

**SATELITE IMAGE METRICS AS INDICATORS OF SOCIOECONOMIC
AND DEMOGRAPHIC CHARACTERISTICS IN SLUM
NEIGHBORHOODS OF ACCRA, GHANA**

A Thesis

Presented to the

Faculty of

San Diego State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Geography

with a Concentration in

Geographic Information Science

Milo J. Vejraska

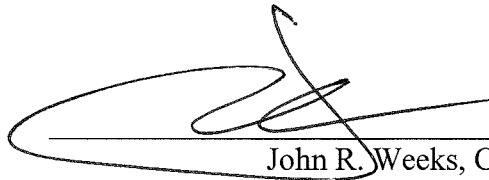
Summer 2013

SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Milo J. Vejraska:

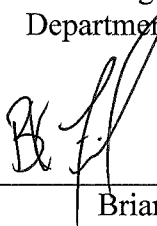
Image metrics as indicators of socioeconomic and demographic characteristics in
slum neighborhoods of Accra, Ghana



John R. Weeks, Chair
Department of Geography



Douglas A. Stow
Department of Geography



Brian K. Finch
Graduate School of Public Health

13 May 2013

Approval Date

Copyright © 2013

by

Milo J. Vejraska

All Rights Reserved

DEDICATION

I would like to acknowledge those who have helped me in any way over the course of my academic career, including, but not limited to: my parents, for never letting me believe anything else growing up besides that I will be heading to college; Dr. John R. Weeks, for opening my mind to the world of demography, spatial statistics, and the natural process of questioning the interconnectedness of the world; Dr. Douglas A. Stow, for all of the extremely valuable training and comments, and for always keeping a light but serious mood; all of my office-mates, for not only the ideas we bounced off of each other, but for keeping a competitive environment whilst also partaking in shenanigans; my graduate class cohort, especially those who were very close to me and critically questioned and evaluated my thought process and work; my friends, who were always there to support me and who allowed me to use them as pseudo-vices as I completed my education; those unmentioned who contributed in any way to this project, especially the members of the Ghana Project team; and those groups that supported this project through monetary funding.

Literacy is a bridge from misery to hope. It is a tool for daily life in modern society. It is a bulwark against poverty, and a building block of development, and essential complement to investments in roads, dams, clinics and factories. Literacy is a platform for democratization, and a vehicle for the promotion of cultural and national identity. Especially for girls and women, it is an agent of family health and nutrition. For everyone, everywhere, literacy is, along with education in general, a basic human right.... Literacy is, finally, the road to human progress and the means through which every man, woman and child can realize his or her full potential.

– Kofi Annan, Ghanaian diplomat, seventh secretary-general of the United Nations,
2001 Nobel Peace Prize winner

Revolutions are brought about by men, by men who think as men of action and act as men of thought.

– Kwame Nkrumah, first president & prime minister of Ghana,
1963 Lenin Peace Prize winner

ABSTRACT OF THE THESIS

Satellite Image Metrics as Indicators of Socioeconomic and
Demographic Characteristics in Slum Neighborhoods of
Accra, Ghana

by

Milo J. Vejraska

Master of Science in Geography with a Concentration in
Geographic Information Science
San Diego State University, 2013

This study explored the connection between remotely sensed imagery and ground-based housing and welfare survey data in slum neighborhoods in Accra, Ghana. Specific household-level variables reflective of housing quality and demographics from the 2009-2010 Housing and Welfare Study (HAWS) of Accra and the 2003 UN-Habitat Accra Slum Survey (AccraSS) were regressed against measures extracted from high spatial resolution Quickbird satellite imagery captured in 2002 and again in 2010. Samples from the two surveys for 37 census enumeration areas (EAs) within the Accra Metropolitan Area (AMA) were analyzed. An exhaustive regression analysis was run to measure the covariation between individual survey data variables and metrics derived from the imagery. A spatial regimes approach explored spatially autocorrelated data in the discontinuous data set, and these results were compared with a geographically weighted regression approach. The goal was to establish “proxy” variables from satellite remote sensing data that are indicative of household health and welfare characteristics over time by combining spatially homogenous predictors in multivariate regression models. By generating proxies of the built environment, we may be able to infer or extrapolate socioeconomic and health statuses for each respective EA and the surrounding neighborhoods at other dates (e.g., between surveys and censuses). Specifically, I test the hypotheses that (1) socioeconomic and demographic characteristics of slum areas can be inferred from spatial variations in vegetation and texture as derived from satellite imagery; and (2) dynamics of socioeconomic and demographic characteristics can be quantified from changes in the image metrics. Since one in six residents of the world is estimated by the UN to be living in a slum, it is important to understand how these neighborhoods might transform over time as the population of a major city in a developing nation increases at a high rate.

TABLE OF CONTENTS

	PAGE
ABSTRACT.....	vi
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
CHAPTER	
1 INTRODUCTION.....	1
2 BACKGROUND.....	4
2.1 Applications of Remotely-Sensed Imagery.....	4
2.2 The VIS Model & Spectral Classifications.....	5
2.3 Connecting Land Cover and Socioeconomic Data.....	6
3 DATA & METHODS.....	10
3.1 Study Site & Survey Data.....	10
3.2 Survey-Derived Variables.....	12
3.3 Imagery-Derived Variables.....	13
3.4 Exploratory Regression Modeling.....	17
3.5 Spatial Heterogeneity & Regime Analysis: Structural & Spatial Instability.....	20
3.6 Comparing Model Results to a Geographically Weighted Regression.....	22
3.7 Evaluating Temporal Robustness using the Bootstrap Procedure.....	23
3.8 Investigating Change Over Time: The Search for Proxies.....	23
4 RESULTS.....	25
4.1 Exploratory Bivariate Regression Analysis & OLS Modeling.....	25
4.1.1 2002 Models.....	25
4.1.2 2010 Models.....	32
4.1.3 Δ Models.....	36
4.2 Estimating Indicators of SEDs Through Spatial Regression Modeling.....	43
4.3 Comparing Results to a GWR and a Spatial Regimes Approach.....	47
5 DISCUSSION & CONCLUSIONS.....	48

REFERENCES50

APPENDIX

 A DISTRIBUTION OF ROOF TYPES IN THE AMA56

 B SURVEY VARIABLE CHANGES BY EA.....58

 C R PROGRAM OF LACUNARITY CODE63

LIST OF TABLES

	PAGE
Table 1. List of variable inputs into the exploratory regression model	17
Table 2. 2002 OLS model results, including R ² values, model predictors variables, and tests for residual normality and heteroskedasticity.	28
Table 3. Diagnostics for spatial dependence and spatial heterogeneity for the t ₂₀₀₂ models – the Global Moran's I and Lagrange Multiplier statistics and probabilities.	30
Table 4. 2010 OLS model results, including R ² values, model predictors variables, and tests for residual normality and heteroskedasticity.	34
Table 5. Diagnostics for spatial dependence and spatial heterogeneity for the t ₂₀₁₀ models – the Global Moran's I and Lagrange Multiplier statistics and probabilities.	35
Table 6. Δ OLS model results, including R ² values, model predictors variables, and tests for residual normality and heteroskedasticity.	38
Table 7. Diagnostics for spatial dependence and spatial heterogeneity in Δt models – the Global Moran's I and Lagrange Multiplier statistics and probabilities.	39
Table 8. % Muslim results for the OLS and spatial lag global models.	44
Table 9. Δ % trash dumped offsite results for the OLS and spatial lag global models.	44
Table 10. Δ % bednet results for the OLS and spatial error global models.	45
Table 11. AICc values for both OLS and global spatial error models for 2002 and Δt ₂₀₀₂₋₂₀₁₀ dates. A decrease in the AICc value of > 3 is considered an indication of improved model specification. Values in green represent improved models, where values in yellow represent error or lag models that did not meet the criteria to be considered an improvement of model fit. A value in blue denotes the specification of an error model.	46

LIST OF FIGURES

	PAGE
Figure 1. The enumeration area (EA) boundaries for the Accra Metropolitan Area, overlaid on a QuickBird 2010 Panchromatic image. EAs selected for the analysis are highlighted in bright green.....	11
Figure 2. A flowchart of the methods used in the analysis of the imagery and survey data.....	24
Figure 3. Selected image metrics included in models with the best results. EA values are displayed in a diverging color scheme with higher values displayed in red & lower values in blue. Images from which each metric was derived from is displayed in each respective backdrop.....	29
Figure 4. Standardized residuals for selected OLS models. The 2002 panchromatic image is displayed in the background.....	31
Figure 5. A side-by-side comparison of image lacunarity in EA 205020 (Sodom & Gomorah) for (a) 2002 and (b) 2010 images in the neighborhood of Sodom & Gomorah. An increase in lacunarity is observed between 2002-2010, directly related to the corresponding increase in building density per EA. Areas outside the boundary are displayed for comparison purposes..	40
Figure 6. A comparison of the 2nd principal component (PC2), derived from QuickBird multispectral images for (a) 2002 and (b) 2010 images in the Gbegbeyise (west of the river) and Chokor neighborhoods. Both dates revealed that PC2 carries a large portion of spectral vegetation data and were considered vegetation proxies.....	41
Figure 7. A comparison of the Blue/Green (B/G) band ratio, derived from QuickBird multispectral images for (a) 2002 and (b) 2010 images in the Gbegbeyise (west of the river) and Chokor neighborhoods. The B/G image metric was a significant predictor in five of the highest performing models.	42

CHAPTER 1

INTRODUCTION

Ghana is currently undergoing a massive rural-to-urban transition, and its capital city of Accra has been a major receiving area for migrants from other parts of the country. Census data over the past 30 years show that the Greater Accra region has experienced high population growth rates, mainly due to a burgeoning youth population in rural areas that exceeds the ability of those areas to create jobs. According to the latest provisional data from the Ghanaian 2010 census, the Greater Accra Region now has a population of 4.4 million, accounting for 18 percent of Ghana's total population. The Accra Metropolitan Area (AMA) is estimated in 2009 to have approximately 2.3 million of these people residing within its boundaries, with 58% of the population living in neighborhoods classified as slums (UN Habitat 2009). UN-Habitat defines slums based on their limited access to safe water, sanitation and sewage infrastructure, the poorer structural quality of housing, higher number of residents per housing unit, and more limited ownership for housing tenure (United Nations Human Settlements Programme 2003). Rural areas surrounding the capital city of Accra, along with the city's metropolitan center, have also experienced very rapid urbanization since the 1980s. These areas experiencing rapid growth in population consist of mostly informal neighborhoods with a lack of urban planning or sufficient infrastructure (Moller-Jensen and Knudsen 2008). The people residing within these informal areas generally belong to lower socio-economic classes. Due to the lack of access to a clean water supply (Stoler et al. 2011), poor sewage systems, and crowded housing, among many other factors, they are at a greater health risk and susceptible to higher levels of disease contraction and mortality (Weeks et al. 2011).

The urban environment serves as a location for the accumulation and integration of social, economic and cultural forces over time (Moudon 1997). People living in an urban setting act as agents of change, creating a dynamic interrelationship between the population and environment. Urban ecosystems are dramatic manifestations of human's impact on the environment (Ridd 1995), and it is important to evaluate and understand the urban

environment and its relationship to social qualities of life. Like a local ecosystem, the functions of the population of a neighborhood have a reciprocal effect on the natural and built features. The physical properties of a neighborhood's built environment are reflective of the local social and spatial contexts that are influenced by the individual characteristics and behaviors of its respective and surrounding populations (Entwisle 2007). These features of the built environment, including a range of built infrastructure, vegetation, agriculture, and other land cover and land use types, can be remotely sensed and proxy variables derived from remotely sensed images may be indicative of the urban lifestyle (Weeks 2003). In order to pragmatically examine urban morphology, however, not only should the physical components of a city be explored, but the temporal aspect should also be accounted for. Understanding the change over time of a place along an urban-to-rural gradient allows for an understanding of processes of urban change that both are affected by and have an effect on social processes and human behavior.

The goal of this study was to examine the degree of co-variability between household level survey variables and metrics derived from high spatial resolution satellite imagery through multivariate regression. The main objective is to derive proxy variables of housing and welfare attributes from satellite, census, and health survey data for Accra, Ghana. The existing literature has demonstrated that there can be significant correlations between land cover metrics classified from high resolution imagery and health and wealth indicators derived from census or survey data. This study sought to explore how different texture measures, spectral band indices, and land cover metrics could be exploited to provide a quintessential component in classifying the variation in demographics and socioeconomic status within a slum neighborhood. It is important to find which image metrics best identify differences in health and wealth indices.

A secondary objective of this study is to evaluate the robustness of the regression models over time, comparing the correlations between the changes in imagery and household variables from 2002/2003 to 2009/2010. Understanding the significance and degree of co-variability between land cover change and quality of life is an integral step in modeling the urban gradient of developing cities in Sub-Saharan Africa. The specific research questions that stemmed from the objectives were as follows:

1. How well can socioeconomic and demographic characteristics within slum areas be quantified through variations in vegetation and textural indices derived from high spatial resolution satellite imagery?
2. How well can changes in socioeconomic and demographic characteristics over time be characterized through changes in metrics derived from high spatial resolution satellite image data?
3. Are statistical correlations computed from both global and localized linear regression models sufficiently high to enable their use as proxies for slum conditions?

CHAPTER 2

BACKGROUND

2.1 Applications of Remotely Sensed Imagery

The benefits of the application of remotely sensed imagery to complement research in the public health sector have been widely examined (Kelly et al. 2011) and such imagery is being integrated into many studies due to its ability to provide measures of factors within the human environment that affect the health of a population. For example, remote sensing products based on the normalized difference vegetation index (NDVI) and object-based image analysis (OBIA) were demonstrated in Kelly et al. (2011) to assess locations, quantities and extents of vegetation, agriculture, water resources, infrastructure, and other land use, along with the temporal scales of change across these components of the human landscape. Image texture has also been demonstrated to be useful in the processing of high-resolution imagery. Image texture can be defined as the spatial arrangement of the gray levels of pixels in a specific window (Herold, Liu & Clarke 2003; Bharati, Liu & MacGregor 2004), or more specifically “a spatial relationship between intensity values of neighboring pixels, repeated over an area larger than the size of the relation” (Raghu et. al. 1995). Statistical measures are most commonly used to characterize the spatial variability of pixel gray levels within an image (Wang & Liu 1999). Specifically, second-order statistics such as measures derived from Gray Level Co-Occurrence Matrices (GLCMs) have assisted in differentiating gaps between land cover types, densely-settled urban areas, and vegetation (Jensen 1996; Herold, Liu & Clarke 2003; Kelly et al. 2011). Whereas first-order statistics are based upon simple statistical measures of gray level variability without being related to a pixel’s context, second-order statistics describe the relationship of a pixel to its neighbors within a defined region (Raghu et al 1995). Local properties or statistics that repeat over the defined region are referred to as texture elements (Arivazhagan & Ganesan 2003). Satellite imagery has been used in Accra to identify areas of health risk from low elevations prone to flooding (Rain et al. 2011) and in Kenya and elsewhere in Sub-Saharan Africa to create

distance thresholds for areas at risk of the vector born disease of malaria (Hay et al. 2001; Tatem & Hay 2004) (Stoler, Weeks, Getis, and Hill 2009).

2.2 The VIS Model & Spectral Classifications

When incorporating imagery into a study, it may be useful to classify land use in order to identify how land is being used by its inhabitants. Ridd (1995) explored the application of a V-I-S model for the analysis and characterization of land cover and land use within urban ecosystems in a study in Salt Lake City, UT. Combinations of vegetation (V), impervious surface (I) and bare soil (S) are considered to be the fundamental components of the urban environment. The V-I-S model was proposed in order to build a conceptual framework for inter-urban ecosystem comparison both spatially and temporally, and has been expanded upon since (Phinn et al. 2002; Rashed et al. 2003; Rashed et al. 2005). V-I-S analysis classifies each pixel within an image as either vegetation, impervious surface, or bare soil through either a hard or soft classification method, depending on the heterogeneity and spatial resolution of an image. Hard classification logic produces a map that consists of discrete categories, whereas fuzzy set (soft) classification logic considers image heterogeneity as a reality, producing a thematic output in which each pixel contains membership probabilities for m number of categories (Jensen 2005). Pixels composed of a homogeneous land cover type are considered “pure” pixels and can be identified through hard classification methods based upon the selection of spectral endmembers (pixels with uniform land cover). Pixels that are not comprised of a homogenous land cover type can be specified as either V, I, or S using a fuzzy classifier. Pixels are then aggregated into spatial eco-units from which land use is derived. The detection and monitoring of urban morphology can also be monitored through V-I-S modeling, by means of which issues of land use and land cover change are being explored for the city of Accra, Ghana (Stow et al. 2007; Stow et al. forthcoming).

A major criticism of the V-I-S model has been the issue of classifying mixed pixels in the urban scene, thought to be mostly driven by improper combinations of spatial and spectral resolution. Pixels may contain high proportions of all three endmembers, and are therefore considered mixed elements (Rashed et al. 2003). If an urban scene is treated as a continuous model, where these mixed pixels are treated as the sum of the spectral

interactions between land cover types within a single pixel instead of discrete elements, spectral mixture analysis (SMA) can estimate the component parts of mixed pixels by predicting the proportion of a pixel that belongs to a particular class or feature based on the discrete spectral characteristics of its endmembers. SMA is a type of fuzzy classification and is considered a “soft” classification method (Rashed et al. 2001; Weeks 2003). Rashed et al. (2003) applied SMA to an urban scene of Los Angeles County and provided an improved measure of the elements of land cover and land use that adequately characterized the urban environment. Also, the value of using spectral mixture analysis to monitor temporal compositions in urban land cover change was demonstrated in Cairo (Rashed et al. 2005). By using the components of V-I-S plus shade (to capture pixels in the shade of tall buildings) at two different dates, multiple endmember fraction images were subtracted from one another, revealing the direction (increase or decrease), magnitude, and categories of change in land cover, demonstrating in Greater Cairo V-I-S changes cascading out from the urban center to the peri-urban fringe regions of the city.

2.3 Connecting Land Cover & Socioeconomic Data

Weeks et al. (2007) classified high spatial resolution Quickbird satellite multispectral imagery using the V-I-S model to characterize land use in neighborhoods within the Accra Metropolitan Area (AMA). By quantifying the proportional abundance of each of land cover surface material, different land use categories were distinguished and then were statistically correlated to a Census-derived slum index for neighborhoods. Factors used to create the index are based on the operational definition provided by UN Habitat. Neighborhoods categorized as “slums” were hypothesized to contain a combination of impervious surfaces and bare soil, both indicative of residential areas in sub-Saharan African developing cities, but relatively low levels of vegetative cover relative to other residential areas. GLCM texture measures were also included in the analysis of the imagery, where slum areas with very dense settlements and little variability in building materials between residences could be further distinguished. Slums were shown to be associated with less vegetation and less variability in land cover, supported by an R^2 value showing that 38% of the variability from one neighborhood to another in the slum index was explained by the proportional abundance of vegetation. The proportional abundance of bare soil also accounted for a great deal of

variability, where neighborhoods composed of lesser amounts of bare soil tended to be those with higher proportions of crowded housing structures.

A recent study explored whether or not variations in the urban landscape of Accra depicted by image classification of satellite image data relate to variability in health and wealth indicators (Engstrom et al. 2011). The authors used a combination of Landsat and high resolution imagery to extract measures of vegetation and impervious surface from the V-I-S model using decision trees. Inputs to the decision trees included original bands, band ratios, vegetation indices, and grey level co-occurrence (GLCM) texture measures. The study was performed on two scales of analysis; first, at the census tract enumeration area (EA) level, and second, at the neighborhood level, where EAs were aggregated based on local vernacular knowledge of where neighborhood boundaries occurred. Regression results yielded moderate relationships between the percent built-up area per neighborhood with both female education levels and population density, with R^2 values of 0.39 and 0.58, respectively. As the percentage of built-up area within an agglomerated neighborhood increased, the proportion of women with at least a secondary education decreased. The inverse relationship between these two variables was suggested to be a response to the implication of having a dual income household allowing a family to acquire a single-family dwelling and attain a higher socio-economic status. Conversely, a positive relationship was discovered between population density and built-up area proportions by neighborhoods, as one would normally assume an increase in urban building density to be a response to an increasing population. Percent vegetation cover was also negatively correlated with the use of charcoal as a cooking fuel ($R^2 = 0.65$). This can be explained by the fact that charcoal is the cheapest source of fuel and is used mostly by those people with a lower socioeconomic status (Engstrom et al. 2011).

Proportions of vegetation per unit area have also proven to be useful in the application of classifying land use and in delineating neighborhood boundaries through the use of object based approaches in Accra, Ghana (Stow et al. 2007; Stow, Lippitt & Weeks, 2010). In analyzing the changes between high spatial resolution imagery from 2002 and 2010, Stow et al. (forthcoming) found strong correlations between housing quality and socio-economic variables. Proportions of vegetation were derived using a simple threshold-based classification of normalized difference vegetation index (NDVI) values in order to examine

the statistical relationships between vegetation cover and a census-derived housing quality index (HQI) at the neighborhood level. Ordinary least squares regression found a very significant degree of spatial covariation between the HQI and vegetation abundance, with an $R^2 = 0.73$ and 0.76 for 2002 and 2010, respectively. Also, high socio-economic status neighborhoods tended to have the highest proportions of landscaped vegetation, while low socio-economic areas and “slums” exhibited the lowest amounts of vegetative cover. In addition, low socio-economic neighborhoods showed the greatest relative decrease in abundance during the eight-year period, analogous to the increase in building density for these neighborhoods. To complement the fluctuations of vegetation proportions, changes in building density were shown by Tsai et al. (2011) to be correlated to socio-economic status in Accra, but to a lesser extent. A statistically significant inverse relationship between new building density and housing quality index was determined at the neighborhood level, yielding an $R^2 = 0.31$, where the delineations of new buildings in the Accra Metropolitan Area (AMA) tended to be located in lower socio-economic neighborhoods.

Herold, Liu & Clarke (2003) applied spatial metric techniques and image texture calculations in an urban environment (Goleta and Santa Barbara, California) to explore the link between structures, land cover heterogeneity, and dynamic changes in urban land uses. Evaluating the quantitative descriptors of spatial urban organization allowed for the discovery of relationships between the physical and spectral properties of objects and the socio-economic, demographic, and ecological characteristics of individual based land cover objects. Again, vegetation metrics had the highest contribution for best average separability between land cover objects, and therefore was what a majority of the classification was based upon.

Differences between formal and informal neighborhoods in cities of developing countries were exploited with a grey level co-occurrence matrix, various texture measures and spectral band ratios by Grasser et al. (forthcoming), which provided multiple explorations into methods of neighborhood classification and characterization. Informal neighborhoods are defined as unplanned and unauthorized housing settlements located in hazardous areas of urban agglomerations with inadequate infrastructure and low availability of services, and distinctly contrast with formal settlements and structures in high resolution imagery. Lacunarity, a texture metric that represents the spatial distribution of gap sizes

between pixels of similar brightness values, was also an integral element in characterizing neighborhoods where building density is generally higher (Grasser et al. forthcoming). Lacunarity measures the deviation of a spatial structure in an image from translational symmetry, where objects that have high lacunarity are heterogeneous to the surrounding pixels and therefore are said to have a higher “gappiness” of geometric structure (Myint & Lam 2005). Lacunarity classification metrics have been demonstrated to enhance the accuracy of texture measurements beyond the capability of GLCMs (Dong 2000) and also can assist in yielding more accurate textural classifications of land cover and land use in an urban region using very high spatial resolution imagery (Myint, Mesev & Lam 2006).

The identification and distribution of linear features in urban environments through a textural analysis can also contribute to a more accurate image classification (Graesser et al. forthcoming, Unsalan & Boyer 2004). Thresholding the responses of various convolution filters can extract the characteristics of linear features that are a unique characteristic of anthropogenic structures and have been shown to indicate structural differences between neighborhoods of different socioeconomic status (Graesser et al. forthcoming).

CHAPTER 3

DATA & METHODS

3.1 Study Site & Survey Data

The research draws upon data collected in the 2003 UN-Habitat Accra Slum Survey (Accra SS), a supplement to the 2003 Ghana Demographic and Health Survey (DHS), which interviewed women in 37 randomly selected enumeration areas (EAs) that met specific criteria that categorized them as slum neighborhoods. The UN-Habitat operationally defines slums as neighborhoods that include some or all following characteristics: (1) inadequate access to clean, potable water, (2) inadequate access to improved sanitation and sewage infrastructure, (3) poor structural durability of housing, (4) overcrowding within housing structures, (5) insecure housing tenure (United Nations Human Settlements Programme 2003). EA boundaries were designed and delineated to encompass approximately 1,000 persons and are the Ghana Statistical Survey's (GSS) equivalent to a US Census Block Group. EAs are generally too small in area and population to be recognized as neighborhoods, but GSS has aggregated them to form larger administratively defined units known as localities, containing approximately 40 EAs each. The survey consisted of questions regarding the availability of basic household infrastructure, such as sewerage and water facilities, the presence of household possessions that might be reflective of a household's socioeconomic status, various demographic characteristics of the residents currently residing in the household, and also contained questions pertaining to a variety of health measures for both women and children in the household.

The second dataset consists of primary data collected in 2009-2010 as part of the Housing and Welfare Study (HAWS) of Accra. The HAWS survey is a representative household survey conducted by the Harvard School of Public Health and University of Ghana with assistance from San Diego State University. The sampling frame for the HAWS was designed to replicate the 2003 UN-Habitat study slum selections, consisting of household surveys collected within the same 37 EAs to form a comparable dataset. The

Accra Metropolitan Area

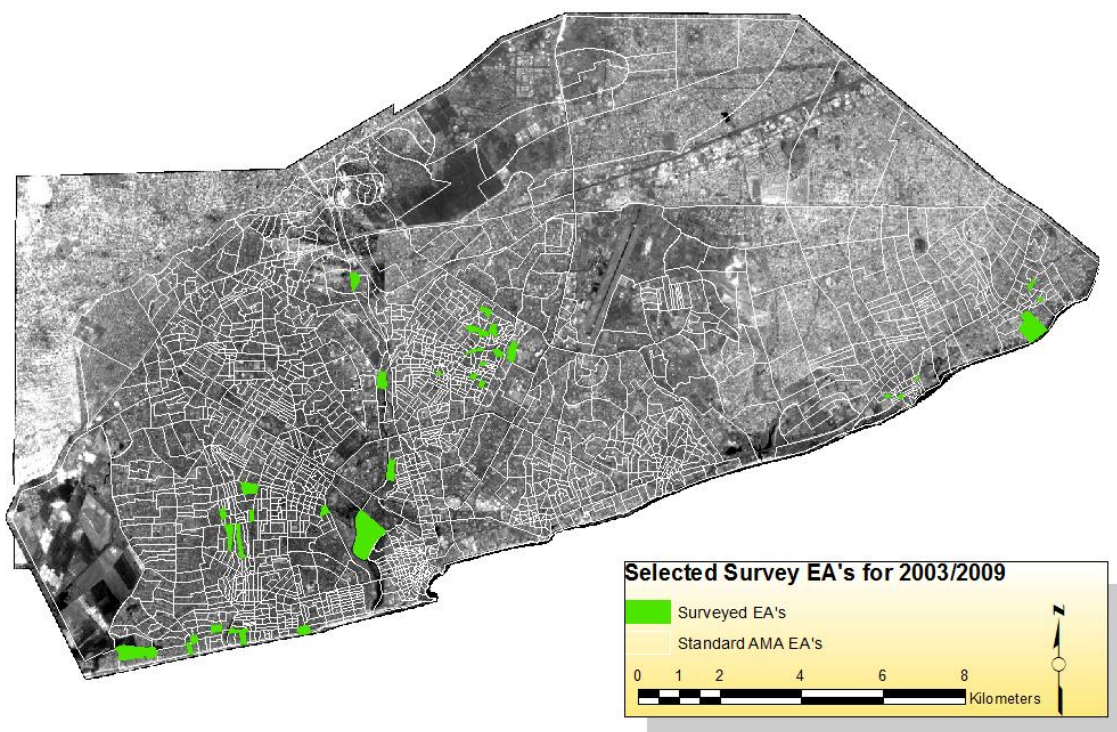


Figure 1. The enumeration area (EA) boundaries for the Accra Metropolitan Area, overlaid on a QuickBird 2010 Panchromatic image. EAs selected for the analysis are highlighted in bright green.

HAWS collected similar data to the Accra SS, and was intended as a follow-up data set to assess changes in the selected EAs time.

There are many housing and welfare variables within the Accra SS and HAWS that can represent differing levels of socioeconomic status of a household. The variables that this study focused upon (described below and listed in Table 1) were selected based on not only their consistency between both surveys, but more importantly, their characteristic ability to have a predictable and intuitive relationship with a household's health and welfare of living. The purpose of studying housing and welfare quality in Accra, Ghana is to link the environmental components of a neighborhood to their social environment, providing a better understanding of a population's health. Accordingly, the selected components of the two surveys were housing characteristics that were reflective of a household's health and well-being. These variables served as the *dependent variables* used in the regression models. A

set of dummy variables was created for all nominal/categorical scale variables in order to identify the presence or absence of a specific household characteristic. The proportions of these characteristics were computed for each EA and were assigned as representative EA values.

3.2 Survey-Derived Variables

The physical components of households selected for regression analyses against image-derived variables included the source of drinking water, type of toilet facility, cooking fuels used, access to electricity, and methods of waste disposal. These were specifically chosen to represent household infrastructure. The access to and adequacy of these characteristics are assumed to be basic household services – the presence of which would significantly improve the housing quality, health, and welfare of household residents.

From more demographic and cultural standpoints, housing tenure, migration, and ethnic and religious composition variables were selected from data collected in Accra from the 2000 Census of Ghana due to the unavailability of these data in the Accra SS and HAWS surveys. These are considered household demographic variables, and contribute where physical characteristics of a household might fall short in explaining differences in housing quality and health between slum neighborhoods. Ethnic residential patterns have been used as predictors of intra-urban health in Accra to show that cultural beliefs and social structures of specific ethnic groups have an effect on the levels of child mortality within neighborhoods (Weeks 2006). Therefore, the assumption was that ethnicity will also have an effect on the composition of the physical environment that will be evident from the imagery. The religious composition of an area has also been shown to affect the clustering or segregation of cultural groups and tends to be somewhat linked with ethnicity, providing potential for greater spatial heterogeneity. Housing tenure was used to represent the stability of a household; if a household has a stable ownership or lease, then it is less likely that there will be much change over time from outside sources such as other tenants, owners, or community or government agencies. However, the scarcity of housing and jobs in Accra motivates incoming migrants to temporarily live in the home of a relative, with whom they will live until they can establish themselves (Weeks et al., forthcoming). This implies that the head of

the household may not change, yet the composition of the household itself may and must be controlled for with the migration variable. The combination and interaction of social forces and the formation of social relationships in specific places is theorized to have the ability to produce unique effects upon a region of interest (Massey 1994) and were considered as vital components of the analysis.

Various *indices* derived from the Accra SS, 2000 Census and HAWS were also incorporated. A *possessions index*, based upon the household ownership of various items considered luxuries in slum areas, such as a refrigerator, television, or automobile, complemented household characteristics. A *slum index* created by a prior study in Accra (Weeks 2007) was rescaled to compare the variability of “slumness” between the 37 selected slum-areas. A *housing quality index* (HQI) created by Weeks et al. (2012) and utilized by Stow et al. (forthcoming) was utilized in addition to the slum index. The HQI was derived through a principal components analysis of demographic data from the 2000 census of Accra using dummy variables for characteristics of housing, infrastructure, and measures of household occupant density and is hypothesized to provide comparable results to the household utilities and configuration regression results.

Since it is important to link these household variables to the representative health of the people residing within the selected slum neighborhoods, a body mass index (BMI) was calculated from the height and weight measurements in the respective surveys. Since we are interested in the portion of the population at risk to poor health, the proportion of the sample population that are underweight (BMI < 18.5) and the proportion of people over weight (BMI = 25+) were computed. Although the BMI is criticized for not taking into account other factors influencing height and weight, such as body muscle, the index furnishes a good indication of whether or not variation in more refined indexes may be explained by image metrics.

3.3 Imagery-Derived Variables

High spatial resolution QuickBird satellite images covering Accra, Ghana were captured in April, 2002, and January, 2010. Panchromatic and multispectral image data in blue, green, and red and near infrared spectral bands were collected for both dates. The

spatial resolutions of the images are 0.6 and 2.4 m, respectively, and cover a 121 km^2 portion of the Accra Metropolitan Area. The scenes cover approximately 80% of the AMA region, along with 83% of the region's population. Both images have been georeferenced independently to the Universal Transverse Mercator map projection by a third-party company (i-cubed) at Digital Globe's (QuickBird image vendor) standard processing level (CE90 = 23 m; RMSE = 14 m). Ocean and inland waters were masked prior to image analyses. An empirical line normalization approach was used to radiometrically normalize the two dates of imagery (Yuan and Elvidge, 1996).

The regression analyses were conducted at the EA level due to the spatial constraints of the data collected in the survey. Since the size of an EA is only constrained by the number of residents within its boundaries and the AMA is known to contain neighborhoods with an extreme range of population density, the size of an EA may vary greatly. Therefore data based on EA units may be vulnerable to the ecological fallacy phenomenon known as the Modifiable Area Unit Problem (MAUP) (Openshaw 1984). To somewhat counter the effect the MAUP might have on the dataset, areas that contain large areas of non-built land cover such as bare soil patches and water were masked out during image metric computation. There was enough uninhabited area within EAs along the coast, lagoon shores, rivers, and canals to where image metrics based on image brightness would be skewed from pixels that do not correspond or relate to any data collected in the surveys. A manually digitized mask was created that modified the boundaries of these EAs to exclude these excess non-built areas.

Spectral and texture measures were calculated at the pixel level for each image using ENVI, ERDAS Imagine, and ArcGIS image processing and analysis software packages. These measures were computed for both the 2002 and 2010 scenes. All areas outside the 37 EAs for both images were masked out in order to reduce computational intensity. Image-derived measures for EAs were calculated and extracted using the zonal statistics of pixel values within each EA. The mean and standard deviation of each image metric was computed for each specified zone, providing a representative average value along with a measure of the variance of the image metrics within each EA.

The panchromatic band of the QuickBird imagery was used to generate the second-order texture statistics from the gray level co-occurrence matrix (GLCM). A rotation-

invariant gray level co-occurrence matrix (GLCM) similar to that used by Graesser et al. (forthcoming) and Herold, Liu & Clarke (2003) was used in this analysis. A function is said to have rotational invariance if the calculated values are not subject to variation when arbitrary rotations are applied to the argument. The texture metrics computed from the GLCM included variance, dissimilarity, entropy, contrast, correlation and homogeneity. Kernel sizes used to calculate the GLCM were specified as 3x3 and 5x5 pixels to take advantage of the high spatial resolution of the images. These sizes were specified in order to capture more accurate representations of object edges, as slum neighborhoods tend to be very densely settled. Convolution filters were applied to the panchromatic and near infrared bands of the QuickBird imagery to capture the magnitude and directions of linear features on the ground, which are unique characteristics of the built environment. Sobel, Roberts, Gaussian, and directional filters were applied.

“Proxies” of building density were measured by extracting the linear feature distributions for EAs and calculating the lacunarity of each EA. Since lacunarity is a textural landscape measure, it can serve as a surrogate measure of the variations in building density. Lacunarity was coded and computed using the R programming language software. Linear features were extracted by thresholding the convolved images mentioned above to create a binary image of straight-line features. A threshold value of 250 was applied to generate linear features most representative of building boundaries and edges. The edges of building boundaries were expected to contrast significantly with the surrounding unpaved roads and pathways of the neighborhoods. More densely-built EAs were expected to have lower lacunarity values and a more crowded and clustered linear feature distribution.

Various spectral bands and band ratios were derived and also included in the regression analyses. Individual bands, band ratios, and indices such as the Normalized Difference Vegetation Index (NDVI) were derived from the multispectral images and mean values of all pixels within an EA were calculated for each spectral measure for data consistency. Vegetation proportions per EA were calculated from a simple threshold-based NDVI classification that was demonstrated to be highly correlated with relative housing quality (Stow et al. forthcoming). Proportions of impervious surface per EA were derived from a previously developed object-based vegetation-impervious surface-soil (VIS) classification of the AMA. Vegetation proportions were also extracted from the object-based

VIS classification and compared to other vegetation metrics in order to find the most suitable vegetation predictors. A principal components analysis (PCA) was also performed on both dates of multispectral images to generate spectral transform measures that tend to relate to brightness/albedo and relative image greenness. This technique uses an orthogonal transformation to convert a set of spectral bands containing groups of possibly correlated brightness values (pixels) into a set of values of linearly uncorrelated spectral variables called principal components. A PCA allows for the definition of a minimal set of non-redundant channels (components), proper for discrimination studies in image processing (Ceballos & Bottino 1997). There were a total of three principal components included in the analysis: 1) first principal component (PC1), representing a spectral transform image representative of image brightness, albedo, and reflectiveness; 2) second principal component (PC2), from which the transform extracted information on vegetation brightness through the identification and discrimination of spectral vegetation reflectance, primarily contained within the Green, Red, and Near-Infrared bands; and 3) third principal component (PC3), which identified regions of homogeneous spectral values in the form of densely-built housing structures.

To serve as additional predictors, binary classifications of vegetation and impervious surface were analyzed in Fragstats, a spatial pattern analysis program designed to compute a wide variety of landscape metrics for categorical map patterns. Patch richness and patch density landscape metrics were generated using the binary classification products. Stow, Lippitt & Weeks (2010) demonstrated that mean vegetation patch size and the proportion of vegetation patches per EA are valuable in the delineation of neighborhood boundaries in Accra, and therefore were explored to see whether they had a significant impact on the responses of survey variables within EAs. All aforementioned imagery-derived metrics (IMs) will serve as *independent* or *predictor variable* inputs in the regression analysis.

Regression Model Inputs			
Image Metrics		Survey Variables	
Spectral Metrics	Texture Measures	Slum Survey Data	2000 Census Data
<i>Bands, Ratios, Veg Indices</i>	<i>GLCM (3x3, 5x5)</i>	<i>Infrastructure</i>	<i>Socioeconomic</i>
Blue	Mean	Electricity	Ethnicity
Green	Variance	Cooking Fuel	Migration
Red	Homogeneity	Disposal of Waste	Religion
Near Infrared (NIR)	Contrast	Drinking Water Supply	<i>Indices</i>
NDVI	Dissimilarity	Type of Toilet	Housing Quality (HQI)
NIR/Red	Entropy	<i>Socioeconomic</i>	Slum Index
NIR/Blue	Second Moment	Housing Tenure	
Red/Green	Correlation	Bednet Presence	
Green/Blue	<i>Convolution Filters</i>	<i>Indices</i>	
<i>Principle Components</i>	Sobel	Possessions Index	
Bands 1-4	Roberts	Telephone	
<i>Vegetation Binary</i>	Directional (0-90)	Refrigerator	
<i>Impervious Binary</i>	Gaussian	Radio	
<i>Landscape Metrics</i>	<i>Linear Feature Density</i>	Television	
Patch Density	<i>Lacunarity</i>	Automobile	
Mean Patch Size		<i>Under-5 Indices</i>	
Euclidean N-N Distance		Height-Weight	
<i>Elevation (DEM)</i>		Weight-Age	

Table 1. List of variable inputs into the exploratory regression model.

3.4 Exploratory Regression Modeling

The study was largely exploratory in nature in that there were a total of 29 different image-derived metrics used in multiple combinations to attempt to explain 27 different household data characteristics. The EA, or enumeration area, comparable to an American census tract, was used as the spatial unit of analysis for both the 2003/2002 and 2009/2010 datasets. An exploratory bivariate regression was conducted using the IBM SPSS (Statistical Package for the Social Sciences) and the spatial statistics toolset in ArcGIS software. Pearson product-moment correlations were computed so as to explore all possible combinations of image metrics (IMs) and survey variables (SED) for a total of 783 correlation analyses. Pearson's r measures the strength of a correlation (linear dependence) two variables X and Y . It is calculated using the equation

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

where \bar{X} and \bar{Y} represent the means of the two variables and the denominator can be described as the product of the sample standard deviations for X and Y . All bivariate

correlations that were statistically significant at the $\alpha = 0.05$ level were used in exploratory ordinary least squares (OLS) multivariate regression models in both SPSS and ArcGIS. The OLS general model was then computed for each combination and is specified in the following equation:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

where x_i is an independent or predictor variable, y_i is the dependent or response variable, and ε_i is the error term. The β coefficients express the average change in y_i for each unit change in x_i . Using this model, the independent variable inputs were metrics derived from the imagery, where the dependent variables included each variable from the survey data sets (see Figure 2). The version of exploratory regression that was implemented is similar to a stepwise regression, but rather than looking for models with high R^2 values, it determines which models pass specific regression diagnostics that are set as model parameters along with the basic assumptions of an OLS model. Various thresholds for a minimum acceptable R^2 value and a minimum coefficient p-value cutoff enable the exploration of various model fits. A maximum VIF (variance inflation factor) value, quantifying the severity of multicollinearity in the OLS regression, was set at 4 to sift out over-specified models where ≥ 2 IMs might have been highly intercorrelated. A threshold for the minimum Jarque-Bera (JB) statistic p-value was set to $\alpha = 0.1$, testing for goodness-of-fit with respect to skewness and kurtosis. The Jarque-Bera statistic is calculated as follows:

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4} (K - 3)^2 \right)$$

where n is the number of observations (degrees of freedom, df), S is the sample skewness, and K is the sample kurtosis (see Jarque & Bera 1987). In classical statistics, a statistically significant Jarque-Bera probability would indicate the presence of a non-normal distribution of error terms in the form of skewness, kurtosis, and heteroskedasticity. However, since this study is trying to explore the inter-slum variability in household characteristics through spatial regression techniques, a non-normal distribution of error terms provides a valid indication of spatial effects. Therefore, models with significant JB probabilities will not be discarded, but rather explored through a switch in model specification. A minimum

acceptable p-value for Global Moran's I was also used to explore initial spatial autocorrelation in the residuals, another indicator of the need for a spatial model specification.

Each model was ranked according to the model fit, measured by their respective adjusted coefficient of determination (R^2) and Akaike Information Criterion (AIC_c) values. The adjusted R^2 value accounts for the number of predictor variables input into the model along with the sample size (number of EAs) of each, and will decrease below the original R^2 value if an explanatory variable is included that does not aid in the prediction of the dependent variables. R^2 is calculated as

$$R^2 = 1 - \frac{n-1}{n-k} (1 - R^2)$$

where R^2 is the proportion of total variation of the dependent variable explained by the predictor variables, n is the sample size, and k is the number of predictor variables. The higher the R^2 value, the more explanatory power the model has. The AIC_c is another goodness of fit measure and is an asymptotically unbiased estimator of the information lost when model g is used to estimate model f (Akaike 1973, 1974). The AIC_c is a function of a model's maximized log-likelihood (ℓ), the number of estimable parameters (K), and a second-order term defined by Hurvich and Tsai (1989) that accounts for the observation sample size. AIC_c is represented mathematically as

$$AIC_c = -2\ell + 2K + \frac{2K}{n-K-1}$$

and is designed to estimate the predictive accuracy of competing model hypotheses when the sample size is small compared to the number of parameters, as is the case with our dataset (Posada & Buckley 2004).

3.5 Spatial Heterogeneity & Regime Analysis: Searching for Structural & Spatial Instability

In the presence of spatial heterogeneity and autocorrelation within the Accra EAs, a combined spatial error/lag and spatial regimes regression technique similar to that demonstrated by Curtis, Voss & Long (2012) was used to account for what Anselin (1996) referred to as “the intrinsic uniqueness of each location.” Spatial autocorrelation and heterogeneity have the potential to affect inferences in a cross-section of spatial units, and furthermore can result in a nonspherical error variance, nullifying standard hypothesis tests (Anselin 1990).

Results from the exploratory regression were transferred into OpenGeoDa in order to determine the correct spatial model selection using a series of Lagrange Multiplier (LM) tests. One directional LM tests (LM[lag] and LM[error]) and their more robust forms (Robust LM[lag] and Robust LM[error]) were computed for each multivariate model in order to form a probability-based decision rule as follows:

- a) if neither LM test rejected the null hypothesis ($\alpha \geq 0.05$), no spatial regression was run;
- b) if one of the LM tests is significant, proceed with the spatial model specified by the respective LM test;
- c) if both LM tests reject the null hypothesis, refer to the significance of the Robust LM tests and proceed with the respective spatial model indicated by the most significant Robust LM probability.

A spatial weights matrix was generated for the spatial regression techniques using an inverse Euclidean distance band, assisting in the quantification of the spatial relationships that exist between EAs. An inverse Euclidean distance band weight computes a kernel weight that decays exponentially with increasing Euclidean (straight-line) distance between features. Row standardization was used to account for the spatial sampling bias of the EAs. A bandwidth of 3.6 km selected as the critical distance that ensured all EAs had at least one neighbor. This weights matrix was used in the spatial error/lag models to identify any spatial dependency of the values of the dependent variable on either the “neighboring” EAs (spatial lag) or a spatial dependency revealed within the error terms (spatial error). A spatial lag

model incorporates spatial effects by including a spatially lagged dependent variable in the regression as an additional predictor, modifying the OLS model to

$$y_i = \rho W y + \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

where Wy is the spatially lagged variable for the weights matrix W and ρ is the spatial coefficient. The lag model treats spatial correlation as a process or effect of interest, where the values of y in one unit of analysis are directly influenced by the values of y in neighboring units. Conversely, the spatial error model examines the spatial autocorrelation between the residual error terms of adjacent areas. The error model is specified as

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

with

$$\varepsilon_i = \lambda W \varepsilon_i + \xi$$

where λ is the spatial error coefficient and ξ is a vector of uncorrelated error terms. This model treats spatial autocorrelation as a nuisance and disregards the idea that spatial correlation may reflect some meaningful process and treats error as an effect of the model misspecification of independent variables. AIC_c values were calculated and used to distinguish whether the global spatial regression was an improved model fit. A decrease in the AIC_c by at least 3 was considered to be the threshold indicative of an improved model (Fotheringham, Brundson & Charlton 2002).

The spatial regimes approach involved regressing independent variables with the dependent variable whose regression coefficients designated distinct “regimes” to create a structurally stable model. Regimes were characterized by clustered ranges of coefficient values in particular spatial units – in this case, EAs. This separated the dataset into discrete regions made up of spatially-grouped EAs for which an OLS regression was run separately, using the same predictors. In considering a two-regime model where the observations composing the regimes have been considered a priori into one of the two regimes, the model notation would be expressed as

$$\begin{matrix} y_i \\ y_j \end{matrix} = \begin{matrix} X_i & 0 \\ 0 & X_j \end{matrix} \begin{matrix} \beta_i \\ \beta_j \end{matrix} + \begin{matrix} \mu_i \\ \mu_j \end{matrix},$$

where i and j designate the distinct regimes. The spatial regime method allowed for the testing of the overall model fit as well as the specific stability in the residual estimates when the unit of analysis is theoretically, or in this case, also physically bounded (Curtis, Voss & Long 2012). We hypothesized that if each regime was associated with a particular region of slum neighborhoods, the spatial regimes approach would then essentially become a test for regional homogeneity (Anselin 1990). The combination of an appropriate model specification for a process as well as a set of carefully chosen covariates has been demonstrated to provide enough to explain the variation in the differential values of a spatial variable for a specific region of analysis (Anselin 1996).

3.6 Comparing Model Results to a Geographically Weighted Regression

Spatial regression results were compared to results provided by a geographically weighted regression (GWR) model. In a GWR model, the relationships between the unit under observation and its neighboring spatial units are considered based upon a specified contiguity matrix. The GWR model specification provides a local version of spatial regression or process by fitting a model to each unit of analysis, incorporating explanatory variables that fall within a specified bandwidth or kernel definition. Local coefficients are estimated for each independent variable included in the regression, of which the magnitude is indicative of the degree of contribution in explaining local variation in the dependent variable (Fotheringham, Brundson & Charlton 2002). The GWR model is specified as

$$y_i = \rho(u_i v_i) W y + \beta_0(u_i v_i) + \beta_1(u_i v_i) x_{1i} + \varepsilon_i(u_i v_i)$$

where a unique set of parameters is computed for each observation i at a set of geographic coordinates $(u_i v_i)$. Due to the lack of a completely contiguous dataset and for the ease of comparison with the global models, the areas in between EAs in this case were considered empty space. Therefore, neighbors will be determined in this case by the bandwidth specified by the spatial error/lag regression previously described. Comparing global spatial model results to a local GWR model was hypothesized to provide an additional measure of

the robustness of the results. Clustering of residuals was evaluated at the local level using local Moran's I.

3.7 Evaluating Temporal Robustness using the Bootstrapping procedure

After the best statistically significant multivariate correlations were established between imagery metrics and household variables, the robustness of the regression models was evaluated. The same modeling procedures run for the 2000-2003 data (t_{2002}) were used to build multivariate models for t_{2010} and Δt . When comparable passing models were constructed, a bootstrapping procedure elucidated by Efron (1979) was utilized to create a distribution of R^2 values for the t_{2002} , t_{2010} and Δt models in order to observe the range of model fits for each multivariate regression. The bootstrap approach is based upon random sampling with replacement, where specific observations may or may not be sampled multiple times within a model. The advantage of bootstrapping is that the range of R^2 values may be analyzed to test for model robustness and stability (Efron 1979). Smaller ranges in R^2 values infer that models are more robust in their prediction of values and have a strong potential to indicate socioeconomic conditions and demographic characteristics based on metrics derived from high-resolution satellite imagery.

3.8 Investigating Change Over Time: The Search for Proxies

The final step in the analysis of image metrics as indicators of ground survey data was to examine how strongly correlated changes in image metrics (ΔIMs) were to changes in the socioeconomic and demographic data ($\Delta SEDs$). The main goal of this component was to investigate the “predictive” (indicative) ability of ΔIMs with respect to $\Delta SEDs$, creating proxy measures of the household environment using imagery. The same processes run for the static analysis of the image metrics were utilized, replacing the temporally static

components with Δ variables in the regression. The models then followed the formula

$$\Delta y_i = \beta_0 + \beta_1 \Delta x_{1i} + \varepsilon_i,$$

where Δy_i and Δx_{1i} would represent ΔSED_i and ΔIM_i , respectively. The bootstrapping approach was applied to reveal the most robust multivariate models, indicating which ΔIMs had the highest potential for use as proxy measures of ground data.

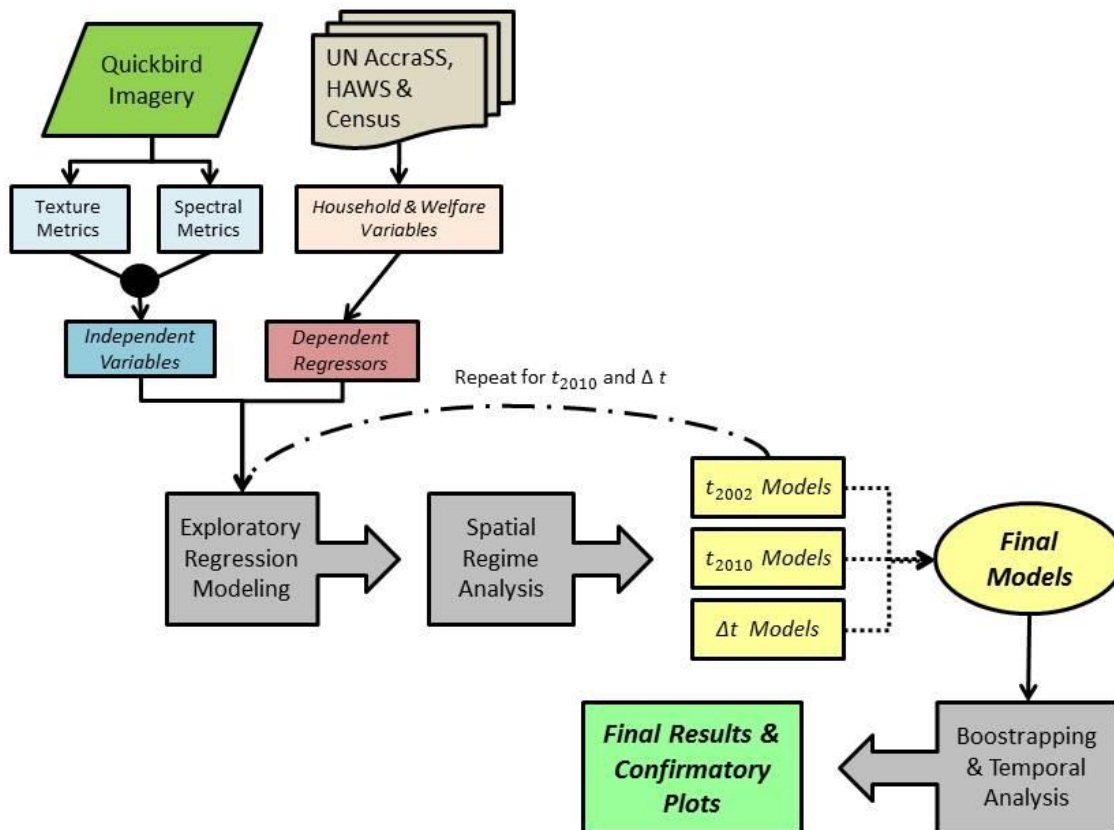


Figure 2. A flowchart of the methods used in the analysis of the imagery and survey data.

CHAPTER 4

RESULTS

Given that the study was conducted primarily in an exploratory manner, the results presented here will focus largely upon the models that provided the strongest and most significant results. These results will be discussed in a stepwise method, following the methodological modeling procedures outlined in the prior sections.

4.1 Exploratory Bivariate Regression Analysis and OLS Modeling

The exploratory bivariate correlations indicated that there are a large variety of image metrics that are able to explain proportions of the variance in demographic and housing quality variables. Table 2 presents the results from the multivariate models formulated through the exploratory combinatory multivariate analysis for the 2002 image. All models for each temporal segment were examined for multicollinearity through the comparison of VIFs (variance inflation values) and condition numbers (CN). Models with any existing multicollinearity between two or more independent variables were discarded. The Breusch-Pagan and Jarque-Bera tests were performed to explore whether each selected model contained residuals with a non-normal distribution or any evidence of heteroskedasticity, indicators of the need for spatial model specifications.

4.1.1 2002 MODELS

Eleven of 19 2002 OLS multivariate models were able to explain over 20% of the variance in their respective survey variables using 14 different image metrics as the predictors. All 19 models were significant at the confidence level of $\alpha = 0.05$ with all predictor variables meeting the same α -level. The *disposal of trash through collection* ($R^2 = 0.78$) and *local offsite dumping* ($R^2 = 0.76$) had the highest correlations. Over 75% of the variance in trash disposal methods can be accounted for using mean values of three metrics – a) first principal component (PC 1) of the 2002 multispectral image, which upon visual

interpretation is a representation of scene brightness (i.e., albedo); b) third principal component (PC 3), a spectral proxy representation of vegetation amounts in this case; and c) average elevation, derived from a high spatial resolution DEM (digital elevation model). The coefficient relationships for trash disposal models suggest that EAs that mainly undergo systematic trash collection are located within less densely settled neighborhoods with more vegetation, containing houses constructed with less reflective rooftop building materials (e.g. – non-metal materials) and built at relatively higher elevations. EAs containing residents that primarily dispose of household trash in an offsite location tend to have an inverse relationship. Traditionally, EAs situated at higher elevations belong to residents of a higher socioeconomic status and are a few kilometers inland from the coast. These areas have primarily been settled by the Ga ethnic group, which was historically known as a coastal fishing culture that tends to live toward more low-lying coastal regions more at risk to flooding and sewage runoff, putting these populations at a higher risk for disease (Rain et al. 2011)(Weeks et al. 2006).

Moderately strong correlations were found between brightness values in the Blue band of the multispectral imagery and proportions of households in each EA whose primary disposal of sewerage is either through a Kumasi ventilated pit latrine (KVIP; $R^2 = 0.34$) or through a method other than a flushing toilet or a KVIP ($R^2 = 0.38$). An inverse relationship between average Blue band brightness and the proportion of households that dispose of their sewerage through a method other than a flushing toilet or KVIP, where if an increase in average EA Blue band brightness values is observed, a decrease in % *Toilet (KVIP)* would be expected. Other methods of sewerage disposal may include traditional pit toilets, buckets or pans, or no formal toilet facility at all. Higher Blue band brightness in the urban slum neighborhoods of Accra correspond to rooftops that are constructed from highly reflective building materials like slate, most commonly observed in the coastal slum regions of Accra, whereas less reflective building materials are more commonly found in the older, more established inland slum regions surrounding the neighborhood of Nima (see Appendix 1), where most roofs are made of corrugated metal that has rusted or faded over time. Almost intuitively, an inverse relationship was observed between the proportion of households that use KVIPs as the primary type of toilet facility and Blue band brightness. These results provide a very good indication of a severe health risk within EAs that do not primarily use

flushing or KVIPs for sewerage, as the stagnant decomposition of sewerage is closely related to various communicable diseases, insects that may act as vectors of these diseases, and poor sanitation conditions for food preparation and sleeping quarters.

Approximately 45% of the variance in the proportions of households of the Ga ethnic group in each EA was explained using a combination of the spatial variance of pixels in the near-infrared (NIR) band brightness values and average pixel values per EA derived from a Laplacian-filtered panchromatic image. Both predictors were statistically significant at the $\alpha = 0.01$ and $\alpha = 0.05$ confidence levels. With an increase in the variance of NIR brightness values and a decrease in the average Laplacian-filtered pixel value, more simply stated as a decrease in building density, we would expect to observe an increase in the proportion of households belonging to the Ga ethnicity. Promising results were also apparent in the models for both the % of the sample population of both the Christian and Islamic populations with $R^2 = 0.59$ and $R^2 = 0.71$, respectively. The textural metric of image lacunarity was included as a statistically significant predictor at the $\alpha = 0.01$ significance level in both OLS models. EAs with higher average lacunarity values are considered to be regions that contain pixels that are more heterogeneous to their respective surrounding pixels and provide a measure of image “gappiness”, another proxy measure of building density. These results show that with an increase in building density comes an increase in the % of households per EA belonging to the Christian religion – the inverse standing for the % of Muslim households per EA. This result is consistent with prior studies that have demonstrated that EAs in Accra with higher building density are moderately correlated with lower housing quality (Tsai et al. 2012), and neighborhoods with higher concentrations of the population that are non-Christian are at a larger risk for high levels of child mortality (Weeks et al. 2006).

The regression results also indicated very strong correlations between various textural transformations and the spatial distribution of religious groups (% of residents that are of the Christian and Islamic faith, $R^2 = 0.59$ and $R^2 = 0.71$, respectively) within the specific slum neighborhoods of Accra. Through a combination of the variance in image pixel correlation (a component of the GLCM), and mean EA values of both lacunarity, a textural transformation that provides additional information on building density through what is

Ordinary Least Squares (OLS) Regression Results & Diagnostics - 2002 Models

Dependent Variable	OLS R-squared	Model Predictors	JB p-value	BP p-value
Electricity	0.14	-0° Filter**	0.40	0.89
Charcoal Use	0.16	-Lacunarity**	0.82	0.12
Trash (collected)	0.78	-PC 1*** Elevation**	0.78	0.07
Trash (dumped offsite)	0.76	PC 1*** -Elevation***	0.59	0.06
Trash (burned or buried)	0.16	Red/Green**	0.00	0.00
Toilets (KVIP)	0.34	-Blue***	0.28	0.61
Toilets (no flush or KVIP)	0.38	Blue**	0.37	0.48
Bednet Use	0.42	-Entropy*** 0° Filter***	0.39	0.89
Possessions Index	0.16	-45° Filter**	0.43	0.73
Ethnicity (2000 Census)				
% Akan	0.22	Green/Blue***	0.99	0.36
% Ga	0.45	NIR***	0.35	0.83
% Mole-Dagbani	0.19	-PC 1*** -Laplacian**	0.00	0.01
Religion (2000 Census)				
% Christian	0.59	-Correlation*** Lacunarity***	0.59	0.83
% Muslim	0.71	Dissimilarity*** -Blue***	0.25	0.33
Migrants(%)	0.14	Entropy** -Lacunarity***	0.00	0.05
2000 Slum Index	0.24	Homogeneity** NDVI**	0.48	0.75
2000 Housing Quality Index	0.24	NDVI***	0.65	0.16
% Underweight (BMI <18)				
% Overweight (BMI = 25+)				

** Significant @ $\alpha = 0.05$
 *** Significant @ $\alpha = 0.01$
 roman - EA mean value
 italics - EA standard deviation
 JB - Jarque-Bera test for residual normality
 BP - Breusch-Pagan test for heteroskedasticity

Table 2. 2002 OLS model results, including \bar{R}^2 values, model predictors variables, and tests for residual normality and heteroskedasticity.

Selected Image Metrics – 2002 Models

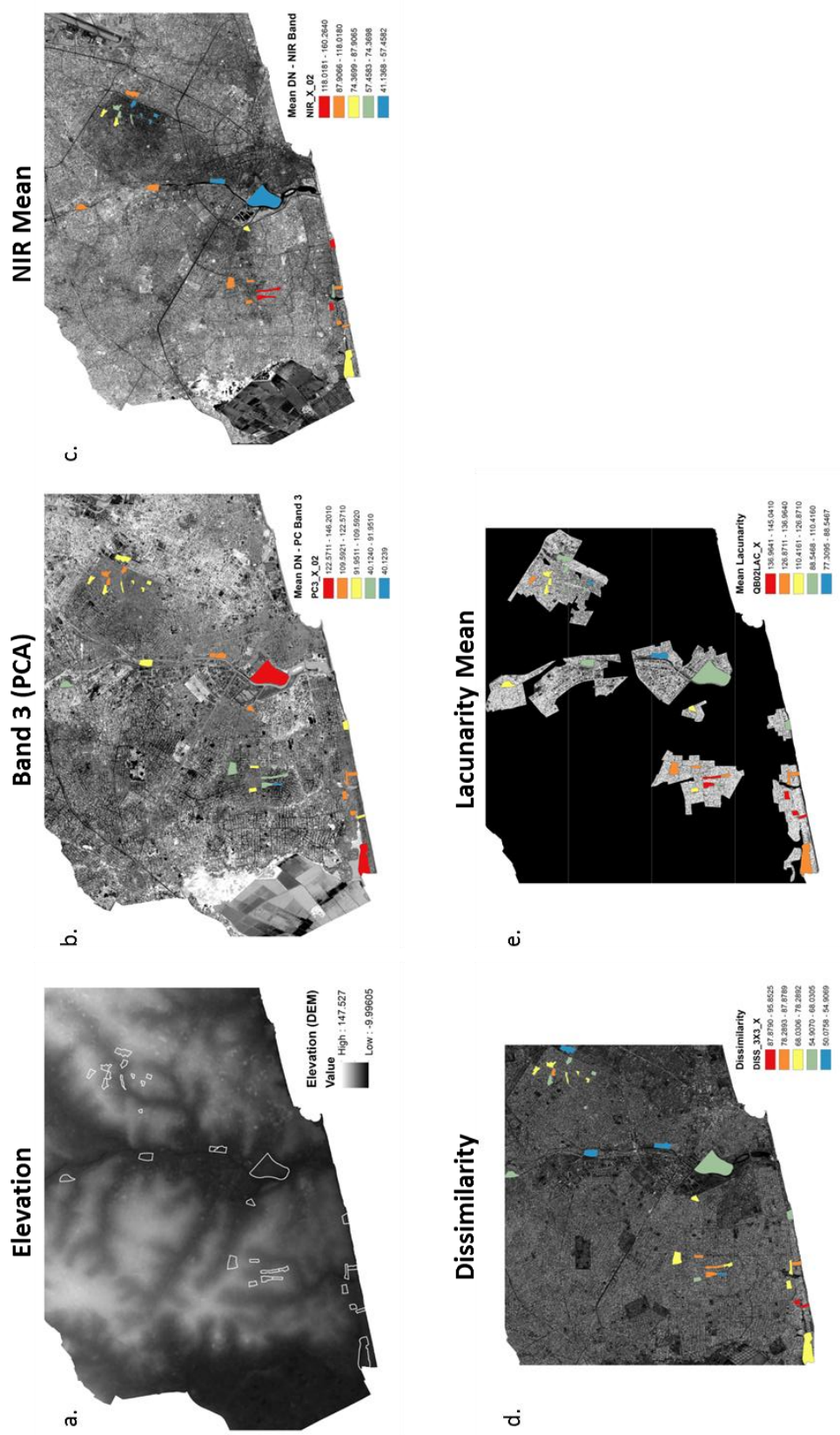


Figure 3. Selected image metrics included in models with the best results. EA values are displayed in a diverging color scheme with higher values displayed in red & lower values in blue. Images from which each metric was derived from is displayed in each respective backdrop.

Diagnostics for spatial dependence in 2002 OLS model - Lagrange Multiplier Tests

Dependent Variable	Global Moran's I		Moran's I		1-directional LM	
	Value	p-value	Lag	Error		
Electricity	-0.05	0.82	0.39	0.46		
Charcoal Use	-0.05	0.83	0.39	0.47		
Trash (collected)	-0.01	0.16	0.13	0.87		
Trash (dumped offsite)	-0.03	0.35	0.53	0.70		
Trash (burned or buried)	0.01	0.29	0.28	0.79		
Toilets (KVIP)	-0.08	0.45	0.34	0.27		
Toilets (flush or KVIP)	-0.10	0.28	0.23	0.20		
Bednet Use	-0.01	0.65	0.91	0.83		
Possessions Index	0.06	0.09	0.54	0.44		
Ethnicity (2000 Census)						
% Akan	-0.01	0.21	0.86	0.91		
% Ga	-0.07	0.66	0.74	0.33		
% Mole-Dagbani	-0.02	0.31	0.60	0.81		
Religion (2000 Census)						
% Christian	-0.03	0.63	0.32	0.68		
% Muslim	0.01	0.10	0.02	0.93		
2000 Slum Index	-0.05	0.97	0.32	0.52		
2000 Housing Quality Index	-0.01	0.61	0.66	0.95		
% Underweight (BMI <18)						
% Overweight (BMI = 25+)						

Table 3. Diagnostics for spatial dependence and spatial heterogeneity for the t_{2002} models – the Global Moran's I and Lagrange Multiplier statistics and probabilities.

OLS Standardized Residuals – 2002 Models

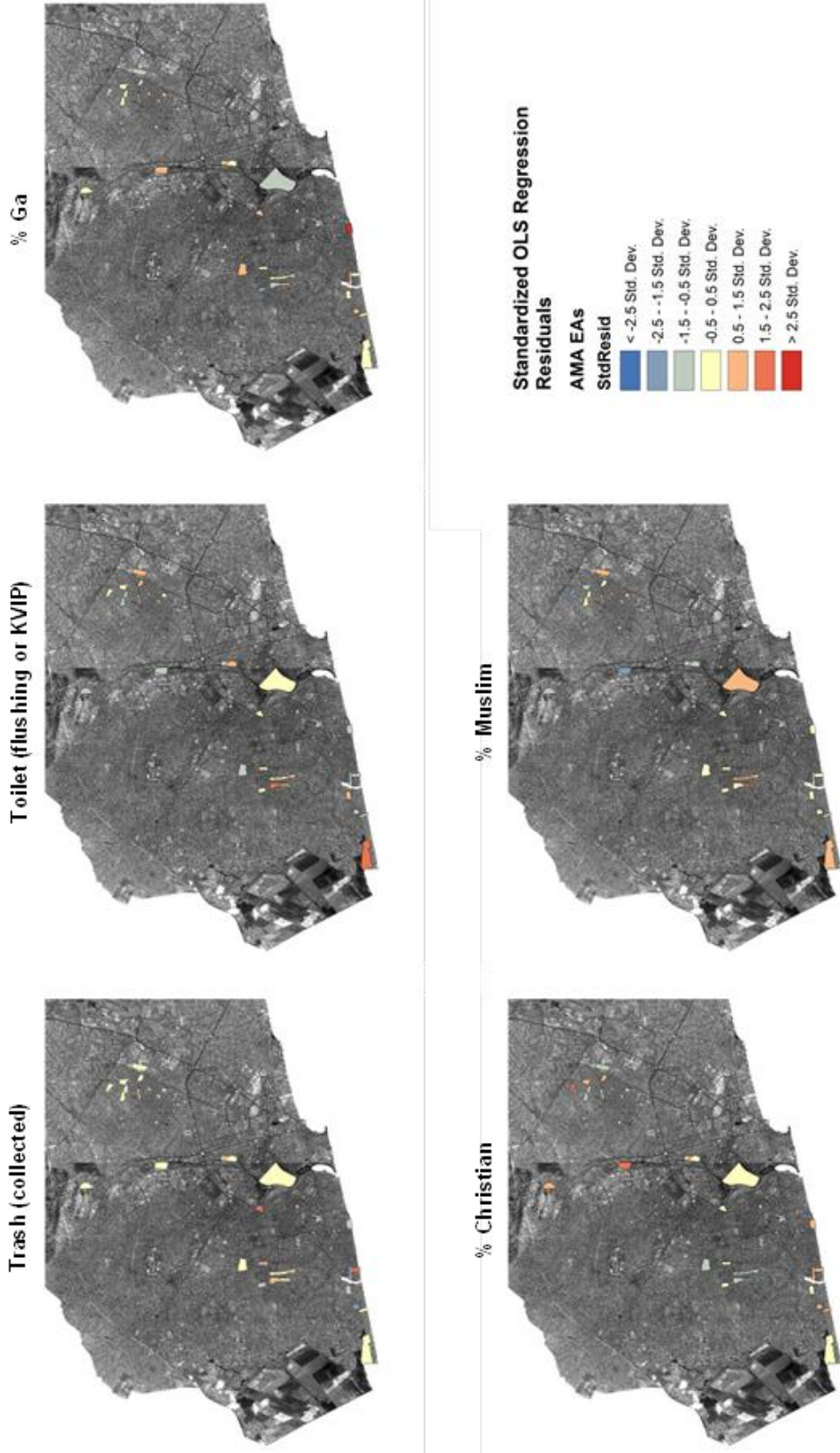


Figure 4. Standardized residuals for selected OLS models. The 2002 panchromatic image is displayed in the background.

known as “image gappiness” (Myint & Lam 2005), and a convoluted panchromatic image using a Sobel edge-detection filter to capture a proxy indicative of building density.

A majority of independent variables had regression coefficients that were below 0.01, indicating that only large increases or decreases in image metric values would have significant relationships with the variability in the changes in survey variables. Standardized residuals for a selection of models with the strongest correlations are presented in Figure 2. Severe over- and under-prediction of standardized residuals was rarely observed. A series of Lagrange Multiplier (LM) tests were performed to detect the presence of spatial dependence of spatial heterogeneity, with the results presented in Table 3. Both one-directional and robust versions of the LM [lag] and LM [error] tests were run to test if the models were missing a spatially lagged dependent variable or if any spatial error dependence existed. For the 2002 series, only one model indicated the presence of spatial effects, % *Muslim*. A statistically significant LM [lag] value suggested that a spatial lag model be specified to account for the effects of neighboring EAs on the dependent variable.

4.1.2 2010 MODELS

For the 2010 series, only 11 models passed all diagnostic tests, with only six of those 11 accounting for > 20% of the variance in their respective survey variables using the image metrics. Both the *collected trash* and *trash burned or buried onsite* models produced evidence of heteroskedasticity and had non-normally distributed residuals, failing both the Jarque-Bera and Breusch-Pagan significance tests. The % *Ga* per EA model exhibited a moderately strong correlation as for the 2002 data, with the combination of average Red band and Simple Ratio vegetation index (SR) brightness values explaining 48 % of the variance. Both independent variables established positive relationships between brightness and the % *Ga* per EA, confirming the inverse relationship between the proportion of *Ga* and the amount of vegetation per EA established in the 2002 model. Models for the % *Christian* and % *Muslim* per EA produced very strong results, with $R^2 = 0.41$ and $R^2 = 0.56$, respectively. A relatively high amount of variance in both models was explained using average values from the Green/Blue band ratio alone. The % of *one room households* per EA, a variable not present in the Accra SS dataset, was included in the 2010 series of models based on the idea that EAs with a higher proportion of one-room households are hypothesized to be built in

overcrowded neighborhoods with higher building density. About 21 % of the variance in the % of one-room households was explained by the variance in EA vegetation, demonstrating the positive relationship between vegetation amounts and densely populated EAs.

A moderately strong relationship was also produced for the model for % *overweight* ($BMI = 25+$), with a Laplacian-filtered image (related to building density) accounting for 29% of the variance. As mean EA Laplacian values increase, we can expect to observe an corresponding increase in both building density and the proportion of the population with BMIs of greater than 25. A standardized β - coefficient of 1.91 for the Laplacian values variable show that as for each 1% increase in mean Laplacian values, the percentage of people considered overweight increase by about 2%. Both Jarque-Bera and Breusch-Pagan tests were statistically significant ($p < 0.000$) and were examined under a GWR model that is discussed later. A positive relationship between vegetation proportions and % *underweight* was established, with EA vegetation proportions explaining 13% of the variance in the distribution of the underweight population of the AMA. These results invite further investigation into the relationship between health variables, building density, and vegetated land cover.

Overall the indicative power of high-spatial-resolution image metrics from 2010 were lower than those for the 2002 image metrics. This may be attributed to the densification of settlements shown in Tsai et al (2012) and loss in vegetation proportions demonstrated in Stow et al. (forthcoming), creating a more heterogeneous physical environment to be observed within the slum neighborhoods. This could also be attributed to the sampling error and uncertainty between the two studies. Only 6 of 12 survey variables from the 2009-2010 HAWS dataset that were comparable between surveys produced passing models, compared to 11 of 12 from the 2003 Accra SS. This may indicate that even though the sampling design and questionnaire from the HAWS was modeled directly after the Accra SS, the effects of selection bias may be present and skewing the results.

Once again, most predictor variables had miniscule coefficient values below 0.10. Diagnostics for spatial dependence indicated that all passing models were subject to spatial heterogeneity. The Global Moran's I statistics and probabilities provide no indication of spatial autocorrelation in the dataset. However, statistically significant probabilities for the

Ordinary Least Squares (OLS) Regression Results & Diagnostics - 2010 Models

Dependent Variable	OLS		Model Predictors		JB		BP	
	R-squared				p-value		p-value	
Electricity	0.11		-Dissimilarity**		0.66		0.43	
Charcoal Use	x				x		x	
Trash (collected)	0.13		-Red**		0.00		0.06	
Trash (dumped offsite)	x				x		x	
Trash (burned or buried)	0.13		-Red**		0.00		0.06	
Sachet Use	0.24		-Correlation***		0.66		0.96	
Toilets (KVIP)	x				x		x	
Toilets (flush or KVIP)	x				x		x	
Bednet Use	x				x		x	
Possessions Index	x				x		x	
Ethnicity (2000 Census)								
% Ga	0.45		Red***		0.28		0.07	
% Ewe	0.13		Sobel/Filter**		0.68		0.42	
Religion (2000 Census)								
% Christian	0.41		Green/Blue***		0.63		0.86	
% Muslim	0.44		-Green/Blue***		0.62		0.87	
Migrants(%)	0.12		-45° Filter**		0.12		0.11	
% One Room Households	0.21		-Vegetation***		0.15		0.87	
% Underweight (BMI <18)	0.13		Vegetation**		0.45		0.92	
% Overweight (BMI = 25+)	0.29		Laplacian Filter***		0.00		0.00	

x - No passing models
 ** Significant @ $\alpha = 0.05$
 *** Significant @ $\alpha = 0.01$
 roman - EA mean value
 italics - EA standard deviation
 JB - Jarque-Bera test for residual normality
 BP - Breusch-Pagan test for heteroskedasticity

Table 4. 2010 OLS model results, including \bar{R}^2 values, model predictors variables, and tests for residual normality and heteroskedasticity.

Diagnostics for spatial dependence in 2010 OLS model - Lagrange Multiplier Tests

Dependent Variable	Global Moran's I Value	Moran's I p-value	1-directional LM probability		Robust LM probability	
			Lag	Error	Lag	Error
<i>Trash (collected)</i>	0.01	0.27	0.36	0.96		
<i>Trash (burned or buried)</i>	-0.03	0.69	0.36	0.69		
<i>% One Room Households</i>	0.12	0.01	0.09	0.08	0.01	0.00
<i>Ethnicity (2000 Census)</i>						
<i>% Ga</i>	-0.07	0.51	0.70	0.29		
<i>Religion (2000 Census)</i>						
<i>% Christian</i>	0.02	0.05	0.31	0.75		
<i>% Muslim</i>	-0.02	0.06	0.27	0.78		
<i>Migrants(%)</i>	-0.12	0.10	0.08	0.10		
<i>% Underweight (BMI <18)</i>	-0.08	0.34	0.20	0.24		
<i>% Overweight (BMI = 25+)</i>	-0.06	0.63	0.33	0.50		

Table 5. Diagnostics for spatial dependence and spatial heterogeneity for the t_{2010} models – the Global Moran's I and Lagrange Multiplier statistics and probabilities.

one-directional LM [error] and Robust LM [error] tests (displayed in Table 5) signify that spatial heterogeneity is present in the form of heteroskedasticity.

4.1.3 Δ MODELS

Fourteen of 19 Δ models had the indicative power to explain > 20% of the variance in the changes in survey variables between 2003-2010 using satellite image metrics, with 11 of 19 multivariate models accounting for > 30% of the variance alone. A table is presented in Appendix 2, listing the amount of change observed in each EA included in the study. Model results and diagnostics for residual normality and heteroskedasticity are presented in Table 6. A total of five models had non-normal residual distributions, providing an indication of the possible spatial effects that might be in play and suggesting further investigation of model specification through a more localized model.

As with the 2002 models, the changes in image metrics were moderately correlated with changes in sewerage variables, more specifically, the proportion of households that primarily utilize KVIPs and the proportion of households whose disposal of sewerage is through neither a flushing toilet nor a KVIP. As the % vegetated land cover and average image brightness decrease, an increase in households that do not dispose of sewerage through flushing toilets or KVIPs is observed ($R^2 = 0.38$). EAs with higher proportions of the population not using either a KVIP or flushing toilet are located in the older, more established slum regions surrounding the neighborhood of Nima, where the addition of formal sewage infrastructure may not be possible. These slums are also very densely built with little to no vegetation existing, supporting the results of the Δ model.

Change in % of households with collected trash and *% of households with trash dumped offsite* reappeared on the list of models with moderately strong correlations ($R^2 = 0.32$, $R^2 = 0.39$, respectively), with *change in % trash dumped offsite* model including the third principal component (PC3) as a predictor variable. Moderately strong results were produced by the Δ % *Christian* ($R^2 = 0.35$) and Δ % *Muslim* ($R^2 = 0.31$) models once again. The Green/Blue band ratio was included as an independent predictor, yet with less statistical significance. The remaining predictors (second principal component, % vegetation) establish a relationship between religious affiliation and EA vegetation proportions, not yet produced by model results in the previous two time series datasets. The positive relationship between

% Christian and % vegetation suggest that EAs with higher proportions of vegetated land cover have higher proportions of the population that are Christian. This is confirmed by the inverse relationship between Δ % Christian and Δ second principal component (PC2) brightness values, where lower Δ PC2 values are representative of higher amounts of vegetated land cover. The relationship between Δ % Muslim and Δ PC2 indicate the exact opposite connection to EA vegetation fractions.

Figures 5-7 present three image metrics that were included as significant model predictors that best represent the changes in the imagery between 2002-2010. There was a general increase in values for image lacunarity (Figure 5) over the 8-year time span, reflecting the densification of the built environment within slum neighborhoods, mostly driven by an increase in population that demands the construction of new, informal housing structures. The neighborhood of Sodom & Gomorah (Figure 5, outlined), along with the surrounding Korle Lagoon area, has more recently been a focal point of incoming migrants from the more northern regions of Ghana and surrounding countries in the Sahel region of West Africa (Rain et al. 2011). This is demonstrated by an increase in the composition of migrants in Sodom & Gomorah from 32% to approximately 79% of the population, a 146% increase over the course of 10 years. Migrants in Sub-Saharan Africa have a higher health risk than the rest of the population, as they tend to live in overcrowded neighborhoods at lower elevations that are prone to flooding and other health hazards (Rain et al. 2011; Sverdlik 2011; UNHSP 2003). New building development and building densification was found to correlate moderately to housing quality (Tsai et al. 2012). Therefore, it is vital to track the changes in the composition of migrants in slum neighborhoods over time in order to monitor and manage levels of morbidity and poor socioeconomic status.

Figure 6 displays the 2nd principal component (PC2) of the multispectral images at both dates. Lower brightness values in PC2 correspond to vegetated land cover, of which there is a visible decrease from 2002-2010. Neighborhood levels of vegetation and other environmental factors within slum neighborhoods have been demonstrated to be connected to differing levels of child mortality (Jankowska et al. 2013), housing quality, and socioeconomic status (Stow et al. 2012). The results of modeling the changes over time in Accra, presented in Table 6, now indicate that vegetation proportions are also connected to different demographic characteristics of the population such as the percentages of the

Ordinary Least Squares (OLS) Regression Results & Diagnostics - Δ Models

Dependent Variable	OLS		Model Predictors		JB		BP	
	R-squared				p-value		p-value	
Electricity	0.21		<i>-Δ Green/Blue</i> ***			0.01		0.44
Charcoal Use	0.21		<i>-Δ Green</i> ***			0.61		0.83
Trash (collected)	0.32		<i>-Δ Green</i> **	<i>-Δ PC 3</i> ***		0.29		0.17
Trash (dumped offsite)	0.39		<i>-Δ Green/Blue</i> **	Δ 45° Filter**		0.56		0.22
Trash (burned or buried)	0.80		<i>-Δ NIR/B</i> ***	Δ NIR/B***	Δ Green/Blue***	0.93		0.11
Toilets (KVIP)	0.31		<i>-Δ Homogeneity</i> *	<i>-Δ PC 1</i> **	<i>-Δ PC 3</i> ***	0.81		0.34
Toilets (no flush or KVIP)	0.38		<i>-Δ PC 1</i> ***	<i>-Δ Vegetation %</i> **		0.94		0.24
Bednet Use	0.19		<i>-Δ Sobel thres.</i> **			0.00		0.05
Possessions Index	0.40		<i>-Δ NDVI</i> ***	<i>-Δ 45° Filter</i> **		0.64		0.54
Ethnicity (2000 Census)								
% Akan	0.47		Δ Contrast***	<i>-Δ Red</i> **	<i>-Δ Vegetation %</i> **	0.13		0.95
% Ga	0.14		<i>-Δ Green</i> **			0.01		0.13
% Ewe	0.24		<i>-Δ Correlation</i> *	Δ Vegetation %**		0.46		0.73
Religion (2000 Census)								
% Christian	0.35		Δ Green/Blue*	<i>-Δ PC 2</i> **	Δ Vegetation %**	0.37		0.34
% Muslim	0.31		Δ Green/Blue*	Δ PC 2**		0.00		0.30
Migrants(%)	0.12		<i>-Δ Laplacian</i> **			0.00		0.50
2000 Slum Index	0.70		Δ Homogeneity***	Δ NIR/B***	Δ PC 1***	0.42		0.32
2000 Housing Quality Index	0.55		<i>-Δ Homogeneity</i> ***	<i>-Δ NIR/R</i> **	<i>-Δ Vegetation %</i> ***	0.83		0.79
% Underweight (BMI <18)								
% Overweight (BMI = 25+)								

* Significant @ $\alpha = 0.10$ roman - EA mean value

** Significant @ $\alpha = 0.50$ italics - EA standard

*** Significant @ $\alpha = 0.01$ deviation

JB - Jarque-Bera test for residual normality

BP - Breusch-Pagan test for heteroskedasticity

Table 6. Δ OLS model results, including \bar{R}^2 values, model predictors variables, and tests for residual normality and heteroskedasticity.

Diagnostics for spatial dependence in Δ OLS model - Lagrange Multiplier Tests

Dependent Variable	1-directional LM			
	Global Moran's I Value	Moran's I p-value	Lag	Error probability
Δ Electricity	0.44	0.00	0.631	0.271
Δ Charcoal Use	0.95	0.43	0.241	0.261
Δ Trash (collected)	0.15	0.88	0.771	0.492
Δ Trash (dumped offsite)	0.15	0.00	0.017	0.034
Δ Trash (burned or buried)	0.54	0.07	0.999	0.475
Δ Toilets (KVIP)	0.62	0.25	0.459	0.176
Δ Toilets (flush or KVIP)	0.70	0.76	0.778	0.456
Δ Bednet Use	0.55	0.00	0.138	0.047
Δ Possessions Index	0.84	0.78	0.413	0.487
<i>Ethnicity (2000 Census)</i>				
Δ % Akan	0.55	0.50	0.167	0.296
Δ % Ga	0.56	0.70	0.476	0.761
Δ % Ewe	0.59	0.43	0.478	0.256
Δ % Mole-Dagbani	0.07	0.28	0.109	0.979
<i>Religion (2000 Census)</i>				
Δ % Christian	0.93	0.03	0.497	0.888
Δ % Muslim	0.63	0.67	0.777	0.377
Δ Migrants(%)	0.17	0.09	0.097	0.078
2000 Slum Index	0.19	0.25	0.122	0.174
2000 Housing Quality Index	0.84	0.76	0.467	0.359

Table 7. Diagnostics for spatial dependence and spatial heterogeneity in Δ models – the Global Moran's I and Lagrange Multiplier statistics and probabilities.

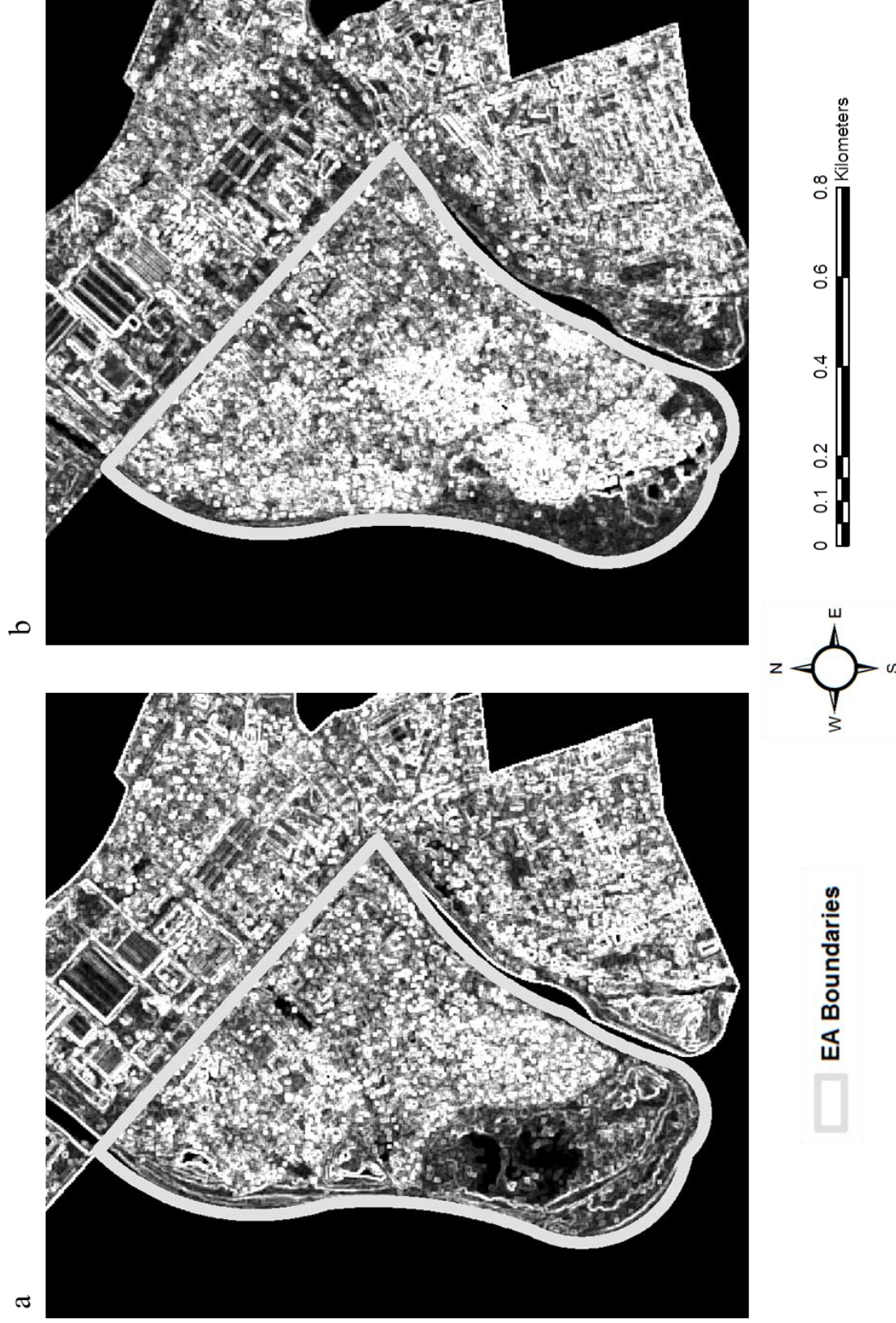


Figure 5. A side-by-side comparison of image lacunarity in EA 205020 (Sodom & Gomorah) for (a) 2002 and (b) 2010 images in the neighborhood of Sodom & Gomorah. An increase in lacunarity is observed between 2002-2010, directly related to the corresponding increase in building density per EA. Areas outside the boundary are displayed for comparison purposes.

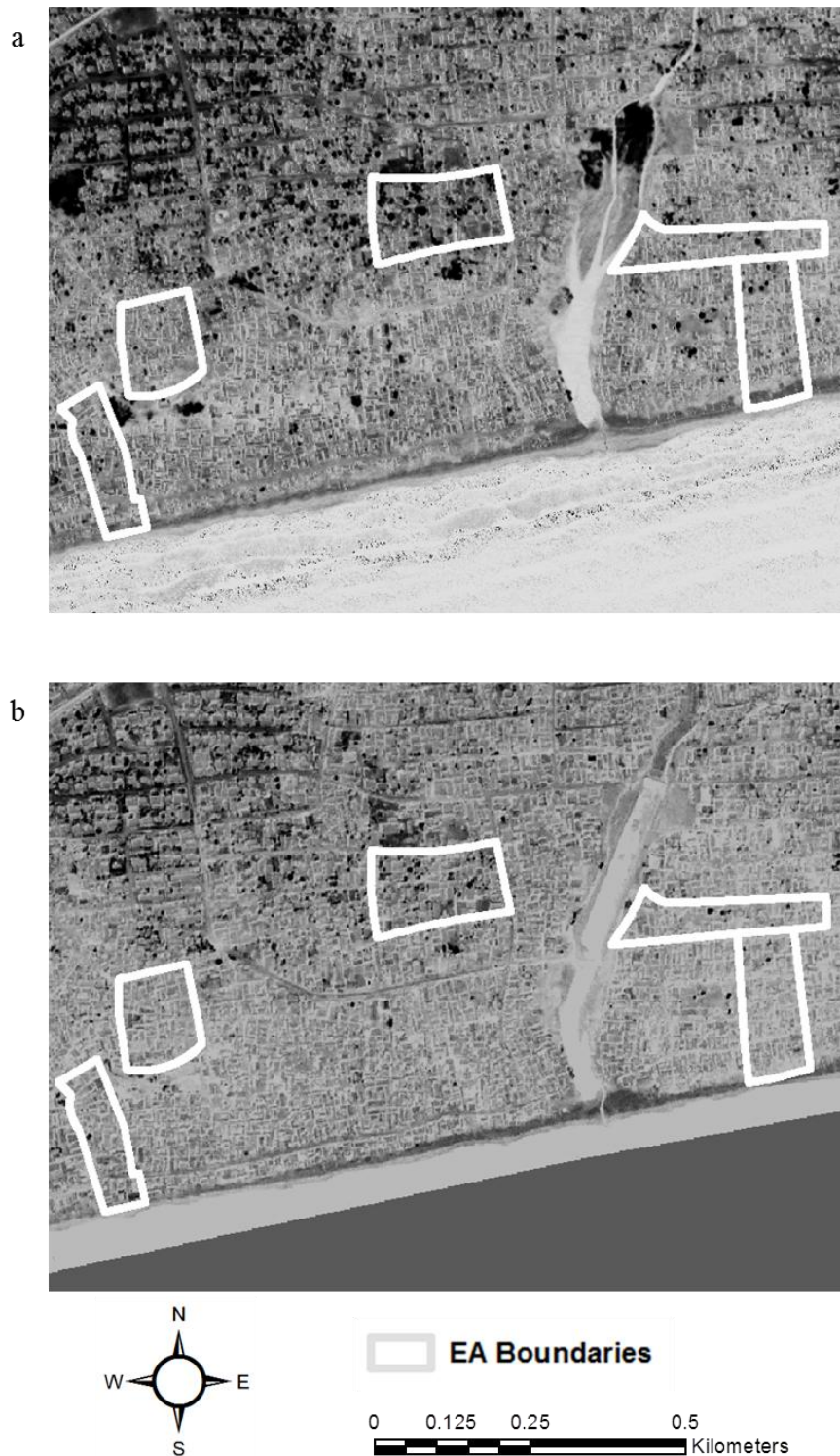


Figure 6. A comparison of the 2nd principal component (PC2), derived from QuickBird multispectral images for (a) 2002 and (b) 2010 images in the Gbegbeyise (west of the river) and Chokor neighborhoods. Both dates revealed that PC2 carries a large portion of spectral vegetation data and were considered vegetation proxies.

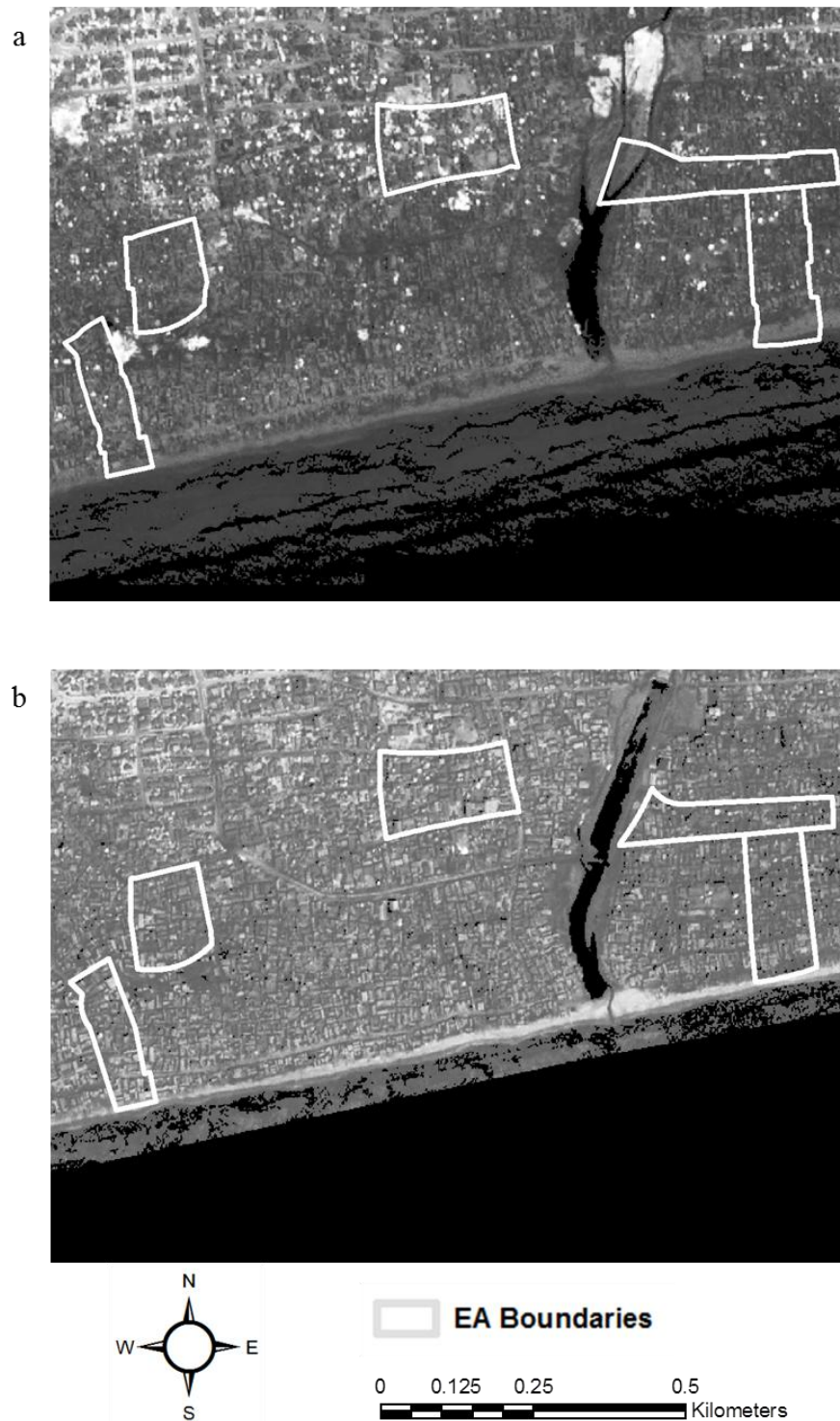


Figure 7. A comparison of the Blue/Green (B/G) band ratio, derived from QuickBird multispectral images for (a) 2002 and (b) 2010 images in the Gbegbeyise (west of the river) and Chokor neighborhoods. The B/G image metric was a significant predictor in five of the highest performing models.

population that belong to the Christian and Islamic faiths, ethnic groups like the Akan, Ga, and Ewe, and also physical household characteristics such as trash disposal and sewage facilities.

4.2 Estimating Indicators of SEDs through Global Spatial Regression Modeling

The LM tests for the Δ models indicated that a spatial lag model should be specified for the Δ *trash dumped offsite* and a spatial error model for Δ *bednet use*. Global Moran's I was statistically insignificant in all but five models, suggesting that spatially autocorrelated residuals may not be the number one concern within this dataset. The LM test results also do show the need for a spatial process model specification on most occasions.

For the 2002 model series, a global spatial lag model was fit to Δ % Muslim, based upon the statistically significant result from the LM[lag] test. A decrease in the AIC_c value from -44.57 to -51.74 was observed along with an increase in log-likelihood value from 26.28 to 30.87, signifying an improved model fit. This is confirmed by the coefficient of the spatially lagged variable ($\rho = -0.77$), which is highly significant ($p = 0.0007487$). The magnitude of all other estimated coefficients slightly increased from the classical OLS model (see Table 8), indicating that a portion of explanatory power of the independent variables originally attributed to their EA values can now be accounted for by values of the respective independent variables in “neighboring” EAs. An increase in the statistical significance of all predictor variables resulted from the switch in model specification, and the value of the likelihood ratio test was also highly significant ($p = 0.0025$), indicating the importance of the spatial autoregressive term.

Two spatial process models were specified in the $\Delta t_{2010-2002}$ series – Δ *trash dumped offsite* and Δ *household bednet use*. Both LM [lag] and LM [error] values were statistically significant, but upon examination of the Robust LM values, the Robust LM [lag] test had a lower probability, indicating the need for a lag model fit. Moran's I ($I = 0.15$) was also highly significant ($p < 0.0000$) for Δ *trash dumped offsite*, signaling the presence of moderate spatial autocorrelation in the form of clustering. The spatial lag model produced a decrease in AIC_c value from 22.54 to 20.62, indicating an improved model fit, confirmed by

□□□□□ - % Muslim Spatial Lag Model Results

Variable	Coefficient		Std. Error		Probability	
	OLS	Lag	OLS	Lag	OLS	Lag
Intercept	-1.087	1.652	0.128	0.193	0.0000	0.0007
Dissimilarity	0.001	0.012	0.003	0.002	0.0026	0.0000
Blue band	-0.004	-0.006	0.001	0.0009	0.0002	0.0000
Lacunarity	-0.009	-0.012	0.002	0.0015	0.0000	0.0000

OLS - $AIC_c = -44.57$

Spatially lagged - $AIC_c = -51.74$

OLS $R^2 = 0.71$

Likelihood Ratio 9.169

Table 8. % Muslim results for the OLS and spatial lag global models.

Δ Trash Dumped Offsite – Spatial Lag Model Results

Variable	Coefficient		Std. Error		Probability	
	OLS	Lag	OLS	Lag	OLS	Lag
Intercept	0.660	0.353	0.716	0.119	0.000	0.003
45° Filter	0.191	0.191	0.067	0.058	0.008	0.001
- Green/Blue	-0.014	-0.015	0.004	0.003	0.001	0.000
Rho ρ		0.595		0.205		0.003

OLS - $AIC_c = 22.54$

Spatially lagged - $AIC_c = 20.62$

OLS $R^2 = 0.38$ Spatial lag pseudo $R^2 = 0.52$

Likelihood Ratio 3.911

Table 9. Δ % trash dumped offsite results for the OLS and spatial lag global models.

the significance of the likelihood ratio test ($p = 0.47$). A decrease in independent variable coefficients was observed along with a decrease in the standard error values for each predictor, demonstrating the importance of the spatial autoregressive term in modeling the effects of “neighboring” EAs on the variability in the % of households that dump household trash offsite. Lambda was highly significant with a value of $\rho = 0.595$.

A spatial error model was fit to Δ bednet use, providing mixed indications of model fit improvement (Table 11). The AIC_c value decrease from -23.67 to -25.81, but the likelihood ratio test was not statistically significant at the $\alpha = 0.10$ level. Standard errors increased in each predictor variable for the error model, yet the variable coefficients slightly decreased, attributing some of their significance to the λ coefficient ($\lambda = 0.489$).

A dataset with a small sample size has the ability to produce a Type-II false-positive indication of spatial autocorrelation or spatial heterogeneity within the distribution. This may be attributed to a misrepresentation of the population through sampling bias and also has the potential to skew model results due to having low degrees of freedom.

Δ Bednet Use – Spatial Error Model Results						
Variable	Coefficient		Std. Error		Probability	
	OLS	Lag	OLS	Lag	OLS	Lag
Intercept	0.552	0.547	0.104	0.112	0.000	0.000
Sobel Threshold	-1.446	-1.450	0.516	0.521	0.009	0.005
Lambda λ		0.489		0.253		0.054

OLS - $AIC_c = -23.67$

Spatially lagged - $AIC_c = -25.81$

OLS $R^2 = 0.19$ Spatial lag pseudo $R^2 = 0.29$

Likelihood Ratio 2.498

Table 10. Δ % bednet results for the OLS and spatial error global models.

AICc Values (Assessments of Model Fit)

<i>Dependent Variable</i>	<i>2002 Models</i>		<i>Δ Models</i>	
	<i>OLS</i>	<i>Lag</i>	<i>OLS</i>	<i>Lag</i>
<i>Electricity</i>	-80.58		-73.51	
<i>Charcoal Use</i>	-48.41		-32.74	
<i>Trash (collected)</i>	-15.87		22.21	
<i>Trash (dumped offsite)</i>	-14.93		24.00	19.25
<i>Trash (burned or buried)</i>	-98.47		-92.69	
<i>Sachet Use</i>	0.55		0.55	
<i>Toilets (KVIP)</i>	-3.75		1.87	
<i>Toilets (flush or KVIP)</i>	-3.42		1.88	
<i>Bednet Use</i>	-87.62		-23.67	-26.81
<i>Possessions Index</i>	-67.37		-60.99	
<i>Ethnicity (2000 Census)</i>				
<i>% Akan</i>	-39.62		-59.85	
<i>% Ga</i>	-53.65		-62.89	
<i>% Mole-Dagbani</i>	-63.03		-65.39	
<i>Religion (2000 Census)</i>				
<i>% Christian</i>	-33.52		-34.46	
<i>% Muslim</i>	-44.57	-51.74	-42.20	
<i>Migrants(%)</i>	-101.42		-22.11	
<i>2000 Slum Index</i>	2.94		-18.16	
<i>2000 Housing Quality Index</i>	5.61		0.91	

Table 11. AICc values for both OLS and global spatial error models for 2002 and $\Delta t_{2002-2010}$ dates. A decrease in the AICc value of > 3 is considered an indication of improved model specification. Values in *green* represent improved models, where values in *yellow* represent error or lag models that did not meet the criteria to be considered an improvement of model fit. A value in *blue* denotes the specification of an error model.

4.3 Comparing Results to a GWR Model & a Spatial Regimes Approach

Results from the global regression models were compared to the results provided by a geographically weighted regression (GWR). A majority of models contained severe design problems and would not compute due to the small sample size and low degrees of freedom.

The premise of the spatial regimes approach is to divide the dataset into relatively homogeneous regions and then model each region separately to create a dual specification that would assist in accounting for the effects of spatial heterogeneity. A major issue was encountered due to a fatally low amount of degrees of freedom in each model. In order to proceed with the spatial regimes approach, a much larger sample of EAs would need to be gathered, increasing the degrees of freedom for each model and allowing for a proper model fit for each spatial regime.

CHAPTER 5

DISCUSSION & CONCLUSIONS

The relationships between neighborhood ethnic proportions and image metrics have demonstrated to be significant products of this study, due to the link between neighborhood ethnic compositions, neighborhood religious composition and differing levels of child mortality (Weeks et al. 2006). As previously stated, a person who belongs to a non-Christian faith or is a member of the Ga ethnic group will tend to have a higher risk of bad health. If certain image metrics such as vegetated land cover fractions or building density proxies can correctly indicate regions where specific ethnic or religious interest groups tend to live based on the physical properties of their environment, policy makers and health officials may find it easier to dispatch the necessary aid or resources to help investigate and combat high levels of child mortality and other forms of disease. Sverdlik (2011) has underlined that informal settlements across sub-Saharan Africa are undergoing many emerging urban health risks and inequalities, many of which are reflected by the environment they reside in. Sverdlik (2011) also denotes that these communities are currently facing the “double burden” of both communicable and non-communicable diseases, and society must create interventions to ensure that populations living in these informal regions may obtain a higher health and socioeconomic status.

An emphasis must be placed on the fact that the relationships within this study are only limited to the aforementioned slum regions. In order to comprehensively understand the dynamic relationships between the environment and the characteristics of the populace within, a call must be made to examine non-slum areas. Limiting our examination of population – environment relationships to only slum regions has likely limited the range of variables that may be explored, with the possibility of skewing the interpretation or understanding of the dynamics of the urban environment and its morphology. The comparison of the ability of image metrics to indicate survey characteristics must be compared between slum and non-slum areas. Understanding areas with a variety of socioeconomic characteristics may enable different relationships to be discovered and contrasted with slum regions that could possibly be at more or less of a disadvantage than previously believed.

It has been demonstrated that there are moderately strong, significant relationships between remotely-sensed variables and household attributes gathered from ground survey in slum neighborhoods of Accra, Ghana. The exploitation of remote sensing metrics as proxies for socioeconomic and health conditions opens up new pathways in the fields of social and public health research. The creation of remote proxy variables for health and welfare characteristics allows for a nuanced method of data collection in developing countries. This study takes a step forward in advancing the increasing capabilities of remote sensing in the public health, socioeconomic, and demographic sectors, and will aid in the analysis of further data in Ghana. Individual and household surveys normally costing tens of thousands of dollars and large investments of time could potentially be parsimoniously streamlined. Observing changes in urban morphology within a sub-Saharan African developing city is becoming increasingly important, specifically within slum neighborhoods. It is in these cities where most of the world's population increase will occur in the next 50 years. Distinguishing where these changes are taking place is the first step in understanding how to predict the health and well-being of the residents in these neighborhoods. The results carry implications from a policy creating standpoint, as a healthy city is a happy, efficient and more prosperous city.

REFERENCES

- Aikaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest.
- Aikaike, H. 1974. A new look at statistical model identification. *IEEE Transactions on Automated Control* 19: 716-723.
- Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht. 119-135.
- Anselin, L. 1990. Spatial dependence and spatial structural instability in applied regression analysis. *Journal of Regional Science* 30:185-207.
- Arivazhagan, S. and Ganesan, L. 2003. Texture classification using wavelet transform. *Pattern Recognition Letters* 24: 1513-1521.
- Bharati, M., Liu, J. & MacGregor, J. 2004. Image texture analysis: methods and comparison. *Chemometrics and Intelligent Laboratory Systems* 72:57-71.
- Ceballos, J. & Bottino, M. 1997. The discrimination of scenes by principal components analysis of multi-spectral imagery. *International Journal of Remote Sensing* 18(11): 2437-2449.
- Curtis, K., Voss, P. and Long, D. 2012. Spatial variation in poverty generating processes: Child poverty in the United States. *Social Science Research* 41:146-159.
- Dong, P. 2000. Test of a new lacunarity estimation method for image texture analysis. *International Journal of Remote Sensing* 21 (17): 3369-3373.
- Efron, B. 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7(1):1-26.
- Entwisle, Barbara. 2007. Putting People Into Place. *Demography* 44 (4) November 2007: 687-703.
- Engstrom, R., E. Ashcroft, H. Jewell, and D. Rain. 2011. Using Remotely Sensed Data to Map Variability in Health and Wealth Indicators in Accra, Ghana. In *Joint Urban Remote Sensing Event (JURSE), 2011 Joint*, 145-148. IEEE. doi:10.1109/JURSE.2011.5764740.
- Fitch, D., Stow, D., Hope, A. & Rey, S. 2010. MODIS vegetation metrics as indicators of hydrological response in watersheds of California Mediterranean-type climate zones. *Remote Sensing of Environment* 114:2513-2523.
- Fotheringham, A., Brudson, C. & Charlton, M. 2002. *Geographically Weighted Regression: The analysis of spatially varying relationships*. Hoboken, NJ. Wiley.
- Giroux, S. 2008. Child Stunting Across Schooling and Fertility Transitions: Evidence from Sub-Saharan Africa. *DHS Working Papers, No. 57*. Calverton, Maryland, USA: ICF Macro.

- Graesser, J., A. Cheriyyadat, R. Vatsavai, V. Chandola, J. Long, E. Bright. Forthcoming. Image Based Characterization of Formal and Informal Neighborhoods in an Urban Landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Forthcoming.
- Hay, Simon I., David J. Rogers, G.Dennis Shanks, Monica F. Myers, and Robert W. Snow. 2001. Malaria Early Warning in Kenya. *Trends in Parasitology* 17 (2) (February 1): 95-99. doi:10.1016/S1471-4922(00)01763-3.
- Herold, M., X. Liu, K. Clarke. 2003. Spatial Metrics and Image Texture for Mapping Urban Land Use. *Photogrammetric Engineering & Remote Sensing* 69 (9) (September 2003): 991-1001.
- Hurvich, C.M. & Tsai, C. L. 1989. Regression and time series model selection in small samples. *Biometrika* 76: 297-307.
- Jankowska, M., Benza-Fiocco, M., Weeks, J.R. 2013. Estimating spatial inequalities of urban child mortality. *Demographic Research* 28(2): 33-62.
- Jarque, C., Bera, A., 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55, 163{172.
- Jensen, J.R. 1996. Introductory Digital Image Processing, 3rd Edition, Prentice Hall Inc., 187-192.
- Kelly, M., Blanchard, S.D., Kersten, E., & Koy, K.. 2011. Terrestrial Remotely Sensed Imagery in Support of Public Health: New Avenues of Research Using Object-Based Image Analysis. *Remote Sensing* 3 (11) (October 27): 2321-2345.
- Massey, Doreen. 1994. Space, Place and Gender. University of Minnesota Press, Minneapolis, MN.
- Mbuya, M., M. Chideme, B. Chasekwa, and V. Mishra. 2010. Biological, Social, and Environmental Determinants of Low Birth Weight and Stunting among Infants and Young Children in Zimbabwe. *Zimbabwe Working Papers, No.7*. Calverton, Maryland, USA: ICF Macro.
- Moller-Jensen, L. 1997. Classification of Urban Land Cover Based on Expert Systems, Object Models and Texture. *Computers, Environment and Urban Systems* 21 (3-4): 291-302.
- Moller-Jensen, Lasse, and Michael H. Knudsen. 2008. Patterns of Population Change in Ghana (1984–2000): Urbanization and Frontier Development. *GeoJournal* 73 (4): 307-320.
- Moudon, A. 1997. Urban morphology as an emerging interdisciplinary field. *Urban Morphology* 1 (1997): 3-10. International Seminar on Urban Forum, 1997.
- Myint, S.W. and Lam, N. 2005. A study of lacunarity-based texture analysis approaches to improve urban image classification. *Computers, Environment and Urban Systems* 29: 501-523.

- Myint, S.W., Mesev, V., and Lam, N. 2006. Urban Textural Analysis from Remote Sensor Data: Lacunarity Measurements Based on the Differential Box Counting Method. *Geographic Analysis* 38: 371-390.
- Openshaw, S. 1984. Ecological fallacies and the analysis of areal census data. *Environment and Planning A* 16: 17-31.
- Phinn, S., M. Stanford, P. Scarth, A. T. Murray, and P. T. Shyy. 2002. Monitoring the Composition of Urban Environments Based on the Vegetation-impervious Surface-soil (VIS) Model by Subpixel Analysis Techniques. *International Journal of Remote Sensing* 23 (20): 4131-4153.
- Posada, D. & Buckley, T. 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Aikake Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Society of Systematic Biologists* 53(5): 793-808.
- Raghu, P.P., Poongodi, R. and Yegnanarayana, B. 1995. A combined neural network approach for texture classification. *Neural Networks* 6:975-987.
- Rain, D., Engstrom, R., Ludlow, C., & Antos, S. 2011. Accra Ghana: A City Vulnerable to Flooding and Drought-Induced Migration. *The Global Report On Human Settlements 2011*: Chapter 4 (Climate Impacts on Urban Areas).
- Rashed, T., Weeks, J.R., Roberts, D., Rogan, J., & Powell, R. 2003. Measuring the Physical Composition of Urban Morphology Using Multiple Endmember Spectral Mixture Models. *Photogrammetric Engineering and Remote Sensing* 69 (9): 1011-1020.
- Rashed, Tarek, John R. Weeks, M. Saad Gadalla, and Allan G. Hill. 2001. Revealing the Anatomy of Cities Through Spectral Mixture Analysis of Multispectral Satellite Imagery: A Case Study of the Greater Cairo Region, Egypt. *Geocarto International* 16 (4): 7-18.
- Rashed, T., Weeks, J.R., Stow, D., & Fugate, D. 2005. Measuring Temporal Compositions of Urban Morphology Through Spectral Mixture Analysis: Toward a Soft Approach to Change Analysis in Crowded Cities. *International Journal of Remote Sensing* 26 (4): 699-718.
- Ridd, M. K. 1995. Exploring a V-I-S (vegetation-impervious Surface-soil) Model for Urban Ecosystem Analysis Through Remote Sensing: Comparative Anatomy for Cities. *International Journal of Remote Sensing* 16 (12): 2165-2185.
- Tatem, A. J., & Hay, S. I. 2004. Measuring Urbanization Pattern and Extent for Malaria Research: A Review of Remote Sensing Approaches. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 81 (3): 363-376.
- Tsai, Y.H., Stow, D., & Weeks, J.R. 2011. Comparison of Object-Based Image Analysis Approaches to Mapping New Buildings in Accra, Ghana Using Multi-Temporal QuickBird Satellite Imagery. *Remote Sensing* 3 (12): 2707-2726.
- Stoler, J., Weeks, J.R., Getis, A., & Hill, A.G. 2009. Distance Threshold for the Effect of Urban Agriculture on Elevated Self-reported Malaria Prevalence in Accra, Ghana. *American Journal of Tropical Medicine & Hygiene* 80:547-554 PMID: PMC2714825.

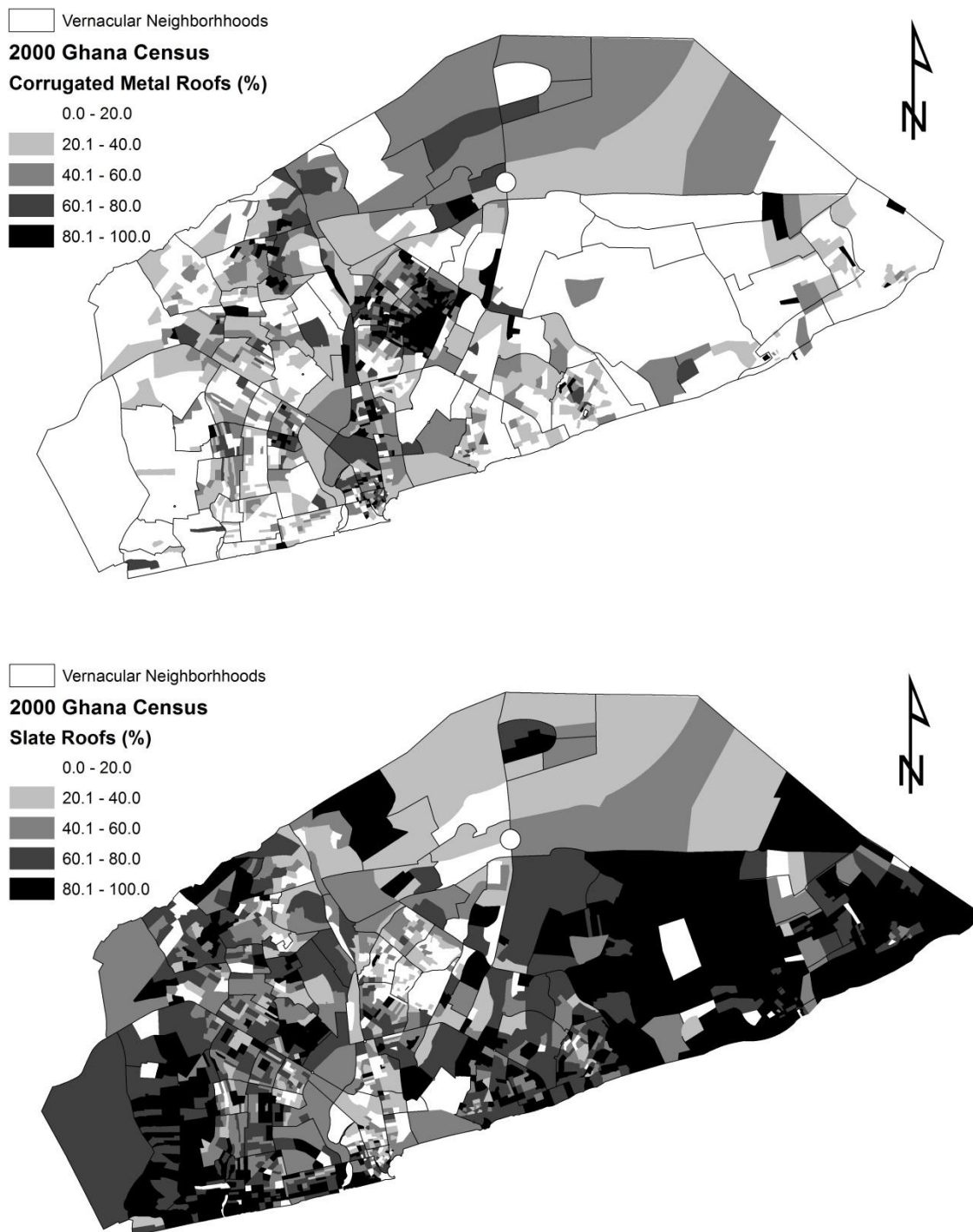
- Stoler, J., Gunther, F., Weeks, J.R., Appiah Otoo, R., Ampofo, J.A., Hill, A.G. 2011. When urban taps run dry: Sachet water consumption and health effects in low income neighborhoods of Accra, Ghana. *Health and Place* 18 (2012): 250-262.
- Stoler, J., Daniels, D., Weeks, J.R., Stow, D., Coulter, L., & Finch, B. 2012. Assessing the Utility of Satellite Imagery with Differing Spatial Resolutions for Deriving Proxy Measures of Slum Presence in Accra, Ghana. *GIScience & Remote Sensing* 49 (1): 31-52.
- Stow, D., Lopez, A., Lippitt, C., Hinton, S., & Weeks, J.R. 2007. Object-based Classification of Residential Land Use Within Accra, Ghana Based on QuickBird Satellite Data. *International Journal of Remote Sensing* 28 (22): 5167-5173.
- Stow, D. A., Lippitt, C. D., & Weeks, J. R. 2010. Geographic Object-based Delineation of Neighborhoods of Accra, Ghana Using QuickBird Satellite Imagery. *Photogrammetric Engineering and Remote Sensing* 76 (8): 907-914.
- Stow, D., Weeks, J.R., Toure, S., Lippitt, C., Coulter, L., Ashcroft, E. (Forthcoming). Urban vegetation cover and change in Accra, Ghana: Connection to quality of life. *Professional Geographer*.
- Sverdlik, A. 2011. Ill-health and poverty: a literature review on health in informal settlements. *Environment and Urbanization* 23: 123.
- United Nations Human Settlements Programme (UNHSP), 2003. The Challenge of Slums: *Global Report on Human Settlements*. Earthscan, London, UK.
- Unsalan, C. and Boyer, K.L. 2004. Classifying land development in high resolution panchromatic satellite images using straight-line statistics. *IEEE Transactions on GeoScience and Remote Sensing* 42: 907-919.
- Wang, L. and Liu, J. 1999. Texture classification using multiresolution Markov random field models. *Pattern Recognition Letters* 20:171-182.
- Weeks, J.R. 2003. "Using Remote Sensing and Geographic Information Systems to Identify the Underlying Properties of Urban Environments" in Champion, T., & Hugo, G., *New Forms of Urbanization: Beyond the Urban-Rural Dichotomy*. Aldershot, UK. Ashgate Publishing Co., 2003.
- Weeks, J.R., Hill, A.G., Getis, A., Stow, D. 2006. Ethnic Residential Patterns As Predictors of Intra-Urban Child Mortality Inequality in Accra, Ghana. *Urban Geography* 27 (6): 526-548.
- Weeks, John R. 2010. "Defining Urban Areas." in T. Rashed and C. Jurgens, *Remote Sensing of Urban and Suburban Areas*. Remote Sensing and Digital Image Processing 10, Springer Science Business Media B.V. 2010
- 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 (1-2)
- Weeks, J., Getis, A., Stow, D., Hill, A., Rain, D., Engstrom, R., Stoler, J., Lippitt, C., Jankowska, M., & Lopez, A. 2012. Connecting the dots between health, poverty and place in Accra, Ghana. *Ann. Assoc. Am. Geogr.*

Weeks, J., Hill, A., Stoler, J., Zvoleff, A. Forthcoming.

Weiss, L., Ompad, D., Galea, S., & Vlahov, D. 2007. Defining Neighborhood Boundaries for Urban Areas. *American Journal of Preventative Medicine*, 2007: 154-159.

APPENDIX 1

DISTRIBUTION OF ROOF TYPES IN THE AMA



Appendix 1. Distribution of roof materials for both corrugated metal roofs (top) and slate roofs (bottom). Households with primarily slate roofs tend to be located in the older slum regions surrounding Nima, whereas corrugated metal roofs tend to dominate the lower-lying coastal regions of the AMA.

APPENDIX 2
SURVEY VARIABLE CHANGES BY EA

Proportional Decrease in Survey Data by EA (2003-2010)

Variable	101005	102001	102009	102014	103003	103015	103026	105010	107018
Δ Electricity	0.037	0.067	0.051	-0.050	-0.007	0.037	0.010	0.000	0.107
Δ Charcoal Use	0.386	-0.094	0.153	0.233	0.103	-0.103	-0.220	0.100	-0.239
Δ Eviction Possibility	0.415	0.346	0.373	-0.167	0.117	0.166	-0.425	0.125	0.153
Δ Bednet Use	0.957	0.968	0.941	0.754	0.963	0.800	0.850	0.850	0.812
Δ Possessions Index	0.038	0.232	0.143	0.241	0.115	0.267	0.252	0.110	0.397
Δ Trash Disposal									
Collected	-0.056	0.000	0.000	-0.167	-0.263	0.030	0.000	-0.900	-0.889
Dumped Offsite	0.012	-0.080	0.000	0.117	-0.007	-0.363	-0.030	0.880	0.889
Burned or Buried	0.043	0.080	0.000	0.000	0.280	0.333	0.030	0.030	0.000
Δ Water Source									
Piped	0.170	0.048	-0.353	0.048	0.275	0.233	-0.042	0.075	0.292
Piped (inside)	-0.064	0.016	-0.412	-0.429	-0.013	-0.167	-0.292	-0.150	-0.021
Piped (outside)	0.234	0.032	0.059	0.476	0.288	0.400	0.250	0.225	0.313
Sachet Use	0.468	0.661	0.294	0.095	0.500	0.267	0.167	0.425	0.333
Δ Toilet Facility									
Flushing	-0.056	0.000	-0.052	-0.294	0.025	0.033	-0.050	-0.250	0.007
KVIP	-0.492	-0.056	-0.056	-0.174	0.038	-0.133	-0.067	-0.200	-0.653
Flush or KVIP	-0.547	-0.056	-0.108	-0.468	0.063	-0.100	-0.117	-0.450	-0.646
Δ Ethnicity (%)									
Akan	-0.143	0.103	0.158	-0.166	0.049	0.017	0.005	0.066	0.110
Ga	0.330	0.047	-0.027	0.044	-0.033	0.063	-0.053	-0.035	-0.021
Ewe	-0.107	-0.037	-0.095	0.198	0.051	0.074	0.021	0.013	-0.049
Mole-Dagbani	-0.023	-0.027	0.000	0.079	0.036	-0.022	0.040	0.030	-0.044
Δ Religion (%)									
Christian	0.094	0.084	0.100	-0.081	-0.030	0.100	0.018	-0.045	0.098
Muslim	0.000	-0.004	-0.080	0.025	-0.015	-0.113	-0.030	0.030	-0.081
Δ Migrants (%)	0.169	0.239	0.655	0.537	0.433	0.120	0.507	0.545	0.380

Proportional Decrease in Survey Data by EA (2003-2010)

Variable	107021	107026	107027	107029	107038	107049	205020	301034	301035
Δ Electricity	0.010	0.040	0.035	-0.015	-0.060	-0.034	0.028	0.258	-0.033
Δ Charcoal Use	0.090	-0.040	-0.047	0.110	-0.056	0.223	0.126	0.112	0.183
Δ Eviction Possibility	-0.207	-0.171	0.074	0.455	0.168	0.252	0.897	0.780	0.467
Δ Bednet Use	0.814	0.843	0.876	0.784	0.958	0.879	0.143	0.561	0.472
Δ Possessions Index	0.200	0.407	0.149	0.125	0.123	0.157	0.092	0.342	0.178
Δ Trash Disposal									
Collected	-0.950	-0.810	-0.544	-0.938	-0.397	-0.869	-0.139	-0.109	-0.950
Dumped Offsite	1.000	0.840	0.755	1.000	0.456	0.980	0.002	0.114	0.850
Burned or Buried	-0.050	-0.020	-0.211	-0.063	-0.059	-0.111	0.137	-0.016	0.050
Δ Water Source									
Piped	-0.179	-0.114	0.119	-0.182	0.375	0.217	-0.021	0.565	0.050
Piped (inside)	-0.143	-0.200	-0.048	-0.273	-0.042	-0.065	-0.007	0.022	-0.100
Piped (outside)	-0.036	0.086	0.167	0.091	0.417	0.283	-0.014	0.544	0.150
Sachet Use	0.464	0.543	0.429	0.455	0.389	0.391	0.901	0.044	0.000
Δ Toilet Facility									
Flushing	-0.150	0.050	-0.063	0.046	0.010	0.031	-0.105	-0.324	-0.122
KVIP	-0.407	-0.343	-0.302	-0.273	-0.082	-0.183	-0.474	-0.611	-0.333
Flush or KVIP	-0.557	-0.293	-0.365	-0.227	-0.071	-0.152	-0.579	-0.935	-0.456
Δ Ethnicity (%)									
Akan	0.105	0.170	-0.049	0.030	-0.001	0.032	-0.138	-0.280	-0.142
Ga	-0.151	-0.065	-0.060	0.011	-0.015	-0.016	-0.029	0.064	-0.138
Ewe	0.026	0.010	0.010	0.031	0.064	-0.033	-0.051	-0.074	-0.019
Mole-Dagbani	-0.038	0.003	0.080	0.082	-0.038	0.028	0.092	-0.014	0.050
Δ Religion (%)									
Christian	0.184	-0.120	0.004	0.023	-0.104	-0.050	-0.208	0.104	0.010
Muslim	-0.034	0.121	-0.018	0.006	0.044	0.088	0.208	-0.019	-0.010
Δ Migrants (%)	0.509	0.220	0.479	0.274	0.497	0.519	0.467	0.305	0.350

Proportional Decrease in Survey Data by EA (2003-2010)

Variable	501054	502018	502029	502051	502062	503007	504014	504020	504025
Δ Electricity	0.105	0.003	-0.007	-0.038	0.000	0.000	-0.014	0.000	-0.160
Δ Charcoal Use	0.031	-0.082	-0.110	-0.275	0.118	-0.039	-0.023	0.120	0.102
Δ Eviction Possibility	0.306	0.236	0.301	0.487	-0.412	0.213	0.296	0.641	0.467
Δ Bednet Use	1.000	0.810	0.617	0.900	1.000	0.864	0.870	0.932	0.920
Δ Possessions Index	0.255	0.215	0.219	0.095	-0.141	0.120	0.156	0.168	0.175
Δ Trash Disposal									
Collected	0.000	-0.670	-1.000	-0.738	-1.000	-0.250	-0.834	-0.900	-0.920
Dumped Offsite	0.000	0.570	0.890	0.737	1.000	0.250	0.834	0.900	0.920
Burned or Buried	0.000	0.100	0.110	0.000	0.000	0.000	0.000	0.000	0.000
Δ Water Source									
Piped	-1.727	-0.286	-0.500	-0.250	-15.000	-0.208	-0.074	-0.136	-0.040
Piped (inside)	-0.546	-0.048	-0.278	-0.100	-2.000	-0.167	-0.037	-0.046	-0.040
Piped (outside)	-1.182	-0.238	-0.222	-0.150	-13.000	-0.042	-0.037	-0.091	0.000
Sachet Use	1.000	0.381	0.444	0.450	0.000	0.417	0.444	0.636	0.360
Δ Toilet Facility									
Flushing	0.220	0.042	0.000	0.000	-0.059	0.134	0.111	0.000	0.000
KVIP	-0.253	-0.642	-0.836	-0.250	-0.941	-0.789	-0.500	-0.259	-0.633
Flush or KVIP	-0.034	-0.599	-0.836	-0.250	-1.000	-0.655	-0.389	-0.259	-0.633
Δ Ethnicity (%)									
Akan	0.128	-0.124	-0.011	-0.026	-0.208	-0.066	-0.005	-0.103	-0.091
Ga	0.044	0.016	-0.041	0.048	0.000	-0.065	0.001	-0.019	-0.020
Ewe	-0.025	0.010	0.029	-0.091	0.000	-0.055	-0.127	-0.112	-0.081
Mole-Dagbani	-0.092	-0.012	-0.087	-0.115	-0.375	-0.211	-0.178	-0.051	-0.121
Δ Religion (%)									
Christian	0.096	-0.100	-0.057	0.040	-0.500	-0.382	-0.239	-0.204	-0.240
Muslim	-0.096	0.100	0.001	-0.030	0.500	0.318	0.202	0.201	0.240
Δ Migrants (%)	0.536	0.499	0.486	0.440	-0.130	0.255	0.560	0.587	0.340

Proportional Decrease in Survey Data by EA (2003-2010)

Variable	504031	504036	505050	603005	Mean EA Δ
<i>Δ Electricity</i>	0.056	-0.097	0.182	0.050	0.018
<i>Δ Charcoal Use</i>	0.114	-0.094	-0.111	0.020	0.026
<i>Δ Eviction Possibility</i>	0.200	0.057	0.222	0.118	0.224
<i>Δ Bednet Use</i>	1.000	0.910	0.791	0.927	0.831
<i>Δ Possessions Index</i>	0.097	0.112	0.242	0.238	0.178
Δ Trash Disposal					
<i>Collected</i>	-0.870	-0.427	-0.734	-0.550	-0.543
<i>Dumped Offsite</i>	0.870	0.427	0.734	0.520	0.518
<i>Burned or Buried</i>	0.000	0.000	0.000	0.000	0.021
Δ Water Source					
<i>Piped</i>	-1.067	-0.370	0.250	0.227	-0.568
<i>Piped (inside)</i>	-0.333	-0.111	0.042	-0.023	-0.196
<i>Piped (outside)</i>	-0.733	-0.259	0.208	0.250	-0.372
<i>Sachet Use</i>	0.867	0.630	0.000	0.318	0.412
Δ Toilet Facility					
<i>Flushing</i>	-0.100	-0.183	0.195	0.000	-0.029
<i>KVIP</i>	0.245	-0.125	-0.333	-0.377	-0.339
<i>Flush or KVIP</i>	0.144	-0.308	-0.139	-0.377	-0.368
Δ Ethnicity (%)					
<i>Akan</i>	-0.036	-0.019	0.051	0.092	-0.016
<i>Ga</i>	-0.089	0.180	-0.006	-0.002	-0.001
<i>Ewe</i>	0.051	0.004	-0.057	-0.003	-0.014
<i>Mole-Dagbani</i>	-0.011	-0.042	-0.036	0.005	-0.033
Δ Religion (%)					
<i>Christian</i>	0.000	-0.088	-0.088	0.031	-0.050
<i>Muslim</i>	0.020	0.111	0.047	-0.026	0.055
Δ Migrants (%)					
	0.583	0.533	0.443	0.459	0.413

Variable	Mean EA Δ
Δ Electricity	0.018
Δ Charcoal Use	0.026
Δ Eviction Possibility	0.224
Δ Bednet Use	0.831
Δ Possessions Index	0.178
Δ Trash Disposal	
Collected	-0.543
Dumped Offsite	0.518
Burned or Buried	0.021
Δ Water Source	
Piped	-0.568
Piped (inside)	-0.196
Piped (outside)	-0.372
Sachet Use	0.412
Δ Toilet Facility	
Flushing	-0.029
KVIP	-0.339
Flush or KVIP	-0.368
Δ Ethnicity (%)	
Akan	-0.016
Ga	-0.001
Ewe	-0.014
Mole-Dagbani	-0.033
Δ Religion (%)	
Christian	-0.050
Muslim	0.055
Δ Migrants (%)	0.413

APPENDIX 3
R PROGRAM OF LACUNARITY CODE

```
#####
# Lacunarity.R
# By Milo Vejraska & Alex Zvoleff
#
# Computes lacunarity for a 3 band image
# - Input image must be separated into 3 bands. This script runs
#   lacunarity for bands 4, 3 & 2 or a multispectral or
#   pan-sharpened image.
# -
#
#####

library(rgdal)
library(raster)
library(nnet)
library(snow)
library(ggplot2)

#set working directory
setwd(c("G:/Ghana/Imagery/lacunarity"))
getwd()
dir()

# Open raster bands
band4 <- raster("qb02psms_unreg_band4.tif")
band3 <- raster("qb02psms_unreg_band4.tif")
band2 <- raster("qb02psms_unreg_band4.tif")

# image(band4)
# image(band3)
# image(band2)

# Calculate 3x3 focal min & max for each band
band4min <- focal(band4, w=3, fun=min, na.rm=TRUE)
band3min <- focal(band4, w=3, fun=min, na.rm=TRUE)
band2min <- focal(band4, w=3, fun=min, na.rm=TRUE)

band4max <- focal(band4, w=3, fun=max, na.rm=TRUE)
band3max <- focal(band4, w=3, fun=max, na.rm=TRUE)
band2max <- focal(band4, w=3, fun=max, na.rm=TRUE)

# Calculate pixel relative heights
rhband4 <- band4max - band4min - 1
rhband3 <- band3max - band3min - 1
rhband2 <- band2max - band2min - 1

# Layerstack relative height bands
rhstack <- stack(rhband4, rhband3, rhband2)

# Compute pixel mass
mass <- sum(rhstack)
# writeRaster(mass, "qb02ms_mass.tif")
# mass <- raster("qb02ms_mass.tif")
image(mass)

# Create lookup table.
lut_vals <- unique(mass)
lut <- matrix(c(lut_vals, rep(0, length(lut_vals))), ncol=2)

# Process over blocks (rather than rows) to save processing time.
```

```

pb <- txtProgressBar(style=3)
bs <- blockSize(mass)
for (block_num in 1:bs$n) {
  setTxtProgressBar(pb, block_num/bs$n)
  this_block <- getValues(mass, row=bs$row[block_num], nrows=bs$nrows[block_num])
  lut_pos <- match(this_block, lut[,1])
  for (i in 1:length(lut_pos)) {
    lut[lut_pos[i],2] <- lut[lut_pos[i],2] + 1
  }
}

close(pb)

# Replace values with occurrences
nboxes <- nrow(mass) * ncol(mass)
lut[,2] <- lut[,2]/nboxes

# Create qmr raster
out <- writeStart(raster(mass), "qb10ms_qmr_test.tif")
pb <- txtProgressBar(style=3)
bs <- blockSize(mass)
for (block_num in 1:bs$n) {
  setTxtProgressBar(pb, block_num/bs$n)
  this_block <- getValues(mass, row=bs$row[block_num], nrows=bs$nrows[block_num])
  # Define Q(M,r) probability function
  lut_pos <- match(this_block, lut[,1])
  this_block <- lut[,2][lut_pos]
  qmr <- this_block
  writeValues(out, qmr, bs$row[block_num])
}

out <- writeStop(out)
close(pb)

# Square mass image (M2)
m2 <- mass * mass

# Compute M2 x Q(M,r)
m2qmr <- m2 * qmr

# Compute M x Q(M,r)
mqmr <- mass * qmr

# Mean sq deviation of mass distribution fluctuations
num <- m2 + qmr

# Square mean of mass dist fluncts
denom <- (mass + qmr) + (mass + qmr)

# Compute lacunarity
lac <- num/denom
writeRaster(lac, "qb10_lacunarity.tif")

#####
# END SCRIPT.....
#####

```