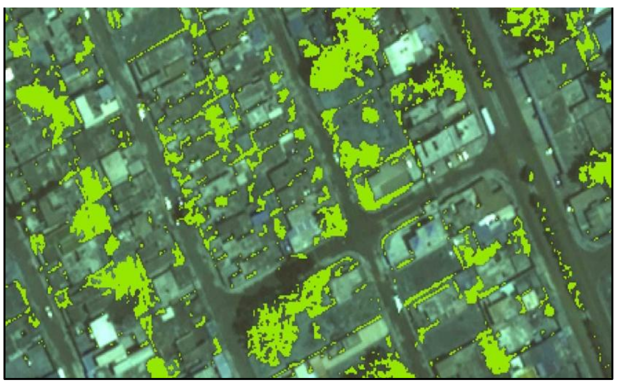


**Figure 22 Formal Settlement vegetation example**

Vegetation percent was calculated from the proportion of vegetation-classified pixels to non-vegetated pixels. Vegetation patches were converted to polygons based on the algorithm resident in the ENVI software. The patch size was then calculated as an area measure in ArcGIS (ESRI, 2009). Figure 22 and Figure 23 graphically depict the difference in appearance of vegetation from two settlements, one informal and one formal.

In shape analysis, the compactness ratio of a feature is simply the square root of the area of polygon divided by the area of circle with circumference equal to perimeter of the polygon (O'Sullivan & Unwin, 2003a, p. 178; Cockings, Fisher, & Longford, 1997). The following equation was used to calculate circularity:

$$C_i = \sqrt{\frac{A_i}{B_i}}$$

where $A_i$ = Area of $\text{Patch}_i$, and $B_i$ = Area of circle with same circumference as $\text{Patch}_i$. Figure 24 is a depiction of compactness ratio, also known as circularity ratio.

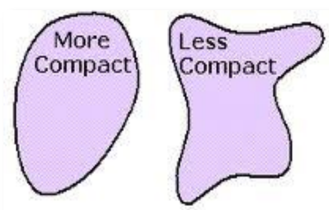



**Figure 24 Visual Example of Compactness Ratio of Polygons**

In the informal communities, vegetation patches were expected to be more compact, assumed to be unkempt and not used as elongated border-like ornamental plantings such as may be more prevalent in a planned community.

The Spectral Angle Mapper (SAM) method was used to classify soil in the entire image and then evaluate the percent soil composition for each masked polygon using ENVI (Exelis, 2009; Kruse et al., 1993). As implemented in ENVI, SAM is a type of spectral classification that uses an *n*-dimensional angle to match pixels to reference spectra from known soil areas. The SAM algorithm determines the spectral similarity between two spectra by calculating the angle between the spectra and treating them as vectors in a space with dimensionality equal to the number of bands (Kruse et al. 1993). More soil than vegetation was expected, due to less orderly movement of people and vehicles, lack of landscaping (greenery), and expected greater presence of dirt roads. Dirt/soil pixels were transferred to intersecting roads to determine the percent road coverage with dirt/soil. The hypothesis was that infrastructure development in the informal areas was low, and that more dirt roads would be present. The SAM method was used to classify asphalt in the entire image scene. Then, asphalt pixels were transferred to intersecting roads pixels to determine the percent road coverage with asphalt in the same manner as soil.

### *Roads Measures*

**Table 14 Roads Measures - Expected Informal Values and Methods**

| Indicator Name | Expected Informal Values | Methods and Tools |
|---|---|---|
| **Road Measures** | | |
| Connected Node Ratio | Lower than formal | [intersections / (intersections – dangles)] (Song & Knapp, 2004); (Dill, 2004)<br>Intersections defined as ≥ 3 connected nodes. |
| Mean Node Valence | Lower than formal | Node valence = # of intersecting roads. This measures the # of joining nodes at each intersection in the sample. |
| Percent 4-way Intersections | Lower than formal | Ratio of 4-way intersections to all intersections of ≥ 1 node (Zhang & Kukadia, 2005). |
| Dangle Ratio | Higher than formal | Of all nodes, % dangles (dead ends) within sample (Yi, 2008). |

| | | |
|---|---|---|
| Road Density per Unit Area | Higher than formal | Expected positive correlation with dwelling size and density – expect dwellings more densely packed. However, we use sample land area instead of dwelling counts (Dill, 2004; Lee & Ahn, 2003) |
| Asphalt Road Content | Lower than formal | Transfer SAM-classified asphalt pixels to road pixels. For asphalt, max spectral angle of *0.06 radians* was used. Polyline roads buffered to 1m from centerline and rasterized. Intersection of SAM-classified result with 1m buffered roads using ArcGIS "Extract By Mask" and 0.6m pixel size. Raster attribute table for roads gives count of each surface type. |
| Dirt Road Content | Higher than formal | Transfer SAM-classified soil pixels to road pixels. For soil/dirt, the max spectral angle of *0.07 radians* was used. Same methods as Asphalt Road Content. |
| Unpaved: Paved Road Ratio | Higher than formal | Expected a higher ratio of Unpaved-to-paved roads in the informal settlements. Uses a pixel × pixel transfer function between 1m road buffer (2m total width upon centerline) and surface feature image classification. Where no roads were present, missing values were excluded from the comparison. |

All types of roads visible within each settlement boundary were digitized at a 1:6,000 spatial scale from the pan-sharpened orthorectified Quickbird imagery using ArcGIS (ESRI, 2009). For network analysis including node valence, connected node ratio, and dangle ratio, the intersecting nodes were calculated, and adjusted to remove false dangles where the settlement boundary bisected a road that actually continued on to a larger road network. If the bisected road was a dirt road, it was considered a dangle - such roads did not generally extend into the road network for an adjacent settlement; if the bisecting node intersected any other surface, it was removed from the dangle list for that settlement. Further descriptions of each road measure follow.

Connected node ratio (CNR) is defined by the number of street intersections divided by the number of (intersections + dangles) (Tresidder, 2005). Values are expressed as $0.2 \leq CNR \leq 1$. Values closer to 1 indicate higher connectivity, fewer cul-de-sacs or dead ends. A lower connected node ratio was expected in the informal settlements. Mean Node Valence is a measure of the number of nodes connected to the given node. A 4-way intersection has a valence of 4. Here, the node valence is computed as the total nodes/total intersections per settlement expressed as a mean node valence per settlement. Percent 4-Way Intersections represents the percent of intersections that are 4-way as a ratio of all intersections. In this study area there was only one 5-way intersection in the settlement of *Piñares*. Informal settlements were expected to have less 4-way intersections than formal, planned communities.

The dangle ratio is the ratio of dangle nodes (dead end streets or cul-de-sacs) to all nodes (Yi, 2008). Settlements with less connectivity will have more dead ends. The expectation was the informal settlements would have a higher dangle ratio as an indicator of less connectivity.

Asphalt road content represents the percent of road surface covered by asphalt. Surface asphalt was classified using SAM. The result was threshholded and saved as a binary image classified as (asphalt / no asphalt). The result was transferred to the intersecting road image using Python code that called the ArcGIS Geoprocessor, a raster attribute table was created, and then queried to count asphalt/non-asphalt pixels (1, 0). More asphalt was expected for the formal settlements.

Dirt road content represents the percentage of road surface covered by dirt or soil. The same methods used to classify asphalt road content were used to measure dirt road content.  Road surface was expected to be comprised of more dirt vs. asphalt in the informal areas.

A high road density can indicate higher building density, greater transportation infrastructure density, or higher connectivity. This measure cannot be used to quantify connectivity without also measuring connected node ratio or node valence. The linear length of each road segment within the sample was summed and divided by the area of the sample.  High road density per area was expected in the informal settlements simply due to the higher building density.

This research hypothesized that informal settlements would have a higher dirt-to-asphalt road surface ratio based on expected lower infrastructure spending by government, and reduced centrality to the city center. There are many other road surface types in Guatemala, but dirt and asphalt were easiest to distinguish using the reference orthoimage, and were sufficient to measure the hypotheses based upon the underlying settlement type.

### *Texture Measures*

**Table 15 Texture Measures – Expected Values and Methods**

| Indicator Name | Expected Informal Values | Methods and Tools |
|---|---|---|
| **TEXTURE MEASURES** | | |
| GLCM Contrast | Higher than formal | Gray-Level Co-occurrence matrix (Haralick et al., 1973); (Shamir, Wolkow, & Goldberg, 2009) is: $M_d(i,j) = \left\|\left[\left((r,s),(t,v)\right): I(r,s) = i, I(t,v) = j\right]\right\|$ $where\ (r,s),(t,v) \in N \times N\ and\ (t,v) = (r+dx, s+dy)$ where *i*,*j* of the co-occurence matrix $M_d$ is the number of occurrences of gray levels *i* and *j* such that the distance |

| | | |
|---|---|---|
| | | between them is *d* pixels.<br>High contrast is a large difference in intensity of neighboring pixels (or high local variation) and greater expected variety of surface materials, shapes, and sizes.<br>$p(i,j) = (i,j)th\ entry\ in\ normalized\ quantization - level\ spatial\ dependence\ matrix$<br>$N = No.of\ distinct\ quantization\ levels\ in\ quantized\ image$<br>$\displaystyle\sum_{i,j=0}^{N-1} P_{i,j}\,(i-j)^2$<br>When i and j are equal, the cell is on the diagonal and *(i−j)=0*. These values represent pixels entirely similar to their neighbor, so they are given a weight of 0. If i and j differ by 1, there is a small similarity, and the weight is 1. If i and j differ by 2, contrast is increasing and the weight is 4. The weights continue to increase exponentially as *(i−j)* increases. (ref: http://mipav.cit.nih.gov/documentation/HTML%20Algorithms/FiltersSpatialHaralickTexture.html) |
| Entropy | Higher than formal | Entropy quantitatively measures randomness of the gray-level distribution. Defined as:<br>$\displaystyle\sum_{i,j=0}^{N-1} P_{i,j}\left(-ln\ P_{i,j}\right)$<br>$P_{ij}$ is a probability measure, where $0 \leq P_{ij} \geq 1$ so $ln(P_{ij})$ will always be 0 or negative, so we take the *–ln* to give us positive number<br><br>• Smaller values of $P_{ij}$ mean appearance of a given pixel combination is less common<br>• The smaller the value of $P_{ij}$, the larger the absolute value of $ln(P_{ij})$<br>• Max entropy occurs when derivative of *(p* ln(p))* with respect to p equals 0. *P = 1/e* gives maximum entropy value, which is P=1/2.718 = 0.378 see: http://mipav.cit.nih.gov/documentation/HTML%20Algorithms/FiltersSpatialHaralickTexture.html); (Sudhira et al., 2004) |
| GLCM Correlation | Lower than formal | GLCM Correlation measures the linear dependence of grey levels on neighboring pixels. GLCM Correlation is independent of the other texture measures.<br>GLCM Correlation is calculated from the mean and variance of neighboring pixel pairs<br>$\displaystyle\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right]$ |

| | | |
|---|---|---|
| | | When $P_{ij}=0$ the result is *0*, which is evidence of complete image uniformity. |
| GLCM Contrast on Roads | Higher than formal | Same formula as scene contrast as implemented in ENVI (2008) software (Haralick et al., 1973). |
| Entropy on Roads | Higher than formal | Same formula as scene entropy as implemented in ENVI (2008) software (Haralick et al., 1973) |

Image texture measures the variation in image tone (brightness values) in a variable-sized, contiguous matrix of pixels in the image, and identifies repeating patterns of local variation in intensity. It is convenient to measure texture properties using the grayscale 0.6m (panchromatic) band (vs. the 2.4m multispectral bands) from the Quickbird data due to the smaller pixel size. Shades of gray in values from 0-255 are represented in the 8-bit grayscale image. Much research has been conducted on the statistical properties of texture to extract and classify image regions. Research is lacking that compares how the texture measures vary by image object or feature type (e.g., roads), since texture is generally computed on a rectangular matrix and not clipped to a feature boundary.

The grey-level co-occurrence matrix (GLCM) is computed as a first step in the texture measures. The GLCM is a two-dimensional array, *P*, in which both rows and columns represent the set of all possible brightness values. It is defined by specifying a displacement vector $d=(dx,dy)$ and counting pairs of pixels separated by *d* having specific gray levels *i* and *j* such that

$P_d[i,j] = n_{ij}$ (Haralick et al., 1973), (Clausi & Jernigan, 1998)

In this research, the displacement vector (*1x*, *1y*) of 45° was used allowing a 3x3 matrix to systematically shift over the image to calculate the central pixel value. $n_{ij}$ is the number of occurrences of pixel values *(i,j)* at distance *d* in the image, and the matrix $P_d$ has dimension $n \times n$ where $n$ is 256 for the panchromatic band. The normalized co-occurrence matrix is:

$$N[i,j] = \frac{P[i,j]}{\sum_i \sum_j P[i,j]}$$

(Puetz & Olsen, 2006).

This normalizes the co-occurrence values so they lie between 0 and 1 as joint probabilities. Once the GLCM is computed, the rest of the texture measures can be created.

The measure of contrast was chosen because it characterizes local variation. Contrast is expected to be high in the informal areas due to inconsistency of building materials, smaller structures, and general haphazardness of features. Correlation was also chosen to determine the linear dependence of grey levels on neighboring pixels.

The Entropy value from the GLCM is a measure of information content resulting from the randomness of brightness values. High entropy means there are no preferred gray level value pairs in the distance vector *d*. Informal settlements were expected to have high entropy values relative to their planned and more affluent counterparts. This is anticipated from results found in prior research (Yeh & Li, 2001). The texture measures were calculated using the following steps:

- Select only the PAN band (0.6m, grayscale)
- Compute Entropy using 3x3 matrix and 64 bit quantization (≥ 5x5 matrix contains less discriminatory power)
- Load Vector of settlement and create .evf (ENVI vector format file)
- Create image subset using this vector resulting in rectangular matrix of values
- Create mask from non-rectangular settlement boundary
- Apply mask to ensure pixels outside boundary contain NaN's and are excluded from computations
- Compute first or second order texture measure excluding outside pixels

A novel approach of limiting the contrast and entropy calculations to the buffered road surface was applied. The roads raster was created from the original polyline feature layer that was buffered 1m in both directions from the centerline. It was determined that 2m in each direction from the centerline incorrectly included building edges of more narrow streets, and there was no additional discriminatory power in the results between informal and formal settlement types by extending the buffer to 2m.

The GLCM contrast measure was calculated for the entire image the same way as Entropy, then clipped to the settlement boundary. In this way, edge effects were minimized. Contrast was expected to be higher in the informal settlements. The same settlement contrast result was masked by the road vector to produce the Contrast on Roads measure. Similar to the expected high GLCM Contrast for informal settlements, Contrast on Roads was expected to be higher on the informal than the formal roads, due to the accumulation of varied surface materials and debris found upon the road surface.

Correlation was calculated for the entire image, and assumed to be lower in the informal areas than the formal areas. Entropy on Roads was calculated from the settlement entropy then masked and evaluated the same as Contrast on Roads. Entropy on informal roads was expected to be higher than formal roads.

### *Scale-related measures*

**Table 16 Scale-related measures - expected values and methods**

| **Indicator Name** | **Expected Informal Values** | **Methods and Tools** |
|---|---|---|
| **Scale-Related Measures** | | |
| Lacunarity | Higher than formal | Create binary image of only pseudo built-up areas. Image-J calculates lacunarity of images. Expected higher lacunarity for informal settlements reported in prior literature. (Filhos and Sobreira, 2007). $$\Lambda(r) = \left(\frac{\sigma}{\mu}\right)^2$$ Which represents the mean coefficient of variation (*CV)* in foreground pixels/box over all grid locations using the sliding box algorithm for a given box size (Karperien, 2007) |
| Fractal Dimension | Higher than formal | Expect higher $D_\beta$ for more compact areas and lower $D_\beta$ for more dendritic elongated shapes. Use Image-J with FracLac Plugin. $$D(B) = -lim\left(log\frac{N(\varepsilon)}{log\ (\varepsilon)}\right)$$ Where *D(b)* is the negative limit of the ratio of the log of the number of boxes at a certain scale (box size/image size) over the log of that scale (Karperien, 2007). |

Lacunarity is a measure of "gappiness" or translational or rotational invariance in an image scene, while Fractal D is a measure of complexity showing how the pseudo-builtup areas fill the space at a given scale (for example, 200). Overall mean Fractal D was expected to be higher for informal settlements – an indication of higher complexity (Cooper, 2005). Similar results were anticipated in this research, that lacunarity values would be higher in the informal settlements than formal settlements. This was important to test since other research showed a different result of lower fractal values in informal heterogeneous settlements on city peripheries (De Keersmaecker et al., 2003). Lacunarity or Fractal Dimension are not intuitive to measure with respect to settlement shape, so further detail regarding methods and algorithms are included here.

### Lacunarity of built-up areas

Lacunarity of built-up areas was performed with ImageJ, an image analysis software package written in Java, and developed at the National Institutes of Health (*Image J*, 2010). The lacunarity plug-in, FracLac (Karperien, 2007), was used to calculate lacunarity of both binary and grayscale images and was tested a variety of box sizes to determine if one or the other performed better at discriminating settlement type. In order to simulate the built-up nature of settlements as represented by non-vegetated structure, the inverse of NDVI classified areas was used, so that non-vegetated areas became foreground pixels for the sliding box algorithm. This method is reasonable given the extreme urban nature of the communities used for this study in Guatemala City. The pseudo built-up classification yielded an overall accuracy of 97% which was verified from visual interpretation of random samples from each settlement area using half-meter orthoimagery. NDVI values follow: $(-1) \leq NDVI \geq (+1)$. The range of −1 to 0.09 was used to represent the non-vegetated (built-up) areas, while any value > 0.09 was considered vegetation based on a visual comparison with the image scene. This was reasonable due to the fact that the delineated urban scenes, in terms of human habitation, consisted of two main classes: built-up or vegetated land. Figure 26 GTowns2 settlement with masked NDVI result is a floating point NDVI image from the formal *Granai Townson 2* settlement, abbreviated as “GTowns2”.

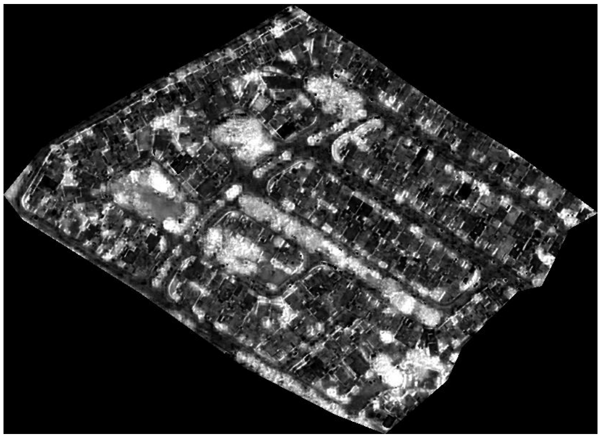


**Figure 26 GTowns2 settlement with masked NDVI result**

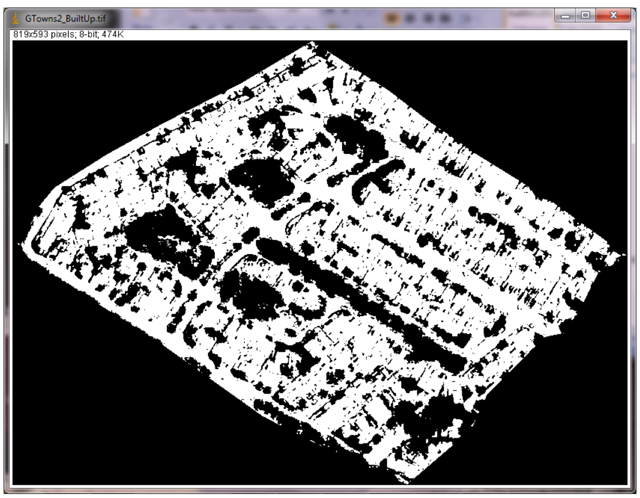



**Figure 25 GTowns2 pseudo Built-up Binary image**

After applying a threshold to select values > 0.09, the resulting binary *(*vegetated=0, built-up=1*)* classification result is shown as a tagged interchange file format (TIFF) in Figure 25 with all foreground pixels displayed in white, while vegetation is black. The threshold value was determined from visual qualitative comparison with the reference orthoimage.

To calculate lacunarity, assume $r$ = box size and $S$ = box mass (# foreground pixels). A frequency distribution of box masses $n(S,r)$ is then converted to a probability distribution $Q(S,r)$ by dividing each frequency value by the total number of gliding boxes of a given size $N(r)$. Then calculate: $Z(1) = \sum S\,Q(S,r)$ and $Z(2) = \sum S^2\,Q(S,r)$. Lacunarity, $\Lambda(r)$, is then expressed as:

$$\Lambda(r) = Z(2) / [Z(1)]^2$$ (Filho & Sobreira, 2007).

In the case of FracLac, for each $r$, we calculate $(\Lambda(r) = (\sigma / \mu)^2)$ representing the average coefficient of variation in foreground pixels/box over all grid locations (Karperien, 2007). The box size or scale is represented as $\frac{1}{image\ size}$. The sliding box method averages out the results over the entire scanned image, producing a single value for $\Lambda(r)$. It is also important to note the final $\Lambda(r)$ value is transformed such that the result is expressed as $\Lambda(r)+1$ to avoid undefined results in a completely homogenous image[5].

[5] The reason the scaled value for lacunarity is $\mathbf{F\lambda=1+(\sigma/\mu)^2=1+\lambda}$ is that when the slope of the *ln-ln* plot of the regression line for each ε grid size (scale) and the foreground pixel count is calculated, an "undefined" value could result for a completely homogeneous image. This means pixels per box would not vary, so that the standard deviation σ, for a box count at some ε will be 0. This means that $\lambda = (\sigma/\mu)^2 = 0$, and ln of λ and therefore the slope of the *ln-ln* regression line for λ and ε would be undefined. The transformation is λ+1, so that a completely homogeneous image has a slope of 0, corresponding intuitively to the idea of no rotational or translational invariance and no gaps.

Fractal dimension of built-up areas

ImageJ was also used to calculate fractal dimension, using the formula:

$$D_B = -lim[logN\varepsilon/log\varepsilon]$$

which is the negative limit of the ratio of the log of the number of boxes at a certain scale (e.g., box size / image size) over the log of that scale (Karperien, 2007). This also represents the slope of the regression line for the log-log plot of scale (box size / image size) and foreground pixel count.

The sub-sampled cells from the pseudo built-up image were converted to binary using an image threshold of 0.09. Figure 27 depicts the difference between a sampled

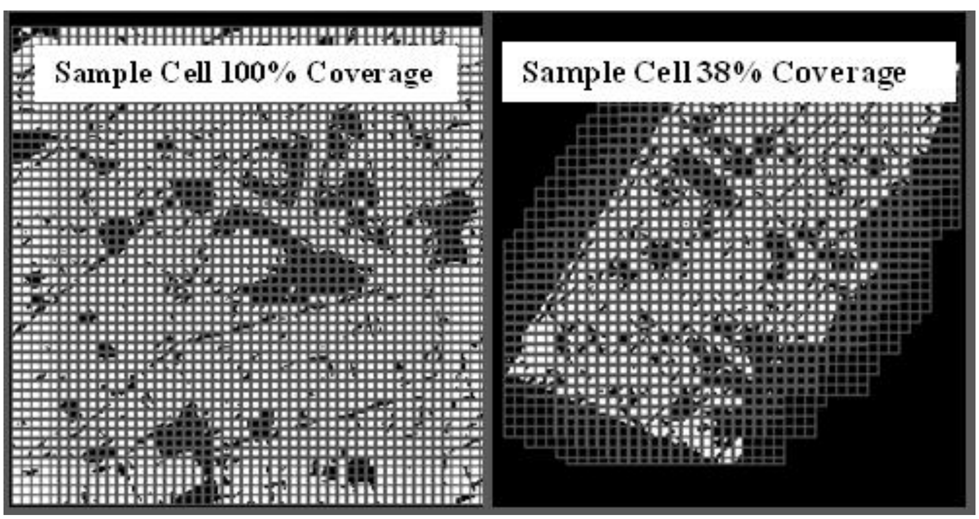



**Figure 27 Gliding box lacunarity intersecting 100% or 38% with sample grid**

cell with 100% area intersection within the Peronia2 informal settlement boundary, while the sample cell to its right only covers 38% of the settlement boundary – an artifact of random sampling.

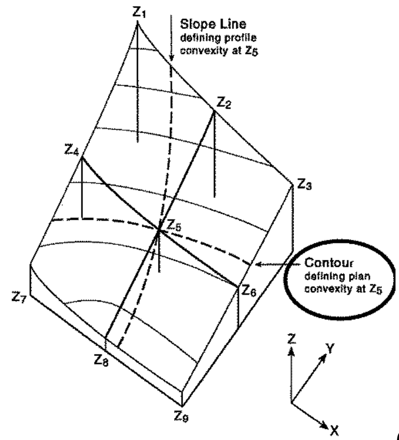
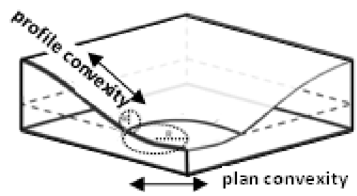
Fractal dimension was calculated on each binary sample over all grid sizes. Due to the smaller size of each sample compared to the initial testing of the entire settlement as a sampling unit, smaller box sizes were needed. For the settlement as a sampling unit, the statistically significant difference in lacunarity and fractal dimension occurred at a box size of 200. However, results could not be computed on box sizes of $\geq$ 100 for the 150$m^2$ sub-samples as there was not enough foreground texture included in the calculation, especially considering some sample grid cells only intersected 25% of the settlement. Therefore, smaller box sizes of 25, 50, and 75 were computed for the sliding box lacunarity on all sub-sampled cells. The following side-by-side images in Figure 28 depict the sliding box lacunarity grid overlay at box sizes of 25, 50, and 75 for a sub-sample in the formal settlement of Balcones, showing how background NaN areas in the box size of 25 and 50 are not included if the foreground-to-background pixel ratio is less than 0.45.

**Figure 28 Sliding Box Lacunarity - differences in box sizes (25, 50, 75)**

### *Topographic Measures*

**Table 17 Topographic Measures – expected values and methods**

| Indicator Name | Expected Informal Value | Methods and Tools |
|---|---|---|
| **Topographic Measures** | | |
| Slope on Roads (μ deg°) | Higher than formal | Rasterize road vector polyline to 0.6m of centerline, intersect with resampled ASTER DEM (30m) elevation data from which degree slope was calculated with a 3x3 kernel (ENVI, 2009). Compute μ of all degree-slope pixels intersecting raster road pixels where slope is 100*tan(θ). |
| Mean Slope | Higher than formal | Degree Slope expressed as arctan (Δelevation/distance). Derived from ASTER 30m elevation data. |
| Plan Convexity | Smaller (less negative) value than formal | Plan Convexity measures variability in horizontal contour, or, the rate of change of aspect. It is a $2^{nd}$ derivative geomorphology measure of surface roughness. Closest to 0 on steep slopes, with extreme negative values in floodplains. If informal areas located on floodplains, this measure will have extreme negative values. |

| | | |
|---|---|---|
| | |  (Evans & Cox, 1999:17)<br> (Ioannilli and Parmegiani, 2008)<br>Plan convexity affects the depth of hydrologic flow, impacting the area drained based on the slope position in the terrain surface. (also see: JP Wood's Thesis at: http://www.soi.city.ac.uk/~jwo/phd/04param.php) |
| Profile Convexity | Smaller (less negative) value | Profile Convexity measures rate of change in a vertical downslope direction. It is a 2nd derivative geomorphology measure of surface roughness, and affects acceleration of surface flow (Evans and Cox, 1999:16). Closest to 0 on steep slopes, with extreme negative values in floodplains. If informal areas located on floodplains, this measure will have extreme negative values. See diagram under Plan Convexity, above. |

Topography and geomorphology play a large role in deciding where to build a home, and is often correlated with living costs related to rent or property values. Notwithstanding the unusual case of high-value properties located on steep cliffs overhanging scenic views, many areas in less developed countries with steep terrain pose a landslide risk and are rejected by only the most desperate of home builders. This presumption underlies the hypotheses related to topography in this study. The assumption is that steep slopes in mountainous Guatemala are difficult to build upon, are not well-served by suitable buttressing or paved infrastructure, and therefore are selected by economically disadvantaged residents who require less-expensive land for their homes, similar to the hazardous slopes upon which slums (favelas) in Rio de Janeiro are built (Vieira & Fernandes, 2004).

The second topographic hypothesis is that the degree of surface curvature, also described as roughness, would characterize extremes of economic well-being of residents, such that more extreme rough or curved surfaces in either the vertical (profile convexity), or the horizontal (plan convexity) direction would also be found in more informal, unplanned neighborhoods. Plan curvature, including convexity and concavity in Figure 30 or profile curvature in Figure 29 create either a divergence or convergence of

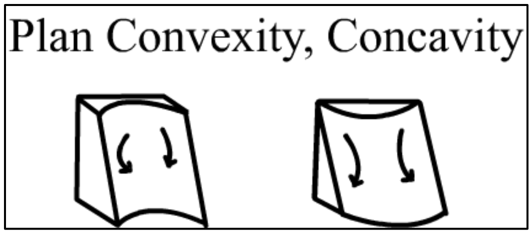



**Figure 30 Plan Convexity, Concavity (horizontal direction) (Owen, W. 2011)**

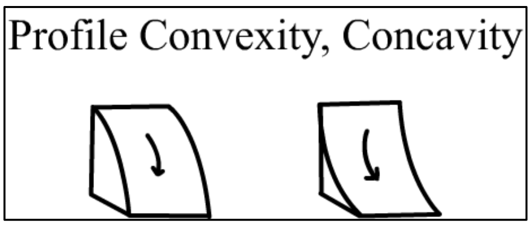



**Figure 29 Profile Convexity, Concavity (vertical direction) (Owen, W. 2011)**

flow (Depraetere & Riazanoff, 2004), which can cause either a deceleration (convexity) or acceleration (concavity) of runoff in extreme precipitation events (Depraetere & Riazanoff, 2004). Increased surface roughness can be more difficult to build dwellings upon and also impacts friction or resistance to hydrologic flow, which impacts soil saturation. Use of plan and profile measures could also help predict where lava flows and extreme rain events causing landslides and erosion could impact population living near or amidst those hazards. Despite the ambiguity in the interpretation of what plan and profile convexity measures reveal about their underlying settlements given the many other variables that impact slope stability (Vieira & Fernandes, 2004), the simplicity of testing this topographic feature allowed it to be incorporated into the current research. Therefore plan and profile convexity were pursued as additional topographic measures to determine if their values differed significantly between settlement types.

For the topographic measures, a 3x3 kernel (window size) was used, and computed from the topographic modeling tools in ENVI™ against the ASTER DEM for the scene. The result was masked to each settlement boundary. As background, elevation is a 0 order differential, while slope is a $1^{st}$ order differential, and plan and profile convexity are $2^{nd}$ order differential surface geomorphology measures. In areas susceptible to flooding from rivers or shallow land forms that attract low-income dwellers, one might expect plan convexity to be highly negative. One might also expect highly negative profile convexity to represent concave surfaces where water flow during rainfall pools and accelerates in the downslope direction, and also where roughness of surface in aspect make building conditions less favorable. The following diagrams show side-by-side the original ASTER DEM, Degree Slope, and Profile Convexity. Clearly visible are the hydrologic features from the elevation data that propagate through the other topographic images.

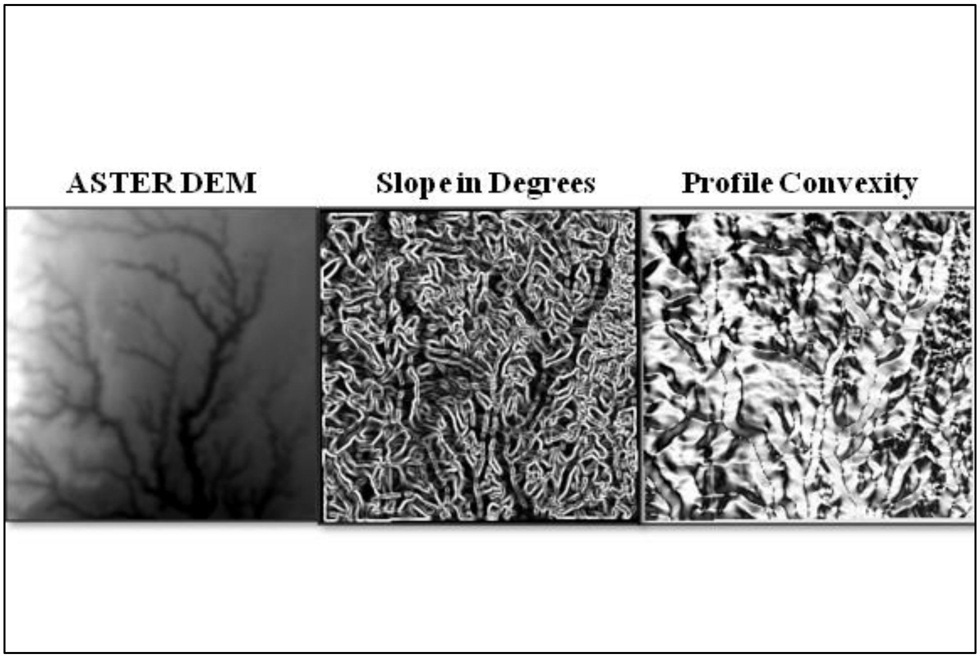


**Figure 31  ASTER-derived Topographic and Geomorphology Measures**

To illustrate the degree slope variability, a transect for two settlements in the direction of the major axis of the settlement minimum bounding rectangle for two settlements, GTowns2 (formal), and Peronia2 (informal) was created. A transect distance of 750m was chosen because it was of sufficient length to cover the smallest of the two settlements' bounding rectangle major axis.  Figure 31 shows the reduced degree slope variability in the formal settlement GTowns2 compared to Figure 33 for the informal settlement Peronia2.

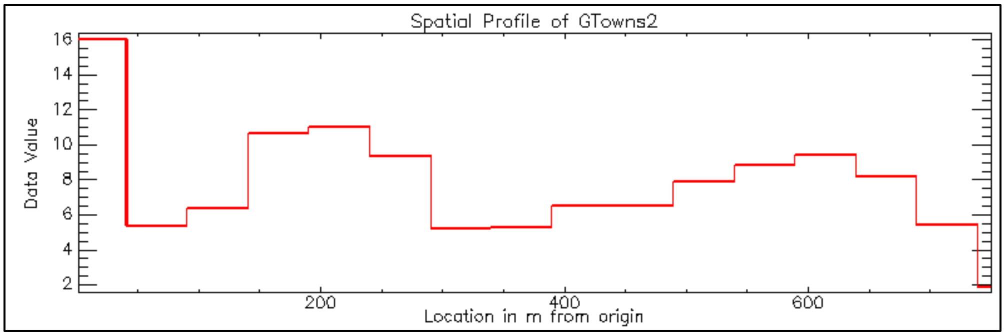



**Figure 32 Example of Degree Slope spatial profile in formal settlement**

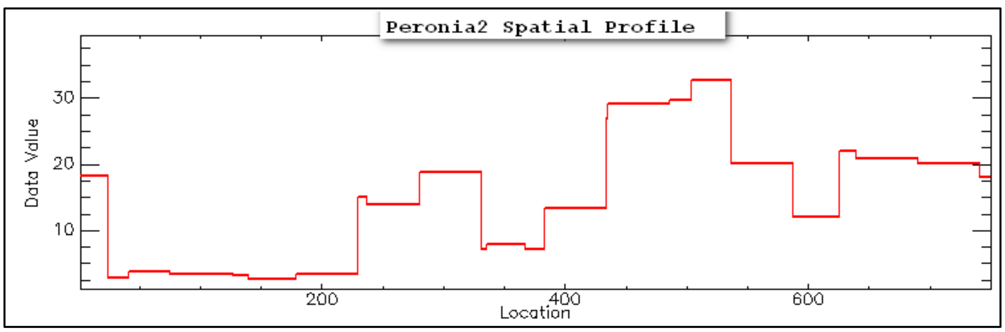



**Figure 33 Example of Degree Slope spatial profile in informal settlement**

The slope varies from nearly 0 degrees to only 16 degrees in the formal settlement, and from 2 to approximately 35 degrees in the informal settlement. The mean degree slope calculation was intended to capture this type of variability.

Slope on Roads underlying the road vectors was computed from the mean of slope pixel values in degrees. The hypothesis was that the roads would exhibit greater slope in the informal areas, intuitively due to less planning and resources invested in the process of leveling a road surface for paving. Greater accuracy in this metric could have resulted if LIDAR data was available with better resolution than 30m, instead of relying upon the ASTER DEM product.

Plan Convexity was calculated from the topographic modeling tools in ENVI™. The plan convexity (intersecting with the XY plane) measures the rate of change of the aspect along the plane. As a measure of surface curvature, plan convexity is orthogonal to profile convexity, with positive values representing convexity and negative values concavity. Plan convexity is in the direction of minimum gravity effects (Exelis, 2009).

Profile Convexity was calculated from the same topographic modeling tool. The profile convexity (intersecting with the plane of the z axis and aspect direction) measures the rate of change of the slope along the profile. Profile convexity is measured in the direction of maximum gravity effects (Exelis, 2009), and similar to the Plan Convexity metric, negative values are concave and positive values are convex.

## Classification Accuracy

The following table summarizes the classification accuracy of the NDVI, spectral angle mapper (SAM), and pseudo-builtup methods from a sample of 84 random points. Ground truth was measured through visual interpretation of the orthorectified true color and false color composites. Sample points were selected in order to exceed the binomial probability theory requirements according to the equation:

$$N = \frac{Z^2(p)(q)}{e^2}$$

where $p$ is the expected percent accuracy of the entire classification, $q=100-p$, $e$ is the allowable error and $Z=2$ (+/− 2 standard deviations capturing all values with 95% confidence) (Jensen, 2005, p. 501). With an expected mapping classification accuracy of

85% and allowing for an error of 8%, *N=79*. Given that 79 points are the minimum required to measure accuracy to the specified standard, seven points were selected randomly from each of the twelve settlements, resulting in 84 sample points for the accuracy assessment.

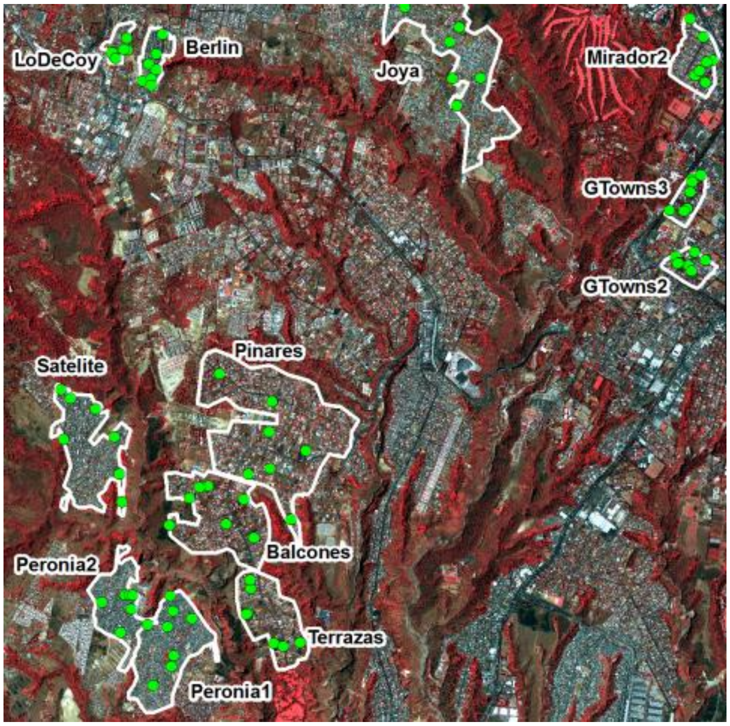



**Figure 34 Sample Point Locations for Classification Accuracy Assessment**

In Figure 34 the green dots represent the randomly placed sample point locations used to assess accuracy. Table 18 lists the accuracy results for the sample points.

**Table 18 Accuracy Results from Vegetation, Asphalt, Soil and Pseudo Built-up classes**

| Error Matrix | | | | |
|---|---|---|---|---|
| | VEG (NDVI) | ASPHALT (SAM) | SOIL (SAM) | BUILTUP (PSEUDO)[6] |
| Correct | 18 | 19 | 10 | 54 |
| Actual (Ground Truth) | 20 | 23 | 12 | 58 |
| Commission (Users) | 2 | 6 | 3 | 2 |
| Omission (Producers) | 3 | 4 | 2 | 2 |
| Overall % Accuracy | 0.90 | 0.83 | 0.83 | 0.93 |
| Omission % Error | 0.15 | 0.17 | 0.17 | 0.03 |
| Commission % Error | 0.10 | 0.26 | 0.25 | 0.03 |

Overall accuracy for NDVI was 90%, for asphalt 83%, for soil 83% and for pseudo built-up 93%. These classifications were performed using different methods appropriate for the feature or material being extracted. Settlement feature comparison and the numerous methods needed to develop metrics using both remote sensing and GIS data precluded estimation of Kappa which is normally performed when an image is completely segmented into all known classes in order to account for chance error. In this research, not all classes were mutually exclusive. In other words, Asphalt and Dirt Roads could both be considered components of the Pseudo Built-up class. Therefore, the reported accuracy uses a modified approach.

[6] Pseudo Built-up can include asphalt, so not all categories are mutually exclusive because different classification algorithms were used based on the settlement attribute to be measured. For example, soil and asphalt can be considered a component of a built-up area.

The next section explains the principles behind the modeling approaches selected to perform this multivariate assessment of each variable's contribution to distinguishing informal and formal settlements.

## Settlement Indicator Modeling Methods

This section describes the modeling methods used to differentiate settlement type using sample data. Three primary methods were applied. First, the t-test of independent means determines statistical significance of group membership between informal and formal communities at the settlement level. The t-test evaluates the hypothesis that two samples have the same means. The reported probability is the likelihood that the two samples tested are sufficiently similar to be from the same population. Therefore, a very low (p) value is preferred if a measured variable is hypothesized to produce significantly different results between the two samples tested. The unpaired *t-statistic* measures the ratio of the difference between the two means and the standard error of the difference of means from the 'pooled variance' which is the bottom part of the equation:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\left[\frac{Var(X_A)}{df'_A} + \frac{Var(X_B}{df'_B}\right]^{\frac{1}{2}}}$$

where $t$ represents student's t-statistic, $\bar{X}_A$ represents the mean of group $A$ (e.g., informal), $df'_A$ is the degrees of freedom of the informal group, $df'_B$ is the degrees of freedom of group $B$ (the formal group). With unequal variances, $s$ of both samples is used to compute *df′* where $n_A$ is the number of samples in group *A* and $n_B$ is the number of samples in group *B* as follows:

$$df' = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{\left(\frac{s_A^2}{n_A}\right)^2}{n_A - 1} + \frac{\left(\frac{s_B^2}{n_B}\right)^2}{n_B - 1}}$$

This method of computing *df′* creates non-integer degrees of freedom for some variables. Despite the small sample size, the value for a variable may be the mean (μ) of up to several thousand points depending on whether the variable being measured is road accessibility-related, actual pixel values of surface materials, or vegetation polygons, for example.

The second method was Discriminant Function Analysis (DFA), also known as Discriminant Analysis (DA). DA creates a linear function that differentiates samples in informal or formal communities through the appropriate mix of variables and loadings, similar to factor analysis. The variables selected by DA are those most likely to predict group membership to determine whether a sample is from an informal/slum area, or a formal/planned community. DA is most appropriate for the following:

- Classifying cases into binary groups via a discriminant prediction equation
- To determine the most parsimonious way to distinguish among groups
- To determine the percent of variance in the dependent variable explained by the independents
- To discard variables which are little related to group distinctions (Garson, 2012)

Discriminant Function Analysis is a form of multivariate analysis of variance (MANOVA) that expects a categorical dependent variable instead of continuous value.

The discriminant function is tested for significance in discriminating between the two parent populations of settlement type. The result uncovers the most powerful linear combination of predictor variables that can be constructed to maximally separate the group means (Griffith & Amrhein, 1997). A discriminant function *D* is created as a linear combination of discriminating (independent) variables such that:

$$D = b_1x_1 + b_2x_2 + ... + b_nx_n + c,$$

where the *b*'s are discriminant coefficients, the *x*'s are discriminating variables (i.e., the indicators), and *c* is a constant. The coefficients maximize the distance between the means of the criterion variable (settlement type), and the criterion variable is categorical rather than interval/ratio (Griffith & Amrhein, 1997). The following diagrams visualize this separability.

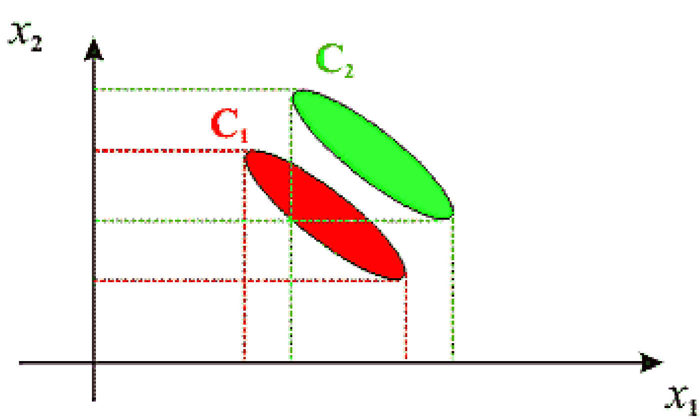



**Figure 35 Two-classes projected in 2D space (http://www.aiaccess.net)**

Figure 35 shows the overlap in response when the bounds of Class $C_2$ and $C_1$ are both projected onto the $X_2$, or $X_1$ axis.  In Figure 36, the discriminant function *D* represents a linear function that maximizes the separability between the two classes.

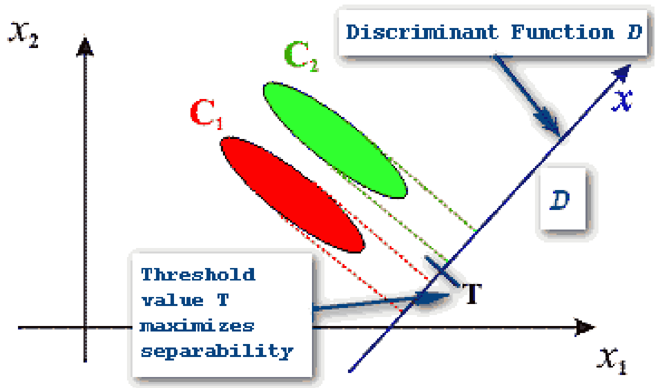



**Figure 36  Linear Discriminant Function that maximizes class separability (adapted from http://www.aiaccess.net)**

There are some notable assumptions in the DA model, including a Gaussian distribution of variables, equality of covariance matrices, error independence, and non-zero variance within samples (Griffith & Amrhein, 1997). DA also assumes similar numbers of samples in the dependent groups.  Another limitation is that DA excludes the entire sample even if only one variable for that sample had a missing.  An example is in the connectivity of nodes in the road network, the variable Connected Node Ratio (CNR). Some samples had *0* nodes, resulting in a null value for CNR for those samples. This resulted in 18/127 samples being excluded entirely from the analysis as a result of several variables having missing values.

Some of the assumptions in DA are not always strictly enforceable with real data, so a non-parametric method called Decision Trees was tested as well. Decision Trees, also known as Classification and Regression Trees (CART), do not make any assumptions regarding the statistical distribution of the dependent or independent variables, handling Gaussian and multimodal data equally as well (Sugumaran, Pavuluri, & Zerr, 2003). Also, the predictor variables can be a mixture of categorical, interval-ratio, or continuous datatypes. CART methods are not affected by outliers or collinearities, heteroscedasticity, or distributional errors affecting parametric methods like DA (Yohannes & Hoddinott, 1999). CART models create a series of decision branches that include certain values for variables that provide a cut-off to assign a case, or sample, to a given class (informal or formal). The successive branches of the tree achieve a series of exhaustive and exclusive partitions among the set of samples the decision maker wants to classify (Diaz Martinez, Fernandez Menendez, & Vargas, 2004). The main concept behind CART algorithms is to choose the variable that provides sufficient information to realize the appropriate partition each time a new branch is created, and ultimately to classify the training set with the variables contributing most to the classification.

The main benefits of the classification and decision tree approach for the researcher are ease of interpretation and the ability to produce a “rule set” clearly listing the variable value combinations and numerical differences making up the decision rule at each branch. In the present research, the rule will explain whether a sample should be classified as informal or formal based on the rule applied.

A limitation of CART is that it is not based on a probabilistic model that could, for example, predict confidence intervals for the classification results (Yohannes & Hoddinott, 1999). Another cited limitation is that CART performs local and not global feature selection, so that each node split is a local decision based on the values of the remaining samples not used for prior splits (Friedl & Brodley, 1997). If a globally optimal solution for all variables is preferred, the CART method may theoretically not solve the researcher's problem.

# RESULTS

This section discusses the results of the research that are intended to (1) define the set of statistically significant indicators that distinguish settlement type (informal from formal), (2) determine if imagery and elevation data alone are sufficient to distinguish informal from formal settlements, and (3) to perform a multivariate evaluation of indicators from roads, vegetation, soil, image texture, and geomorphology. The results are first reported by settlement, where each of the 12 settlements is treated as a sampling unit. This was done in order to winnow an acceptable subset from the 23 indicators to include ones that do not co-vary. After discussing the results at the settlement level, the results at the grid cell (sub-sampling unit) are reported.

## Results from Sampling by Settlement.

The results in this section explain indicator performance at the settlement level before random sub-sampling was conducted. Sampling by settlement treats each settlement as a single observation, meaning that within-settlement variation is not recognized. Recall that 6 informal and 6 formal settlements were delineated in the image scene from expert knowledge and according to certain principles that focus the measures on residential areas. The first phase of the research explored whether the proposed indicators produced different results between the two types of settlements - informal or formal - using each settlement as the sample unit.  If an indicator did not show significant difference when a t-test of independent means was applied, it could theoretically be excluded from the model.  Significance was considered at the 90% confidence level. According to Bartlett et al. (2001, p. 45), an $\alpha$ of 0.1 is acceptable if the researcher simply wishes to identify differences or as a precursor to other statistical testing. These conditions correctly describe the current requirement.

The second phase of the research involved proportional stratified random sub-sampling with $150m^2$ grid cells overlaid on the image scene, then selecting 50% of settlement-intersecting grid cells randomly from each settlement to achieve a sample size large enough to test statistical significance of the model (Beyer, 2004).  Of the 50% selected, only grid cells with $\geq$ 25% overlapping coverage of the settlement boundary were included in the sample.

### Scale-related Results by Settlement

Lacunarity and fractal dimension were tested as scale-related measures from prior research applied to informal settlements to evaluate structure and pattern.  They were also measured to evaluate performance between both formal and informal settlement types, since the results from the literature review were somewhat contradictory.   The current research shows the built-up areas comprised of man-made materials yield differences that appear at certain spatial scales, yet are more random at others.

### *Lacunarity and Fractal Dimension*

Lacunarity (*Λ)* is considered scale-dependent because the result depends on the box size used. Earlier, it was mentioned that higher heterogeneity is expected to produce higher *Λ* values. Figure 37 is an example of how a single sub-scene of a formal settlement was transformed from its false-color composite to an NDVI image from a density slice of all values < 0.09 and then inverting the Black/White result to a pseudo built-up (foreground) binary image to compute lacunarity. The term 'pseudo built-up' is used to avoid confusion with more formalized methods that measure built-up areas, such as the normalized difference built-up index (NDBI) also found in the literature, which requires the mid-IR band such as that available from the Landsat Thematic Mapper sensor. However, Landsat-based indices have a resolution of 28.5m$^2$ pixel sizes, which is too large to meaningfully identify smaller settlement areas. Figure 37 displays an example of the image transformation required to generate the pseudo built-up binary image used for Lacunarity and Fractal measurement.

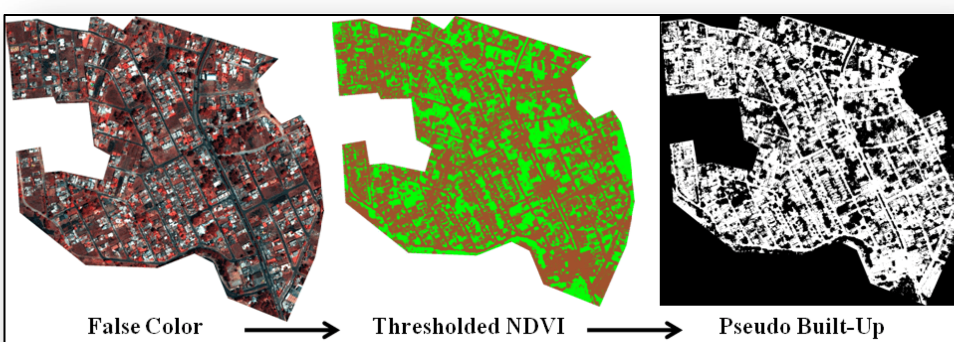



**Figure 37 Image Transformation for lacunarity and fractal analysis**

It was also reported earlier that Fractal Dimension for a given box size ($D_B$ ) has been studied in several contexts in the literature, from measuring street edges in relation to the building outlines and sizes, to measuring housing patches in sprawl areas, to measuring entire cities, with mixed results. No effort specifically focused on fractal dimension of informal settlements, prompting the test as a measure for this research. In Figure 38, $D_B$ and $\Lambda$, respectively, are listed for the pseudo built-up settlements in the study area. The box size of 200 was selected because it had the lowest (p) value, indicating better

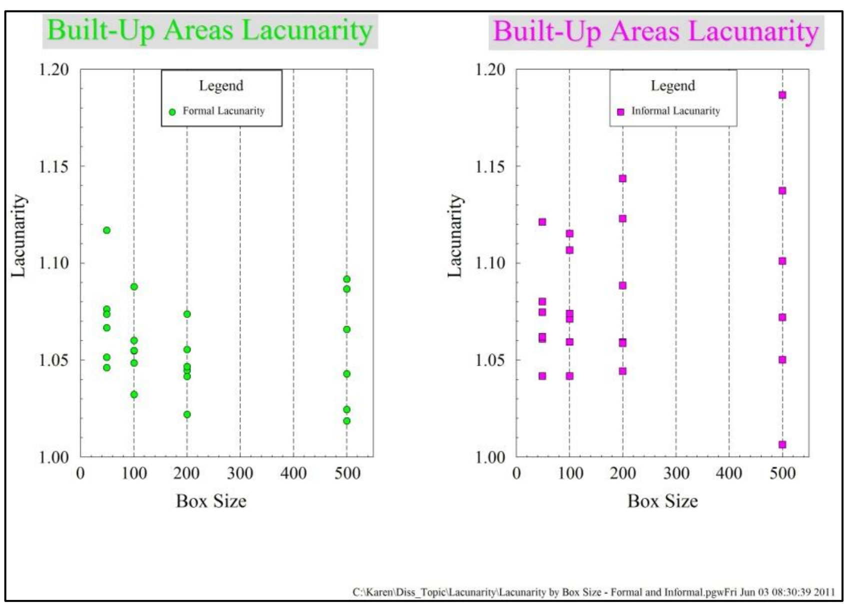



**Figure 38 Lacunarity of Formal and Informal settlements by Box Size**

performance than other box sizes chosen (50, 100, and 500). The Lacunarity results for various box sizes in the formal and informal settlements are reported in Appendix C.

Figure 38 shows Lacunarity (*Λ)* for the 6 formal and 6 informal settlements by different scales or box sizes. The variance for the box size of 500 is wider, and the box sizes of 50 and 100 did not exhibit useful differentiation between settlement types.

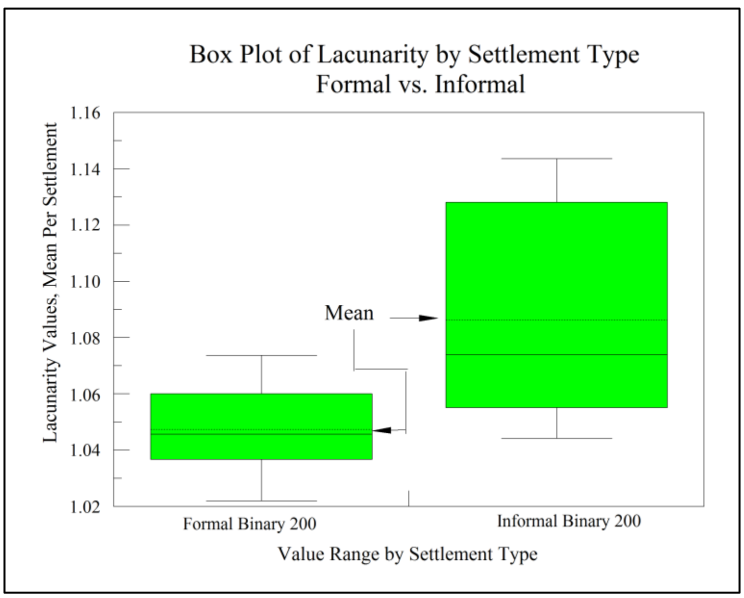



**Figure 39 Box Plot of Lacunarity by Settlement Type (formal vs. informal)**

The box size of 200 shown in the box plot in Figure 39 has a smaller range of lacunarity among formal settlements (less variation), and a slightly larger range for the informal values. If the box size is selected specifically to capture feature variation for features of a given size, then a box size of 200 should capture dwellings that are up to 8.4m x 8.4m, or up to approximately 27ft per side, whereas a much larger box size could capture several dwellings together. A larger box size would therefore not capture the heterogeneity of settlements with predominantly smaller sized shacks and houses and the spaces separating them. The reduced significance exhibited by the experiments with a box size of 100 (p=0.149) and a box size of 500 (p=0.24) demonstrated this fact. Table 20 reports the consolidated results for both Lacunarity and Fractal Dimension with settlements as sampling units.

**Table 19 Fractal Dimension and Lacunarity, box size 200, by settlement**

| | **Foreground $\Lambda=1+cv^2$ Box Size 200, Binary** | **Count of Boxes with FG Pixels** | **Fractal $(D_B)$, Binary** |
|---|---|---|---|
| ***Informal*** | | | |
| Satelite | 1.088 | 3187 | 1.877 |
| Berlin | 1.059 | 536 | 1.827 |
| Joya | 1.123 | 4792 | 1.806 |
| LoDeCoy | 1.059 | 295 | 1.665 |
| Peronia1 | 1.044 | 2933 | 1.907 |
| Peronia2 | 1.144 | 2157 | 1.833 |
| ***Formal*** | | | |
| Terrazas | 1.045 | 2285 | 1.913 |
| Balcones | 1.022 | 3042 | 1.917 |
| Mirador2 | 1.056 | 1178 | 1.94 |
| GTowns2 | 1.047 | 592 | 1.89 |
| GTowns3 | 1.042 | 510 | 1.806 |
| Pinares | 1.074 | 7856 | 1.908 |

.

At the settlement level, Lacunarity and Fractal Dimension were somewhat negatively correlated with a Pearson's product moment correlation coefficient of −0.253.

With n=12 and p=0.427, this correlation was not significant. Lacunarity was generally higher in the informally settled areas, and based on the total number of sliding boxes for which $\Lambda$ was computed, there was a statistically significant difference between the informal and formal settlement types. There was only an 87% chance the parent means (using the unpaired t-test) from the grayscale built-up images were from different underlying populations, while the binary image performed significantly better with *p(0.052)*, leading to the conclusion that lacunarity on binary image built-up areas may differentiate these types of settlements in Guatemala. Table 21 shows the results of a t-test of unpaired means with the results of the box size 200 for both fractal dimension and lacunarity outlined in black.

**Table 20 Correlation of Scale-related measures, settlement level**

| Pearson Correlation Coefficient | Fractal Dimension |
|---|---|
| Lacunarity | −0.253 |

**Table 21 Significance of scale-related measures by settlement**

| Image Type | (*p*) value | *t*-score |
|---|---|---|
| $\Lambda(r_{50})$ , Binary | 0.92 | 0.1076 |
| $\Lambda(r_{100})$ , Binary | 0.14 | 1.5821 |
| $\Lambda(r_{200})$ , Binary | 0.05 | −2.2 |
| $\Lambda(r_{500})$ , Binary | 0.23 | 1.278 |
| $\Lambda(r_{100})$ , Grayscale | 0.15 | 1.57 |
| $\Lambda(r_{200})$ , Grayscale, | 0.85 | 0.194 |
| Fractal D – binary built-up, box size 200 | 0.079 | 1.9 |
| Fractal D – grayscale built-up, box size 200 | 0.58 | .59 |

The results are also consistent with the Filho & Sobreira (2007) findings of higher lacunarity in the informal versus formal areas. Mean $\Lambda(r_{200})$ is 1.047 for the formal and 1.086 for the informal settlements.

Fractal dimension is calculated using the formula: $D_B = -\lim[\log N\varepsilon/\log\varepsilon]$, the negative limit of the ratio of the log of the number of boxes at a certain scale over the log of that scale (Karperien, FracLac for ImageJ, V2.5). Fractal dimension of the binary image using the sliding box method and box size (scale) of 200 produced the only promising result from the unpaired t-test ($p=0.079$) that fractal dimension of built-up areas in the informal vs. the informal communities yield different results. No significant difference was detected on the grayscale image ($p = 0.58$). Table 21 summarizes the results in tabular format.

As an interesting exercise and because of the simplicity of creating a binary representation of all asphalt roads following classification, $D_B$ was also computed on asphalt roads in the image subsets, yielding no statistically significant difference. This is likely because the same underlying structural process contributed to their creation regardless of whether the settlement was rich or poor, and the same government infrastructure resources would have been used.

Fractal dimension and lacunarity are not intuitive to measure and interpret but were used to determine if the spatial scale of analysis impacts the results. In this research, the $\Lambda(r_{200})$ and $D_{B(200)}$ represent a box of 8.4m$^2$ and provide the best discriminatory power. This is likely due to the range of smaller buildings and built-up structures that can still be detected surrounding the vegetation and the gaps. In the informal areas, lacunarity increases with box size, but is most significantly different than formal and exhibits the smallest σ and range for $\Lambda(r_{200})$. The results for lacunarity at the settlement level were the expected results, but fractal dimension results were inconsistent with Cooper (2005) where higher fractality implied higher urban fragmentation.

### Topographic-related results by settlement

Underlying topography is a fixed aspect of the landscape and is related to geomorphology. The terrain that dwellings are built upon impacts residents' living conditions related to weather, climate, and natural (or man-made) hazards. The results of the four topographic-related measures are reported here: slope on roads, degree slope, plan convexity (curvature), and profile convexity (curvature). Neither plan curvature nor profile curvature are correlated with degree slope, but degree slope is highly correlated with slope on roads.

**Table 22 Correlation of topographic measures, settlement level**

| Pearson Correlation Coefficient | Degree Slope on Roads | Degree Slope | Profile Convexity |
|---|---|---|---|
| Degree Slope | 0.988 | | |
| Profile Convexity | −0.119 | −0.127 | |
| Plan Convexity | −0.362 | −0.377 | 0.116 |

**Table 23 Significance of topographic measures by settlement**

| Indicator | p-value | t-score | Degrees of Freedom |
|---|---|---|---|
| Slope (degrees) | 0.03 | 2.52 | 9.9 |
| Profile Convexity | 0.044 | −2.3 | 9.93 |
| Plan Convexity | 0.17 | −1.47 | 9.44 |
| Degree Slope (on roads) | 0.065 | −2.07 | 9.77 |

### *Slope in Degrees*

Guatemala is a very mountainous country, so the slope indicator was selected because higher slopes in more difficult to build locales were expected in the informal areas. The mean degree slope was calculated for the 12 settlements using a 3x3 kernel from the 30m ASTER data over the entire region. Figure 40, and Table 23 show mean degree slope for each settlement type.

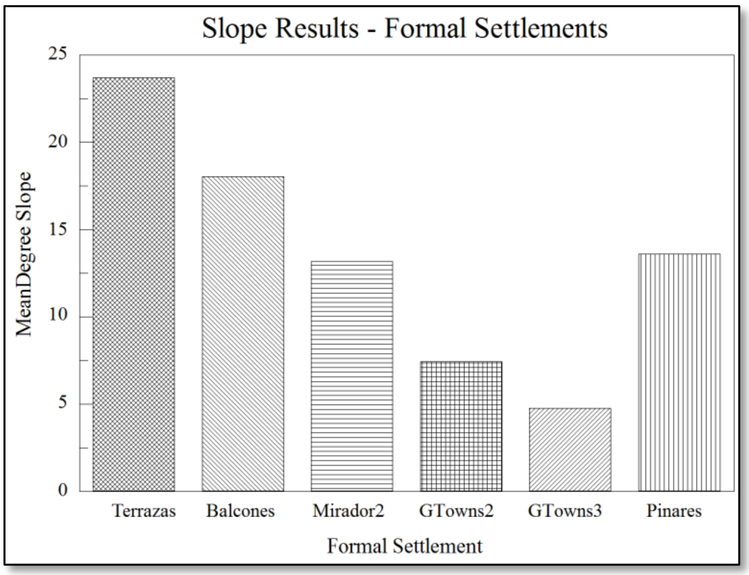



**Figure 40 Degree Slope - Formal Settlements**

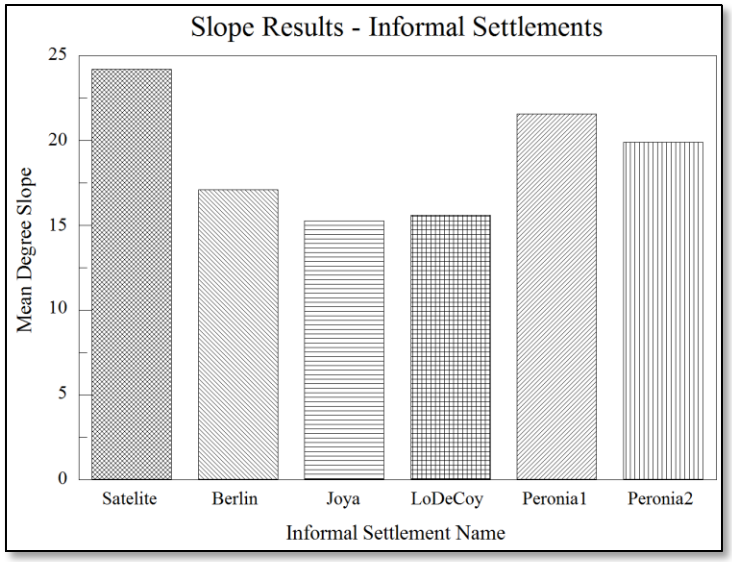



**Figure 41 Degree Slope - Informal Settlements**

The mean degree slope is higher in the informal settlements and lower in the formal settlements, except for the outlier Terrazas which means "Terraces" (in Spanish) - a formal settlement built for its views. The p-value (0.03) reported in Table 29 from the t-test reveals that slope is a good discriminator of settlement type.

### *Profile and Plan Convexity*

The results of the t-test of means for plan convexity are displayed in Table 23 displays plan convexity overlaid with settlement boundaries. Formal settlements are magenta and informal settlements are in red. Highly negative numbers (more concave) are black, and less negative numbers are white.

Plan convexity, which represents the rate of change of the angle of aspect, measures the curvature of contours. A contrasting measure is profile convexity

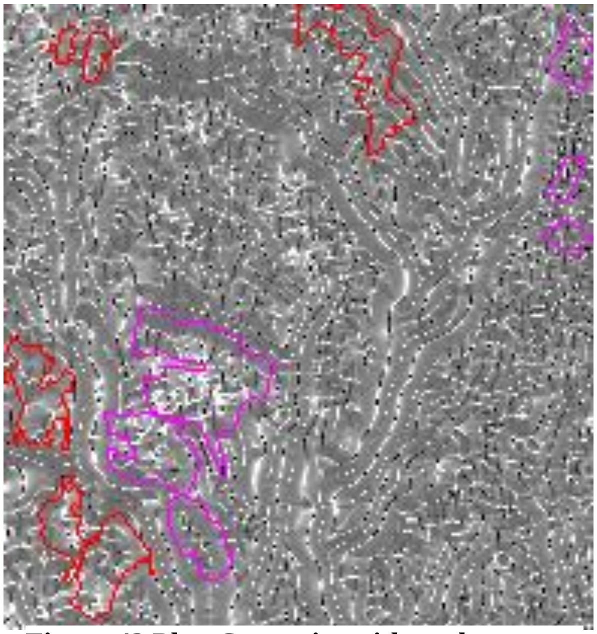


**Figure 42 Plan Convexity with settlement boundaries superimposed**

(sometimes referred to as profile curvature) which calculates the downhill or uphill rate of change in slope in the gradient direction. Negative values are concave upward, indicating an accelerated flow of water over the surface, while positive values are convex upward indicating reduced flow over the surface. One should expect highly negative profile convexity values could cause water to pool. Plan convexity should approach zero on very steep slopes and have extremely low values in floodplains (Evans & Cox, 1999). In the entire image, plan convexity ranged from -1093 (divergent flow over the surface) to 10,308 (highly convex) over all settlement cells. In the horizontal plane where effects of gravity are minimized (Wood, 1996) the informal areas were slightly less convex (μ 350 vs. μ 442 for formal). The differences are highly statistically significant when all pixels in the plan convexity image are evaluated as sampling units (e.g., plan convexity for all informal settlement cells are then compared to plan convexity for all formal settlement cells). However, by comparing the results for plan convexity at the settlement level using the mean of all pixels within the settlement boundary as the sampling unit, the t-test shows an insignificant difference (t = −1.476, $p = 0.17$) between the formal and informal settlements.

In profile convexity (the vertical plane), gravity effects on flow are maximized (Wood, 1996). The mean profile convexity is lower, at −0.704 for all informal cells and −0.545 for all formal cells. Despite steeper slopes, there is geomorphically more concavity in the vertical direction in the informal terrain. The differences are highly statistically significant when all pixels from the profile convexity image are treated as samples for a given settlement type. When profile convexity is averaged for each discrete settlement, the t-test results are still statistically significant, t= −2.3, $(p) = 0.044$. The recommendation is to use profile convexity and not plan convexity for subsequent testing.

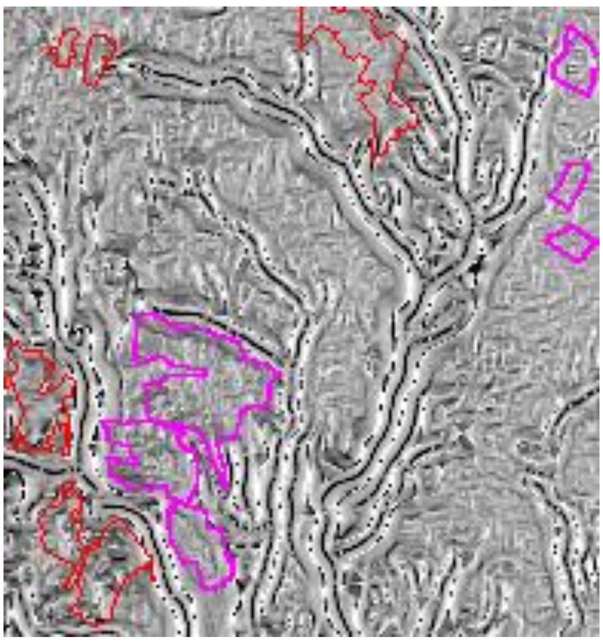


**Figure 43 Profile Convexity with settlement boundaries superimposed.**

### Spectral-related results by settlement

Spectral information from imagery is especially useful to understand and classify surface materials. The spectral properties of soil, vegetation, and asphalt from the imagery were used to evaluate vegetation percent, vegetation patch size, compactness ratio, soil percent, and surface materials on roads (soil or asphalt) with results shown in Table 25.

**Table 24 Correlation of Spectral-related measures, settlement level**

| Pearson Correlation Coefficient | Compactness Ratio | Soil Percent | Vegetation Percent | Asphalt Road Content | Dirt Road Percent |
|---|---|---|---|---|---|
| Vegetation Patch Size (area) | −0.501 | −0.119 | 0.796 | 0.242 | −0.107 |
| Compactness Ratio | | 0.074 | −0.621 | −0.339 | 0.17 |
| Soil Percent | | | −0.101 | −0.268 | |

**Table 25 Significance of spectral-related measures, settlement level**

| Indicator | p-value | t-score | Degrees of Freedom |
|---|---|---|---|
| Vegetation Percent | 0.0006 | −4.9 | 10 |
| Mean Vegetation Patch Size | 0.08 | −1.94 | 7.84 |
| Mean Vegetation Patch Compactness Ratio | 0.000073 | −7.55 | 7.84 |
| Soil Percent | 0.32 | 1.067 | 9.44 |
| Asphalt on Roads | 0.037 | −2.39 | 9.81 |
| Dirt on Roads | 0.35 | 0.97 | 9.68 |

It was hypothesized there would be more vegetation in the formal areas, given that presence of vegetation has been associated with higher quality of life in prior research. The percent of vegetation is significantly lower in the informal areas and this difference is highly statistically significant ($p = 0.0006$). Mean percent vegetation in the informal settlements was 6.5% compared to 13.9% in the formal settlements. A related measure is Mean Vegetation Patch Size. Vegetation patches were hypothesized to be smaller in the informal settlements. Mean vegetation patch size was slightly statistically significant ($p = 0.08$), with an overall mean patch size of $8.2m^2$ in the informal settlements, which was smaller than the mean of $10.4m^2$ in the formal settlements.

Using the Vegetation Patch Compactness Ratio measure, vegetation patches were hypothesized to be more circular in the informal areas representing a lack of purpose for vegetation, compared to informal areas where vegetation could be used to line streets and land parcels or as ornamental borders. The mean compactness ratio for informal settlements was 0.74 and 0.71 for the formal settlements, yielding a highly statistically significant difference ($p=0.000073$) where vegetation is more elongated and less circular in the formal areas.

The difference in soil percent at the settlement level between formal and informal settlements was not statistically significant ($p = 0.31$). However, it will become clear later that the internal variability hidden by summarizing the entire settlement by using the percent soil coverage actually obscures important differences that appear in the smaller formal vs. informal samples.

Mean asphalt percent coverage of roads in the formal areas was higher:  42% vs. 18% in the informal areas. At *p=0.03*, the asphalt road coverage exhibits a statistically significant difference between formal and informal settlements.  For Dirt Road coverage, however, the hypothesis that significantly more dirt-covered roads would be found in informal settlements was not proven. The results shown in Table 25 reveal that dirt road content was not a significant discriminator.

### Road network-related results by settlement

Roads are a direct representation of the movement of people across a landscape, and their composition, shape, and topology are necessary to understand how settled residential areas are interconnected.  The roads measures that do not include surface materials in their definition are instead used to evaluate accessibility.  Informal settlements are hypothesized to be less accessible, with more dead ends, simple intersections, and less interconnectedness.  Before reviewing the results of the road accessibility measures, a description of measure correlation is included.  Connected Node Ratio and Mean Node Valence are highly positively correlated (*Pearson r= 0.7687*). Connected Node Ratio and Percent Dangles are negatively correlated (*Pearson r= −0.6698*). Mean Node Valence is slightly positively correlated with Pct 4Way Intersections (*Pearson r=0.4493*) and slightly negatively correlated with Percent Dangles (*Pearson r= −0.4072*).

**Table 26 Correlation of road network-related measures, settlement level**

| Pearson Correlation Coefficient | Connected Node Ratio | Mean Node Valence | Pct 4Way Intersections | Pct Dangles | Road Density |
|---|---|---|---|---|---|
| Mean Node Valence | 0.7687 | | | | |
| Pct 4Way Intersections | 0.1721 | 0.4493 | | | |
| Pct Dangles | −0.6698 | −0.4072 | -0.1324 | | |
| Road Density | 0.1405 | 0.1000 | 0.0744 | 0.0041 | |
| Unpaved-to-paved Ratio | −0.0503 | -0.1065 | −0.0877 | 0.1015 | −0.1523 |

**Table 27 Significance of Road Connectivity measures by settlement**

| Indicator | p-value | t-score | Degrees of Freedom |
|---|---|---|---|
| Connected Node Ratio | 0.0082 | 3.28 | 8 |
| Ratio of 4-way Intersections | 0.0004 | 5.29 | 8 |
| Dangle Ratio | 0.0052 | 3.56 | 8 |
| Road Density per Area | 0.016 | 2.87 | 9.93 |
| Unpaved-to-paved Road Ratio | 0.337 | 1.05 | 5.09 |
| Mean Node Valence | 0.166 | 1.49 | 9.96 |

Connected Node Ratio is [*(intersections–dangles)/intersections*], with intersections defined as ≥ 3 connected nodes (Song & Knapp, 2004; Dill, 2004). Connected Node Ratio was statistically significant. This is likely due to the fact that this ratio does not measure the degree of connectedness of all intersections, only their rate. The ratio of 4-Way intersections to all intersections, representing a magnitude of connectedness, was statistically significant. There were more 4-way intersections in the formal areas, indicating higher connectivity of the street network. The dangle ratio is a ratio of dead-ends or cul-de-sacs to all nodes and is negatively correlated with the connected node ratio (*Pearson* $r = -0.6698$) and positively correlated with mean node valence (*Pearson r = 0.7687*) because to a certain extent, all of these measures rely on intersection valence.

The last road accessibility measure was Mean Node Valence. This measures the mean intersection size among all 2-way, 3-way and 4-way intersections (there was only one 5-way intersection in the dataset). There was only slight statistical significance for this measure at a confidence of only 84% taking the mean of all node valence values for all intersections within a settlement. If, however, all node valence values are combined together into their representative settlement type (e.g., all formal valence values and all informal valence values) the result is then highly statistically significant: ($p = 1.5*e^{-8}$, $t = 5.73$, $df = 606$). In this case, the mean node valence for all formal settlements is 1.47 but the informal settlements were less-connected with a mean valence of only 1.13.

### Texture-related results by settlement

Higher texture measures of entropy, contrast and lower GLCM correlation[7] (Haralick et al., 1973) were expected in informal settlements compared to planned communities. GLCM Correlation is a measure of image linearity or linear dependence of the intensity values of pairs of pixels in the panchromatic band. The texture measures of Contrast (a measure of local variations in an image) and entropy (a measure of the randomness of intensity distribution of pairs of grey-level pixel values) were also produced for road surfaces at the settlement level. The expected higher values for

[7] The texture measure of 'GLCM correlation' is different than the Pearson Correlation Coefficient, which is a standard statistic of correlation between measured variables.

Contrast and Entropy in the informal areas are due to the wider variety of building products and often discarded materials used to patch and extend the roofs of structures. When constrained to the road surface, these expected higher values are due to the more variable condition of road surfaces, with debris and discarded materials found more often on roads in informal areas. Prior efforts stated that *correlation*, *entropy*, *mean*, and *variance* are not well-correlated with each other (Warnick et al., 2008). However with settlements as samples, Entropy and GLCM Correlation were highly negatively correlated, and GLCM Mean and Entropy were positively correlated. The hypothesis that entropy is a useful measure of heterogeneity and would be higher in the informally settled areas was also tested (Gong & Xu, 2006). The following table lists the correlation coefficients for settlement-level values of each of the texture measures attempted.

**Table 28 Correlation of texture-related measures by settlement**

| Pearson Correlation Coefficient | Contrast | GLCM Mean | GLCM Correlation | Entropy | GLCM Variance | Dissim-ilarity | Contrast Roads |
|---|---|---|---|---|---|---|---|
| GLCM MEAN | 0.411 | | | | | | |
| GLCM Correlation | −0.502 | −0.883 | | | | | |
| Entropy | 0.916 | 0.547 | −0.701 | | | | |
| GLCM Variance | 0.994 | 0.359 | −0.444 | 0.882 | | | |
| Dissimilarity | 0.926 | 0.261 | −0.421 | 0.779 | 0.924 | | |
| Contrast Roads | 0.608 | 0.674 | −0.555 | 0.722 | 0.558 | 0.347 | |
| Entropy Roads | 0.551 | 0.775 | −0.612 | 0.683 | 0.510 | 0.256 | 0.949 |

Many of the texture measures are highly correlated. The texture measures of GLCM Correlation, Entropy, Contrast, GLCM Mean, Variance, and Dissimilarity were calculated by settlement. GLCM Contrast and Entropy constrained to the buffered road network were also calculated.  GLCM Contrast and Entropy on Roads were highly positively correlated.  The mean texture measure of GLCM Correlation was lower for all informal settlements merged together −40.3 vs.  −35.47 for all formal settlements merged.

**Table 29 Significance of Texture related measures by settlement**

| Indicator | p-value | t-score | Degrees of Freedom |
|---|---|---|---|
| **GLCM Correlation** | 0.016 | 3.008 | 8.19 |
| **Entropy** | 0.039 | 2.59 | 6.1 |
| **Entropy on Roads** | $4 * e^{-5}$ | 7.35 | 9.1 |
| **Contrast** | 0.119 | 1.72 | 8.8 |
| **Contrast on Roads** | $5.3 * e^{-6}$ | 8.75 | 10 |
| **GLCM Mean** | 0.01 | 3.44 | 7 |
| **GLCM Variance** | 0.17 | 1.46 | 9.6 |
| **Dissimilarity** | 0.72 | 0.35 | 8.2 |

In terms of GLCM Correlation, the difference between the two settlement types is significant.  Based on these results, GLCM Correlation, Entropy, Entropy on Roads and Contrast on Roads, and GLCM Mean appear to be good measures to evaluate differences in settlement structure.

## Variable Correlations and Exclusions

In multivariate statistical modeling, prior correlation between included variables contributes to model misspecification by including extraneous variables.  This section and Appendix B describe how variables were evaluated and which ones were eliminated from the model due to the multicollinearity condition. The graphics in Appendix B provide details of bivariate correlations that existed among several of the 23 originally measured indicators.

Pearson's *r* was calculated for all pairs of variables on all 127 grid cells to determine whether one or more correlated variables should be excluded to avoid the multicollinearity condition prior to Discriminant Function and CART modeling. Spearman's $\rho$ was also computed on grouped data to determine whether correlation between the variables differed for informal settlements and formal settlements. The scatterplots in Appendix B show the fitting of a regression line for the grouped data in order to visualize the direction of the relationship, and where further explanation was needed, histograms provided additional visualization of these distributions.

Table 30 lists the variables that were excluded, followed by a narrative summary.

**Table 30 Summary and Explanation for Excluded Variables**

| Variable Excluded | Result | Variable Correlation | Explanation |
|---|---|---|---|
| *Fractal Dimension* | Lacunarity was a marginally better settlement type discriminator | Slightly Negative with Lacunarity | Lacunarity is a measure of gappiness or translational or rotational invariance in an image scene, while Fractal D is a measure of complexity showing how the pseudo-builtup areas fill the space at a given scale (or box size). Overall mean Fractal D was higher for informal settlements – an indication of higher complexity. |
| *Mean Node Valence* | Not significant at the settlement level | Positive with Connected Node Ratio | Higher node valence is an indication of a more connected street network. |
| | Not significant at the settlement level | Negative with Dangle Ratio | Higher node valence is inversely proportional to the dangle ratio in a road network |
| | | | |

| | | | |
|---|---|---|---|
| *Percent Dirt Roads* | Not significant at the settlement level | Positive with Unpaved-to-Paved Road Ratio | In several of the samples, only dirt roads existed, while in others there were no dirt roads at all. Asphalt roads were consistently distributed through all samples with a higher proportion in formal areas. |
| *Unpaved-to-Paved Road Ratio* | Not significant at the settlement level, distribution is independent | Positive with Percent Dirt Roads | The proportion of dirt to asphalt was not a useful measure and naturally was correlated with Percent Dirt Roads. |
| *Vegetation Percent* | Veg. Percent provides less information than shape-based measures of vegetation. | Positive with Mean Vegetation Patch Area. | The more vegetation existed in a sample, the larger the area of each patch. This is an expected relationship. Veg. Percent was excluded in favor of more detailed patch-based measures. |
| | Veg. Percent provides less information than patch-based measures of vegetation. | Negative with Mean Compactness Ratio. | The more informal a settlement, the more compact its vegetation. Formal settlements exhibit more elongated vegetation patches. |
| *Entropy* | High correlation with multiple other variables. Variance Inflation Factor highest for Entropy, so not retained. | High Positive with Contrast | Informal settlements have a wider variety of structure shapes and sizes to indicate randomness of intensity distribution (no preferred gray level pairs in the distance vector measured by entropy). Contrast (as a measure of local variation) is also higher in the informal settlements. |
| | GLCM Correlation is an additional measure that is lower in informal settlements. | Negative with GLCM Correlation. Relationship stronger in formal settlements. | The higher the entropy, the lower will be the correlation between grey levels in the image scene. GLCM Correlation is a measure of image linearity. |
| *GLCM Mean* | Significant correlation with Contrast on Roads and with Entropy | High positive with Entropy on Roads | The higher GLCM mean represented a higher grey level value over all pixels in the selected region for informal areas. |
| *Contrast on Roads* | To improve robustness with increased variety of texture measures, retain Entropy on Roads, since Entropy overall was already excluded. | High Positive with Entropy on Roads. | The higher the entropy on the roads, the higher, too, will be the contrast because these two measures are positively correlated. Contrast captures sum of squares variance, while entropy captures the orderliness of pixel values in a sliding window. Informal settlements have higher contrast on roads. |
| *Degree Slope on Roads* | Using 30m pixels, overall degree slope is a more robust indicator than on roads because of higher N. | Near perfect high positive correlation with Degree Slope | If smaller elevation pixels were available, slope on roads may be preferred, but use of this measure was limited by pixel size. Informal settlements overall have higher slope. |

Lacunarity and Fractal Dimension were slightly negatively correlated. The original hypothesis that Lacunarity would be higher in the informal settlements required lacunarity to be tested for sufficiency as a discriminator and to compare the results in this research to the conflicting results reported in prior research. Due to correlation between lacunarity and fractal dimension and the higher discriminatory power of lacunarity at the settlement level, fractal dimension was excluded. Connected node ratio (CNR) and mean node valence were highly positively correlated, but mean node valence was not a significant discriminator at the settlement level (p=0.166) and was therefore dropped. Percent Dirt Roads and Unpaved-to-paved Road Ratio were positively correlated, but neither alone provided useful discriminatory power to the model. Also, both variables have extreme outliers, so both of these indicators were dropped. Despite the higher significance level of Vegetation Percent compared to Vegetation Patch Size when comparing the means at the settlement level, the additional shape information derived from vegetation patches was preferred. Therefore Vegetation Percent was removed from the model while Vegetation Patch Size was retained.

The texture measures of GLCM Correlation and Contrast were correlated with Entropy. Even though Entropy was more significant in discriminating settlement type at the settlement level, the need to explore additional metrics was preferred for the potential contribution to the science of settlement analysis. Entropy was also correlated with multiple other measures and suffered the highest variance inflation factor (VIF)[8] of all variables (5.38), so Entropy was excluded. Contrast on Roads was highly positively correlated with Entropy on Roads, but since Entropy was already excluded, the potential for increased model robustness and greater variable diversity meant Entropy on Roads was retained while Contrast on Roads was excluded. Page 205 in Appendix B lists the variance inflation factors for all variables.  Last, Degree Slope on Roads was excluded due to the large pixel size of the underlying DEM dataset in comparison to the size and location of some of the road segments, and the potential for better explanatory power using all of the slope pixels in a settlement area instead of just constraining them to the road surface.  Plan Convexity was also removed from the model because it lacked significance (p=0.17 from t-test of independent means).

After excluding the above variables, Table 31 lists the remaining fourteen variables that were used for DA and CART, and the number of missing values.

[8] VIF of a predictor variable $k$ is measured by $\frac{1}{1-R_k^2}$ where $R_k^2$ is the $R^2$ value obtained by regressing the $k^{th}$ predictor variable on the remaining predictors.  VIF shows how much the variance of the estimated regression coefficient $b_k$ is "inflated" by the existence of correlation among the predictor variables in the model. A VIF of 1 means that there is no correlation among the $k^{th}$ predictor and the remaining predictor variables. VIFs exceeding 4 should be investigated further, and VIF's exceeding 10 indicate incorrect predictor specification due to multicollinearity (Simon, 2004)

**Table 31 Un-correlated variables included in discriminant analysis**

| Number | Variable Name | # Missing Values | % Missing Values | Variable Category |
|---|---|---|---|---|
| **1** | Lacunarity | | | Scale-dependent Measures |
| **2** | Connected Node Ratio (CNR) | 6 | 4.7% | Roads Measures |
| **3** | Pct 4-Way Intersections[9] | 17 | 13.3% | |
| **4** | Pct Dangles | 6 | 4.7% | |
| **5** | Asphalt Road Content[10] | 2 | 1.6% | |
| **6** | Road Density | | | |
| **7** | Vegetation Patch Area | | | Spectral Measures |
| **8** | Veg. Compactness Ratio | | | |
| **9** | Soil Coverage | | | |
| **10** | GLCM Correlation | | | Texture Measures |
| **11** | Contrast | | | |
| **12** | Entropy on Roads | | | |
| **13** | Degree Slope | | | Topographic Measures |
| **14** | Profile Convexity | | | |

## Results from Sampling by Grid Cell

This section describes the results from measuring the indicators using the sub-sampling method described earlier. In this section, each sampling unit was 150m$^2$ (or less) and intersected with the settlement boundary. Figure 44 illustrates this point with two side-by-side samples in the informal settlement of Berlin. The left image shows a zoomed-in view of the smallest and largest sized samples as they occurred randomly in the sample design, side-by-side. The smallest sample was only 5,691m$^2$ while the largest

[9] No intersections existed
[10] Roads classified as neither asphalt nor dirt.

wass 22,500m$^2$, or 150x150m as a square randomly-selected sample that fit completely within the settlement boundary.   The smallest sample's area covered at least 25% of the original settlement, but its area was constrained by the outer settlement boundary.

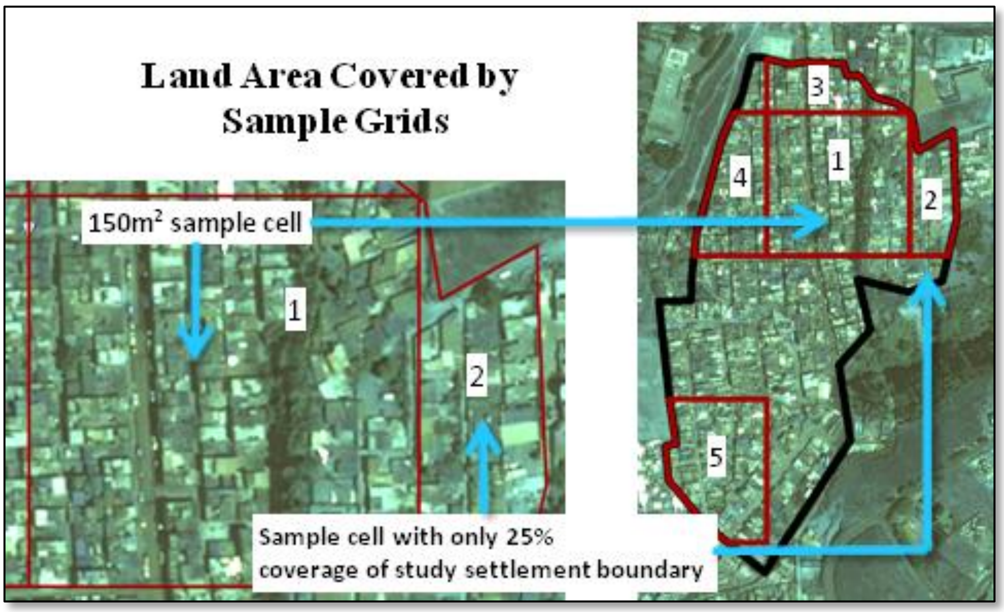



**Figure 44 Sample Grid Coverage over Settlement Boundary**

This is easier to visualize in the image on the right side of Figure 44, where the irregular shaped black line represents the boundary of the settlement of Berlin, so that all random samples in Berlin (there were 5, outlined in red) are clipped to the Berlin boundary. These random samples originally came from the overlaid grid in the Figure 17.  Recall that the final 127 samples that intersected the settlements were randomly and proportionally selected to ensure at least 25% of their area intersected the settlement boundaries, and that the sample size was sufficient to ensure model robustness.

### Discriminant Function Analysis Results

Discriminant Function Analysis (DFA), or DA, was used to determine which linear combination of variables performed best at discriminating settlement type according to the equation:

$$D = b_1x_1 + b_2x_2 + ... + b_nx_n + c,$$

where the $D$ is the discriminant function, $b_n$ coefficients of the independent variables $x_n$ maximize the distance between the means of the criterion variable of settlement type (Griffith & Amrhein, 1997). The Wilks' lambda (*F-test*) tests if the discriminant model as a whole is significant. If the *F test* shows significance for a variable, then the individual independent variables will be assessed to see which differ significantly in mean by group to classify the dependent variables. For each variable $X_j$, Wilk's Lambda is the ratio of the within-group sum of squares and total sum of squares, as follows:

$$\Lambda_j = \frac{within\ group\ sum-of-squares}{total\ sum-of-squares}$$

(El Ouardighi, El Akadi, & Aboutajdine, 2007). The closer the corresponding $\Lambda_j$ for variable $X_j$ is to 0, where $0 \leq \Lambda_j \leq 1$, the better its ability to differentiate, and the more $X_j$ contributes to the discriminant function. $\Lambda_j = 1$ indicates all group means are the same and no differences exist.

In the present research, the removal of variable collinearity creates a DA model that now tests only 14 of the original 23 variables. The included variables are listed in

Table 31. In solving for the Discriminant Function, it is noted that 18 samples have at least 1 missing value in one or more variables (see Table 31), and were excluded from the analysis. The variables with the most missing values were related to road accessibility and road surface metrics due to several samples with minimal road coverage and zero intersections or nodes. The model was run with and without the stepwise constraints. When the stepwise requirement was relaxed the model incorporated 12 variables instead of 6 and only Percent Dangles was explicitly rejected due to a low tolerance value. In SPSS, tolerance is reported as a measure of variable dependence or collinearity, so a low value indicates that other variable(s) still co-vary with it. Having this many variables could result in overfitting, which makes the model less generalizable and is not preferable.

The stepwise method was therefore applied to avoid overfitting. The probability of significance using the F-ratio must be $\leq 0.05$ to be entered/included and $\geq 0.2$ to be excluded. The exclusion F-ratio was somewhat relaxed because of the robust effort to eliminate prior correlated variables. Table 32 lists the 6 variables in the discriminant function that had greatest contribution to the model. The table includes the tolerance or proportion of the variance not accounted for by other independent variables, and the Wilks' Lambda result. The low Wilks' Lambda for each variable in Table 32 shows that Entropy on Roads, Mean Vegetation Patch Area, Compactness Ratio, Profile Convexity, Road Density, and Soil Percent by Grid were needed to distinguish settlement type.

**Table 32 Variables selected by Discriminant Function Analysis**

| Variables Included in the Analysis | | | |
|---|---|---|---|
| | Tolerance | Sig. of F to Remove | Wilks' Lambda |
| Entropy_Roads | .966 | .000 | .442 |
| Mean Veg Area | .915 | .001 | .281 |
| CompactnessRatio | .850 | .000 | .292 |
| ProfileConvexity | .906 | .000 | .289 |
| RoadLenPerArea | .903 | .005 | .275 |
| SoilPctByGrid | .906 | .011 | .271 |

The reported canonical correlation (0.863) between the discriminant scores and the groups they define is suitably high, with the resulting DA function explaining 100% of the variance in the dependent variable. This discriminant function is positively separating samples into the correct groups because the proportion of the total variance in the DA scores not explained by differences between the groups is low. The $\chi^2$ reports the ability of this discriminant function to distinguish settlement types. $\chi^2$ for the function, where $f_i$ is the observed count for $class_i$ and $F_i$ is the expected count for $class_i$ and sample size = 108 (*127 −17* missing values −*2* parameters to be estimated (informal/formal) –*1*) is:

$$\chi^2 = \sum_{i=1}^{2} \frac{(f_i - F_i)^2}{F_i} = 142$$

This result is significant at α=.005.

The separability of informal and formal discriminant function scores (y-values) displayed by their histograms are graphically shown in Figure 45 and Figure 46 where Formal (SettleType=2) and Informal (SettleType=1) are shown with a common x-axis scale of −5 to +5. The mean score for formal settlements is −1.8 and for informal settlements is +1.64 indicating good separability.

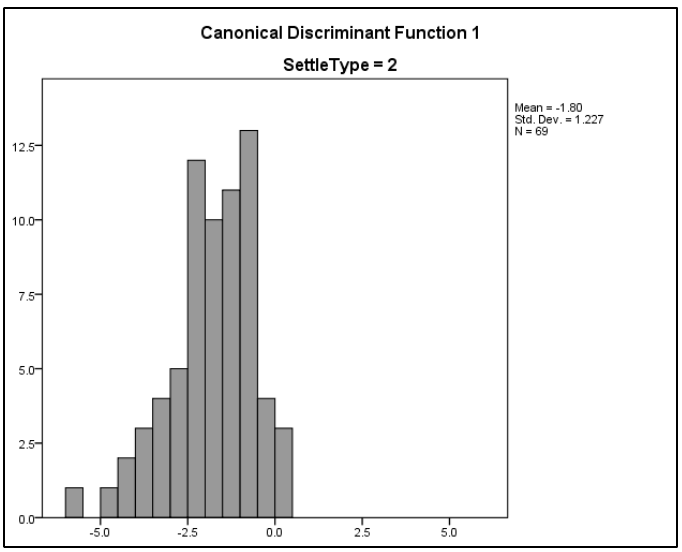



**Figure 45 Histogram of Discriminant Function Scores - Formal Settlements**

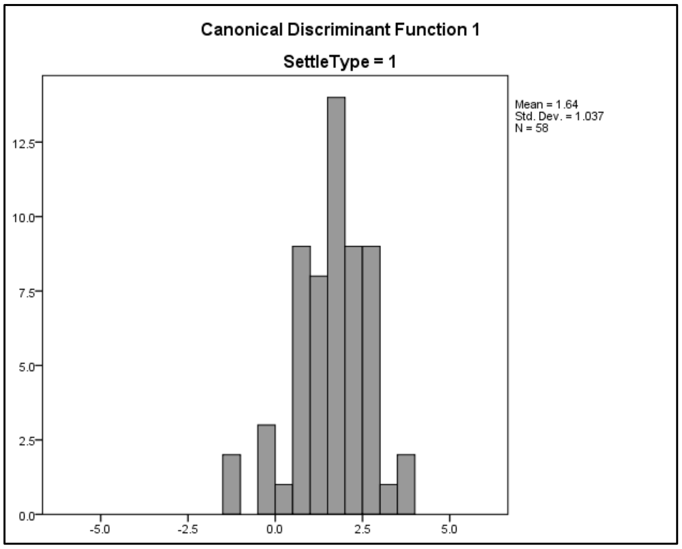



**Figure 46 Histogram of Discriminant Function Scores - Informal Settlements**

After cross-validation, 92.1% of the cases were correctly classified. The cases that were not correctly classified were distributed fairly evenly across settlements (3 cases misclassified Informal as Formal - 1 from Satélite, 1 from Joya, 1 from Peronia; and 5 cases misclassified Formal as Informal - 1 from GTowns2, 2 from Piñares, 1 from Balcones and 1 from Terrazas). This implies the DA function is effectively differentiating settlement type. The standardized canonical discriminant function *beta* coefficients are shown in Table 33.

**Table 33 Standardized Canonical Discriminant Function Coefficients**

**Standardized Canonical Discriminant Function Coefficients**

| | Function |
|---|---|
| | 1 |
| RoadLenPerArea | .336 |
| Mean Veg Area | -.373 |
| CompactnessRatio | .452 |
| SoilPctByGrid | .304 |
| Entropy_Roads | .767 |
| ProfileConvexity | -.418 |

Based on the results in Table 33, the final discriminant function is written as:

$$\begin{aligned} D_{(formal,\ informal)} = {} & (0.336 \times \mathit{RoadLenPerArea}) + (-0.373 \times \mathit{MeanVegArea}) + \\ & (0.452 \times \mathit{CompactnessRatio}) + (0.304 \times \mathit{SoilPctByGrid}) + \\ & (0.767 \times \mathit{EntropyRoads}) + (-0.418 \times \mathit{ProfileConvexity}) + c). \end{aligned}$$

The group centroids are listed in Table 34. If the canonical coefficient multiplied by the variable's measured value is summed for all variables using the above $D_{(formal,\ informal)}$ equation, any sample with a result closer to 1.6 means the settlement is most likely informal. If the result is closer to −1.7 the settlement is most likely formal. The cut score of −0.0155 is half the centroid difference. This means if the result for a new sample is < −0.0155 the settlement would be classified as formal, while a result > −0.0155 would be classified as informal.

**Table 34 Discriminant Function Group Centroids**

**Functions at Group Centroids**

| SettleType | Function 1 |
|---|---|
| 1 | 1.681 |
| 2 | -1.712 |

**Table 35 Discriminant Analysis Cross Validation Results**

**Classification Results[a,c]**

| | | SettleType | Predicted Group Membership 1 | 2 | Total |
|---|---|---|---|---|---|
| Original | Count | 1 | 53 | 5 | 58 |
| | | 2 | 3 | 66 | 69 |
| | % | 1 | 91.4 | 8.6 | 100.0 |
| | | 2 | 4.3 | 95.7 | 100.0 |
| Cross-validated[b] | Count | 1 | 53 | 5 | 58 |
| | | 2 | 5 | 64 | 69 |
| | % | 1 | 91.4 | 8.6 | 100.0 |
| | | 2 | 7.2 | 92.8 | 100.0 |

a. 93.7% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 92.1% of cross-validated grouped cases correctly classified.

It is notable that at least one variable was selected in the DFA from four of the five major categories of Texture, Topography, Roads, Spectral-related measures, and Scale-related measures. This means capturing a broad spectrum of settlement structural characteristics is required to differentiate informal and formal neighborhoods.

### *Variables Required for Classification using Discriminant Analysis*

Lacunarity measures 'gappiness' of image texture and translational or rotational invariance, while Fractal Dimension measures self-similarity across scales described as complexity of image texture. During sub-sampling, lacunarity was the only Scale-related metric evaluated, and it did not contribute meaningfully to the DA function. As a way to confirm the related measure of fractal dimension was also not a useful differentiator, a t-test of means was applied to all 127 samples for Fractal Dimension. The result failed to discriminate between settlement types, confirming that fractal dimension is not useful in capturing patterns of built-up areas in small area samples ($\leq 150m^2$ sample cells). Fractal Dimension and Lacunarity should only be used at the settlement scale.

Entropy on roads was a novel approach to constraining a texture metric to an anthropogenic feature common to all residential areas. Entropy, as a measure of disorder in grey level tonal value, was significantly higher in the informal areas. Mean Vegetation Patch Size is smaller in the informal areas. This is a direct indication of how little greenspace is preserved in such settlements, and its minimal size and shape. Compactness ratio was a vegetation patch shape-related indicator derived from spectral information that was higher in the informal settlements. This means in addition to being small in size the vegetation patches are not elongated in the informal areas, possibly contrasting with border or ornamental plantings found in more affluent or planned areas. Soil Percent was an unexpected discriminator, with the result that excess soil was found in the informal settlement samples. It is not known if the reason for the excess soil is from building construction areas exposed on the imagery capture date, more unpaved parking lots, or fallow garden soil. The result could underscore the high soil content where informal settlements in a constant state of building and regeneration, described in the literature as *in-situ accretion* or *densification* (J. X. Barros, 2004; Griffin & Ford, 1980; Gilbert, 1996).

Profile Convexity was the only significant topographic metric needed for the result. The informal settlements had more negative Profile Convexity values, and were therefore built on more concave terrain in the gradient (upward, downward) direction. This results in a minimization of gravity effects of flow over the surface, which implies a more disturbed ground shape, and less desirable locations for building and habitation. The more negative profile convexity in the informal settlements is an indication of terrain that may be susceptible to pooled water. From a terrain perspective, builders often wish to build on ridgelines and not beneath them. Building in the concave parts of the landscape appears to be a good indicator of the informal settlement type. Further modeling of drainage profiles, surface and subsurface hydrologic regimes, and soil infiltration rates are needed to go beyond this initial speculation, but the result is the more negative the profile convexity measure, the less desirable may be the land for human habitation.

Road Density was also a very significant variable in predicting settlement type, with informal settlements having higher road density. Bias was avoided by summing asphalt and dirt road lengths together instead of only including asphalt roads, yet it seems apparent that the more densely populated a region, the more roads there will be, and urban informal settlements are more densely populated by their nature. An additional attribute of road width could improve this measure by adding to the understanding of road structure, but higher resolution imagery would be needed to capture accurate road widths. This indicator could also possibly contribute to current efforts measuring population density from imagery. Without the added certainty of building outlines, care must be taken when interpreting its meaning. Also, this indicator may not be useful where multi-story informal settlements predominate, such as in Rio de Janeiro in Brazil.

### Classification and Regression Tree Results

This section describes the results from the CART approach. CART was selected for its potential to overcome the restrictions of the DA model, namely the inability to incorporate any samples with missing data, prohibition of multicollinearity among variables, and the preference for variables normally distributed. CART is a non-parametric approach that efficiently selects from among a large number of variables the ones most able to explain the dependent variable class (e.g., informal or formal). It uses a recursive partitioning method that divides the entire sample space into binary subsamples by seeking variable splits that produce homogenous subgroups. For the purposes of this discussion, *samples* (150m$^2$ grid cells) and *cases* are used synonymously.

The important variables in the CART model differed only marginally from DA, and fewer categories were needed to group samples by settlement type. The contribution of GLCM Correlation (Texture) and Profile Convexity (Topographic) were almost negligible in the CART analysis (3% and 1% contribution, respectively). This means only the categories of Texture, Spectral Properties and Road Networks were needed in order to classify settlements, unlike DA which also required Topographic measures. Greater parsimony is obtained when using CART for classification because fewer variables must be added to the model. The variable importance is summarized in Table 36, and also includes surrogate splitters that are used when missing values impact the splitting criteria. Asphalt Road Content, Connected Node Ratio, and Road Density were not actually used in the final decision tree, but still contribute significant information content when considered as surrogate splitters.

**Table 36 Variables deemed important in CART compared to DA**

| **Variable Name** | **Variable Group** | **% Info. per variable, CART** | **Used in DA Result?** |
|---|---|---|---|
| **Entropy on Roads** | Texture | 100% | Y |
| **Mean Vegetation Patch Size** | Spectral | 75% | Y |
| **Vegetation Patch Compactness** | Texture | 58% | Y |
| **Asphalt Road Content** | Spectral | 53% | N |
| **Connected Node Ratio** | Road Accessibility | 52% | N |
| **Road Density** | Spectral | 37% | Y |
| **GLCM Correlation** | Texture | 3% | N |
| **Profile Convexity** | Topographic | 1% | Y |

Variables not used by the DA model but with > 50% information content to the CART model were Asphalt Road Content and Connected Node Ratio. The fact that Entropy on Roads, Mean Vegetation Patch Size, and Vegetation Compactness comprised three of the top four variables in both the CART and DA models reveals the importance of these categories of indicators when differentiating settlements.

### *Variables Required for Classification using CART*

The optimized CART result produced a 7-node classification. Figure 47 displays the tree. The tree shows Entropy on Roads is the most important criteria, with 65 cases having entropy values ≤ 1.51, of which nearly 90% are informal. Vegetation Patch Size is the next most important criteria as the splitting variable used for both sides of the top level split. For the right side of the tree, Vegetation Patch Size of ≤ $21m^2$ contains 45 cases, of which 100% are informal. The vegetation patch size for the left side of the top split contains 50 cases > $14.6m^2$ giving a good indication that mean vegetation patch size for 61 of the 69 formal settlements is > $14.6m^2$. The third split on the left side of the diagram is somewhat counterintuitive because there were 7 informal and 3 formal settlements with mean vegetation patch size of > $12m^2$. The tree helps visualize the variability that still exists among the informal cases, with some having vegetation patches as large as their formal counterparts. Continuing down the left side of the tree, the final split occurs with two formal cases having a GLCM Correlation of > −33.68, while 90% of the remaining cases are informal with a lower GLCM Correlation. On the right side of the tree, after separating into two groups based on Vegetation Patch Size of $21.2m^2$, the remaining 17 cases greater than the cutoff are further divided into formal and informal based on compactness ratio, with 100% of informal samples classified as having compactness > 0.69. Recall that a circle has compactness of 1, so the informal settlements can be described as having small, compact vegetation patches lacking complex shapes.

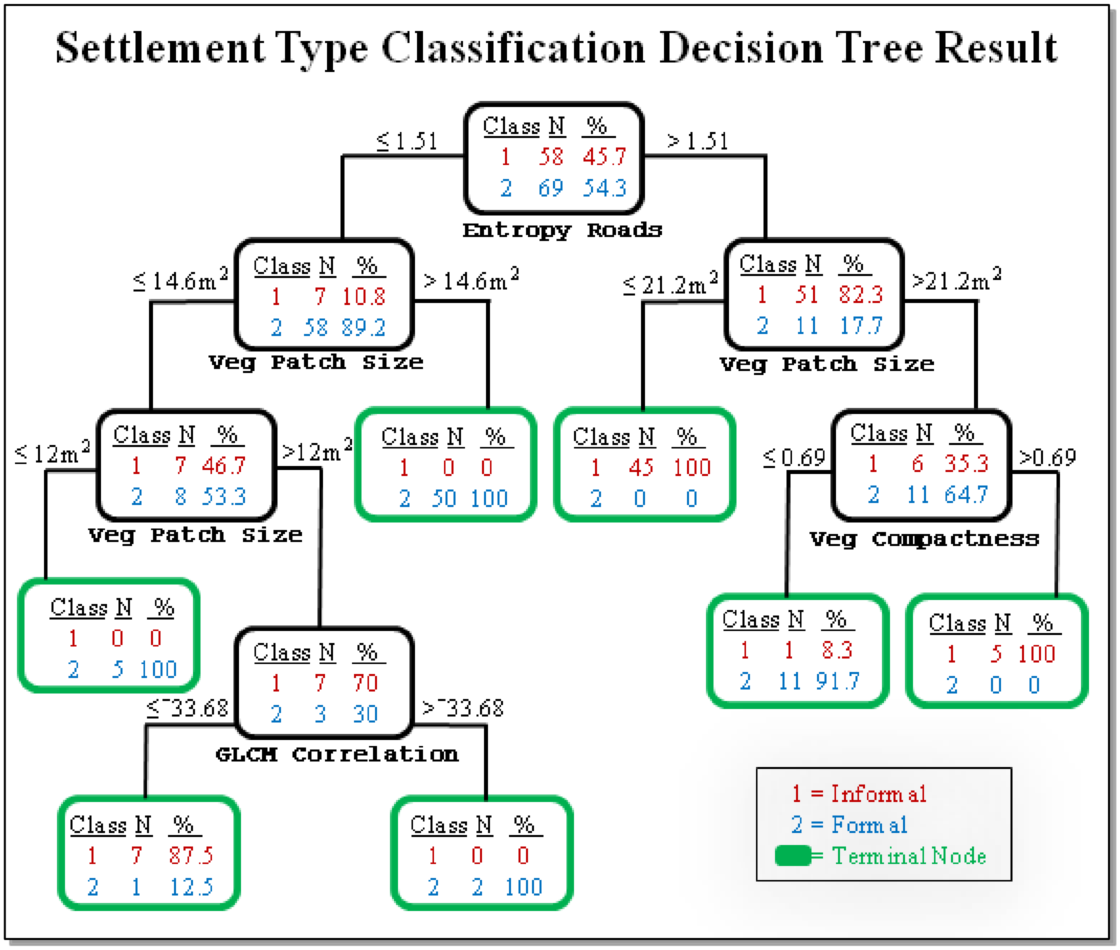



**Figure 47 CART Decision Tree Diagram**

For CART, the impact of missing data was minimal. Of all variables used in the decision tree (primary or surrogate splitters), Connected Node Ratio (a Roads measure) was missing 4.7% of values and Asphalt Road Content was missing 1.6% of values.  The model was unaffected by other variables that were missing data, such as Percent 4-Way Intersections and Percent Dangles because these were not used as splitting criteria. When cross-validation was performed, the estimate of overall classification accuracy was 87.5%. The misclassification rate for test data by class is shown in Table 37, where 7 informal and 9 formal cases were misclassified on cross-validation.

**Table 37 Misclassification for test data, CART model**

| Class | N Cases | N Misclassified | Pct Error |
|---|---|---|---|
| **1 (Informal)** | 58 | 7 | 12.07 |
| **2 (Formal)** | 69 | 9 | 13.04 |

Classification accuracy was ≥ 87% with only four variables (Entropy on Roads, Vegetation Patch Size, Vegetation Patch Compactness, and GLCM Correlation) used as primary splitters. Although the CART model is less accurate than the DA model, which reported 92% accuracy after cross validation, it requires fewer variables to complete. Entropy on Roads, Vegetation Patch Size, and Vegetation Compactness were the only measures needed to classify 92% of cases (117/127). The added splitting criteria of GLCM Correlation completed the classification with an overall 87.5% accuracy, which still exceeds the 80% lowest acceptable margin of error suggested by the international Expert Working Group on Slum Mapping (Sliuzas et al., 2008, p. 15).

# CONCLUSIONS

The research objectives, reiterated here, were:

1. To test the hypothesis that foundational measures derived from roads, vegetation, soil, image texture, and geomorphology evaluated with multivariate methods can distinguish settlement structure of residential areas.

2. To determine if imagery and elevation data alone are sufficient to distinguish informal from formal settlements.

3. To develop a small set of statistically significant indicators that distinguish settlement type to be used by organizations such as the UN HABITAT and urban planners without the need for field work or surveys.

This research proved that a limited set of indicators derived from remote sensing image analysis and GIS techniques were required to correctly distinguish informal from formal settlements in an urban scene. Although field reconnaissance was conducted to define the areas for testing purposes, no household surveys or questionnaires were used.

Figure 48 graphically depicts the relative importance of each indicator variable by modeling method. For the CART result, the y-axis represents the percent of that variable's information content that is useful in deciding whether to classify a sample as informal or formal. For DA, the y-axis represents the absolute size of the variable's correlation with the discriminant function as a measure of agreement.

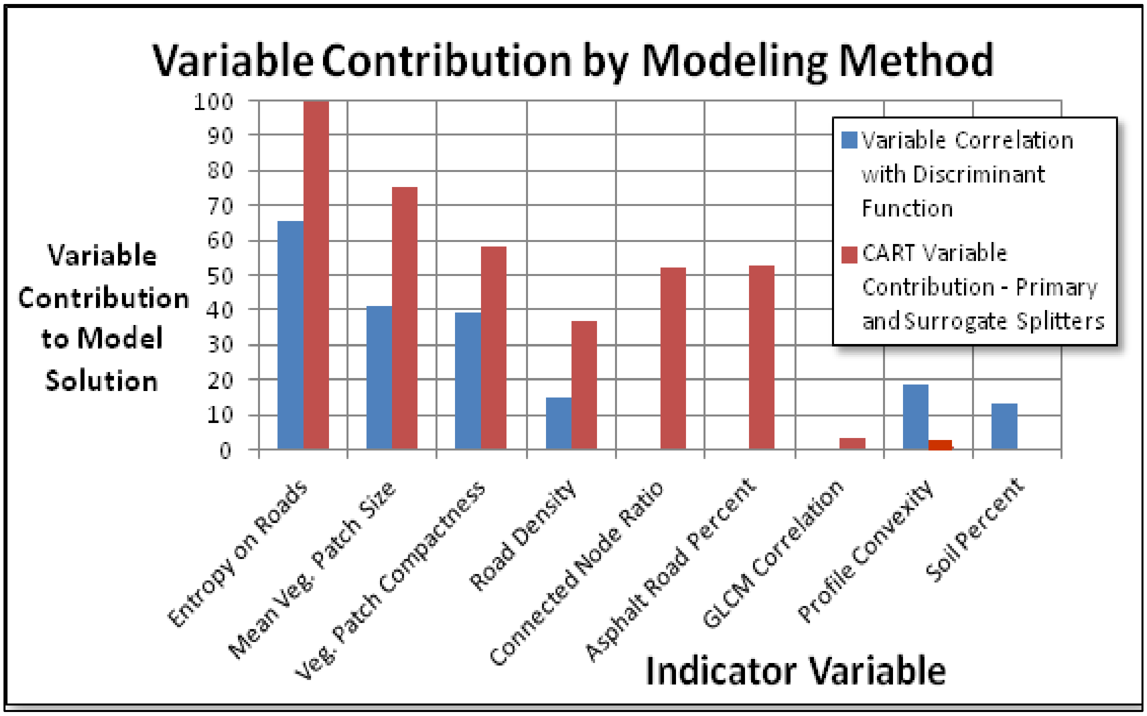



**Figure 48  Relative variable importance by modeling method**

### Reducing variable dimensionality.

Limiting the measured indicators to texture, vegetation characteristics derived from spectral information, and geomorphology from elevation data is a more economical way to parsimoniously pre-process large amounts of imagery to reveal significant differences in settlement types. The settlement boundary must be delineated to mask the results, and roads must be vectorized and buffered (for Entropy on Roads), but road accessibility measures may not be needed for the CART approach. However, the trade-off for parsimony is accuracy. The predictive accuracy of the CART methods tested through cross-validation was high (87.5%), but not as high as the DA method (92%). Using DA, the additional pre-processing required to include road accessibility enjoys the added benefit of increased predictive accuracy, and both methods fall comfortably above the suggested 80% minimum margin of error for measuring informal settlements.

### Differentiating informal from formal settlements.

The diversity of land features, mixed-use settlements, topographic variation, and the heterogeneity of building materials and neighborhood structure in informal settlements worldwide serve to limit universal applicability of the proposed measures. Based on the contribution of this research, however, new indicators could easily be developed for extremely arid climates, low-lying, littoral or riparian areas proximate to oceans or large water bodies, or settlements with snow or frost during large parts of the year. These characteristics impact settlement structure and where the impoverished choose to live. Despite regional variability, many of the metrics that were significant in Guatemala City will be relevant elsewhere. Road accessibility, topography, texture, and spectrally-derived materials such as asphalt and vegetation provide the information needed to quantify a settlement's structure. The conclusions for each category are summarized here.

#### *Road Accessibility*

All settlements have roads or pathways – paved and unpaved. Road topology, surface materials and road distribution can lead to an understanding of settlement quality and road accessibility. Intensely urban areas with warrens of dead-ends and many road fragments typify the world's worst slums (Davis, 2007). This research has contributed a variety of roads measures that could be applied anywhere to evaluate accessibility by showing that informal settlements exhibit lower road accessibility than formal areas.

### *Topography and Geomorphology*

Topographic indicators derived from elevation data that have been constrained to the settlement boundaries of residential areas provide an understanding of how terrain impacts residents. From elevation, detailed geomorphology can be ascertained. The choice and cost of home sites is influenced by slope, curvature (profile and plan convexity), and other geomorphic properties that also impact hydrologic or volcanic flow and drainage in residential areas. In other parts of the world, instead of steep slopes, informal settlements might be found in floodplains susceptible to inundation. This research proves that aspects of geomorphology can discriminate between informal and formal settlements. A broad geomorphic profile that includes soils and hydrologic flow regimes would provide even more detail.

### *Texture*

The classic texture measures of Entropy and GLCM Correlation differ significantly between informal and formal areas. When constrained to roads, entropy is the strongest settlement type discriminator of all, and is an indication that road variability is high in informal areas. GLCM correlation, as a measure of image linearity or linear dependence of the intensity values on neighboring pixels is higher in the formal settlements.  This represents a level of evenness and consistency of intensity values over such settlement areas. Texture can be measured from most advanced remote sensing image processing tools, but the ability to constrain the results to residential areas and specifically to the roads is a new approach that could be useful in settlement analysis worldwide.

### *Spectral-related measures*

Spectral properties of dominant materials such as vegetation, asphalt, and soil extracted from imagery and then quantified demonstrate significant differences between informal and formal settlements. Vegetation was the most important feature material used to differentiate settlements. When NDVI-derived vegetation is transformed into vector features from pixels, it was demonstrated that compact shape and small size of the resulting vegetation patches were among the top three most significant metrics.  Given that local materials based on natural resources are often a component of road surfaces or roofs, selecting some key materials such as asphalt (or concrete) and soil proved extremely indicative of the dissimilarities in settlement types. Asphalt was found more often on roads of more upscale formal communities, while soil coverage over the entire area and not just of the road surface was higher in informal rather than formal settlements.

***Scale-related measures of built-up areas***

The Scale-related category that includes the measures of Lacunarity and Fractal Dimension was previously reported in the literature to measure the structural qualities of slums in other parts of Latin America, yet those published results were inconsistent. The conclusion of this research is that higher Lacunarity (but not Fractal Dimension) does characterize a settlement’s built-up areas as informal versus formal. However, the smaller the sampling unit, the less useful this indicator becomes. Lacunarity is not a straightforward indicator and requires significant image preparation and pre-processing. Therefore, its practical applications are limited.

To summarize, this research proves that the structure of a settlement can be quantified by the spectral properties of dominant materials, road surfaces and accessibility, image texture, geomorphology from elevation data, and especially vegetation shape, size and distribution and that significant differences exist between informal and formal settlements using these characteristics. Settlement types can be correctly differentiated with 87-92% accuracy without the need to extract individual dwelling footprints. Relying only on multispectral VHR imagery and digital elevation data provides a means to measure settlements without sophisticated ancillary datasets or household surveys. This is a major step toward scientific measurement of indicators that can be associated with quality of life and economic wellbeing from imagery and provides simplified methods for annual monitoring of slum conditions.

## Limitations

The objective in this research was to apply multivariate analysis using a variety of possible remote sensing and GIS-based indicators to assign image areas to one member of a binary group (formal or informal) in a part of the world containing urban slums, and often lacking usable data. Household surveys and other ancillary data were excluded for reasons of safety, cost and lack of availability. As a result, the socio-economic indicators of quality of life and economic status, which are known to be lower in slums and informal settlements, would still require survey data for validation. This research did not use any ancillary data, such as census or real estate records. However, even if such data were available in small spatial units such as neighborhoods, their boundaries would not coincide precisely with measurable physical settlement boundaries of residential areas like the initial sampling units created here. The modifiable areal unit problem (MAUP) would likely arise, so that instead of measuring a neighborhood, the measurement would be applied to an administrative boundary, and the results would likely differ (Wong, 2004; Wong & Lee, 2005; O'Sullivan & Unwin, 2003). In this way, the limitations of this research are also an advantage.

Another limitation is that single-story buildings were assumed to be dwellings. This is because the extraction of multi-story buildings from imagery requires additional algorithms to distinguish shadows and may need to incorporate alternative sensors such as LIDAR or RADARSAT. If the distortion of spectral information caused by shadows was accounted for, some indicators such as road surface may have slightly different results. This research was therefore applicable to neighborhoods of mostly single or two-story dwellings.

Part of the criteria for selecting a settlement boundary was its residential status. In this part of Guatemala, family-run businesses that adjoin residences are common, as are other mixed-use structures. The intentional exclusion of much-larger buildings as commercial (e.g., industrial facilities, apartment buildings, or office buildings) was a heuristic that was not systematically derived. This could present a slight conflict given the mixed-use nature of property in urban areas of Latin America. Visual reconnaissance and settlement boundary demarcation conducted in the early stages of the research was intended to exclude these non-residential areas. It should also be mentioned that refugee camps (Giada et al., 2003) are a distinctly different settlement type often with regular patterns and self-similar shapes that were not the focus in this research. The regularized structure of refugee camps could possibly lead to a different set of indicators.

A final limitation concerns the absence of any indicators based on dwelling footprints. Extracting accurate building outlines is non-trivial, and though attempted, was not completed for reasons visually apparent in Figure 21. Much research is currently dedicated to the building extraction problem, but the test areas are small (usually < 100 dwellings). The scope of those efforts would not scale upward to the larger settlement areas measured here (4km$^2$ total area) because shape and image variability increase dramatically with study area size in urban communities. Intrinsically, this research has overcome the limitations caused by difficulties in using dwelling footprint metrics because it reveals an alternative way to successfully differentiate informal from formal settlements.

## Recommendations for Future Research

The most practical next step is to apply the significant indicators from this research (Entropy on Roads, Vegetation Patch Area, Vegetation Patch Compactness, Asphalt Road Content, Road Density, Connected Node Ratio, GLCM Correlation, Profile Convexity, and Soil Percent) to study areas with different climatic regimes, historical settlement patterns, and building vernacular in order to generalize indicators to other regions. Measuring improvement or deterioration of slum conditions over time could then easily be performed from imagery using change detection techniques, which is considerably less costly and devoid of the risks inherent in annual field surveys.

## APPENDICES

## APPENDIX A – STATISTICS FOR ALL VARIABLES

### Statistical significance of the difference in means for all variables

| **Independent Samples Test - Stratified Proportional Random Sampling by Grid Cell – All 127 Samples** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | t-test for Equality of Means - unequal variances | | | | | | |
| | | | | | | 90% Confidence Interval of the Difference | |
| **Variable Name** | t | df | Significance (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| FractalD | 2.872 | 102.12 | .005 | .037 | .013 | 0.016 | .058 |
| Lacunarity | -.583 | 109.55 | .561 | -.003 | .006 | -0.013 | .006 |
| CNR | -5.576 | 69.47 | .000 | -.191 | .034 | -0.248 | -.134 |
| MeanNodeVal | -4.369 | 104.65 | .000 | -.440 | .101 | -0.606 | -.273 |
| Pct4Way | -2.464 | 76.59 | .016 | -.111 | .045 | -0.185 | -.036 |
| PctDangl | 5.576 | 69.47 | .000 | .191 | .034 | 0.134 | .248 |
| AsphRdPct | -6.164 | 121.24 | .000 | -.236 | .038 | -0.299 | -.173 |
| DirtRdPct | 2.792 | 112.52 | .006 | .064 | .023 | 0.026 | .102 |
| RoadLenPerArea | 3.391 | 121.61 | .001 | .012 | .004 | 0.006 | .018 |
| Upaved:PavedRdSurface | .125 | 121.10 | .901 | 1.856 | 14.837 | -22.737 | 26.450 |
| Mean Veg Area | -7.931 | 100.83 | .000 | -18.688 | 2.356 | -22.600 | -14.776 |
| Compactness | 8.061 | 113.08 | .000 | .025 | .003 | 0.020 | .030 |
| SoilPctByGrid | 2.451 | 77.06 | .016 | 2.044 | .834 | 0.656 | 3.432 |
| VegPctByGrid | -11.418 | 105.45 | .000 | -10.348 | .906 | -11.852 | -8.844 |
| Mean_Correlation | -5.186 | 109.86 | .000 | -4.463 | .861 | -5.891 | -3.036 |
| Mean_Contrast | 2.786 | 124.12 | .006 | 2.279 | .818 | 0.923 | 3.634 |
| MeanEntropy | 8.020 | 102.20 | .000 | .092 | .011 | 0.073 | .111 |
| Mean_Contrast_Roads | 8.374 | 88.99 | .000 | 9.755 | 1.165 | 7.819 | 11.691 |
| MeanEntropy_Roads | 10.969 | 124.40 | .000 | .417 | .038 | 0.354 | .480 |
| DegreeSlopeRoads | 3.467 | 108.85 | .001 | 6.922 | 1.997 | 3.610 | 10.234 |
| DegreeSlope | 3.379 | 109.34 | .001 | 6.445 | 1.907 | 3.281 | 9.609 |
| ProfileConvexity | -3.331 | 87.05 | .001 | -.279 | .084 | -0.419 | -.140 |
| PlanConvexity | -.815 | 124.98 | .417 | -.549 | .673 | -1.665 | .567 |

## Descriptive statistics for CART and DA modeled variables

| | Valid N | Min | Max | Mean | StdDev | Var | Skew-ness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| AsphRdPct | 125 | .001 | .867 | .31807 | .243869 | .059 | .396 | -1.022 |
| CNR | 121 | 0.000 | 1.000 | .88363 | .202752 | .041 | -1.986 | 3.815 |
| CompactnessRatio | 127 | .634 | .740 | .68580 | .021348 | .000 | .241 | .200 |
| GLCM Correlation | 127 | -49.693 | -23.053 | -37.45168 | 5.509143 | 30.351 | .286 | -.123 |
| Entropy on Roads | 127 | .485 | 2.116 | 1.50651 | .298962 | .089 | -.286 | -.100 |
| Mean Veg Patch Size | 127 | 4.217 | 86.524 | 24.04613 | 16.768155 | 281.171 | 1.408 | 1.906 |
| Profile Convexity | 127 | -2.31 | .73 | -.5783 | .46979 | .221 | -.789 | 1.676 |
| RoadLength PerArea | 127 | .014 | .157 | .04841 | .021191 | .000 | 1.714 | 5.729 |
| SoilPct | 127 | .902 | 30.838 | 7.37421 | 4.521307 | 20.442 | 1.824 | 5.587 |

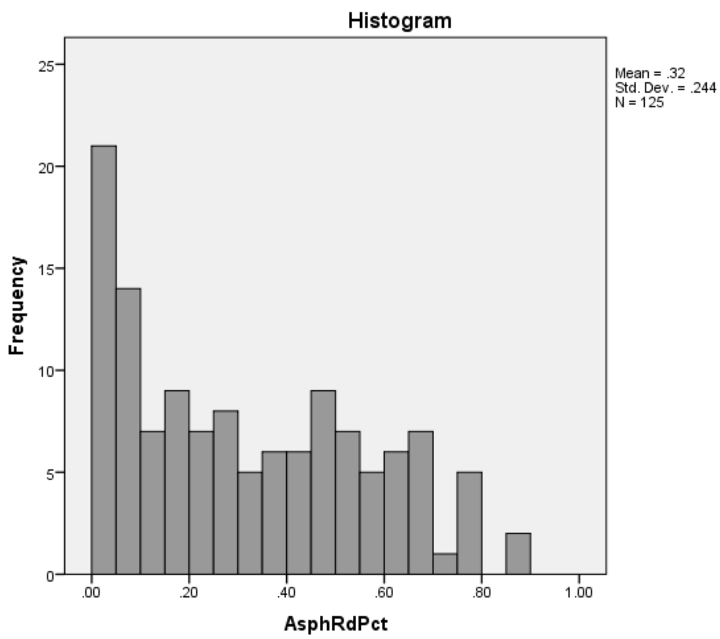
## Descriptive statistics for variables used in DA by settlement type

| Settlement Type | Indicator | Min | Max | Mean | Std. Deviation | Skewness |
|---|---|---|---|---|---|---|
| Formal | RoadLenPerArea | .014 | .128 | .04280 | .020439 | 1.615 |
| | CompactnessRatio | .634 | .724 | .67429 | .016046 | -.293 |
| | Mean Veg Area | 4.755 | 86.524 | 32.580 | 17.3745 | 1.042 |
| | SoilPct | 1.453 | 12.949 | 6.440 | 2.6929 | .190 |
| | Entropy_Roads | .485 | 1.917 | 1.31621 | .225243 | -.523 |
| | ProfileConvexity | -1.38 | .18 | -.4507 | .32311 | -.396 |
| Informal | RoadLenPerArea | .024 | .157 | .05509 | .020269 | 2.436 |
| | CompactnessRatio | .664 | .740 | .69948 | .018700 | .314 |
| | Mean Veg Patch Size | 4.217 | 51.036 | 13.892 | 8.2656 | 2.359 |
| | SoilPctByGrid | .902 | 30.838 | 8.484 | 5.8504 | 1.408 |
| | Entropy_Roads | 1.153 | 2.116 | 1.732 | .2026 | -.810 |
| | ProfileConvexity | -2.31 | .73 | -.7300 | .56585 | -.362 |

## All variable histograms and descriptives

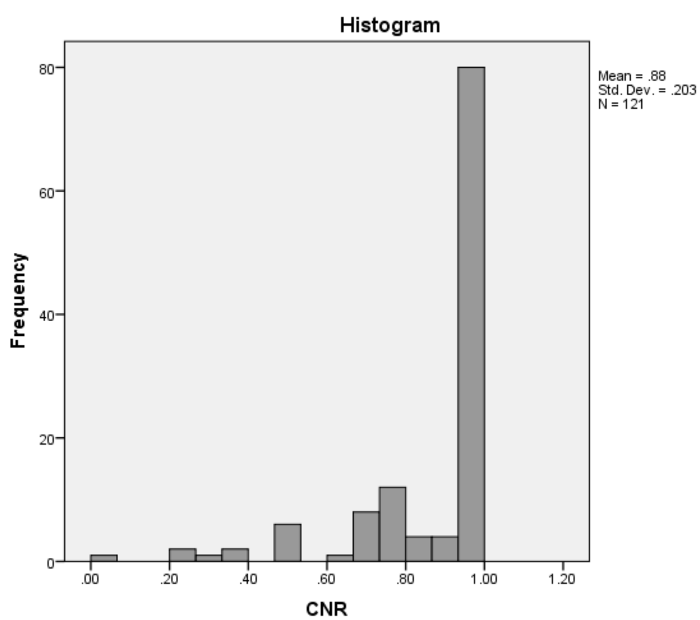
### *Asphalt Road Content*

| Statistics | | |
|---|---|---|
| AsphRdPct | | |
| N | Valid | 125 |
| | Missing | 2 |
| Mean | | .3181 |
| Std. Error of Mean | | .02181 |
| Std. Deviation | | .24387 |
| Variance | | .059 |
| Skewness | | .396 |
| Std. Error of Skewness | | .217 |
| Kurtosis | | -1.022 |
| Std. Error of Kurtosis | | .430 |
| Range | | .87 |
| Minimum | | .00 |
| Maximum | | .87 |
| Percentiles | 25 | .0821 |
| | 50 | .2831 |
| | 75 | .5064 |



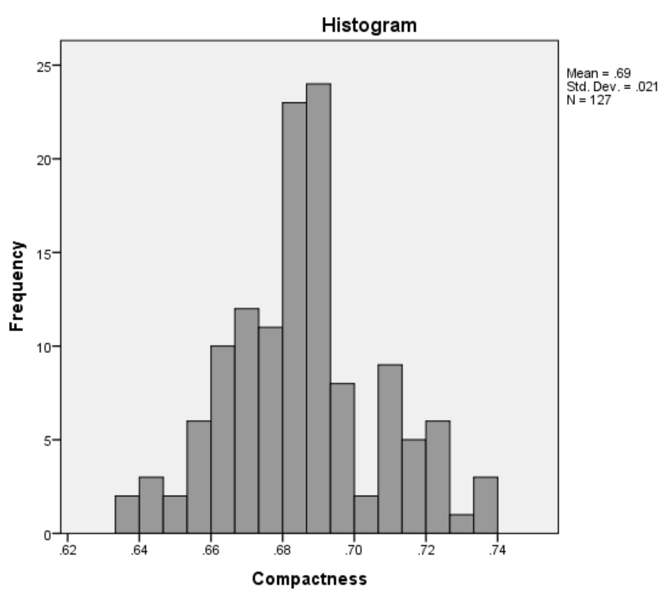
### *Connected Node Ratio (CNR)*

| Statistics | | |
|---|---|---|
| CNR | | |
| N | Valid | 121 |
| | Missing | 6 |
| Mean | | .8836 |
| Std. Error of Mean | | .01843 |
| Std. Deviation | | .20275 |
| Variance | | .041 |
| Skewness | | -1.986 |
| Std. Error of Skewness | | .220 |
| Kurtosis | | 3.815 |
| Std. Error of Kurtosis | | .437 |
| Range | | 1.00 |
| Minimum | | .00 |
| Maximum | | 1.00 |
| Percentiles | 25 | .8000 |
| | 50 | 1.0000 |
| | 75 | 1.0000 |



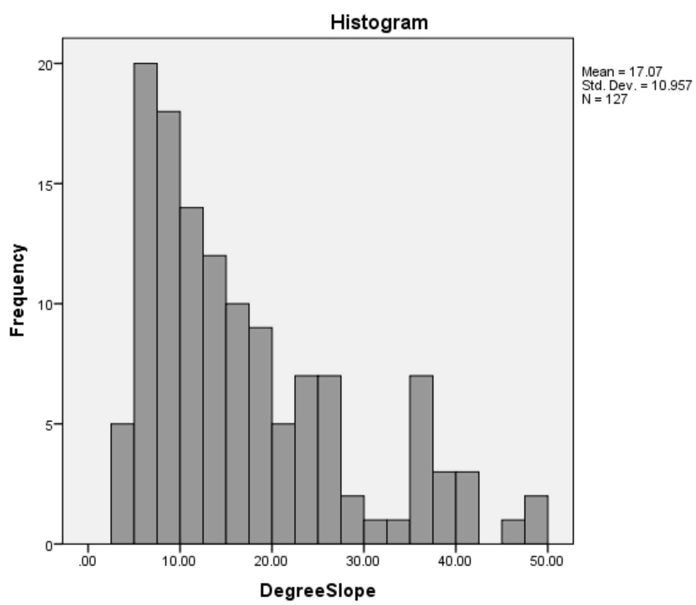
### *Vegetation Compactness*

| Statistics | | |
|---|---|---|
| Compactness | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | .6858 |
| Std. Error of Mean | | .00189 |
| Std. Deviation | | .02135 |
| Variance | | .000 |
| Skewness | | .241 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | .200 |
| Std. Error of Kurtosis | | .427 |
| Range | | .11 |
| Minimum | | .63 |
| Maximum | | .74 |
| Percentiles | 25 | .6723 |
| | 50 | .6855 |
| | 75 | .6949 |



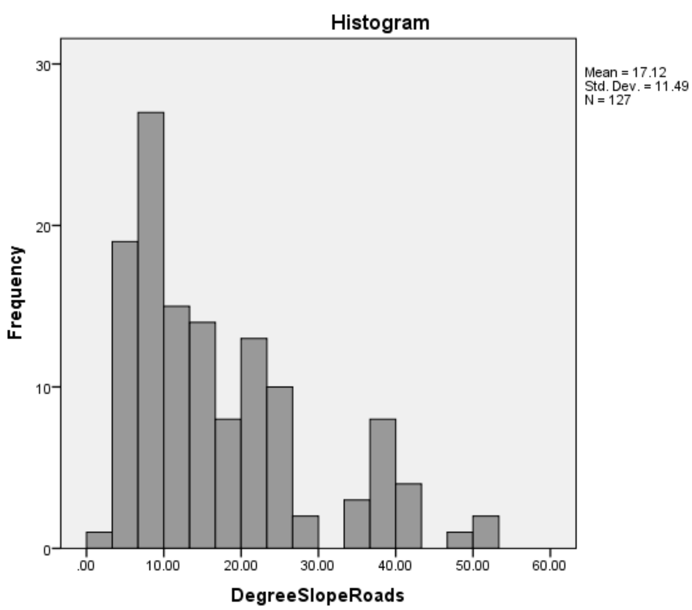
***Slope (in degrees)***

| Statistics | | |
|---|---|---|
| DegreeSlope | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 17.0683 |
| Std. Error of Mean | | .97230 |
| Std. Deviation | | 10.95730 |
| Variance | | 120.063 |
| Skewness | | 1.074 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | .334 |
| Std. Error of Kurtosis | | .427 |
| Range | | 44.40 |
| Minimum | | 3.81 |
| Maximum | | 48.21 |
| Percentiles | 25 | 8.4528 |
| | 50 | 13.4163 |
| | 75 | 23.8344 |



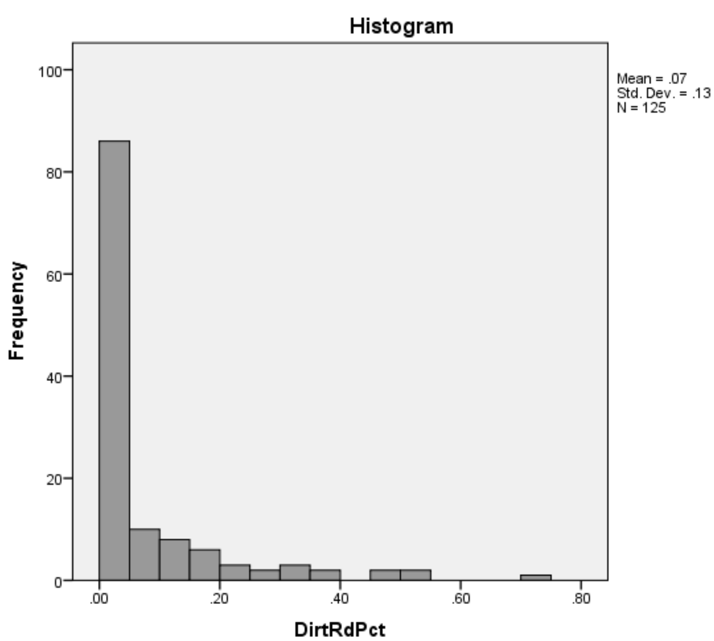
### *Slope on Roads*

| Statistics | | |
|---|---|---|
| DegreeSlopeRoads | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 17.1231 |
| Std. Error of Mean | | 1.01955 |
| Std. Deviation | | 11.48974 |
| Variance | | 132.014 |
| Skewness | | 1.124 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | .510 |
| Std. Error of Kurtosis | | .427 |
| Range | | 47.61 |
| Minimum | | 3.24 |
| Maximum | | 50.85 |
| Percentiles | 25 | 8.0759 |
| | 50 | 13.6938 |
| | 75 | 23.0237 |



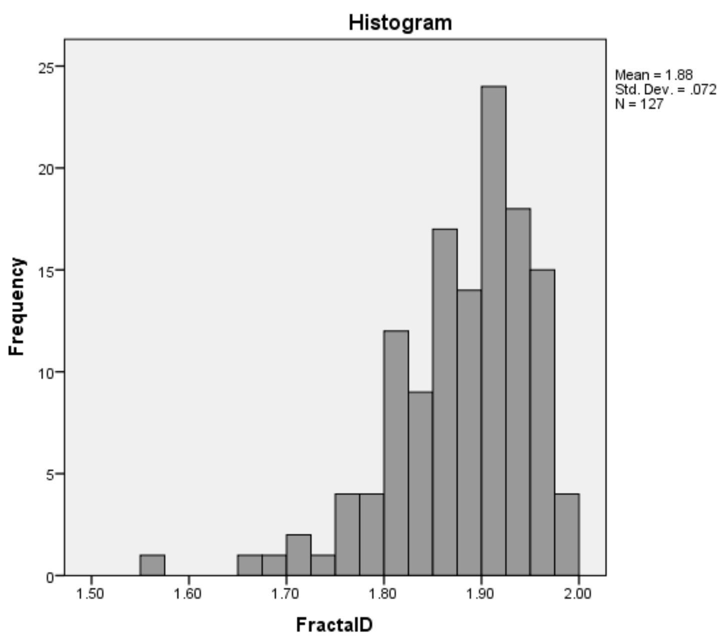
### *Dirt Road Percent*

| Statistics | | |
|---|---|---|
| DirtRdPct | | |
| N | Valid | 125 |
| | Missing | 2 |
| Mean | | .0723 |
| Std. Error of Mean | | .01162 |
| Std. Deviation | | .12995 |
| Variance | | .017 |
| Skewness | | 2.508 |
| Std. Error of Skewness | | .217 |
| Kurtosis | | 6.742 |
| Std. Error of Kurtosis | | .430 |
| Range | | .71 |
| Minimum | | .00 |
| Maximum | | .71 |
| Percentiles | 25 | .0001 |
| | 50 | .0089 |
| | 75 | .0816 |



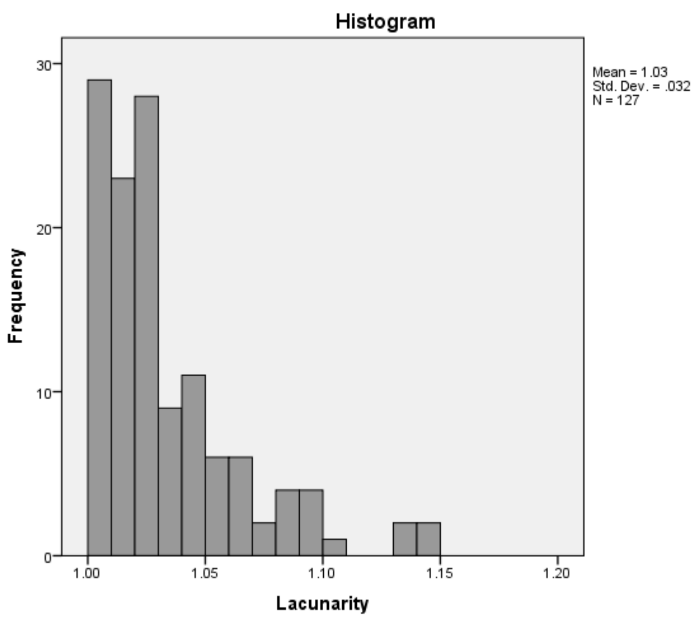
***Fractal Dimension (D)***

| Statistics | | |
|---|---|---|
| FractalD | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 1.8811 |
| Std. Error of Mean | | .00641 |
| Std. Deviation | | .07220 |
| Variance | | .005 |
| Skewness | | -1.309 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | 2.841 |
| Std. Error of Kurtosis | | .427 |
| Range | | .44 |
| Minimum | | 1.56 |
| Maximum | | 2.00 |
| Percentiles | 25 | 1.8399 |
| | 50 | 1.8950 |
| | 75 | 1.9366 |



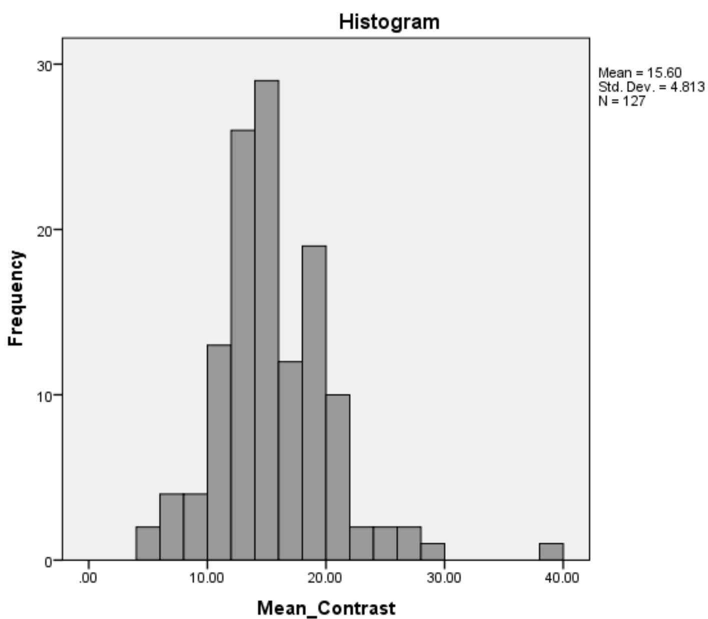
***Lacunarity***

| Statistics | | |
|---|---|---|
| Lacunarity | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 1.0332 |
| Std. Error of Mean | | .00281 |
| Std. Deviation | | .03172 |
| Variance | | .001 |
| Skewness | | 1.675 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | 2.850 |
| Std. Error of Kurtosis | | .427 |
| Range | | .15 |
| Minimum | | 1.00 |
| Maximum | | 1.15 |
| Percentiles | 25 | 1.0113 |
| | 50 | 1.0230 |
| | 75 | 1.0457 |



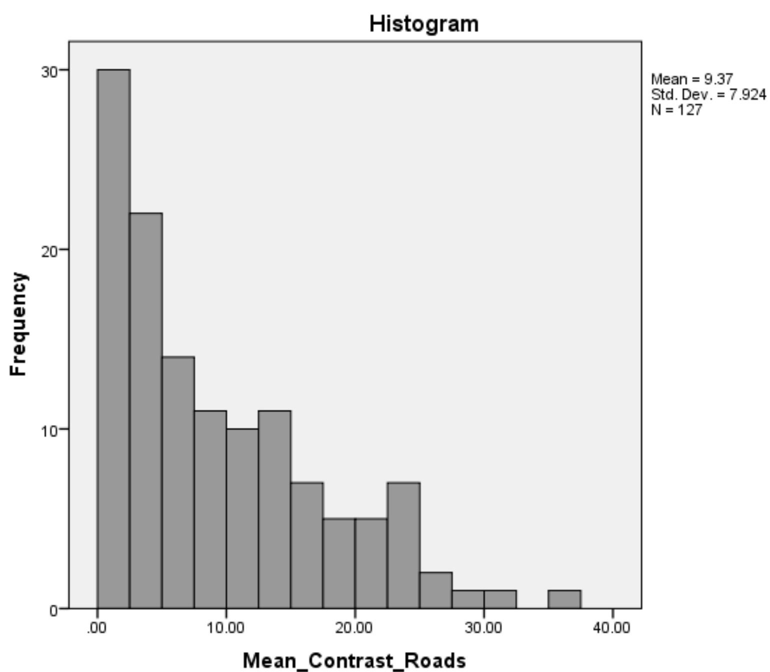
### *GLCM Contrast*

| Statistics | | |
|---|---|---|
| Mean_Contrast | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 15.6028 |
| Std. Error of Mean | | .42706 |
| Std. Deviation | | 4.81275 |
| Variance | | 23.163 |
| Skewness | | .967 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | 3.596 |
| Std. Error of Kurtosis | | .427 |
| Range | | 34.06 |
| Minimum | | 4.31 |
| Maximum | | 38.37 |
| Percentiles | 25 | 12.9139 |
| | 50 | 14.7534 |
| | 75 | 18.4629 |



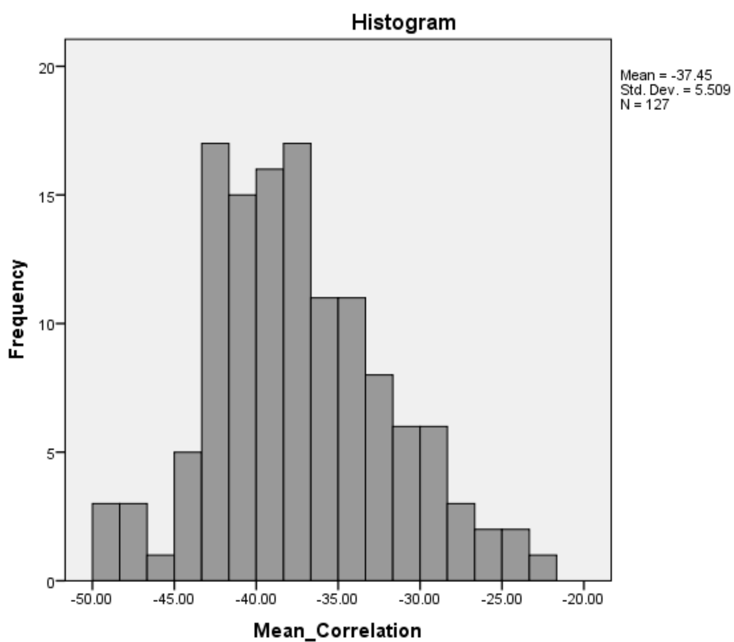
### *Contrast on Roads*

| Statistics | | |
|---|---|---|
| Mean_Contrast_Roads | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 9.3652 |
| Std. Error of Mean | | .70318 |
| Std. Deviation | | 7.92448 |
| Variance | | 62.797 |
| Skewness | | 1.022 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | .387 |
| Std. Error of Kurtosis | | .427 |
| Range | | 35.76 |
| Minimum | | .45 |
| Maximum | | 36.21 |
| Percentiles | 25 | 2.5305 |
| | 50 | 6.7198 |
| | 75 | 14.5160 |



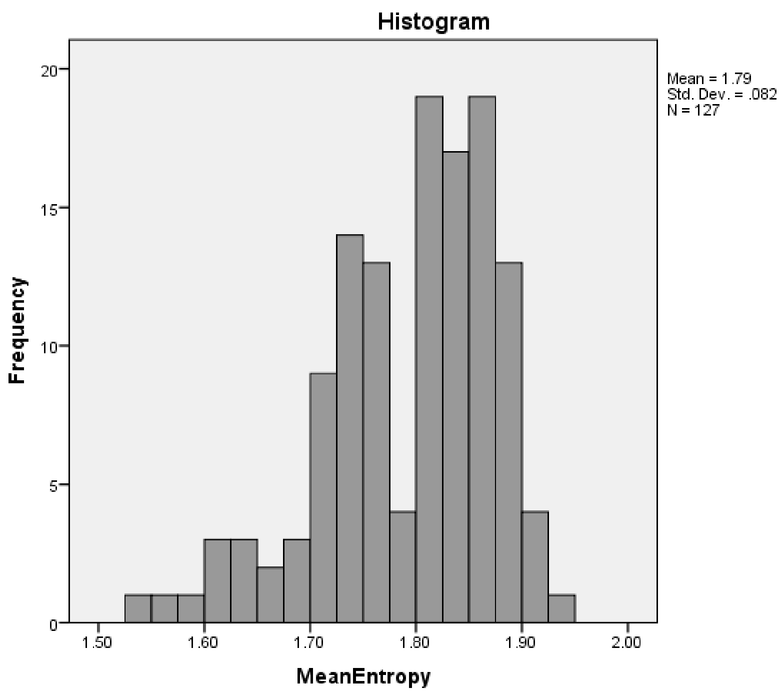
### *GLCM Correlation*

| Statistics | | |
|---|---|---|
| Mean_Correlation | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | -37.4517 |
| Std. Error of Mean | | .48886 |
| Std. Deviation | | 5.50914 |
| Variance | | 30.351 |
| Skewness | | .286 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | -.123 |
| Std. Error of Kurtosis | | .427 |
| Range | | 26.64 |
| Minimum | | -49.69 |
| Maximum | | -23.05 |
| Percentiles | 25 | -41.3726 |
| | 50 | -38.0030 |
| | 75 | -33.8618 |



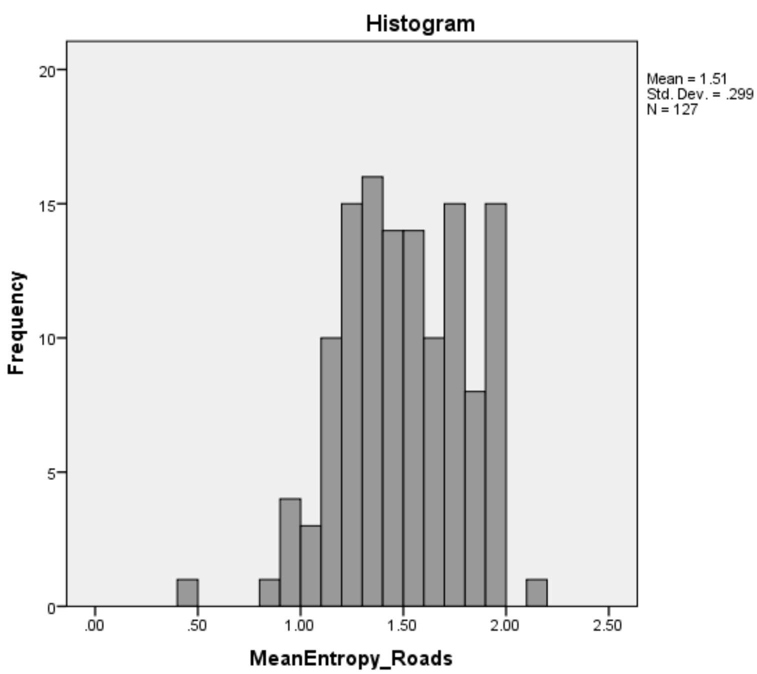
***Entropy***

| Statistics | | |
|---|---|---|
| MeanEntropy | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 1.7936 |
| Std. Error of Mean | | .00726 |
| Std. Deviation | | .08178 |
| Variance | | .007 |
| Skewness | | -.846 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | .350 |
| Std. Error of Kurtosis | | .427 |
| Range | | .39 |
| Minimum | | 1.54 |
| Maximum | | 1.93 |
| Percentiles | 25 | 1.7362 |
| | 50 | 1.8154 |
| | 75 | 1.8544 |



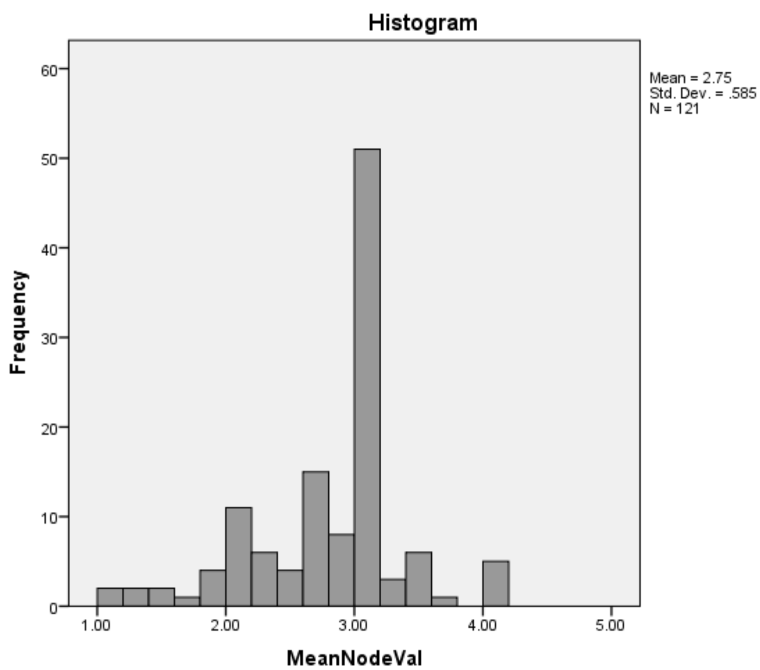
## *Entropy on Roads*

| Statistics | | |
|---|---|---|
| MeanEntropy_Roads | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 1.5065 |
| Std. Error of Mean | | .02653 |
| Std. Deviation | | .29896 |
| Variance | | .089 |
| Skewness | | -.286 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | -.100 |
| Std. Error of Kurtosis | | .427 |
| Range | | 1.63 |
| Minimum | | .49 |
| Maximum | | 2.12 |
| Percentiles | 25 | 1.2696 |
| | 50 | 1.4979 |
| | 75 | 1.7549 |



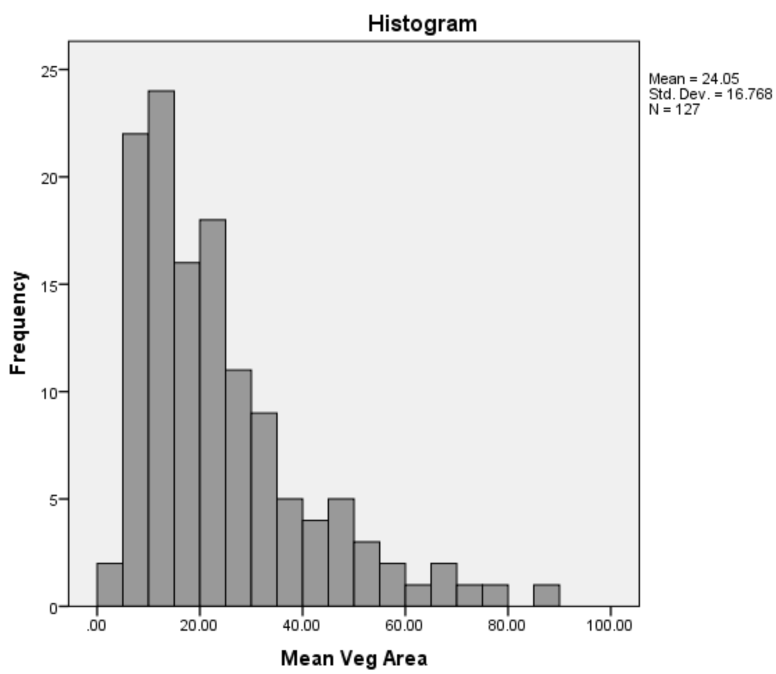
***Node Valence***

| Statistics | | |
|---|---|---|
| MeanNodeVal | | |
| N | Valid | 121 |
| | Missing | 6 |
| Mean | | 2.7547 |
| Std. Error of Mean | | .05317 |
| Std. Deviation | | .58491 |
| Variance | | .342 |
| Skewness | | -.664 |
| Std. Error of Skewness | | .220 |
| Kurtosis | | .989 |
| Std. Error of Kurtosis | | .437 |
| Range | | 3.00 |
| Minimum | | 1.00 |
| Maximum | | 4.00 |
| Percentiles | 25 | 2.5147 |
| | 50 | 3.0000 |
| | 75 | 3.0000 |



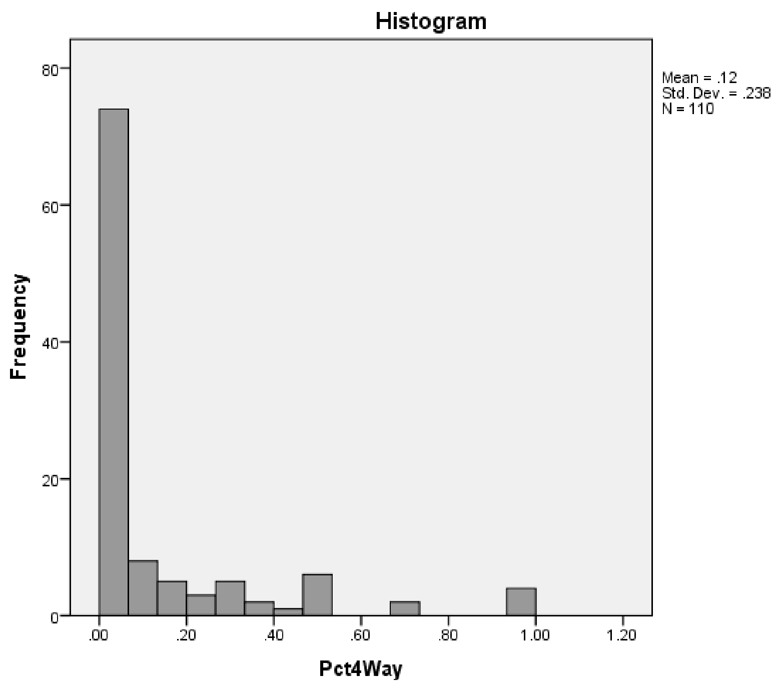
### *Vegetation Patch Size (Area)*

| Statistics | | |
|---|---|---|
| Mean Veg Area | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 24.0461 |
| Std. Error of Mean | | 1.48793 |
| Std. Deviation | | 16.76815 |
| Variance | | 281.171 |
| Skewness | | 1.408 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | 1.906 |
| Std. Error of Kurtosis | | .427 |
| Range | | 82.31 |
| Minimum | | 4.22 |
| Maximum | | 86.52 |
| Percentiles | 25 | 12.1448 |
| | 50 | 19.5893 |
| | 75 | 31.3453 |



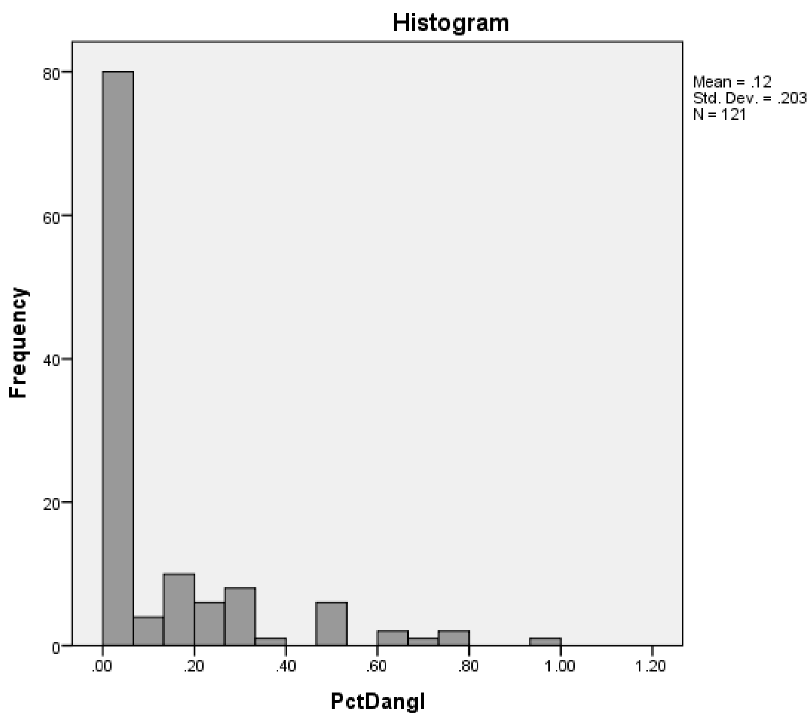
### *4-Way Intersections*

| Statistics | | |
|---|---|---|
| Pct4Way | | |
| N | Valid | 110 |
| | Missing | 17 |
| Mean | | .1243 |
| Std. Error of Mean | | .02271 |
| Std. Deviation | | .23823 |
| Variance | | .057 |
| Skewness | | 2.284 |
| Std. Error of Skewness | | .230 |
| Kurtosis | | 5.023 |
| Std. Error of Kurtosis | | .457 |
| Range | | 1.00 |
| Minimum | | .00 |
| Maximum | | 1.00 |
| Percentiles | 25 | .0000 |
| | 50 | .0000 |
| | 75 | .1488 |



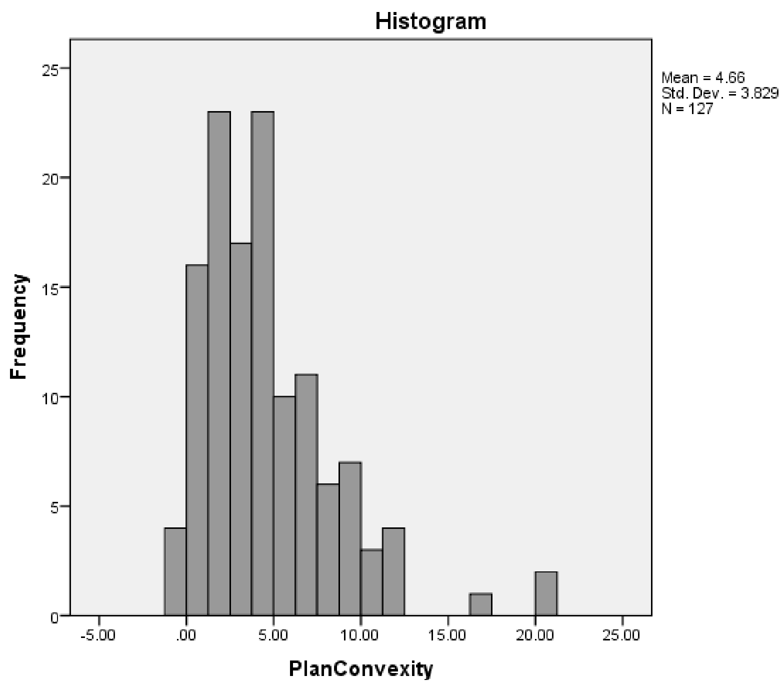
### *Dangle Ratio*

| Statistics | | |
|---|---|---|
| PctDangl | | |
| N | Valid | 121 |
| | Missing | 6 |
| Mean | | .1164 |
| Std. Error of Mean | | .01843 |
| Std. Deviation | | .20275 |
| Variance | | .041 |
| Skewness | | 1.986 |
| Std. Error of Skewness | | .220 |
| Kurtosis | | 3.816 |
| Std. Error of Kurtosis | | .437 |
| Range | | 1.00 |
| Minimum | | .00 |
| Maximum | | 1.00 |
| Percentiles | 25 | .0000 |
| | 50 | .0000 |
| | 75 | .2000 |



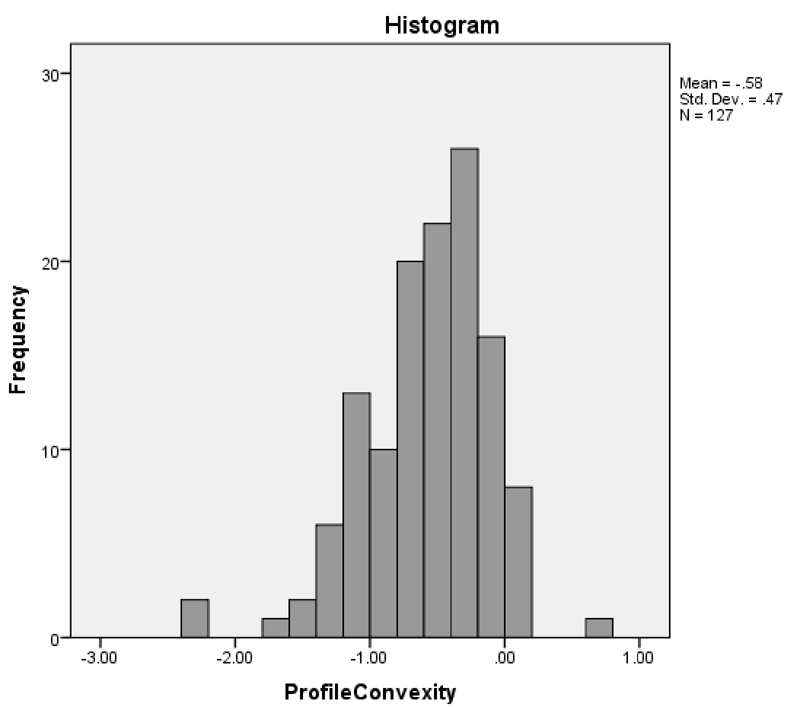
***Plan Convexity***

| Statistics | | |
|---|---|---|
| PlanConvexity | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 4.6628 |
| Std. Error of Mean | | .33975 |
| Std. Deviation | | 3.82879 |
| Variance | | 14.660 |
| Skewness | | 1.645 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | 4.176 |
| Std. Error of Kurtosis | | .427 |
| Range | | 21.65 |
| Minimum | | -.68 |
| Maximum | | 20.98 |
| Percentiles | 25 | 1.8996 |
| | 50 | 4.1526 |
| | 75 | 6.4428 |



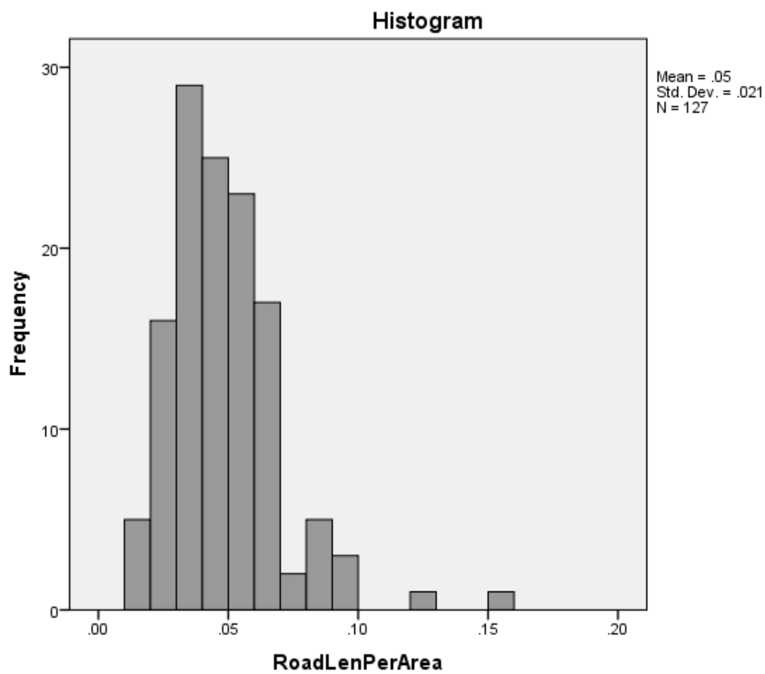
### *Profile Convexity*

| Statistics | | |
|---|---|---|
| ProfileConvexity | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | -.5783 |
| Std. Error of Mean | | .04169 |
| Std. Deviation | | .46979 |
| Variance | | .221 |
| Skewness | | -.789 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | 1.676 |
| Std. Error of Kurtosis | | .427 |
| Range | | 3.04 |
| Minimum | | -2.31 |
| Maximum | | .73 |
| Percentiles | 25 | -.8558 |
| | 50 | -.5283 |
| | 75 | -.2531 |



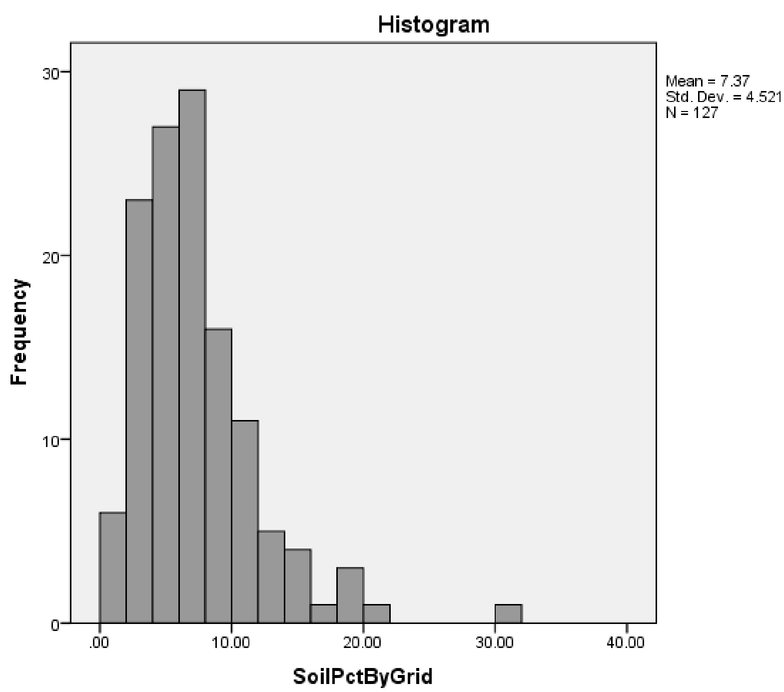
### *Road Density*

| Statistics | | |
|---|---|---|
| RoadLenPerArea | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | .0484 |
| Std. Error of Mean | | .00188 |
| Std. Deviation | | .02119 |
| Variance | | .000 |
| Skewness | | 1.714 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | 5.729 |
| Std. Error of Kurtosis | | .427 |
| Range | | .14 |
| Minimum | | .01 |
| Maximum | | .16 |
| Percentiles | 25 | .0355 |
| | 50 | .0441 |
| | 75 | .0586 |



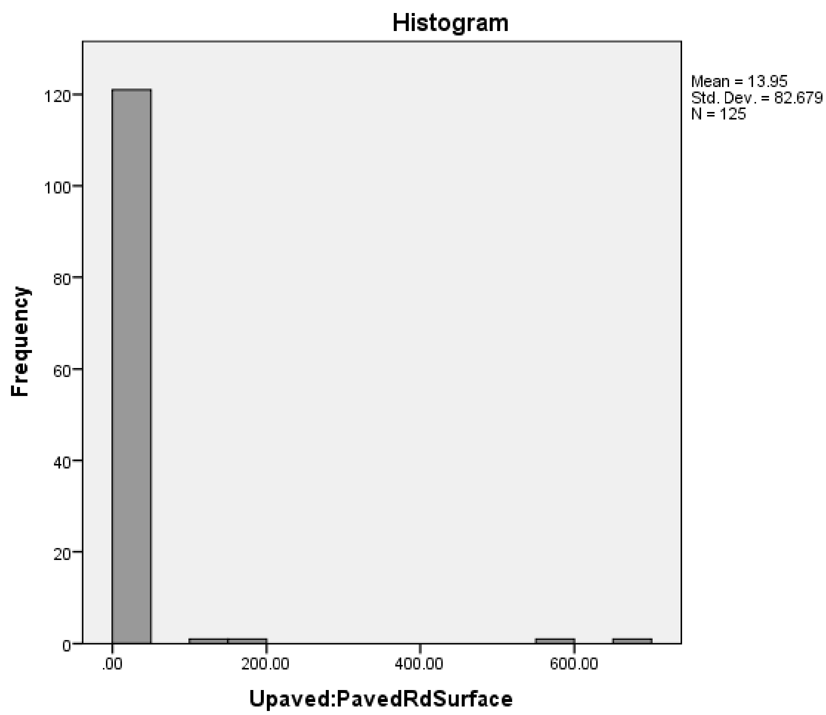
### *Soil Percent*

| Statistics | | |
|---|---|---|
| SoilPctByGrid | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 7.3742 |
| Std. Error of Mean | | .40120 |
| Std. Deviation | | 4.52131 |
| Variance | | 20.442 |
| Skewness | | 1.824 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | 5.587 |
| Std. Error of Kurtosis | | .427 |
| Range | | 29.94 |
| Minimum | | .90 |
| Maximum | | 30.84 |
| Percentiles | 25 | 4.2448 |
| | 50 | 6.6480 |
| | 75 | 9.3461 |



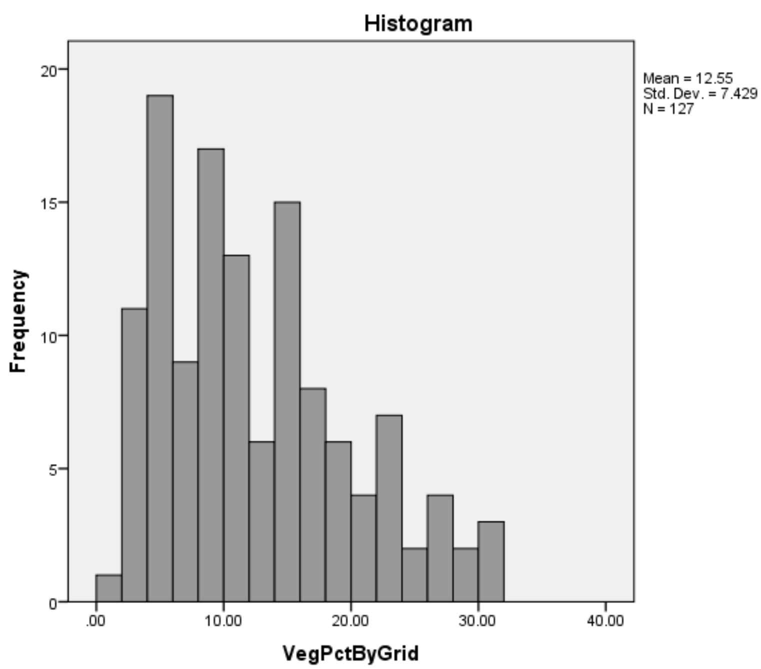
***Unpaved to Paved Road Ratio***

| Statistics | | |
|---|---|---|
| Upaved:PavedRdSurface | | |
| N | Valid | 125 |
| | Missing | 2 |
| Mean | | 13.9481 |
| Std. Error of Mean | | 7.39504 |
| Std. Deviation | | 82.67910 |
| Variance | | 6835.833 |
| Skewness | | 7.218 |
| Std. Error of Skewness | | .217 |
| Kurtosis | | 53.206 |
| Std. Error of Kurtosis | | .430 |
| Range | | 676.50 |
| Minimum | | .00 |
| Maximum | | 676.50 |
| Percentiles | 25 | .0000 |
| | 50 | .0351 |
| | 75 | .9668 |



### *Vegetation Percent*

| Statistics | | |
|---|---|---|
| VegPctByGrid | | |
| N | Valid | 127 |
| | Missing | 0 |
| Mean | | 12.5475 |
| Std. Error of Mean | | .65918 |
| Std. Deviation | | 7.42858 |
| Variance | | 55.184 |
| Skewness | | .703 |
| Std. Error of Skewness | | .215 |
| Kurtosis | | -.272 |
| Std. Error of Kurtosis | | .427 |
| Range | | 30.60 |
| Minimum | | 1.19 |
| Maximum | | 31.79 |
| Percentiles | 25 | 6.0017 |
| | 50 | 11.0610 |
| | 75 | 17.1744 |

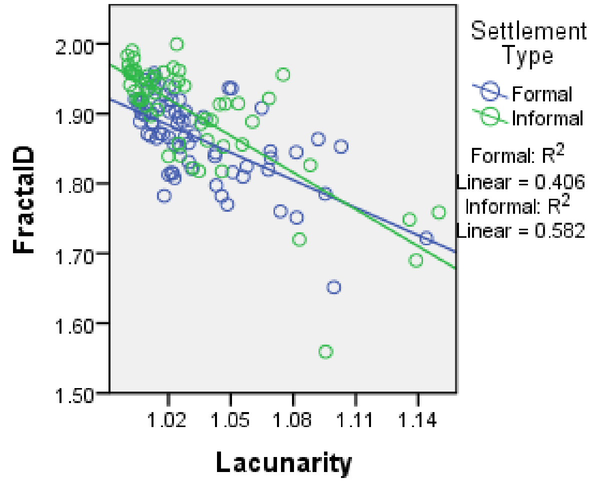


# APPENDIX B – BIVARIATE CORRELATIONS FOR ALL VARIABLES

The graphics in this appendix evaluate bivariate correlations prior to inclusion in the model. Pearson's r (correlation coefficient) for each variable pair is reported in order to help determine whether any of the variables should be excluded from the model due to prior correlation. There were two ways of testing statistical significance used in this research to improve model performance and exclude co-variates. One is the t-test of means at the settlement level. This treats the mean for each settlement as a sample, so there are 6 informal and 6 formal samples. The second way of evaluating significance was the Pearson *r* where all 127 randomly selected sample grid cells represented the sample set. Then, Spearman's $\rho$ was computed on grouped data to determine whether correlation between the variables differed by group (e.g., informal or formal). The scatterplots show the fitting of a regression line for the grouped data in order to evaluate the distributions by group, and where further explanation was needed, histograms provided additional visualization of the distributions.

## FractalD and Lacunarity

FractalD and Lacunarity are negatively correlated with a Pearson *r* of -0.699 for the un-grouped population. Fitting a linear regression line for each group produces a Formal $R^2$ of 0.406 and Informal $R^2$ of 0.582. A t-test of independent means with settlements as samples yields (p)=0.07 for Fractal Dimension and (p)=0.06 for Lacunarity. The discriminating power of Lacunarity at the settlement level was slightly higher than Fractal D, so Lacunarity was chosen for model inclusion. Additionally, Lacunarity was well-cited in the literature as a good differentiator of informal settlements.



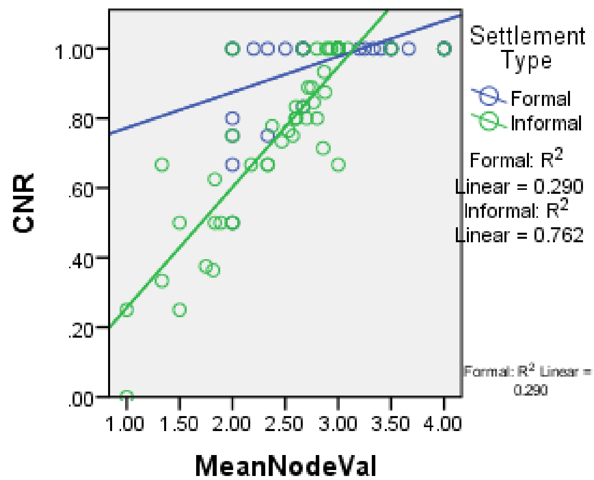

**Correlations**

| Settlement Type | | | | FractalDimension | Lacunarity |
|---|---|---|---|---|---|
| Spearman's rho | Formal | FractalDimension | Correlation Coefficient | 1.000 | -.591** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 69 | 69 |
| | | Lacunarity | Correlation Coefficient | -.591** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 69 | 69 |
| | Informal | FractalDimension | Correlation Coefficient | 1.000 | -.718** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 58 | 58 |
| | | Lacunarity | Correlation Coefficient | -.718** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 58 | 58 |

**. Correlation is significant at the 0.01 level (2-tailed).

### Connected Node Ratio (CNR) and Mean Node Valence

Connected Node Ratio (CNR) and Mean Node Valence are positively correlated with a Pearson *r* of 0.794 for the un-grouped population. Fitting a linear regression line for each group produces a Formal $R^2$ of 0.29 and a much higher Informal $R^2$ of 0.762. At the settlement level, the t-test of means for CNR produces a (p) value of 0.0082 while Mean Node Valence was not significant ((p)=0.166), therefore CNR was selected and Mean Node Valence was dropped from the model.



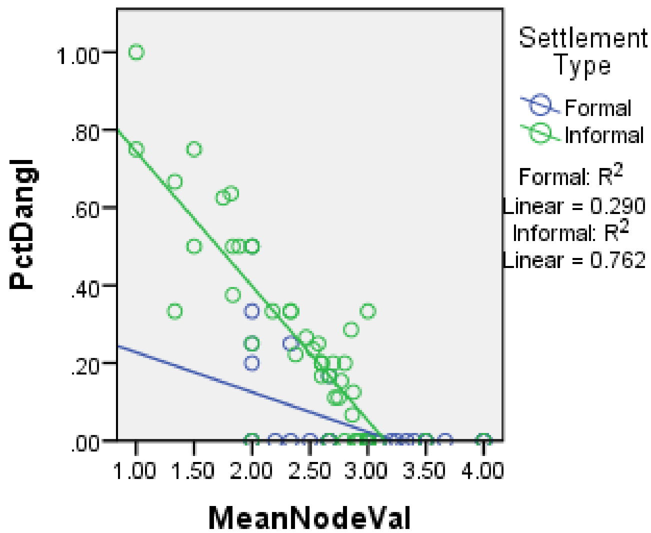

**Correlations**

| | Settlement Type | | | CNR | MeanNodeValence |
|---|---|---|---|---|---|
| Spearman's rho | Formal | CNR | Correlation Coefficient | 1.000 | .513** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 64 | 64 |
| | | MeanNodeValence | Correlation Coefficient | .513** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 64 | 64 |
| | Informal | CNR | Correlation Coefficient | 1.000 | .870** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 57 | 57 |
| | | MeanNodeValence | Correlation Coefficient | .870** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 57 | 57 |

**. Correlation is significant at the 0.01 level (2-tailed).

### Dangle Ratio and Mean Node Valence

Dangle Ratio and Mean Node Valence were negatively correlated with a Pearson *r* of -0.794 for the un-grouped population. This is an expected relationship since the more dangles there are (dead end streets) in a sample, the less will be the mean number of roads intersecting at any given node. Fitting a linear regression line for each group produces a Formal $R^2$ of 0.29 and a much higher Informal $R^2$ of 0.762. At the settlement level, the t-test of means for Dangle Ratio had a (p) value of 0.0052 while Mean Node Valence was not significant. Mean Node Valence was already dropped based on its correlation with Connected Node Ratio, therefore Dangle Ratio was retained.



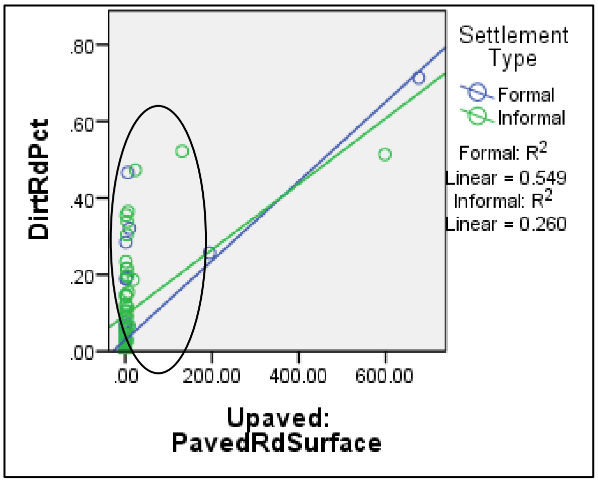

**Correlations**

| | Settlement Type | | | PctDangles | MeanNodeValence |
|---|---|---|---|---|---|
| Spearman's rho | Formal | PctDangles | Correlation Coefficient | 1.000 | -.404** |
| | | | Sig. (2-tailed) | . | .001 |
| | | | N | 64 | 64 |
| | | MeanNodeValence | Correlation Coefficient | -.404** | 1.000 |
| | | | Sig. (2-tailed) | .001 | . |
| | | | N | 64 | 64 |
| | Informal | PctDangles | Correlation Coefficient | 1.000 | -.864** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 57 | 57 |
| | | MeanNodeValence | Correlation Coefficient | -.864** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 57 | 57 |

**. Correlation is significant at the 0.01 level (2-tailed).

### Percent Dirt Roads and Unpaved-Paved Road Ratio

Percent Dirt Road surface is positively correlated with Unpaved-Paved Road Ratio as evidenced by a Pearson *r* of 0.612 for the un-grouped population. This relationship is understood from the context of having more road surface covered by dirt also impacting the ratio of dirt roads to asphalt roads. Fitting a linear regression line for each group produces a Formal $R^2$ of 0.549 and a lower Informal $R^2$ of 0.26. The unpaved-paved road surface ratio contained several distinct outliers that impact this comparison, seen in the following scatter plot. At the settlement level, the t-test of means for Percent Dirt Road had a (p) value of 0.35 while Unpaved-Paved Road Ratio also did not exhibit a significant difference between the informal and formal means ((p) value of 0.337). The two scatterplots show the outliers in the Unpaved:Paved Road Surface ratio, and the histograms demonstrate neither variable is normally distributed. Therefore both of these variables were excluded from the model.



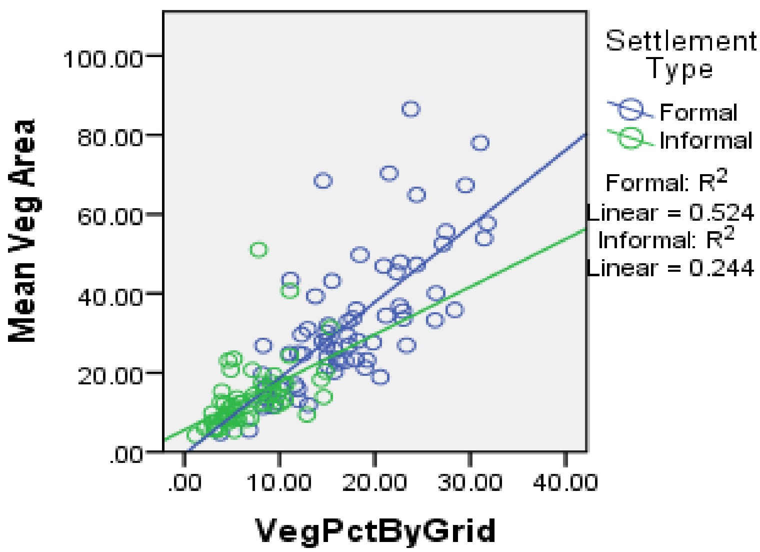

**Correlations**

| | Settlement Type | | | DirtRdPct | Upaved: PavedRdSurface |
|---|---|---|---|---|---|
| Spearman's rho | Formal | DirtRdPct | Correlation Coefficient | 1.000 | .994** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 68 | 68 |
| | | Upaved:PavedRdSurface | Correlation Coefficient | .994** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 68 | 68 |
| | Informal | DirtRdPct | Correlation Coefficient | 1.000 | .805** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 57 | 57 |
| | | Upaved:PavedRdSurface | Correlation Coefficient | .805** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 57 | 57 |

**. Correlation is significant at the 0.01 level (2-tailed).

**Mean Vegetation Patch Area and Vegetation Percent**

Mean Vegetation Patch Area represents the mean area in $m^2$ for vegetation patches intersecting each sample. It is highly positively correlated with Vegetation Percent (Pearson *r* of 0.796 for the un-grouped population, with similar high correlation for the formal and informal grid samples). Fitting a linear regression line for each group produces a Formal $R^2$ of 0.524 and a lower Informal $R^2$ of 0.244. At the settlement level, the t-test of means for Mean Vegetation Patch Area ((p) value of 0.08) was less significant than Vegetation Percent ((p) value of 0.0006) in differentiating between formal and informal means. Vegetation Percent from the NDVI band ratio is a common remote sensing metric developed from a ratio of a multispectral image's red and near infra-red bands. As a purely pixel-based metric, it is simple to calculate. In comparison, Mean Vegetation Patch Area requires creating polygons from clusters of pixels and opens up rich possibilities to measure other shape-dependent values such as feature size and compactness. In this case, shape provides an additional level of detail needed for an object-based approach. Vegetation Percent will therefore be excluded from the model, and Mean Vegetation Area will be included instead.



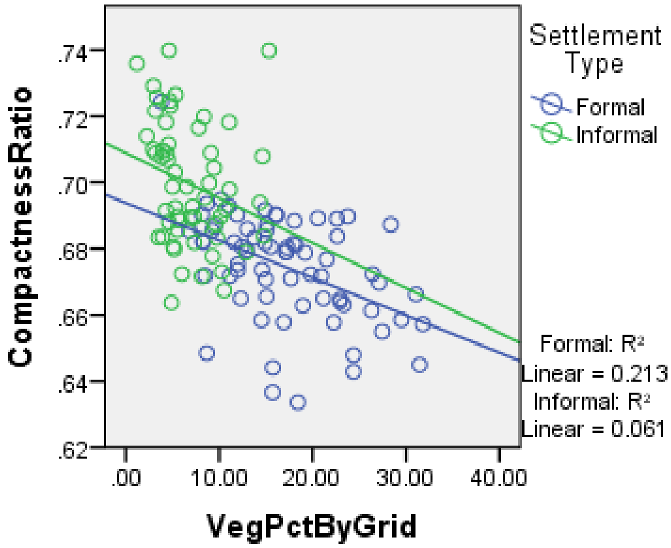
**Correlations**

| Settlement Type | | | | VegPctByGrid | Mean Veg Area |
|---|---|---|---|---|---|
| Spearman's rho | Formal | VegPctByGrid | Correlation Coefficient | 1.000 | .735** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 69 | 69 |
| | | Mean Veg Area | Correlation Coefficient | .735** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 69 | 69 |
| | Informal | VegPctByGrid | Correlation Coefficient | 1.000 | .644** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 58 | 58 |
| | | Mean Veg Area | Correlation Coefficient | .644** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 58 | 58 |

**. Correlation is significant at the 0.01 level (2-tailed).

### Mean Compactness Ratio and Vegetation Percent

Mean Compactness Ratio and Vegetation Percent are highly negatively correlated, with a Pearson *r* of -0.621 for the un-grouped population – the more vegetation in a sample, the less compact (or circular) will be its shape. Fitting a linear regression line for each group produces a Formal $R^2$ of 0.213 and a lower Informal $R^2$ of 0.06. At the settlement level, the t-test of means for Vegetation Patch Compactness Ratio with a (p) value of $<0.00008$ was a slightly stronger differentiator between the informal and formal means than the Vegetation Percent with a (p) value of 0.0006. As mentioned previously the added object-based information provided by the shape metric of compactness can provide additional meaning that can shed light on the reason for the vegetation. For example, elongated vegetation shape may indicate it was planted and is maintained to delineate property boundaries or as ornamental greenery in neat rows. On the other hand, a compact shape could indicate the vegetation was leftover as uncultivated or untended weeds or shrubs when vacant land continued to densify with additional dwellings as occurs somewhat randomly in informal settlements. Vegetation Percent was already excluded from the model so Compactness Ratio is retained.



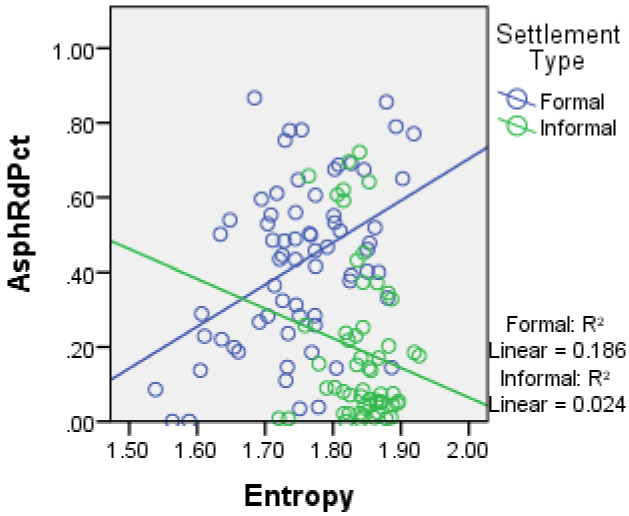

**Correlations**

| | Settlement Type | | | VegPctByGrid | Compactness Ratio |
|---|---|---|---|---|---|
| Spearman's rho | Formal | VegPctByGrid | Correlation Coefficient | 1.000 | -.465** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 69 | 69 |
| | | CompactnessRatio | Correlation Coefficient | -.465** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 69 | 69 |
| | Informal | VegPctByGrid | Correlation Coefficient | 1.000 | -.351** |
| | | | Sig. (2-tailed) | . | .007 |
| | | | N | 58 | 58 |
| | | CompactnessRatio | Correlation Coefficient | -.351** | 1.000 |
| | | | Sig. (2-tailed) | .007 | . |
| | | | N | 58 | 58 |

**. Correlation is significant at the 0.01 level (2-tailed).

### Asphalt Road Content and Entropy Texture Measure

Asphalt Roads are found less often in samples having higher Entropy. An Asphalt Road t-test of independent means at the settlement level has a (p) value of 0.037 - there is ~94% confidence the means are significantly different between formal and informal settlements. Using Spearman's $\rho$ the correlation between Asphalt Road Content and Entropy is significant at the 0.05 level. These two measures are derived from different methods. Asphalt Road percent from spectral properties of bands 1, 2, 3 & 4 of Quickbird imagery, while Entropy is from the GLCM Texture measure using only the Panchromatic Band. The grouped Spearman's $\rho$ is not significant for informal settlements but is positive and significant for the formal settlement type.




**Correlations**

| | Settlement Type | | | Entropy | AsphRdPct |
|---|---|---|---|---|---|
| Spearman's rho | Formal | Entropy | Correlation Coefficient | 1.000 | .335** |
| | | | Sig. (2-tailed) | . | .005 |
| | | | N | 69 | 68 |
| | | AsphRdPct | Correlation Coefficient | .335** | 1.000 |
| | | | Sig. (2-tailed) | .005 | . |
| | | | N | 68 | 68 |
| | Informal | Entropy | Correlation Coefficient | 1.000 | -.152 |
| | | | Sig. (2-tailed) | . | .258 |
| | | | N | 58 | 57 |
| | | AsphRdPct | Correlation Coefficient | -.152 | 1.000 |
| | | | Sig. (2-tailed) | .258 | . |
| | | | N | 57 | 57 |

**. Correlation is significant at the 0.01 level (2-tailed).

Despite the popularity of Entropy as a measure of informality of settlement structure, in this multivariate model, an assessment of multicollinearity resulting from ordinary linear regression reveals that Entropy suffers the highest variance inflation factor (VIF) of 5.4 compared to all remaining included variables, displayed in the following table.

**Coefficients[a]**

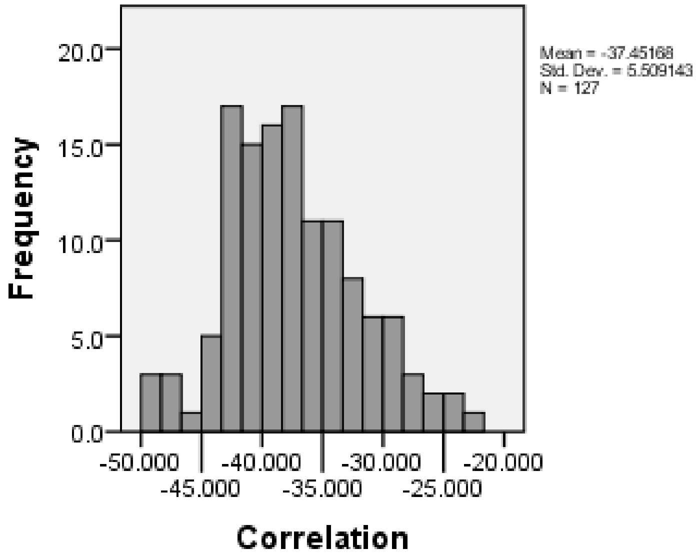
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Tolerance | VIF |
| 1 | (Constant) | 8.482 | 1.935 | | 4.383 | .000 | | |
| | Lacunarity | -.570 | .944 | -.034 | -.604 | .548 | .587 | 1.703 |
| | CNR | .131 | .157 | .049 | .834 | .406 | .541 | 1.849 |
| | Pct4WayIntersections | .132 | .096 | .063 | 1.378 | .172 | .884 | 1.132 |
| | AsphRdPct | -.004 | .130 | -.002 | -.033 | .974 | .450 | 2.222 |
| | RoadLenPerArea | -1.597 | 1.232 | -.068 | -1.296 | .198 | .661 | 1.512 |
| | CompactnessRatio | -2.696 | 1.486 | -.114 | -1.815 | .073 | .464 | 2.155 |
| | SoilPctByGrid | -.018 | .005 | -.167 | -3.214 | .002 | .681 | 1.467 |
| | VegPctByGrid | .027 | .004 | .392 | 6.248 | .000 | .465 | 2.150 |
| | Correlation | -.015 | .006 | -.163 | -2.325 | .022 | .374 | 2.675 |
| | Contrast | .022 | .007 | .201 | 3.125 | .002 | .441 | 2.266 |
| | Entropy | -2.694 | .616 | -.434 | -4.374 | .000 | .186 | 5.383 |
| | Entropy_Roads | -.501 | .135 | -.286 | -3.717 | .000 | .310 | 3.224 |
| | DegreeSlopeRoads | .000 | .002 | -.007 | -.131 | .896 | .743 | 1.346 |
| | ProfileConvexity | .145 | .048 | .138 | 3.019 | .003 | .880 | 1.137 |

a. Dependent Variable: SettleType

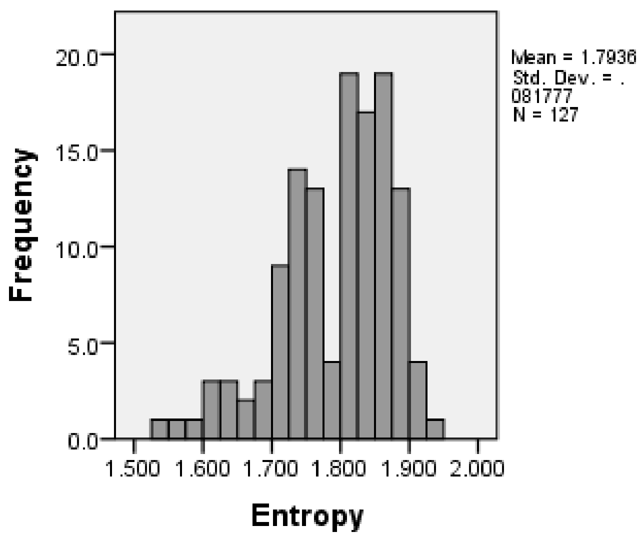
The VIF is a measure of how much the variance of the estimated regression coefficient is inflated by predictor variable correlation. (see: [http://online.stat.psu.edu/online/development/stat501/12multicollinearity/05multico_vif.html](http://online.stat.psu.edu/online/development/stat501/12multicollinearity/05multico_vif.html) ). Generally, a VIF of >4 is cause for concern. This VIF validates the decision to remove Entropy from the model.

### GLCM Correlation and Entropy

The Texture metric of GLCM Correlation was highly negatively correlated with Entropy exhibiting a Pearson *r* of -0.711. The following histogram shows the GLCM Correlation measure is normally distributed.




Entropy is also fairly normally distributed with slight negative skewness (sk = −0.846) in the following figure.




This relationship is not unexpected, since entropy is a measure of disorder in image tone, while GLCM correlation is a measure of linear dependence of grey levels on the grey levels of neighboring pixels in the panchromatic band. The t-test of independent means conducted at the settlement level reveals (p) = 0.035 for GLCM Correlation, with a very similar significance level for the texture metric of Entropy ((p)=0.039).

**Correlations**

| Settlement Type | | | | Correlation | Entropy |
|---|---|---|---|---|---|
| Spearman's rho | Formal | Correlation | Correlation Coefficient | 1.000 | -.735** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 69 | 69 |
| | | Entropy | Correlation Coefficient | -.735** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 69 | 69 |
| | Informal | Correlation | Correlation Coefficient | 1.000 | -.260* |
| | | | Sig. (2-tailed) | . | .049 |
| | | | N | 58 | 58 |
| | | Entropy | Correlation Coefficient | -.260* | 1.000 |
| | | | Sig. (2-tailed) | .049 | . |
| | | | N | 58 | 58 |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

From Spearman's ρ in the previous table, the texture measures of Entropy and GLCM Correlation are highly negatively correlated for Formal settlements but less (negatively) correlated for Informal settlements. Based on the prior decision to remove Entropy from the model due to its relationship with the Asphalt Roads measure, the texture measure of GLCM Correlation will be retained.

### Contrast and Entropy

Contrast (sum of squares variance) and Entropy are highly positively correlated ( r= 0.608). The t-test of means at the settlement level is less significant with (p)= 0.119 for Contrast but more significant ((p) = 0.039) for Entropy. At the sub-sampling level, Entropy and Contrast were also significantly correlated within groups according to Spearman's ρ in the following table.

**Correlations**

| Settlement Type | | | | Contrast | Entropy |
|---|---|---|---|---|---|
| Spearman's rho | Formal | Contrast | Correlation Coefficient | 1.000 | .637** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 69 | 69 |
| | | Entropy | Correlation Coefficient | .637** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 69 | 69 |
| | Informal | Contrast | Correlation Coefficient | 1.000 | .443** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 58 | 58 |
| | | Entropy | Correlation Coefficient | .443** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 58 | 58 |

**. Correlation is significant at the 0.01 level (2-tailed).

There was high correlation also between entropy and contrast on the un-grouped data:

**Correlations**

| | | | Entropy | Contrast |
|---|---|---|---|---|
| Spearman's rho | Entropy | Correlation Coefficient | 1.000 | .620** |
| | | Sig. (2-tailed) | . | .000 |
| | | N | 127 | 127 |
| | Contrast | Correlation Coefficient | .620** | 1.000 |
| | | Sig. (2-tailed) | .000 | . |
| | | N | 127 | 127 |

**. Correlation is significant at the 0.01 level (2-tailed).

The present research demonstrates similar results to other works referenced in the literature that show Entropy is a good differentiator of informal settlements. Due to the contribution to the scientific literature of evaluating alternate GLCM texture measures than Entropy for their performance in differentiating settlement type from imagery, Entropy has been removed from the model and GLCM Contrast is retained at this point.

### Contrast on Roads and Entropy on Roads

Following the same trend as the high correlation between Contrast and Entropy texture measures, Contrast on Roads is also highly positively correlated with Entropy on Roads ($r = 0.685$). At the settlement level the difference in means between formal and informal settlement types using the Contrast on Roads metric is very significant at (p) = 0.000053) while Entropy Roads at the settlement level is also very significant at (p) = 0.0004.

**Correlations**

| Settlement Type | | | | Contrast_Roads | Entropy_Roads |
|---|---|---|---|---|---|
| Spearman's rho | Formal | Contrast_Roads | Correlation Coefficient | 1.000 | .453** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 69 | 69 |
| | | Entropy_Roads | Correlation Coefficient | .453** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 69 | 69 |
| | Informal | Contrast_Roads | Correlation Coefficient | 1.000 | .662** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 58 | 58 |
| | | Entropy_Roads | Correlation Coefficient | .662** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 58 | 58 |

**. Correlation is significant at the 0.01 level (2-tailed).

The grouped data exhibit similar high correlation between Contrast on Roads and Entropy Roads (p=0.74 for ungrouped, (p)=0.66 for the Informal group and (p)=0.45 for the Formal group). Since Entropy has already been excluded from the variable set, robustness could be increased by retaining Entropy on Roads. Therefore Contrast on Roads will be excluded from the model.

**Correlations**

| | | | Contrast_Roads | Entropy_Roads |
|---|---|---|---|---|
| Spearman's rho | Contrast_Roads | Correlation Coefficient | 1.000 | .741** |
| | | Sig. (2-tailed) | . | .000 |
| | | N | 127 | 127 |
| | Entropy_Roads | Correlation Coefficient | .741** | 1.000 |
| | | Sig. (2-tailed) | .000 | . |
| | | N | 127 | 127 |

**. Correlation is significant at the 0.01 level (2-tailed).

### Degree Slope on Roads and Degree Slope

Degree Slope Roads and Degree Slope exhibit near perfect positive correlation because they are both derived from the same underlying elevation data (r = 0.988). At the settlement level, the slope on roads has slightly less discriminatory power ((p) = 0.065) than overall Degree Slope ((p) = <0.001), which is likely due to the reduced number of slope pixels used for roads compared to the number of underlying slope pixels needed for degree slope of the entire sample area. For the same reasons a similar high positive correlation also appears on the ungrouped data (Spearman's $\rho$ = 0.98).

**Correlations**

| | | | DegreeSlope | DegreeSlope Roads |
|---|---|---|---|---|
| Spearman's rho | DegreeSlope | Correlation Coefficient | 1.000 | .981** |
| | | Sig. (2-tailed) | . | .000 |
| | | N | 127 | 127 |
| | DegreeSlopeRoads | Correlation Coefficient | .981** | 1.000 |
| | | Sig. (2-tailed) | .000 | . |
| | | N | 127 | 127 |

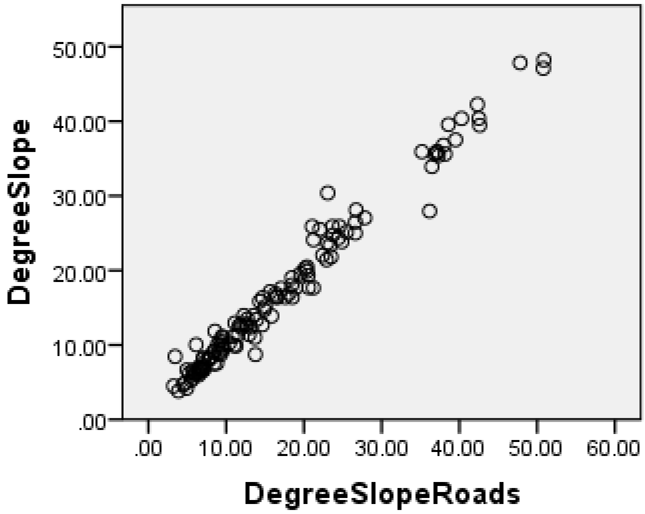
**. Correlation is significant at the 0.01 level (2-tailed).

To investigate further, Spearman's $\rho$ was also computed on the grouped data and also exhibited nearly perfect correlation (0.95 for Formal and 0.99 for Informal). Due to the greater discriminatory power of degree slope for the entire sample area, Degree Slope on Roads will be eliminated from the model.

**Correlations**

| | Settlement Type | | | DegreeSlope | DegreeSlope Roads |
|---|---|---|---|---|---|
| Spearman's rho | Formal | DegreeSlope | Correlation Coefficient | 1.000 | .953** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 69 | 69 |
| | | DegreeSlopeRoads | Correlation Coefficient | .953** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 69 | 69 |
| | Informal | DegreeSlope | Correlation Coefficient | 1.000 | .991** |
| | | | Sig. (2-tailed) | . | .000 |
| | | | N | 58 | 58 |
| | | DegreeSlopeRoads | Correlation Coefficient | .991** | 1.000 |
| | | | Sig. (2-tailed) | .000 | . |
| | | | N | 58 | 58 |

**. Correlation is significant at the 0.01 level (2-tailed).




### Summary of Variable Exclusions

Of the bivariate correlations listed, Fractal Dimension, Mean Node Valence, Vegetation Percent, Entropy, Contrast on Roads, and Degree Slope are the variables excluded because they are already correlated with other variables. For similar reasons, Mean Node Valence (number of roads joining at a nodal point), Unpaved-to-Paved Road Ratio and Percent Dirt Roads were eliminated.

# APPENDIX C - LACUNARITY RESULTS BY SETTLEMENT FOR ALL BOX SIZES

| IMAGE TYPE | Box Size | Settlement | Formal | Lacunarity Formal | Settlement | Informal | Lacunarity Informal |
|---|---|---|---|---|---|---|---|
| BINARY | 49 | Terrazas | F | 1.0762 | Satelite | I | 1.061 |
| BINARY | 49 | Balcones | F | 1.0461 | Berlin | I | 1.062 |
| BINARY | 49 | Mirador2 | F | 1.0515 | Joya | I | 1.1212 |
| BINARY | 49 | GTowns2 | F | 1.0666 | LoDeCoy | I | 1.0747 |
| BINARY | 49 | GTowns3 | F | 1.0737 | Peronia1 | I | 1.0417 |
| BINARY | 49 | Pinares | F | 1.1169 | Peronia2 | I | 1.0801 |
| BINARY | 100 | Terrazas | F | 1.0548 | Satelite | I | 1.0712 |
| BINARY | 100 | Balcones | F | 1.0323 | Berlin | I | 1.0593 |
| BINARY | 100 | Mirador2 | F | 1.0486 | Joya | I | 1.1152 |
| BINARY | 100 | GTowns2 | F | 1.055 | LoDeCoy | I | 1.074 |
| BINARY | 100 | GTowns3 | F | 1.0601 | Peronia1 | I | 1.0417 |
| BINARY | 100 | Pinares | F | 1.0878 | Peronia2 | I | 1.1067 |
| BINARY | 200 | Terrazas | F | 1.0447 | Satelite | I | 1.0884 |
| BINARY | 200 | Balcones | F | 1.022 | Berlin | I | 1.0593 |
| BINARY | 200 | Mirador2 | F | 1.0555 | Joya | I | 1.1229 |
| BINARY | 200 | GTowns2 | F | 1.0467 | LoDeCoy | I | 1.0587 |
| BINARY | 200 | GTowns3 | F | 1.0416 | Peronia1 | I | 1.0442 |
| BINARY | 200 | Pinares | F | 1.0737 | Peronia2 | I | 1.1436 |
| BINARY | 500 | Terrazas | F | 1.0658 | Satelite | I | 1.1011 |
| BINARY | 500 | Balcones | F | 1.0187 | Berlin | I | 1.0501 |
| BINARY | 500 | Mirador2 | F | 1.0917 | Joya | I | 1.1374 |
| BINARY | 500 | GTowns2 | F | 1.043 | LoDeCoy | I | 1.0064 |
| BINARY | 500 | GTowns3 | F | 1.0246 | Peronia1 | I | 1.072 |
| BINARY | 500 | Pinares | F | 1.0866 | Peronia2 | I | 1.1867 |

## REFERENCES

## REFERENCES


Alsayyad, N. (1993). Squatting and Culture - A Comparative Analysis of Informal Developments in Latin America and the Middle East. *Habitat International*, *17*(1), 33–44.

Aplin, P. (2003). Comparison of Simulated IKONOS and SPOT HRV imagery for classifying urban areas. *Remotely Sensed Cities*. London: Taylor & Francis.

Augustijn-Beckers, E.-W., Flacke, J., & Retsios, B. (2011). Simulating informal settlement growth in Dar es Salaam, Tanzania: An agent-based housing model. *Computers, Environment and Urban Systems*, *35*, 93–103.

Baltsavias, E. P., & Mason, S. O. (1997). Automated shack reconstruction using integration of cues in object space. Presented at the 3D Reconstruction and Modeling of Topographic Objects 32(Part 3-4W2), *IAPRS*: Stuttgart.

Barros, J., & Sobreira, F. (2008). City of Slums: self-organisation across scales. *Unifying Themes in Complex Systems IV*. Springer Berlin Heidelberg. Retrieved from http://www.springerlink.com/content/h6772q773g5p9485/

Barros, J. X. (2004, October). *Urban Growth in Latin American Cities, exploring urban dynamics through agent-based simulation.* University College, London, UK.

Barry, M., & Rüther, H. (2001). Data Collection and Management for Informal Settlements Upgrades. Presented at the International Conference on Spatial Information for Sustainable Development, Nairobi, Kenya.

Bartlett, II, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning, and Performance Journal*, *19*(1), 43–50.

Besussi, E., Chin, N., Batty, M., & Longley, P. (2010). The Structure and Form of Urban Settlements. *Remote Sensing of Urban and Suburban Areas*, Remote Sensing and Digital Image Processing (p. 16). Netherlands: Springer.

Beyer, H. L. (2004). *Hawth's Analysis Tools for ArcGIS*. Retrieved from http://www.spatialecology.com/htools

Bharathi, V. L., & Lakshmi, T. V. (2005). 31st International Symposium on Remote Sensing & Environment. Presented at the *Global Monitoring for Sustainability and Security, ISPRS*: St. Petersburg, Russia.

Bhatta, B., Saraswati, S., & Yopadhayay, B. D. (2010). 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*.

Bishop, I., Barry, M., Urquhart, K., Escobar, F., McPherson, E., & Nascarella, J. (2002). Meeting the Need for GIS Skills in Developing Countries: The Case of Informal Settlements. *Transactions in GIS*, *6*(3), 311–326.

Blaschke, T., & Lang, S. (2006). Object-Based Image Analysis for Automated Information Extraction - A Synthesis. Presented at the *MAPPS/ASPRS Conference*, San Antonio, TX.

Busgeeth, K., Brits, A., & Whisken, J. (2008). Potential Application of Remote Sensing in Monitoring Informal Settlements in Developing Countries where Complementary Data Does Not Exist. Presented at the *Planning Africa Conference*, Johannesburg, South Africa.

Cable, D. (2009, May). Partner for Surgery - Private Conversations during Medical Mission to Rural Guatemala, see: http://www.partnerforsurgery.org

Cao, J., Song, Z., Wang, Y., & Wang, T. (2010). Analysis and application of the Optimal Path based on ArcGIS Geodatabase' Network Model. *2010 International Conference on Environmental Science and Information Application Technology (ESIAT)* (Vol. 1, pp. 434–437), IEEE.

Cheng, Q. (1999). The gliding box method for multifractal modeling. *Computers & Geosciences*, *25*, 1073–1079.

Clausi, D. A., & Jernigan, M. E. (1998). A Fast Method to Determine Co-occurrence Texture Features. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(1), 298–300.

Cockings, S., Fisher, P. F., & Longford, M. (1997). Parameterization and Visualization of the Errors in Areal Interpolation. *Geographical Analysis*, *29*(4), 314–328.

Cooper, J. (2005). Assessing urban character; the use of fractal analysis of street edges. *Urban Morphology*, *9*(2), 95–107.

Cowen, D., & Jensen, J. (2001). Extraction and Modeling of Urban Attributes Using Remote Sensing Technology. *People and Pixels: Linking Remote Sensing and Social Science*. Washington, DC: National Academies Press.

Dare, P. M., & Fraser, C. S. (2001). Mapping informal settlements using high resolution satellite imagery. *International Journal of Remote Sensing*, *22*(8), 1399–1401.

Davis, M. (2007). *Planet of Slums*. Verso Books, London.

De Keersmaecker, M.L., Frankhauser, P., & Thomas, I. (2003). Using Fractal Dimensions for Characterizing Intra-urban Diversity: The Example of Brussels. *Geographical Analysis*, *35*(4), 310–328.

Declassified NIMA imagery of British Honduras and Guatemala. (1941, August 25). Film Canister 10201617, ON=043200, 1A-1057, Frame GM4-106, US National Archives and Records Administration, Silver Spring, MD.

Depraetere, C., & Riazanoff, S. (2004). The new Digital Elevation Model data set from the Shuttle Radar Topography Mission: Hydrogeomorphological applications in the Ohrid region. Presented at the BALWOIS Conference on Water Observation and Information Systems for Decision Support, Ohrid, Macedonia.

Diaz Martinez, Z., Fernandez Menendez, J., & Vargas, M. J. S. (2004). *See5 Algorithm versus Discriminant Analysis. An Application to the Prediction of Insolvency in Spanish Non-life Insurance Companies* (p. 21). Universidad Computense de Madrid.

Dickenson, J. P. (1996). *A Geography of the Third World*, Routledge Publishers.

Dill, J. (2004). Measuring Network Connectivity for Bicycling and Walking. Presented at the *Transportation Research Board Annual Meeting*, Washington, D.C.

Dubovyk, O., Sliuzas, R., & Flacke, J. (2011). Spatio-temporal modelling of informal settlement development in Sancaktepe district, Istanbul, Turkey. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*, 235–246.

El Ouardighi, A., El Akadi, A., & Aboutajdine, D. (2007). 3rd International Symposium on Computational Intelligence and Intelligent Informatics. Aqadir, Morocco: IEEE.

Environmental Systems Research Institute (ESRI), *ArcGIS Desktop*. (2009). Redlands, CA.

Epstein, J., Payne, K., & Kramer, E. (2002). Techniques for Mapping Suburban Sprawl. *Photogrammetric Engineering & Remote Sensing*, *63*(9), 913–918.

European Commission Joint Research Centre (JRC), Int'l Centre for Remote Sensing of Environment, & ISPRS (2009). *Proceedings, 33rd International Symposium on Remote Sensing of Environment: sustaining the millenium development goals,* May 4-9, 2009, Stresa, Italy.

Evans, I. S., & Cox, N. J. (1999). Relations between Land Surface Properties: Altitude, Slope and Curvature. Process Modelling and Landform Evolution, *Lecture Notes in Earth Sciences* (Vol. 78, pp. 13–45).

Evertsz, Rick, (2009). Retrieved from: http://rick.evertsz.name/365_moments/la_limonada.html on 18 April 2012. Photographed 29 May 2009.

Ewing, R., Pendall, R., & Chen, D. (2002). *Measuring Sprawl and Its Impact*. Smart Growth America, Washington, DC.

Ewing, R., Schmid, T., Killingsworth, T., Zlot, A., & Raudenbush, S. (2008). Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity. *Urban Ecology*. Springer.

Exelis, Inc., ENVI, (2009). Retrieved from http://www.exelisvis.com/ProductServices/ENVI.aspx

Fang, J., Liu, S., Hong, Y., & Qing, Z. (2007). Measuring urban sprawl in Beijing with geo-spatial indices. *Journal of Geographical Sciences*, 17(4), 469-478.

Filho, B., & Sobreira, F. (2007). A Multiscale Analysis of Socio-Spatial Patterns. Presented at the 10th International Conference on Computers in Urban Planning and Urban Management, Sao Carlos, Brazil.

Filho, B., & Sobreira, F. (2008). Accuracy of Lacunarity Algorithms in Texture Classification of High Spatial Resolution Images from Urban Areas. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Part B3b (Vol. 37). Beijing.

Filho, M. B., & Sobreira, F. (2005). *Assessing Texture Pattern in Slums Across Scales - An Unsupervised Approach* (Working Paper No. 87, p. 12). Center for Advanced Spatial Analysis, University College London.

Forsyth, A., Schmitz, K., Oakes, M., Zimmerman, J., & Koepp, J. (2006). Standards for Environmental Measurement Using GIS: Toward a Protocol for Protocols. *Journal of Physical Activity*, *3*(1), 241–257.

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*(3), 399–409.

Galeon, Florence (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. Vol. 37, Part B4, Beijing.

Galster, G., Hanson, R., Ratcliffe, M., Wolman, H., Coleman, S., & Freihage, J. (2001). Environmental Measurement Using GIS: Toward a Protocol for Protocols. *Housing Policy Debate*, *12*(4), 681–717.

Garson, G. D. (2012). Statnotes: Topics in Multivariate Analysis. Retrieved January 27, 2012, from http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm

Gilbert, A. (1996). *The mega-city in Latin America*. United Nations University Press: New York.

Glaeser, E., & Kahn, M. (2003). *Sprawl and Urban Growth* (Discussion Paper No. 2004). Harvard Institute of Economic Research, Cambridge, MA.

Gleijeses, P. (1992). *Shattered hope*. Princeton University Press.

Gong, P., & Xu, B. (2006). Contextual Image Analysis Methods for Urban Applications. In S. M. D. Jong & F. D. V. der Meer (Eds.), *Remote Sensing Image Analysis: Including The Spatial Domain*, Remote Sensing and Digital Image Processing (Vol. 5, pp. 137–152). Dordrecht: Springer Netherlands.

Griffin, E., & Ford, L. (1980). A Model of Latin American City Structure. *Geographical Review*, *70*(4), 397–422.

Griffith, D. A., & Amrhein, C. G. (1997). *Multivariate statistical analysis for geographers*. University of MI: Prentice Hall

Guisti de Pérez, R., & Pérez, R. (2008). *Analyzing Urban Poverty - GIS for the Developing World*. Redlands, CA: ESRI Press.

Hall, G. B., Malcolm, N. W., & Piwowar, J. M. (2001). Integration of Remote Sensing and GIS to Detect Pockets of Urban Poverty: The Case of Rosario, Argentina. *Transactions in GIS*, *5*(3), 235–253.

Haralick, R., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systmes, Man, and Cybernetics, 3*(6), 610–621.

Hasan, A. (2006). Orangi Pilot Project: the expansion of work beyond Orangi and the mapping of informal settlements and infrastructure. *Environment and Urbanization*, *18*(2), 451–480.

Hasse, J. (2002). *Geospatial Indices of Urban Sprawl in New Jersey* (Dissertation). Rutgers University, New Jersey.

Hasse, J., & Kornbluh, A. (2004). Measuring Accessibility as a Spatial Indicator of Sprawl. *Middle States Geographer*, *36*.

Hasse, J., & Lathrop, R. (2003). Land Resource Impact Indicators of Urban Sprawl. *Applied Geography*, *23*(2-3), 159–175.

Hay, G. J., & Castilla, G. (2006). Object-Based Image Analysis: Strengths, Weaknesses, Opportunities and Threats (SWOT). *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XXXVI*(4/C42). Retrieved from http://www.commission4.isprs.org/obia06/Papers/01_Opening%20Session/OBIA2006_Hay_Castilla.pdf

Hofmann, P. (2001). Detecting Urban Features from IKONOS Data Using an Object-Oriented Approach. In: RSPS 2001, Geomatics, Earth Observation and the Information Society.

Hofmann, P., Strobl, J., Blaschke, T., & Kux, H. (2008). Detecting informal settlements from Quickbird data in Rio de Janeiro using an object-based approach. *Object-based Image Analysis - Spatial Concepts for knowledge-driven Remote-Sensing Applications*. Springer.

Holz, R. K., & Huff, D. L. (1973). Urban spatial structure based on remote sensing imagery. *The Surveillant Science: Remote Sensing of the Environment* (pp. 375–380). Boston: Houghton-Mifflin.

Hoornweg, D., Ruiz Nunez, F., Freire, M., Palugyai, N., Villaveces, M., & Herrera, E. W. (2007). *City Indicators: Now to Nanjing* (Working Paper). World Bank. Retrieved from http://www.worldbank.icebox.ingenta.com.mutex.gmu.edu/search/article?index=7&value1=slums&option1=tka&sort=score&year_from=1900&year_to=2100&database=1&search_type=quick

Hurskainen, P., P., P. (2004). Change Detection of Informal Settlements using multi-temporal aerial photographs - the case of Voi, SE-Kenya. *Proceedings of the 5th AARSE conference, Nairobi, Kenya*. Retrieved from http://www.helsinki.fi/science/taita/reports/aarse_hurskainen.pdf

Ioannilli, M. & Paregiani, A. (2008). Automated Unsupervised Geomorphometric Classification of Earth Surface for Landslide Susceptibility Assessment, Lecture Notes in Computer Science, Vol. 5072:268-283.

Jacobsen, K., Nelson, J., & Owen, K. (2011, Spring). Measuring Access to Radio Health Communications in Rural Guatemala. *Global Studies Review*, *7*(7). Retrieved from http://www.globality-gmu.net/archives/2415

Jain, S. (2007). Use of IKONOS satellite data to identify informal settlements in Dehradun, India. *International Journal of Remote Sensing*, *28*(15), 3227–3233.

Jensen, J. R. (2005). *Introductory Digital Image Processing - a remote sensing perspective* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Jin, X., & Davis, C. (2005). Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural, Contextual, and Spectral Information. *EURASIP Journal on Applied Signal Processing*, *2005*(14), 2196–2206.

Junior, S., & Filho, M.B. (1997). *Enhancing Urban Analysis Through Lacunarity Multiscale Measurement*. London: Center for Advanced Spatial Analysis, University College London, Universidade Federal de Pernambuco Brazil.

Junior, S., & Filho, B. (2005). Multiscale measurements of fragmented cities: Enhancing urban analysis through lacunarity based measures. Presented at the 8th International Conference on Geocomputation, University of Michigan.

Karperien, A. (2007). *FracLac for ImageJ*. Plug-in for Image-J. Australia: Charles Stuart University. Retrieved from http://rsb.info.nih.gov/ij/plugins/fraclac/FLHelp/Introduction.htm

Kellett, P., & Napier, M. (1995). Squatter Architecture? A Critical Examination of Vernacular Theory and Spontaneious Settlement With Reference to South America and South Africa. *Traditional Dwellings and Settlements Review*, *6*(11), 7–24.

Klaric, Matt, Scott, Grant, Shyu, Chi-Ren, & Davis, Curt. (2005). Automated Object Extraction through Simplification of the Differential Morphological Profile for

High Resolution Satellite Imagery (Vol. 2, pp. 1265–1268). Presented at the Geoscience and Remote Sensing Symposium, IEEE.

Kruse, F. A., Lefkoff, J. B., Boardman, J. B., Heidebracht, A. T., Shapiro, A. T., Barloon, P. J., & Goetz, A. F. H. (1993). The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging spectrometer Data. *Remote Sensing of Environment*, *44*, 145–163.

Lang, S., Walz, U., Klug, H., Blaschke, T., & Syrbe, R.-U. (2009). Landscape Metrics - A toolbox for assessing past, present and future landscape structures. *Geoinformation Technologies for Geo-cultural Landscapes - European Perspectives*. London, UK: Taylor & Francis Group.

Lee, C.M., & Ahn, K.H. (2003). Is Kentlands better than Radburn? The American Garden City and the new urbanist paradigms. *Journal of the American Planning Association*, *69*(1), 50.

Lefèvre, S., Weber, J., & Sheeren, D. (2007). *Automatic building extraction in VHR images using advanced morphological operators*. Presented at the IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, Paris, France.

Lemma, T., Sliuzas, R., & Kuffer, M. (2006). *A Participatory Approach to Mapping Slum Conditions - an Example from Ethiopia. In Mapping for change: practice, technologies and communication* ( No. 54). Participatory Learning and Action (pp. 58–66). International Institution for Environment and Development, London.

Li, J., Li, Y., Chapman, M., & Rüther, H. (2005). Small Format Digital Imaging for Informal Settlement Mapping. *Phtogrammetric Engineering & Remote Sensing*, *71*(4), 435–442.

Li, Z., Yan, H., Ai, T., & Chen, J. (2004). Automated building generalization based on urban morphology and Gestalt Theory. *International Journal of Geographical Information Science*, *18*(5), 513–534.

Liu, H., Wang, Y., & Luan, Q. (2007). Semiautomatic extraction of building information basedon mathematical morphology. *Proceeings of SPIE* (Vol. 6753, p. 675329). Presented at the Geoinformatics.

López, E., Bocco, G., Mendoza, M., & Duhau, E. (2001). Predicting land-cover and land-use change in the urban fringe: A case in Morelia City, México. *Landscape and Urban Planning*, *55*, 271–285.

Lopez-Ornclas, E., & Flouzat, G. (2008). Implicit Spatial Information Extraction from Remote Sensing Images, in *Headway in Spatial Data Handling*, Lecture Notes in Geoinformation and Cartography, Springer-Verlag: Heidelberg.

Marangoz, A. M., Oruc, M., & Buyuksalih, G. (2004). Object-oriented Image Analysis and Semantic Network for Extracting the Roads and Buildings from IKONOS Pan-sharpened Images. Presented at the ISPRS 20th Congress, Istanbul.

Martínez, J. (2009). The use of GIS and indicators to monitor intra-urban inequalities. A case study in Rosario, Argentina. *Habitat International*, *33*(4), 387–396.

Martínez, J., Boerboom, L., & Sliuzas, R. (2007). Modelling Slum Improvement Dynamics. Presented at the 10th International Conference on Computers in Urban Planning and Urban Management, Iguassu Falls, Brazil.

Martínez, J., Mboup, G., Sliuzas, R., & Stein, A. (2008). Trends in Urban Slum Indicators across developing world cities, 1990-2003. *Habitat International*, *32*, 86–108.

Mason, S., & Baltsavias, E. (1997). Image-Based Reconstruction of Informal Settlements. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)* (Vols. 1-2, Vol. II, pp. 97–108). Basel, Switzerland: Birkhauser Verlag.

Mason, S. O., Baltsavias, E. P., & Bishop, I. (1997). Spatial decision support systems for the management of informal settlements. *Computers, Environment and Urban Systems*, *21*(3-4), 189–208.

Mayunga, S. D., Coleman, D. J., & Zhang, Y. (2007). A semi-automated approach for extracting buildings from QuickBird imagery applied to informal settlement mapping. *International Journal of Remote Sensing*, *28*(10), 2343–2357.

Mayunga, S. D., Coleman, D. J., & Zhang, Y. (2010). Semi-Automatic Building Extraction in Dense Urban Settlement Areas from High-Resolution Satellite Images. *Survey Review*, *42*(315), 50–61.

Melesse, M. (2005). City Expansion, Squatter Settlements and Policy Implications in Addis Ababa: The Case of Kolfe Keranio Sub-City, Working papers on population and land use change in central Ethiopia, no. 2. Acta Geographica-Trondheim, Norway. Retrieved from http://www.svt.ntnu.no/geo/Doklager/Acta/Serie_A_9_Melesse.pdf

Monkkonen, P. (2008). Using Online Satellite Imagery as a Research Tool - Mapping changing Patterns of Urbanization in Mexico. *Journal of Planning Education and Research*, *28*(2), 225–236.

Myint, S., Mesev, V., & Lam, N. (2006). Urban Textural Analysis from Remote Sensor Data: LacunarityMeasurements Based on the Differential Box Counting Method. *Geographical Analysis*, *38*, 371–390.

NASA and METI, 2010. ASTER Global DEM (GDEM) is a product of the Japanese Ministry of Economy and Trade and the United States National Aeronautics and Space Administration, Available online at: http://lpdaac.usgs.gov/get_data [Accessed 12 Dec 2010].

National Research Council. (2007a). *Tools and Methods for Estimating Population at Risk from Natural Disasters and Complex Humanitarian Crises*. Committee on the Effective Use of Data, Methodologies, and Technologies to Estimate Subnational Populations at Risk. Washington, DC: National Academies Press.

National Research Council. (2007b). Successful Response Starts with a Map: Improving Geospatial Support for Disaster Management. Washington, DC: National Research Council of the National Academies.

Niebergall, S., Loew, A., & Mauser, W. (2007). Object-Oriented Analysis of Very High-Resolution QuickBird Data for Mega City Research in Delhi/India. Paris, France: IEEE, Urban Remote Sensing Joint Event.

O'Sullivan, D., & Unwin, D. J. (2003). *Geographic information analysis*. John Wiley and Sons: Hoboken, NJ.

Ooi, G., & Phua, K. (2007). Urbanization and Slum Formation. *Journal of Urban Health, 84*(0), 27–34.

Owen, K. (2009). Predictive Analysis of Invasive Species - the Case of Phragmites Australis (common reed) along the Rappahannock River Basin. Presented at the ASPRS Annual Conference, Baltimore, MD.

Owen, K. (2011). Settlement Indicators of Wellbeing and Economic Status – Lacunarity and Vegetation. *Proceedings of ASPRS Pecora 18*. Herndon, VA: ASPRS.

Owen, K. (2012). Monitoring Growth in Informal Settlements – Anticipating Conflict and Risk. (Accepted) in *USGIF Monograph Series #1: Challenges of Socio-Cultural Dynamics for Global Security*. Herndon, VA.

Owen, K. K., Obregón, E. J., & Jacobsen, K. H. (2010). A geographic analysis of access to health services in rural Guatemala. *International Health*, *2*(2), 143–149.

Owen, W., (2011), Plan Convexity and Profile Convexity figures, unpublished.

Parker, R. (1995). Lessons from Four Turkish Urban Areas. *Informal Settlements, environmental degradation, and disaster vulnerability - the Turkish Case Study* (Vol. 97). Washington, DC: The World Bank.

Pauchard, A., Aguayo, M., Peña, E., & Urrutia, R. (2006). Multiple effects of urbanization on the biodiversity of developing countries: The case of a fast-growing metropolitan area (Concepción, Chile). *Biological Conservation*, *127*, 272–281.

Pesaresi, M., & Benediktsson, J. A. (2001). A New Approach for the Morphologial Segementation of High-Resolution Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(2), 309–320.

Pesaresi, M., & Ehrlich, D. (2009). A Methodology to Quantify Built-up Structures from Optical VHR Imagery, in *Global Mapping of Human Settlements: Experiences, data sets and prospects*. Taylor and Francis: Boca Raton, FL.

Puetz, A., & Olsen, R. C. (2006). Haralick Texture Features Expanded Into The Spectral Domain. *Proceedings of SPIE* (Vol. 6233). Presented at the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII.

Pumain, D., & Tannier, C. (2005). Fractals in urban geography : a theoretical outline and an empirical example. *Cybergeo : European Journal of Geography*, Systèmes, Modélisation, Géostatistiques. Retrieved from http://cybergeo.revues.org/3275

Qadeer, M. (2000). Ruralopolises: The spatial organisation and residential land economy of High-density Rural Regions in South Asia. *Urban Studies, 37*(9), 1583–1603.

Rapoport, A. (1988). Spontaneous Settlements as Vernacular Design. *Spontaneous Shelter: International Perspectives and Prospects* (ed. C. Patton., pp. 51–57). Philadelphia: Temple University Press.

Rice, J. (2008). The Urbanization of Poverty and Urban Slum Prevalence: The Impact of the Built Environment on Population-Level Patterns of Social Well-Being in the Less Developed Countries. *Care for Major Health Problems and Population Health Concerns: Impacts on Patients, Providers and Policy*, Research in the Sociology of Health Care (Vol. 26). Bingley, UK: Emerald Group Publishing Limited.

Riitters, K., Wickham, R., O'Neill, R., Jones, B., & Smith, E. (2000). Global-scale patterns of forest fragmentation. *Conservation Ecology*, *4*(2), Art. 3.

Rüther, H., Martine, H. M., & Mtalo, E. G. (2002). Application of snakes and dynamic programming optimisation technique in modeling of buildings in informal settlement areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, *56*(4), 269–282.

Shamir, L., Wolkow, C. A., & Goldberg, I. G. (2009). Quantitative measurement of aging using image texture entropy. *Bioinformatics*, *25*(23), 3060 –3063.

Sietchiping, R. (2000). Proceedings of ICSTM2000. Presented at the International Conference on Systems Thinking in Management, Geelong, Australia. Retrieved from http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-72/089%20Sietchiping%20GIS.pdf

Sietchiping, Remy. (2004). *A Geographic Information Systems and Cellular Automata-Based Model of Informal Settlement Growth*. University of Melbourne.

Simon, L. (2004). STAT501-Multicollinearity. Retrieved March 4, 2012, from http://online.stat.psu.edu/online/development/stat501/12multicollinearity/05multico_vif.html

Sliuzas, R. (2008). *Diversity of global slum conditions - is a universal spatial definition of slums feasible?* Presented at the Expert Group Meeting on Slum Mapping, 21 May 2008, Enschede, Netherlands.

Sliuzas, R., Mboup, G., & de Sherbinin, A. (2008). *Report of the Expert Group Meeting on Slum Identification and Mapping* (p. 36). Enschede, Netherlands: International Institute for Geo-Information Science and Earth Observation (ITC), UN-HABITAT Global Urban Observatory, Center for International Earth Science Information Network (CIESIN).

Song, Y., & Knapp, G.J. (2004). Measuring Urban Form: Is Portland Winning the War on Sprawl? *Journal of the American Planning Association*, *70*, 210–225.

Stasolla, M., & Gamba, P. (2008). Spatial Indexes for the Extraction of Formal and Informal Human Settlements From High-Resolution SAR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *1*(2), 98–106.

Stasolla, Mattia, & Gamba, Paolo. (2007). Exploiting Spatial Patterns for Informal Settlement Setection in Arid Environments Using Optical Spaceborne Data. Presented at the Phtogrammetric Image Analysis, Munich, Germany.

Stein, A., Chair. (2008). *Methods for Scaling Data - Breakout Group on Scaling Measurements*. Presented at the Expert Group Meeting on Slum Identification and Mapping, 21 May 2008, Enschede, Netherlands.

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

Sugumaran, R., Pavuluri, M. K., & Zerr, D. (2003). The use of high-resolution imagery for identification of urban climax forest species using traditional and rule-based classification approach. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(9), 1933–1939.

Sulik, J., & Edwards, S. (2010). Feature extraction for Darfur: geospatial applications in the documentation of human rights abuses. *International Journal of Remote Sensing*, *31*(10), 2521–2533.

Sutton, P. (2003). A scale-adjusted measure of “Urban Sprawl” using nighttime satellite imagery. *Remote Sensing of Environment*, *86*, 353–369.

Tannier, C., & Pumain, D. (2005). Fractals in urban geography : a theoretical outline and an empirical example. *Cybergeo - European Journal of Geography*, Systems, Modeling, Geostatistics.Retrieved from http://www.cybergeo.eu/index3275.html?

Thomas, I., Frankhauser, P., & Biernacki, C. (2008). The morphology of built-up landscapes in Wallonia (Belgium): A classification using fractal indices. *Landscape and Urban Planning*, *84*(2), 99–115.

Thomas, I., Tannier, C., & Frankhauser, P. (2008). Is there a link between fractal dimension and residential environment at a regional level? *Cybergeo : European Journal of Geography, Systèmes, Modélisation, Géostatistiques* (413). Retrieved from http://www.cybergeo.eu/index16283.html

Transportation Research Board, National Research Council. (1998). *The Costs of Sprawl - Revisited* ( No. 39). Transit Cooperative Research Program. Washington, DC: Federal Transit Administration.

Tresidder, M. (2005, December). *Using GIS to Measure Connectivity: An Exploration of Issues*. Portland State University.

UN Habitat. (2003). *The Challenge of Slums: Global Report on Human Settlements 2003*. London: Earthscan.

UN Habitat. (2009). *Global Urban Observatory Program* (Table 4 Slum population projection 1990-2020). UrbanInfo Database. United Nations.

UN High Commission on Refugees. (1996). *An Urbanizing World: Global Report on Human Settlements*. Retrieved from: http://www.unhabitat.org/downloads/docs/GRHS.1996.0.pdf, Oxford: Oxford University Press.

UN-HABITAT. (2001). *Global Urban Observatory Statistics*. Retrieved from http://ww2.unhabitat.org/programmes/guo/documents/Table1.pdf

UN-HABITAT. (2003). *Slums of the World: The Face of Urban Poverty in the New Millennium, Monitoring the Millennium Development Goal, Target 11- World-wide Slum Dweller Estimation* (Working Paper). Nairobi, Kenya: Global Urban Observatory Program.

UN-HABITAT. (2006). *The State of the World's Cities Report 2006-2007: 30 years of shaping the habitat agenda*. Nairobi, Kenya. Retrieved from http://www.unhabitat.org/pmss/getPage.asp?page=bookView&book=2916

UN-HABITAT. (2010). *State of the world's cities 2010/2011: bridging the urban divide*. UN-HABITAT.

United Nations. (2010). *The Millennium Development Goals Report 2010* (p. 80). New York, NY. Retrieved from http://www.un.org/millenniumgoals/pdf/MDG%20Report%202010%20En%20r15%20-low%20res%2020100615%20-.pdf#page=64

Valladares Cerezo, C. E. (2003). *Urban Slums Report - the Case of Guatemala City, Guatemala*. Understanding Slums: Case Studies for the UN Global Report on Human Settlements. London: Development Planning Unit, University College London. Retrieved from http://www.ucl.ac.uk/dpu-projects/Global_Report/cities/guatemala.htm, United Nations: New York.

Vicente, M. C., Villarin, J. R., Galgana, G. A., Guzman, A. L. G., & de Mesa, J. (2006, November 3). *The Urban Poverty Morphology Project: RS-GIS Applications for Metro Manila, Philippines*. Presented at the 9th Annual Earth Sciences Congress, Santiago, Chile.

Vieira, B. C., & Fernandes, N. F. (2004). Landslides in Rio de Janeiro: The role played by variations in soil hydraulic conductivity. *Hydrological Processes*, *18*(4), 791–805.

Wallace, J., Morris, B., & Howarth, P. (2004). The Effects of Scale on Fractal Dimension of Topography. *Proceedings of IGARSS*. Ontario, Canada.

Ward, P., & Peters, P. (2007). Self-help housing and informal homesteading in peri-urban America: Settlement indentification using digital imagery and GIS. *Habitat International*, *31*, 205–218.

Warnick, R., Brewer, K., Megown, K., Finco, M., Schwind, B., Ralph Warbington, & Barber, J. (2008). Texture Metric Comparison of Manual Forest Stand Delineation and Image Segmentation. Presented at the RS 2008, Salt Lake City, UT: USDA.

Weber, C., & Puissant, A. (2003). Urbanization pressure and modeling of urban growth: example of the Tunis Metropolitan Area. *Remote Sensing of Environment*, *86*, 341–352.

Weeks, J.R., Hill, A., Stow, D., Getis, A., Fugate, D., (2007). Can we spot a neighborhood from the air? Defining neighborhood structure in Accra, Ghana, *GeoJournal* 69:9-22

Weng, Q. (2002). Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *Journal of Environmental Management*, *64*, 273–284.

Wilson, E. H., Hurd, J. D., Civco, D. L., Prisloe, M. P., & Arnold, C. (2003). Development of a Geospatial Model to quantify, describe and map urban growth. *Remote Sensing of Environment*, *86*, 275–285.

Wong, D. (2004). Comparing Traditional and Spatial Segregation Measures: A Spatial Scale Perspective. *Urban Geography*, *25*(1), 66–82.

Wong, D. (2008). A Local Multidimensional Approach to Evaluate Changes in Segregation. *Urban Geography*, *29*(5), 455–472.

Wong, D. W.-S., & Lee, J. (2005). *Statistical analysis of geographic information with ArcView GIS and ArcGIS*. John Wiley & Sons, Inc.: Hoboken, NJ.

Wood, J. D. (1996). *The geomorphological characterisation of digital elevation models*. PhD Dissertation, University of Leicester, UK. Retrieved from http://www.soi.city.ac.uk/~jwo/phd/

Wu, X. B., & Sui, D. Z. (2001). An initial exploration of a lacunarity-based segregation measure. *Environment and Planning B: Planning and Design*, *28*, 433–446.

Yaakup, A., & Healey, RG. (1994). A GIS approach to spatial modelling for squatter settlement planning in Kuala Lumpur, Malaysia. *Environment and Planning B: Planning and Design*, *21*, 21–34.

Yeh, A. G.-O., & Li, X. (2001). Measurement and Monitoring of Urban Sprawl in a Rapidly Growing Region Using Entropy. *Photogrammetric Engineering & Remote Sensing*, *67*(1), 83–90.

Yi, C. (2008). *Utilizing GIS to measure Street Connectivity and Pedestrian Accessibility: Comparing Cul-De-Sac and Grid Neighborhoods in the Houston Metropolitan Area* (Paper Competition). University of TX, Austin. Retrieved from http://www.gis-t.org/files/H5A8N.pdf

Yohannes, Y., & Hoddinott, J. (1999). *Classification and Regression Trees: An Introduction* (Technical Guide No. #3). Washington, D.C.: International Food Policy Research Institute. Retrieved from http://pdf.usaid.gov/pdf_docs/PNACH725.pdf

Zebardast, E. (2006). Marginalization of the urban poor and the expansion of the spontaneous settlements on the Tehran metropolitan fringe. *Cities*, *23*(6), 439–454.

Zeilhofer, P., & Topanotti, V. P. (2008). GIS and ordination techniques for evaluation of environmental impacts in informal settlements: A case study from Cuiaba, central Brazil. *Applied Geography*, *28*, 1–15.

Zeng, H., Sui, D. Z., & Li, S. (2005). Linking Urban Field Theory with GIS and Remote Sensing to Detect Signatures of Rapid Urbanization on the Landscape: Toward a New Approach for Characterizing Urban Sprawl. *Urban Geography*, *26*(5), 410–434.

Zhang, M., & Kukadia, N. (2005). Metrics of Urban Form and the Modifiable Areal Unit Problem. *Transportation Research Record: Journal of the Transportation Research Board*, *1902*(-1), 71–79.

**CURRICULUM VITAE**

Karen K. Owen received her Bachelor of Arts and Master of Science from The American University in 1984 and 1986, respectively. She received a Master of Science in Telecommunications in 2006 from George Mason University. She is currently employed in Northern Virginia, USA.

## Publications


- Owen, Karen, (2012), Monitoring Growth in Informal Settlements – Anticipating Conflict and Risk, (Accepted*). USGIF Monograph Series #1: Challenges of Socio-Cultural Dynamics for Global Security.* Herndon, VA.
- Owen, Karen, (2011), Settlement Indicators of Wellbeing and Economic Status – Lacunarity and Vegetation, *American Society of Photogrammetry and Remote Sensing William T. Pecora Conference*, 14-17 November 2011, Herndon, VA.
- Owen, Karen, Cart-Obregon, Elizabeth J. and Jacobsen, Kathryn H., (2010), "A geographic analysis of access to health services in rural Guatemala, Journal of International Health, Volume 2, Issue 2 Pages 143-149, June 2010.
- Owen, Karen, (2009), Predictive Analysis of Phragmites Australis – Invasive Species along the Rappahannock River Basin, Virginia, *American Society of Photogrammetry and Remote Sensing Conference Proceedings*, November 2009, Baltimore, MD.