12-2018

Utility of High Resolution Human Settlement Data for Assessment of Electricity Usage Patterns

Pranab Kanti Roy Chowdhury
University of Tennessee, proychow@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Pranab Kanti Roy Chowdhury entitled "Utility of High Resolution Human Settlement Data for Assessment of Electricity Usage Patterns." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Energy Science and Engineering.

Budhendra L. Bhaduri, Major Professor

We have read this dissertation and recommend its acceptance:

Joshua S. Fu, Nicholas N. Nagle, Olufemi A. Omitaomu

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Utility of High Resolution Human Settlement Data for Assessment of Electricity Usage Patterns

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Pranab Kanti Roy Chowdhury

December 2018

To my parents

Drs. Kuntala Roy Chowdhury and Pijush Kanti Roy Chowdhury

for your love, guidance, and the everlasting belief in me.

Acknowledgments

I would like to take this opportunity to thank my advisor, Dr. Budhendra L. Bhaduri, for his time, support, and guidance over the last four years. I am grateful to Dr. Joshua S. Fu, Dr. Nicholas N. Nagle, Dr. Olufemi A. Omitaomu, and Dr. Benjamin L. Preston (former committee member) for their time and guidance. I would like to express my gratitude to Dr. Leo L. Riedinger, Director of the Bredesen Center program for offering me the Energy Science and Engineering Fellowship and for his help and guidance during my time at the program. I am thankful to Wanda Davis, Jessica Garner, and Ben Allen at Bredesen Center for their support. I have received immense help and encouragement from my senior colleagues at Oak Ridge National Laboratory (ORNL), for which I am thankful to Nagendra Singh, Dr. H.M. Abdul Aziz, Dr. Melissa Allen, Jacob McKee, Jeanette Weaver, Dr. Jiangye Yuan, Dr. Christa Brelsford, Dr. Amy Rose, and Dr. Debjani Singh. I would also like to thank Beata Taylor, Ashton Chatman, and Lisa Gorman who provided administrative support at ORNL.

I have been fortunate to receive constant support and motivation from my friends. I am thankful to Dr. Avik Mukherjee, Claire Edel, Alex Pawloski, Eric Muckley, Rani Laha, Jay Jay Billings, Christine Ajinjeru, Taylor Hastings, Lauryn and Braxton Bragg, Adam Alsamadissi, Dr. Paulo Raposo, Dr. J.P. Mahalik, Kristine Cabugao, Dr. Mark Coletti, Dr. Jeremy Auerbach, and Dr. Rahul Bannerjee. Thanks are also due to Dr. Sandeep Maithani and Dr. Tilottama Ghosh for introducing me to the joy of research and encouraging me from the very beginning of my scientific career. I have received valuable support and encouragement from Drs. Nita and Ranjan Ganguly, Sucharita and Dr. Ajay Mukherjee, Drs. Jheelam and Adrija Sharma, Sriparna and Dr. Kaushik Bannerjee, Drs. Priyanka and Sabornee Chatterjee, Poulomi and Dr. Satyabrata Sen; for which I am eternally thankful.

My parents showered me with unwavering support and encouragement throughout my life. I would not be here without their love, affection, patience, and guidance. No word can honestly justify my gratefulness, so I hope that I will be able to justify their faith in me throughout my life and work. This is dedicated to you, Ma and Baba! Thanks are also due to my extended family members in India and elsewhere in the World, who in numerous ways have contributed to my journey.

Abstract

Electricity is vital for modern human civilization, and its demands are expected to significantly rise due to urban growth, transportation modernization, and increasing industrialization and energy accessibility. Meeting the present and future demands while minimizing the environmental degradation from electricity generation pathways presents a significant sustainability challenge. Urban areas consume around 75% of global energy supply yet urban energy statistics are scarce all over the world, creating a severe hindrance for the much-needed energy sustainability studies. This work explores the scope of geospatial data-driven analysis and modeling to address this challenge. Identification and measurements of human habitats, a key measure, is severely misconceived. A multi-scale analysis of high, medium, and coarse resolution datasets in Egypt and Taiwan illustrates the increasing discrepancies from global to local scales. Analysis of urban morphology revealed that high-resolution datasets could perform much better at all scales in diverse geographies while the power of other datasets rapidly diminishes from the urban core to peripheries. A functional inventory of urban settlements was developed for three cities in the developing world using very high-resolution images and texture analysis. Analysis of correspondence between nighttime lights emission, a proxy of electricity consumption, and the settlement inventory was the conducted. The results highlight the statistically significant relationship between functional settlement types and corresponding light emission, and underline the potential of remote sensing data-driven methods in urban energy usage assessment. Lastly, the lack of urban electricity data was addressed by a geospatial modeling approach in the United States. The estimated urban electricity consumption was externally validated and subsequently used to quantify the effects of urbanization on electricity consumption. The results indicate a 23% lowering of electricity consumption corresponding to a 100% increase in

urban population. The results highlight the potential of urbanization in lowering per-capita energy usage. The opportunity and limits to such energy efficiency were identified with regards to urban population density. The findings from this work validate the applicability of geospatial data in urban energy studies and provide unique insights into the relationship between urbanization and electricity demands. The insights from this work could be useful for other sustainability studies.

Table of Contents

Appendices   **142**

Vita   **146**

List of Tables

xiii

List of Figures

Chapter 1

Introduction

Urbanization is a defining signature of human civilization in the post-industrial revolution era. More than half the world's population have been living in urban areas since 2007, reaching to about 55% at present. As per the United Nation's 2018 World Urbanization Prospect report, the proportion of urban dwellers will reach around 68% by 2050. This massive growth will add 2.5 billion new urban dwellers, and almost 90% of this new growth will take place in Africa and Asia [198]. On the other hand, the global rural population is expected to peak in 2030 and decline thereafter [73], marking a point from which all future population growth will be urban. Globally, urban energy consumption accounts for nearly 75% of the global primary energy supply [196], and this causes around 71% of energy-related CO_2 emissions worldwide [178]. The unprecedented urbanization rates and a gradual convergence in the global urban landscape pose severe challenges for energy and environmental sustainability. To reach the current sustainability goals, addressing the challenges to energy security, environmental compatibility of energy systems, and equitable access to clean energy at local to global scales have assumed prime importance [73].

Urban areas are characterized by a high concentration of human activities, which fosters innovation and socio-economic growth, offering pathways to economic prosperity and human development [166]. Consequently, urban areas act as the economic engines of the modern society. The cumulative urban economic output covered more than 80% of the global gross domestic product (GDP) in 2011, the 600 largest cities alone contributed to about 60% of the global output. It is expected that the contribution of largest cities towards global economy

will stay at similar levels through the year 2025 [166, 41]. Urbanization has also been found to have a positive impact on poverty alleviation, a study by the Asian Development Bank highlighted a 73% reduction in the number of urban poor between 1990 and 2008 in East Asia and Pacific, while the corresponding reduction in rural poverty was estimated at 69% [125]. Opportunities of employment in manufacturing, trading, or service jobs, coupled with the higher quality of life attract increasingly more people from the countryside to urban centers, contributing to their rapid growth [144], a trend that is expected to continue in the foreseeable future.

However, with the emergence of urban areas as the hub of human activities, their negative effects on the environment have exacerbated significantly. While occupying only about 3% to 4% of the World's available land area, urban areas exhort a disproportionately massive effect on environmental and human well-being from all around the world [166]. Negative effects of urbanization have been extensively documented in the existing scientific literature. Crucial environmental systems such as global bio-geochemical systems [180], land resources [99], precipitation patterns [103], atmospheric and climatic systems [180], human health [140], and the well-being of urban dwellers [74] were observed to be severely affected by anthropogenic activities originating in the cities, often the detrimental effects occurring far beyond the immediate urban boundaries. Present energy generation pathways negatively affect food, water, and environmental systems [120, 143, 9]; and the associated emission of greenhouse gases severely harms the atmosphere and environment at multiple scales [147]. The unceasing population growth also escalates the demands for goods and services in cities, far exceeding the limits of existing capacities. Consequently, urban infrastructure and provision of essential services are experiencing severe stress [127].

Recognizing this overwhelming control of urbanization over the socio-economic landscape, environment, and human well-being; several recent reports have predicted a shared global future which is almost entirely determined by urbanization dynamics. This has led several scholars to point out that we are already living on an "urban planet" [166, 73]. The crucial role of urbanization as the source and the solution to complex global challenges towards determining the future of humanity has started to be widely recognized. It is now critical to ensure that this inevitable transition occurs in an environmentally sustainable, socially

and economically equitable manner. Hence, sustainable urbanization has become an active research area in recent times. Sustainable urbanization mandates that the socio-economic advancements and physical developments, which are dependent on continuing supply of natural resources, are designed to last for a long time [194]. The European Environment Agency recommends five goals of urban sustainability which aims at the minimization of space and natural resource consumption, efficient management of urban flows, protection of the health of urban population, provision of equal access to services and resources, and preserving social and cultural diversity [36]. While energy consumption is necessary for economic growth and socio-economic equality, the urban energy consumption pathways also cause irreversible damage to the natural environment and the dependence on non-renewable energy fuels accelerates the depletion of natural resources. Understanding of the effects of urban energy usage on environmental systems and socio-economic structures from regional, national to global levels is critical towards attaining urban energy sustainability. However, several pieces of key knowledge in this system such as the accurate measure of urban areas and the trends in urbanization, energy consumption patterns in cities, and effects of urbanization on energy consumption are still in nascent stages.

1.1 Urban energy dynamics

The International Energy Agency indicates that between 1973 and 2015, the global primary energy consumption grew by 101.3%, from 4,661 Mtoe (million tonnes of oil equivalent) to 9,384 Mtoe [92]. Urban area-specific estimates indicate that 240 Exajoules of energy was consumed in 2005 (the equivalent of 5,732 Mtoe). By 2050, global urban energy consumption is expected to reach 730 Exajoules (equivalent of 17,440 Mtoe) [35]. In the world, electricity is the second largest energy fuel by consumption, which accounted for 18.5% of total global usage in 2015. For the OECD (Organization for Economic Co-operation and Development) countries, electricity holds a slightly larger share of 22.2% in the energy portfolio. Compared to 1973, when electricity's share was only 9.4%, this growth is highly significant. It is also notable that the global share of Oil, the largest energy source, fell from 48.3% to 41.0% during the same period. The consumption of electricity grew by 296% globally and 149% for the

OECD countries in this time frame [92], which is almost three times the growth of total energy usage. These statistics unambiguously portray the importance of electricity in present times. Urban areas require significant amounts of electrical energy for operation and maintenance of the built environments that include buildings and associated infrastructure both indoor and outdoors, e.g., appliances, machines, transportation, and lighting infrastructure. Increasing electricity access around the world, especially in rural areas, modernization of existing infrastructure with electrical systems and machine automation, and the introduction of electric vehicles in personal as well as mass transportation are expected to further increase electricity demands in the future.

Urban energy systems are distinct from other energy systems. The high concentration of population and personal income in urban areas contributes to a high density of energy demand. Urban energy demands typically ranges between 10-100 W/m^2, but can reach as high as 1000 W/m^2 as seen in Tokyo, Japan [73]. The high intensity of energy use may increase the emission of air pollutants and waste heat from the energy generation process, contributing to pollution and urban heat island effects. However, the high intensity and diversity of energy usage in urban areas could also present opportunities for better energy management practices, energy recycling via exchange of materials and waste between industries, and clean-energy practices [21, 73]. Along with characteristics such as locational attributes and population size, urban energy consumption is also shaped by its economic functions. Urban areas are the hub of trade, service, and manufacturing,. These activities considerably shape their consumption profiles. For example, urban areas with a significant presence of manufacturing sector could be expected to consume more energy per capita than areas primarily engaged in the service sector [154]. Urban growth also plays a vital role in determining the energy consumption profiles of cities. Prospects of higher employment, wages, and better life draw more people to urban centers. Due to this population growth, the built-environment and support services also expand which further increases the energy demands. In fact, by 2030, the urban infrastructures are projected to expand more than three times their size in 2000 [35], Concerning the urban size, larger, highly populated cities offer more jobs, services, and higher life prospects than smaller ones. Thus they attract people at a much higher rate than their smaller counterparts. This causes the larger cities

to grow and consume energy at much faster rates than the smaller ones. However, the larger urban areas could also be benefited from the co-existence of diversely skilled sets of people in close proximity, leading to heightened socio-cultural interaction which contributes to the technological and innovative edge of large cities [65]. Urban areas, due to the presence of co-located diverse sets of activities and people, present unique potential to lower per person energy demands. The free flow of people, information, capital, and materials also encourage engagement in transitional change, coupled with high-demands, this may help to achieve the returns from economies of scale and lower per person energy consumption. Higher wealth generation in prominent urban areas may also make it feasible to absorb the upfront cost required to implement energy efficiency measures. Efficient building stock, space heating and cooling mechanisms, and lighting technologies may be used to lower electricity demand; while mixed land-use practices, energy-efficient mass transportation may help lower the use of energy in transportation sector. A past study has indicated that the urban areas in the industrial nations indeed consume less per capita energy than their respective national averages [43]. The interplay of social, economic and technological factors determine the relationship between urbanization and energy consumption, setting the urban energy systems apart from other energy-related systems and calling for special attention.

1.2 Challenges to sustainability

The challenges to energy sustainability are well embedded within challenges to the social, economic, and environmental sustainability [73]. The lack of universal energy access poses a critical challenge to energy sustainability goals. Severe inequality still exists regarding energy access between developed, developing, and under-developed nations. International Energy Agency estimates that as of 2016, 1.1 billion people still lacked access to electricity [91]. Africa faces the toughest challenge in terms of energy deficiency, where 16 out of 20 nations with the largest electricity deficit are situated [199]. In determining energy access, the energy affordability plays a critical role. The typical energy expenditure in developing countries often exceeds 10% of the household income, but the household incomes in low and medium income countries are so low that after provisioning for food and other basic necessities there

is very little left for spending towards energy. An average household in Kuwait and the United States respectively consumed 40,142 kWh and 12,305 kWh of electricity in 2014, while an average Indian households consumed about 1,165 KWh, and the corresponding consumption in an Ethiopian household was only 534 KWh [94, 213]. The low affordability of clean energy results in the use of cheap, inefficient fuels among rural and urban poor households. To keep the energy costs low, solid fuels like wood, coal, charcoal, and animal waste are often used for heating and cooking purposes. United Nations estimates that around three billion people around the world are using these cheap yet dirty fuel sources to meet their energy needs. These practices result in severe indoor pollution and degradation to human health. United Nations points out that around 4.3 million people perished globally in 2012 from indoor air pollution-related illness [199]. Existing reports indicate that globally around 30% urban residents lack access to clean fuel and modern energy technologies. Due to the lack of affordable energy supply, urban poor change their fuel mix based on fuel availability, prices and subsidies, season, and cultural beliefs. While these asymmetries are less pronounced in developed nations, they assume much higher importance in low income, developing countries. Inequalities also exist between different income groups within nations or cities regarding energy-related expenditure, physical access to electricity or other energy sources, and choice of fuels meet the energy needs. The expenditure differences between urban high and low-income groups were found to vary by a magnitude of 100 or more [73]. However, lack of appropriate data and monitoring methods hinders analysis of these challenges to energy access.

Environmental challenges from energy consumption mainly occur in the form of air pollution, the pressure on water resources, and land-use change arising from the energy generation process. Within the urban areas, the high energy demand increases the chances of air pollutant emission. Typical energy use in larger urban areas ranges between 10 100 W/m^2, whereas the clean, renewable energy sources can only supply about 0.1-1 W/m^2. This deficit increases the reliance on fossil fuels that has significantly more detrimental impacts on air qualities. Electricity generated through fossil fuel, biomass, and waste burning results in the emission of oxides of carbon, sulfur, and nitrogen; particulate matters; and heavy metals such as Mercury into the atmosphere. EIA estimates that around 67% of the total electricity

generation in the US in 2016 came from these non-renewable fuels. In the United States, power plants were found to contribute to 64% of economy-wide SO_2, 50% of Mercury, and 75% of acid gas emissions [124]. Exposure to air pollution is not solely an urban problem. The rural areas face 66% of the global air pollution exposure. According to the global exposure equivalent [183], the prime cause of this exposure in the developing nations is indoor air pollution. High intensity of energy use and low air movement in urban areas lead to the urban heat islands which makes the cities warmer than surrounding countryside by 1-3ºC, however temperature differences of upto 12ºC have been observed. The high thermal mass of urban buildings and waste heat from energy generation and space cooling processes also contribute to this process; consequently the effects of urban heat islands increases with higher energy consumption [73].

The air pollutants released from energy generation can get into the water resources through wet deposition, and degrade water quality. The hot and chemical laden wastewater from steam power plants can have a detrimental impact on the water quality of the surrounding area upon discharge. Various heavy metals like mercury, selenium, cadmium, and chromium have been found in the power plant discharges. These hazardous materials may enter the food chain and in turn affect the human population in addition to other ecosystem losses. The steam turbine enabled electric power plants account for 30% of toxic waste discharge from all industries in the US [124]. As vast amount of water is needed in mining energy resources, refining processes, and in the energy generation process itself; the energy sector was estimated to be the largest user of water in the US [90]. The power plants also have large footprints in addition to that required for resource mining, transportation, pipelines, transportation systems. Many of these land requirement occur outside the urban boundaries, but the impact of these land-use changes can be felt in the ecosystem and wildlife around the world.

1.3 Lack of energy data and opportunities for geospatial data-driven methods

Notwithstanding the clear urban centric nature of the global energy systems, the current methods of energy statistics are predominantly focused on national or regional scales. The lack of appropriate data severely hinders the studies of urban energy systems [65, 73]. The dearth of standardized data collection and reporting system also renders the available urban datasets useless for comparative analysis [65]. A 2012 meta-analysis [104] of 219 studies related to urban energy modeling revealed that the resolution and reliability of these models were negatively impacted by the data availability in addition to computational constraints. Around 58% of these studies used datasets with a temporal granularity of a year or more, while 44% studies modeled energy systems at a district or a coarser spatial scale. The data scarcity is even worse in the developing world, where urban scale datasets mostly do not exist. Urban authorities in developed countries have deployed intelligent systems such as smart grids and smart meters in place. Such equipment is capable of recording customer behavior and consumption patterns [5, 55]. However, neither all the cities in the developed world nor the cities in developing countries possess such infrastructure or the capital needed to acquire these technologies [83]. As discussed earlier, almost all the new urbanization in next few decades are expected in these data starved regions of the world, this makes the task of developing efficient and generalizable urban level energy data gathering methods extremely critical for studying the present energy dynamics and predicting the future scenarios. On the other hand, the cities in the developed and industrialized nations still consume most of the global energy supply. Past studies focused on such areas have mostly dealt with energy usage and emissions in the transportation sector, but very less has been reported on electricity consumption.

In the absence of statistical data on electricity consumption, proxy-based estimates have been used by researchers. Building energy end-use intensity arising from domestic hot water usage, space cooling, and electricity consumption was estimated and mapped at block level in New York [89]. An interactive interface for this project can be viewed at https://qsel.columbia.edu/nycenergy/. The distribution of electricity consumption in three counties in

the state of Tennessee, US was estimated using a hybrid dasymetric mapping and machine learning approach [141]. However, both studies utilized detailed electricity consumption data, obtained from local administrative offices or utilities. While such studies prove that electricity and broader energy consumption can be modeled at high resolution with the appropriate data, a more generalizable approach is needed for understanding energy usage patterns in data starved regions, or for a large system of cities where detailed data may not be available for all entities.

In the context of developing an effective urban energy data collection and assessment mechanism, the 2012 Global Energy Assessment report by International Institute for Applied Systems Analysis highlights two key areas that need immediate attention: (1) A unified approach to urban boundary and definition of energy systems, and (2) availability of energy data, its quality assurance, and detailed documentation of data generation methods. Under the circumstances where the development of a global urban energy statistics system is still uncertain, geospatial and remote sensing data-driven approaches could be used to study extents and patterns of urbanization and define proxies of urban electricity consumption. Over past three decades, satellite recorded nighttime lights emission from human activities have been at the forefront of such applications. However, a few known caveats have kept these datasets from being reliably applied at urban scales. A detailed overview of the present state-of-art has been provided in the literature review section. With newer high-resolution datasets and increasing computational power, which enables complex analysis of large volumes of data, the boundaries of remote sensing image analysis can now be pushed further. By extracting information on urban configurations, objects and their characteristics, deeper understanding of urban energy dynamics may be generated. Due to its scale dependency, the geospatial analysis is possible at different scales, where at each level the amount of available information determines the scope of such analyses. This has been illustrated in figure 1.1 (adopted from Netzband & Jürgens, 2010). The strength of geospatial data-driven analysis lies in its flexibility to focus on and connect between different spatial scales. With present capabilities, data extraction from satellite images can help the detection of urban structure types [145, 72, 8]. The current challenges remain on how to meaningfully utilize the information on urban configuration and form to predict its function.

9

Scale		Evaluation	Scope
City/Municipality		Urban Land use and land cover analysis.	Land-use changes, suburbanization, urban heat islands, flooding, landslides.
Local districts		Analysis of urban morphology.	Urban configuration, land-use, settlement structures.
Local neighborhoods		Identification of urban structure types.	Degree of impervious land-cover, analysis of mixed and homogenous structures, environmental quality analysis.
Buildings		Identification of urban structure elements.	Urban dynamics at local scales, disaster management, indicators of settlement types, high-resolution energy models.

image from: www.shutterstock.com

Figure 1.1: Relationship between spatial scale and urban analysis *(adopted from Netzband & Jürgens, 2010)*.

For understanding urban electricity consumption patterns in a spatially explicit manner, accurate detection of urban areas is a critical first step. The consistency issues with global human settlement and land cover datasets is a potential challenge for reaching a consensus regarding a generally acceptable measurement of the urban areas in the world. However, the high resolution image analysis could provide a way forward in this regard by enabling minute detection of urban patterns with unprecedented details. The United Nation's sustainable development goals in clean and affordable energy focuses on universal access to sustainable, modern, and clean energy supply, including the least-developed countries by 2030. As many of the low income countries lack reliable statistical data, remote sensing based identification and assessment of urban energy consumption patterns may prove to be crucial in achieving such goals. The advancements reached in the domains of remote sensing and geospatial data analysis provide an opportunity to fill these data voids till a more efficient data collection system can be put in place. After the necessary identification of urban areas and proxies of energy consumption, geospatial data driven modeling of urban energy systems is also needed to analyze the present dynamics and identify the probable future trends. Such efforts can

fulfill the urgent data needs and immensely benefit not only the developing countries but even many developed countries as well.

1.4   Data driven analysis in Geography

The philosophical foundation of the data-driven analysis paradigm can be found in Abductive Reasoning. Other inference approaches in scientific knowledge discovery, such as the deductive approach develops a hypothesis based on the exiting theory; it then tests the hypothesis to either accept or reject it. Thus, such inference mechanisms thereby assume that something must be true. The inductive reasoning on the other hand is more flexible that allows the possibility that the conclusion may even be false and the strength of argument rests on the probability of the conclusion being true. In comparison to other inference approaches, abductive reasoning is a much softer method which assumes that something may be true. Abductive reasoning approach begins with the data, through observation and exploration of such data a direction is obtained. The process ends with a hypothesis which in turn explains the data [134, 135]. Researchers [134] have argued that this approach holds critical importance in scientific discovery especially before one reaches the stage of deductive or inductive methods of knowledge building. Miller & Goodchild [135] points out that for the success of abductive reasoning, various sub-components such as the capabilities of being able to propose fragments of theory, availability of a large knowledge set from which information can be drawn, the ability to search the knowledge base for correlation, patterns, and explanations; and strategies for complex problem solving including analogy, guesses, and approximation are necessary. The importance of geographic visualization could also serve as a core framework upon which different inferential strategies could be applied, leading to a synergy among different knowledge building approaches in Geography [69, 135]. Figure 1.2, adopted from Gahegan 2009 [69], illustrates the relative positions of different inferential approaches in geographic knowledge discovery process.

In the context of geospatial data driven analysis in urban studies, availability of remotely sensed imageries has shaped much of the manner in which the derived datasets and applications have evolved. The availability of high resolution remotely sensed datasets
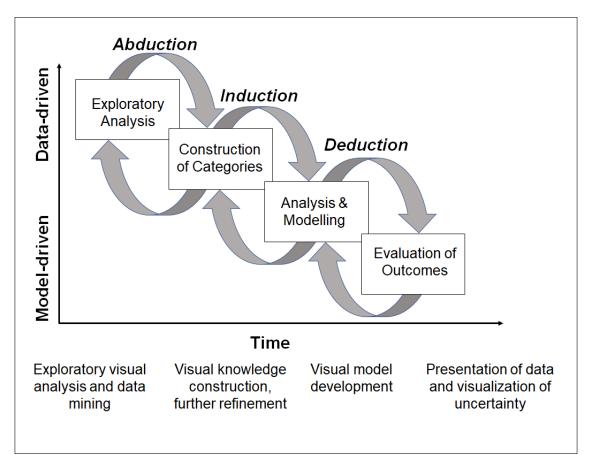
Figure 1.2: Overview of the inference approaches in geographic knowledge discovery process *(adopted from Gahegan, 2009).*

along with efficient computing resources and algorithms in recent times can efficiently process massive datasets to extract information at high spatial resolution. This enables the identification of urban patterns and objects with high detail, which can lead to identification of patterns in urban areas and development of correlations that have not been discovered so far. In the absence of statistical data, new urban metrics could be generated that satisfactorily describe the socio-economic patterns in cities, and thereby lead to hypothesis generation. The subjective background is of high importance in this stage to rule out spurious correlations in the hypothesis building process. The principle of data driven analysis and modeling may be used to study human settlement and energy interaction leading to more efficient and spatially explicit local level electricity usage mapping, especially at data sparse regions. As [135] points out, the geography has moved into data-rich environment that offers variety of near real time data. Thus, these multitude of datasets need to be harnessed to

study urban phenomenas, especially under the scarcity of urban statistical data. Progress in this approach may ultimately enable the planners and policymakers to base their decisions on more reliable information.

1.5   Organization and summary

The ongoing discussion elucidates the role of urbanization in shaping the present and future course of modern human civilization. Given its influence on the world economy and the control over individual prosperity, urbanization is bound to continue at its current pace in the foreseeable future. Electricity, which is gaining market share much faster than any other energy fuel, is mostly utilized in urban areas; which indicates its ever growing importance at present times and for the future. Electricity consumption pathways are known to have significant environmental footprints, negatively affecting air, water, and land resources. Reliance on non-renewable fuel resources aggravates the challenges of their depletion and energy security. Unlike other energy end usage, electricity has been less studied until now. The lack of urban scale data is one of the main hindrances that must be overcome in order to ensure sustainable energy policy development. In the absence of a statistical mechanism on urban level energy consumption, estimates may be generated using geospatial data-driven methods. This dissertation work attempts to address this crucial issue by exploring the application of geospatial data and modeling in understanding electricity consumption patterns over diverse geographies. After a comprehensive review of the present literature, three key knowledge gaps were identified, which have been presented in section 2.4. The current chapter provides an introduction to urbanization, the present trends, and influences on the global socio-economic and environmental systems. Which is followed by a background discussion on urban energy consumption and the challenges it presents for sustainability. The discussion then highlights the issues of data scarcity in urban energy domain, and the opportunities for geospatial data-driven methods to fill the voids. The organization of the rest of this work is detailed in the following.

A comprehensive literature review has been presented in the second chapter, which highlights the present state-of-the-art in three related but key areas pertinent to this study.

(1) The present status and the issues related to global and regional level mapping of human settlements. Reliable identification and measurement of the human habitats is a critical step to address a global level issue like the energy sustainability. Accurate measurements of urban areas and mapping of intricate urban patterns helps in identification of the changes in urban landscapes, find vulnerable sections of population without access to energy and services, and develop correlations between urban configurations and energy indicators. However, the present state-of-the-art is severely discordant regarding the measurement of global urban lands. The degree of such discord can be found to vary with scale and regions of the world. The literature review identifies that despite the development of high-resolution settlement datasets, their performance assessment against the coarse and low resolution datasets at multiple scales and over diverse geographies is severely lacking. This obfuscates the assessment of the benefits obtained from high-resolution mapping compared to the conventional approaches, and hinders a gradual progression towards a convergence in global human settlement mapping efforts. (2) Understanding of urban energy consumption patterns using remote sensing approaches. Studies have noted the capability of high-resolution satellite images to record various morphological traits of urban settlements, which can be analyzed to differentiate between settlement types. The literature indicates that there is a potential of using the functional inventory of human settlement in studying urban socio-economic functions. However, no attempt has been made to explore the usability of these inventories in understanding urban energy consumption patterns. (3) The impact of urbanization on electricity consumption. While there have been many studies that investigated the relationship between urbanization and energy consumption at national and regional levels, decidedly less has been reported at urban levels. As mentioned above, the scarcity of appropriate statistical data severely hinders such endeavors. Even in the United States, the second largest consumer of electricity in the world, urban level datasets are very limited. The limited number of studies that studied the effect of urbanization on electricity at city levels indicate diverging directions of the effect. In some cases, return of scale has been observed, highlighting a reduction in per capita energy use in bigger urban areas. But there are other studies that reported a liner effect of urbanization or cases where urbanization mildy increased per capita electricity consumption in urban dwellers. Thus,

14

further exploration of this issue is required to discern the effects of urbanization on electricity consumption and assess the potential of urbanization in lowering electricity consumption, which may be crucial for future energy sustainability. Finally, the three research objectives of this dissertation, reached upon after through consultation of the existing body of knowledge, has been presented.

The third chapter describes the research done to address the first research objective. Global and regional estimates of the extents of built-up and urban areas vary severely among different medium and low-resolution datasets. Such discord poses a severe challenge for understanding the impacts of urbanization on several global and regional level systems including energy. Identification of urban patterns at local to global scale allows to develop correlations with energy dynamics; urban extents can also be used as a proxy of different socio-economic parameters. This chapter thus presents a study of assessment of the mutual correspondence between four distinct global and regional level land cover and built-up area datasets, ranging between 8 and 500 meter in spatial resolution. The key questions asked in this chapter are:

1. What is the variation in built-up area estimates amongst the datasets at multiple scales and geographies?

2. Whether the observed discords amongst the datasets are associated with morphological traits of the urban landscape, and what are the implications of such association?

The study was conducted at multiple scales in Egypt and Taiwan, offering distinct geographies regarding physiography, economy, and demography that could influence the mutual correspondence among datasets. This study analyzes the influence of landscape complexity on the mutual correspondence amongst the datasets in these two diverse landscapes. Findings from this study indicate that increasing disaggregation and complexity of urban landscapes are associated with an increasing mutual disagreements; while increasing contiguity is associated with an increase in the mutual agreements. The study also calculates the marginal effects of these landscape characteristics on the inter dataset mutual correspondence to assess the degrees of the association. From core urban areas, the mutual agreements were seen to decrease over peri-urban areas, where most complex urban

landscapes are found. Due to lack of spatial resolution in the input data, coarse and medium resolution datasets are deficient in identifying complex urban patterns. This chapter sheds light on the causes of the observed discords amongst datasets, and firmly establishes the potentials of high-resolution urban mapping as a path to convergence in measurements of urban extents.

The absence of statistical data on urban energy is a major obstacle in understanding urban energy consumption patterns. This problem is much more severe in developing nations. Past studies have applied nighttime lights data to study energy consumption patterns at global and regional scales, but the limitations around spatial and spectral resolutions restrict its applications at urban scales. The new VIIRS DNB datasets offer much higher radiometric resolution, which enables the identification of subtle variations in light intensity caused by different energy consumption levels. However, its 500 meter spatial resolution acts as a limiting factor for understanding intra-urban energy consumption patterns. Using a high-resolution metric alongside nighttime lights may help in spatially explicit understanding of urban energy consumption patterns. In the fourth chapter, the following questions related to this issue were addressed:

1. Can a generalizable functional inventory of human settlements be developed for cities in developing nations?

2. Can such inventories be used alongside nighttime lights, used as a surrogate of electricity consumption, to draw insight on urban electricity consumption patterns?

The study was conducted on the three cities of Johannesburg in South Africa, Sana'a in Yemen, and Ndola in Zambia. The regression results between the proportion of settlement types and the corresponding nighttime light emission were found to be statistically significant in all three cities. It was found that in Johannesburg, the high density, large urban buildings had the maximum association with nighttime light emission, which can be interpreted as an indication of high electricity usage in commercial and industrial establishments. In Ndola, the medium sized, low to medium density urban and suburban residential buildings were found to have maximum association with nighttime lights. In Sana'a, highest association with nighttime lights emission was associated with the large buildings as well as the small,

16

non-orthogonal settlements, which are most likely representative of industrial, commercial and administrative establishments and high density, small informal settlements respectively. The results from this study highlights the applicability of functional settlement inventories to study intra-urban socio-economic dynamics. Addition of such information to the existing high-resolution settlement datasets may further enhance their usability in economic and social studies.

The relationship between levels of urbanization and electricity consumption in an urban system indicates the effects of urbanization on energy consumption, indicate probable future scenario with respect to prevailing rates of urban growth, and help in planning and policy making. Urban scaling is an efficient way of capturing such dynamics. However, due to the lack of urban level data, such studies are rare. The study presented in the fifth chapter is focused on the urban areas in the United States, which is the second highest consumer of electricity after China. To overcome the lack of data, an electricity consumption estimation model was developed before conducting the scaling analysis. The specific questions addressed in this chapter are:

1. Can geospatial data based models be used to estimate urban electricity consumption in the United States?

2. What is the effect of urbanization on electricity consumption in the United States, and its implications?

3. What are the implications of the return-of-scale in urban energy consumption in the United States, and where are the limits of such gains?

The urban electricity consumption estimates were validated against the California data, which is the only available available electricity consumption data at county scales. The estimates were showed a mean relative error of 12%. Since the electricity data for the country was only available at the state level, the regression model was used to make county-level predictions, which was beyond the range of the initial data. This caused some systematic error in the estimates, which was evident from the distribution of estimation errors relative to urban size. The larger urban areas showed far less error than the smaller ones. The

subsequent scaling analysis indicated an increase of 77% in electricity consumption associated with each doubling of urban population in the US. The gain of 23% less electricity usage per person corresponding to a 100% urban population growth is indeed an indication that urbanization can offer energy savings in the US. However, limitlessly expanding urban areas may not be a practical policy solution for energy sustainability. On the other hand, larger urban areas are linked to urban heat islands, faster resource depletion, and other associated sustainability challenges. Thus, the relationship between per capita electricity consumption and urban population density was assessed in this analysis. The results revealed that urban areas offer rapid gains in per-capita electricity consumption in the initial stages of growth till the population densities reach about 800 persons/km^2. The gains increasingly slow down between the population densities of 800 to 2000 persons/km^2; however, beyond this point there is no more in per capita consumption savings. Hence, there is an effective higher limit to the effectiveness of urbanization with regards to energy savings. The study presented in this chapter thus not only highlights the applicability of geospatial data modeling to overcome the scarcity of urban data, but also assesses the potential and limits of urbanization as a tool for energy efficiency.

The sixth and the final chapter includes a systematically organized conclusion from the previous chapters, along with a discussion on the uncertainties and the limitations of this work. The chapter ends with a brief outline on the future research direction.

Chapter 2

Literature review

This chapter presents a comprehensive review of the existing body of literature pertinent to this dissertation. The discussion starts with a description of the critical nature of remote sensing based human settlement mapping. The issues associated with human settlement mapping at local to global scale has been described next along with a discussion on the past work done on accuracy assessment. The review then focuses on the existing work related to the application of remote sensing in urban socio-economic analysis, especially on using satellite image-based metrics as proxies of urban socio-economic processes. The final part of this discussion includes the studies related to the application of geospatial data-driven methods in the analysis of urban sustainability, focusing on electricity consumption. As mentioned in the introduction, there is a global scarcity of urban electricity consumption data which hinders researchers from assessing the effects of urbanization on electricity consumption, consequently the number of studies reporting on urbanization and electricity consumption is relatively less.Thus, a broad discussion on the effect of urbanization on general energy consumption and the implications, as noted by past researchers, has been presented before specifically focusing on electricity consumption. This chapter then identifies the knowledge gaps in view of the existing literature, which provides the basis of the research objectives of this dissertation.

2.1 Global and regional human settlement mapping

Urban areas occupy only between 3% to 4% of the world's land area [166], yet the significance of its impacts on global environmental systems and human health are already well-documented [63, 47, 164]. The current projections estimate that urban growth will add around 2.5 million additional people to urban areas by 2050, with the majority of such growth taking place in Asia and Africa. However, the urban areas trajectories of the urban growth to accommodate these additional population is not yet clear. Similarly, the effects of such future growth on the environment is also less understood. Such uncertainties around future urban growth mandate accurate and regular mapping of the urban areas at multiple spatial scales. Past researchers [163, 164] have strongly recommended that the crucial tasks of accurate mapping and measurement of urban areas and associated socio-economic processes must be undertaken in order to identify and understand the influences of urbanization on global and regional scale anthropogenic and environmental systems. Remotely sensed images constitute an essential input data to the settlement mapping process [177]. Benefit analyses of the remote sensing based urban mapping over traditional methods such as census and ground surveys reveal two key advantages. Firstly, unlike remote sensing data, conventional survey methods lack the spatial information which renders them unsuitable for many scientific applications [175]. Secondly, survey-based methods are extremely resource and time consuming, which is why such undertakings generally occur at decadal intervals [168, 45]; whereas remote sensing satellites provide continuous coverage at multiple scales and facilitates regular updates to the human settlement inventories. Recent studies [15, 107, 26] in the application of novel data extraction methods have established that information on human settlements can be methodically extracted from these images at different spatial scales using automated or semi-automated data methods. The efficiency of this process may be further increased using high-performance computing resources capable of unprecedented processing speeds [153]. Such advancements helps efficient analysis of very large volume of data, typically associated with the high and very-high resolution images. The following subsections provide a detailed overview of coarse and medium resolution land-use land-cover datasets and studies that investigate their consistency and indicate their strength and

weaknesses. The new high-resolution datasets are then described along with the current state of their accuracy assessment.

2.1.1 Coarse and medium resolution datasets:

The Global Land Cover Characteristics Database (GLCC) represents the first instance of global land cover mapping using satellite images, it was jointly developed by the United States Geological Survey (USGS) and the European Commission's Joint Research Center (JRC) for the 1992-1993 period [121, 164]. Since then, over the past few decades, a wide range of global and regional land cover and human settlement datasets have been developed. Urban extent datasets from the Global Rural-Urban Mapping Project (GRUMPv1) were provided in both polygon vector and gridded raster formats at 30 arc-second (about 1 km) resolution for the year 1995[30]. The yearly MODIS (Moderate Resolution Imaging Spectroradiometer) land cover type product [68, 24] was generated using images from MODIS sensors on-board Terra and Aqua satellites. This 500 meter data describes global land-cover types at yearly cycles between 2001 and 2013. European Space Agencys Climate Change Initiative released a 300 meter global land cover data [82] for 2008-2012, 2003-2007, and 1998-2002 time periods. These datasets were developed using MERIS (Medium Resolution Imaging Spectrometer) and SPOT (*Satellite Pour l'Observation de la Terre*) vegetation data as input. The CORINE (Coordination of Information on The Environment) land cover project by European Space Agency (ESA) was initiated in 1985 and land cover maps consisting of 44 classes at 250 and 100 meters were produced for 1990, 2000, 2006, and 2012. The CORINE data development process utilized a suite of datasets from Landsat-5, 7, SPOT 4/5, IRS (Indian Remote Sensing Satellite) P6 LISS III (Linear Imaging self Scanner) and RapidEye satellites [57, 155]. Coarse and moderate resolution human settlement datasets have previously been used to study a broad range of scientific areas such as population distribution and modeling, climate change, resource management, urban planning, disaster response, health planning, and energy dynamics [42, 158, 177, 211, 130, 80, 136].

2.1.2 Consistency analysis of low resolution datasets:

Past researchers [173, 163, 164, 80, 174], however, have noted a significant discord among the low and medium resolution datasets, especially regarding the representation and estimation of built-up and urban areas. An urban area focused comparative accuracy assessment [163], involving six global level coarse and medium-resolution datasets around the year 2000, revealed a large range of global urban area estimates. In this study the smallest estimates of urban area of 0.27×10^6 was obtained from the Vector Map Level 0 data [38], while the largest measurement of 3.52×10^6 km^2 was obtained from the GRUMP data. At a regional level, the study found the highest inter-map agreements in North America ($\bar{r} = 0.90$), followed by Europe, central and South America, and Sub-Saharan Africa ($\bar{r} = 0.78$), the lowest inter-map agreement ($\bar{r} = 0.63$) amongst these datasets was seen in Asia.

A follow-up study [164] involving eight global datasets adopted a two-tier approach, utilizing a stratified random sample of 10,000 high-resolution Google Earth images and 140 medium resolution Landsat images of urban areas. While assessing the omission of cities in these datasets using a 247 city sample set, the study noted that most of these datasets either totally missed or only partially mapped some of the cities included in the sample. This rate of omission was generally inversely related the city size and economic levels of the corresponding country. The mean omission rate for European and other developed nations was 0.2% while the same for Africa, South-Central-West Asia, Southeast, and East Asia were 2.8%, 5.5%, and 5.8% respectively. Similarly, the assessment of estimated city size showed a wide range of accuracies. Finally, pixel by pixel comparison was carried out using the 140 city sample, the highest mapping accuracy was obtained for Europe and other developed countries along with Latin America and the Caribbean Islands, while the lowest accuracies were found in Africa, East and South-East Asia.

It should be noted that the scale of representation, whether global or regional, could potentially have a bearing on the variation across these estimates [109], which was evident in these studies. The ambiguity around a universally acceptable definition of urban areas across the datasets also creates a major drawback for their applicability [174, 109]. Region

or site-specific variations in urban area mapping accuracy could be another issue, as region-specific accuracy measures may not necessarily be an indication of a global accuracy. The spectral and spatial heterogeneity of build-up areas also affects the data accuracies. The coarse resolution datasets were found deficient in resolving the spatial complexities of urban settlements [109]. Therefore, such issues could result in unwarranted discrepancies, such as failure to detect built-up area, inaccuracies in derivative datasets, or potentially erroneous and unreliable results and conclusions from the subsequent analysis that use these datasets [58]. As human settlements exhibit complex form and uneven distribution over space, their precise detection is required to accurately convey the urban dynamics through derived products [201].

2.1.3 High-resolution urban data:

The precision of built-up area detection is a function of the input remote sensing image characteristics and methods of information extraction [77]. The spatial resolution of input images determines the amount of extractable information content present in the data. The advantage of high-resolution satellite images (<5m) in capturing complex landscape patterns over medium (80-15m) and low (>100m) resolution datasets have been highlighted in published literature [156, 16, 109]. Recently, the increasing availability of high (<5 meter) and very high-resolution (<0.5 meter) remote sensing images, high-performance computing infrastructure, and associated development of efficient algorithms to exploit high-resolution imageries have paved the way for generating high-resolution human settlement data [27, 203]. The Global Human Settlement Layer (GHSL) was released in 2014 by the European Commission's Joint Research Center as a global data at 38.2 meter (fine) and 305.8 meter (aggregated) resolutions [157]. The datasets were developed applying a symbolic machine learning based classification method [34, 160] and a texture extraction method called PANTEX [109, 159] on cross-platform and multi-sensor image data. Another dataset called Global Urban Footprint (GUF) was produced by the German Aerospace Center at 2.8 (aggregated) and 0.4 (fine) arc-seconds (approximately 84 and 12 meters at the equator) [56]. GUF was developed using local speckle analysis on 3 meter radar data from TerraSAR-X and TanDEM-X missions [109]. In the United States, the Oak Ridge National

Laboratory (ORNL) is using very high resolution (≤ 0.5m) satellite images from WoldView 2 and WoldView 3 satellites to develop a human settlement dataset. This dataset, known as LandScan Settlement Layer (LandScan SL), is generated at approximately 8 meter spatial resolution. All the urban and land cover datasets mentioned above, with the exception of the last three, have been widely used by researchers in the advancement of scientific disciplines worldwide. However, the new high-resolution datasets have now begun to be used in scientific analyses [67, 2, 44], and by virtue of their inherent higher information content, these datasets are expected to provide new insights and push the boundaries of present scientific knowledge.

2.1.4   Evaluation of the advantages of high-resolution datasets:

To assess the true advantages of the high-resolution datasets over their coarse and medium resolution counterparts, comprehensive performance analysis of high-resolution settlement datasets against others is needed. These comparative studies could indicate their advantages in scientific applications, especially regarding the identification and quantification of human settlements across variegated geographies. Important issues that must be addressed are the relationship between input data resolution and information gain at various geographic scales and sites. Given these circumstances, the inconsistencies must be explored to provide a sound basis for using appropriate datasets in downstream studies in the geospatial data analysis paradigm. However, scientific studies addressing the accuracies of high-resolution human settlement and urban area datasets remain scarce. As per a recent analysis [109], about six and eleven studies were reported on single data product accuracies, respectively involving high and low-resolution data. While there have been a few studies reported on multiple data product accuracies on low and medium resolution data, only two were focused on urban areas, and no study was reported on multi high-resolution data product accuracies. In their study [109] tested the consistencies of GUF and GHSL datasets, both at 12 meters, against 500 meters MODIS global urban extent data and 300 meter GlobCover data. The multi-scale accuracy assessment was carried out over two 100 km by 100 km test sites located in Cologne, Germany and Tuscany, Italy; that offer a balanced representation of urban and rural landscapes [109]. The study found that due to the coarse resolution, MODIS and GlobCover data fails to accurately identify the small-scale built-up areas and consequently

show moderate correspondence with GUF and GHSL only in core urban areas. As expected, the low resolution datasets perform poorly in rural landscapes, but the study found that the GUF and GFHSL also exhibit marginal weakness over rural regions. This study successfully highlights the advantages of high resolution datasets in providing precise information on human settlement extents and patterns. Till date, this remains the only work that examined the benefits of high resolution datasets against coarse and medium resolution data. However, as this study was conducted in European settings, the multi-resolution consistency analysis of high and low resolution datasets over variegated geographies that represent distinct economic, demographic, and physiographic characteristics remain unanswered.

2.2   Remote sensing of urban energy consumption

Beyond the mapping of built-up areas, satellite images could also be used to extract proxy measures to understand socio-economic characteristics on the ground. A summary review of the application of remote sensing in studying urban energy consumption has been presented in this section. Nighttime lights data has been used extensively by researchers at regional to global level for this purpose. However, incorporation of information on urban structure, patterns, and morphology could reveal more insights into urban socio-economic and energy processes. Hence, this section also includes a summary review of studies using high-resolution images in characterizing urban landscapes, with special focus on energy and economic applications.

2.2.1   Estimation of urban energy consumption using nighttime lights:

Welch, 1980, [209] authored one of the initial work of utilizing nighttime lights data from Defense Meteorological Satellite Program's Operational Linescan System sensor (DMSP/OLS) to urban energy usage patterns in 18 cities the United States in 1975, the study found a high level of correlation (coefficient of determination (r^2) = 0.89) between total consumed energy and recorded light emissions, measured using microdensitometer from

25

analogue media. This work highlighted the significance of using remote sensing data in urban electricity usage estimation. A similar method was used to study the relationship between nighttime lights and population (r^2=0.88), and electricity utilization (r^2=0.96) for 35 cities in Eastern and Western United States in the same year [210]. Elvidge *et al.* 1997, found significant correlation between the natural log of lit area and natural log of population (R^2=0.85), GDP ((R^2=0.97)), as well as electricity consumption (R^2=0.96) for 21 countries [49]. These initial studies firmly established the usefulness of nighttime satellite observations of light emission from anthropogenic activities as efficient means for monitoring national and regional level urban electricity usage.

Subsequently, Amaral *et al.*, 2005, [4] observed strong linear correlations between DMSP/OLS nighttime lights foci and population as well as consumed electricity in the Brazilian Amazonia in 1999. Spatio-temporal characterization of electricity consumption between 1993-2002 was carried out in India. The authors integrated the observed change in the recorded light levels with demographic data to characterize the urban development in Indian states and major cities. The study observed increasing light intensities along the peripheries of all major cities as well as some extinction of lights in poverty-stricken areas. Correlation analysis between the growths in population, electricity consumption, and nighttime lights returned a moderate correlation, $r^2 = 0.59$ and 0.56 respectively [23]. It needs due mention that several studies have highlighted the known drawbacks of the nighttime lights data from DMSP/OLS sensor, which prevents applications of these datasets beyond national and regional level analysis. Some of the well-known issues are: [1] the datasets have a coarse spatial resolution of 1 km., which negatively affects their scope in smaller urban areas. [2] the six-bit radiometric resolution, with digital number (DN) values ranging from 0 to 63, causes the data to saturate rapidly over urban cores and miss the subtle differences between different lighting levels. [3] The lack of on-board calibration. [4] lack of sufficient spectral channels required for differentiating the thermal sources of observed lights and low light imaging bands which could have enabled discrimination of lighting types [53, 51, 52]. These known deficiencies in the dataset severely limited it's urban-centric applications. Xie & Weng, 2015 [215] analyzed the factors influencing the relationship between electricity

consumption and nighttime lights emission, and noted that these datasets are not applicable in less affluent regions due to the inability of this sensor in the detection of low luminosity.

Several researchers have used region specific methods to rectify the saturation and noise issues. Letu *et al.*, 2010, [114] in a complex series of rectification measures, first used a noise reduction filter to omit the periodic component (arbitrary noise) and extract stable lights from DMSP/OLS data. They applied an additional area correction procedure for the places at higher latitude. A cubic regression model, based on the spatial distribution of the stable lights, was then developed to rectify the saturation problems in Japan. The model was used in Japan and 12 other countries in Asia, including China and India. The results indicated a high correlation between national electricity consumption and cumulative national level nighttime lights emission (for uncorrected data R^2 =0.88, for corrected data R^2 = 0.94). Time-series DMSP/OLS datasets also suffered from inconsistencies in DN values, due to regular change in satellites. Zhao *et al.* [221] applied a set of inter-calibration parameters, developed by [54], to rectify 1995, 2000, and 2005 DMSP/OLS datasets and study the temporal change in electricity consumption in China. More recently, Elvidge *et al.* [51] studied the percentage of the population with access to electric power across 229 countries using DMSP/OLS and LandScan population data, they reported a total of 1.62 billion people with access to electricity which is very close to the International Energy Agency reported estimate (1.58 billion). However, even with the application of these novel rectification methods, DMSP/OLS datasets fails to satisfy the need of present applications.

With the introduction of Visible Infrared Radiometer Suite (VIIRS) Day/Night Band (DNB) data, most of the previous issues with DMSP/OLS, except for the lack of multi-spectral low light imaging capabilities, have been addressed [52]. Compared to DMSP/OLS data (1 km. pixels), the VIIRS DNB data offers higher spatial resolution (500 m pixels) and has a significantly higher radiometric resolution (14-bit). These improvements could therefore enable minute detection of urban socio-economic phenomena at local scales. A comparative study [181] between DMSP/OLS and VIIRS DNB data was carried out to examine the potential of VIIRS DNB data to model GDP and electricity consumption at multiple scales in China. The authors found that linear regressions between the sum of lights obtained from VIIRS DNB showed higher correlation with GDP and electricity consumption

than with DMSP/OLS at both provincial and prefecture levels. Jaturapitpornchai *et al.* [98] applied Otsu thresholding method [149] and k-means clustering approach on VIIRS DNB data to map urban areas in Thailand with considerable success. Ma *et al.* [122] tested the responses of VIIRS DNB data at local and finer scales to draw insight from the correlation between population, GDP, electric consumption, paved road area, and the total lights intensity in Chinese cities. The study found a strong correlation between increase in nighttime lights and linear growth in urban indicators, and suggests a strong applicability of VIIRS DNB in studying socioeconomic indicators at local scales. A recent study [169] used VIIRS DNB data to observe aggregated human behavior through the variations in electricity demand patterns, especially during special social occasions. The energy consumption behavior during these special occasions were found to follow the socio-cultural boundaries at city, district, and country levels. This study demonstrates the capabilities of VIIRS DNB data to monitor lighting demands on a daily basis at urban scales. These applications indicate a strong possibility of utilizing VIIRS DNB data to study urban electricity consumption at local levels. However, for all these existing studies, the city has been the minimum unit of analysis without any attempt to explore the variations within the cities. Till now, no analysis has been reported on exploring the intra-city electricity consumption patterns using the high radiometric capabilities of VIIRS DNB data.

2.2.2 High-resolution images in urban remote sensing

High-resolution day-time images offer unique capabilities to distinguish between various settlement types based on their physical appearances, some of these characteristics observable in high-resolution images are [145, 72]:

1. Complexity in the appearance of settlement shape, heterogeneity within settlements.

2. Heterogeneous building materials and structures of different types of settlements.

3. Variation in building density and size.

4. Differences in street layout, such as narrow and irregular patterns in contrast with planned road network with wide and gridded roads.

These parameters have been used to identify settlement patterns in diverse urban areas, and could therefore be used do develop a settlement database by their visual formality. A link between the forms to functions of different settlement types can then be made. Such functional human settlement inventory may be used for gathering more insights into urban socio-economic patterns, including energy usage.

Multiple researchers have used spectral, texture, or object-based methods to identify and map different types of urban structures [72], including delineation of urban areas, distinction between formal and informal settlements, and classification of different settlement types. Gamba *et al.* [70] applied object-based edge detection method on very-high resolution optical and hyperspectral images in Italy. Texture analysis and local spatial statistics were used to improve the object-based classification of urban areas on QuickBird images [186]. Authors of this study compared among different combinations of moving window sizes and features of grey-level co-occurrence metrics (GLCM) alongside spatial statistics to find optimum combinations for classification accuracy improvement. Comparison of their output with only spectral information based results indicated a significant improvement offered by the hybrid approach. Another study [26] used texture analysis and high-performance computing on high-resolution images from IKONOS satellite, they observed a correlation between urban land cover areas and GLCM and local-edge pattern co-occurrence matrix generated statistical features. Hofmann *et al.* [85] applied a hybrid application of an object-oriented texture classifier and fuzzy logic rule based segmentation method to identify informal settlements in Brazil from QuickBird data. Pesaresi *et al.* [159] presented a texture based index called PanTex for finding presence of built-up areas. The index is created by calculating the anisotropic texture co-occurrence properties from panchromatic images. The method was found to provide a compact and rotation invariant built-up area index, which is robust to contextual and seasonal variations in the data. Statistical properties of different land use classes in high-resolution 1 meter images were carried out to explore the applicability of computer vision approach to distinguish among various land covers in satellite images [204].

Researchers have also attempted to identify urban socio-economic characteristics via indicators such as settlement types. Past studies have been reported in identification of informal settlements from high-resolution images. A semi-automated object-based

classification was applied on high-resolution QuickBird images of Delhi, India to identify the urban land cover characteristics [145]. The classified image was then linked to ground surveyed socio-economic data on population distribution and water consumption related parameters to identify the informal settlements. To develop a generalizable model for mapping of settlement types, applicable in diverse urban landscapes, Graesser *et al.* [72] introduced a new method to differentiate between informal settlement and other structures such as industrial, formal residential, and commercial buildings. The study utilized different texture measures to characterize local neighborhoods of informal and formal settlements, and non-settlement types in Caracas, Venezuela; Kabul, Afghanistan; and La Paz, Bolivia. They achieved a high overall accuracy ranging between 85% to 92% . Benzhaf & Hofer [8] used color infrared orthophotos, object-based image analysis, and image segmentation to identify and discriminate between settlement structures on the basis of their tone, shape, texture, and contextual characteristics in Leipzig, German.

However, almost all presently available remote sensing datasets represent urban or built-up areas as a single contiguous class, without any further characterization. In the past, researchers have indicated that separation of settlement types could lead to the identification of urban dynamics patterns that were hitherto unseen [152, 190]. Recent advancements in remote sensing capabilities regarding higher resolution input data as well as sophisticated algorithms have paved the way for the identification of different urban settlement types [217, 72]. Development of these settlement typology classifiers may reduce our dependence on ground surveys to a large extent and find useful applications in understanding socio-economic dynamics in data sparse regions.

2.3   Urbanization and energy consumption

The dual role of urban areas as the cause and also the potential solution to the challenges of environmentally sustainable urban growth is a widely accepted fact [178, 65, 223]. The relationship between urbanization and energy consumption has gained attention since the 1990s among economists, Dahl & Erdogan [37] explained the rise of oil consumption in developing countries using income, price, and population growth resulting from urbanization

as explanatory variables. Burney [19] used a cross-sectional data of 93 countries to study the effect of urbanization on national electricity consumption while controlling for the effects of per capita income and share of the industrial sector in GDP. A more recent study [113] explored the effect of income growth on per capita energy requirements. The study observed a differential effect of increasing income on energy requirements across Australia, Brazil, Denmark, India, and Japan. For all countries except for Brazil, the elasticity of per capita energy usage was observed to be less than one. However, in-spite of the general trend, the controls of various socio-economic and demographic explanatory variables over energy consumption differed significantly across countries, leading the authors to propose that universal energy management strategies may not be ideal for energy usae reduction, instead country-specific policies are needed.

A positive relationship between the rate of urbanization and energy use has been evidenced in many econometric analysis. However, the reported growth in energy use has varied considerably among these works. Jones [101] used a cross-sectional data of 59 developing countries for 1980 to find the elasticity of per capita aggregate energy use to vary between 0.35 and 0.48, while controlling for per capita income and extents of industrialization. Parikh & Shukla [151] used a sample of national-level data from developed and developing countries for 1965-1987 period to find an elasticity of per capita energy use with respect to urban population to be 0.28, while holding all other variables fixed. Going further with their analysis on a subset of the sample, consisting of developing countries, they found an elasticity of total electricity usage to be 0.43, thereby indicating that the urbanization was less energy efficient in developing countries than in developed countries. A study [112] was done on Canadian cities between annual per capita electricity usage and demographic variables like inhabitant age and urban density, economic characteristics such as land wealth and expenditure as well as meteorological data; the study revealed a negative correlation between urbanization and per capita electricity consumption. This study also revealed that urban density was one of the main factors in determining urban energy use. The authors also pointed out that identification of good indicators for energy consumption was a challenging issue. Pachauri & Jiang [150] investigated household energy transitions in India and China to find that in both countries the rural households use more total energy than

their urban counterparts, mostly due to the dependence on inefficient solid fuel sources that constituted more than 85% of the energy requirements in the rural households. Along with urbanization, the authors also found the income, energy access, price, and local availability of fuels to be key drivers of the transition to modern energy.

Several studies have used time series data and the Granger causality test to identify the causal relationship between urbanization and energy consumption. In support of the theory that urbanization causes energy consumption, two plausible arguments have been presented: the first argument suggests that urbanization offers greater access to electricity which encourages more consumption; and the second one states that the formerly rural households are likely to increase their energy use as they transition into urban households due to higher usage of appliances [86, 75, 118]. Along these lines, Mishra et al. [137] found a short run granger from urbanization to energy use in Pacific Island nations. However, for the long run, they found Granger causality between electricity consumption and urbanization to GDP, as well as from GDP and urbanization to electricity consumption. An alternative theory supports energy consumption causes urbanization, which has found support in the work of Liddle & Lung [115]. The authors tested the long-run causal relationship between urbanization and electricity consumption using data from 105 countries during 1991-209, and found a long-run Granger causality running from electricity consumption to urbanization. Though the study did not reject the causality relation in the opposite direction. Energy consumption fosters GDP growth, employment generation, and better life qualities and this may cause more people to move to urban areas [118, 115]. However, this is a long-term effect and could therefore be seen mostly in established urban areas. Liu [118] explored the relationship between urbanization and energy consumption in China between 1978-2008 to find a long run relationship among GDP, population, energy consumption, and urbanization. Most importantly, he found a unidirectional causal relationship from urbanization to total energy consumption . Wang *et al.* [206] analyzed data from ASEAN (Association of Southeast Asian Nations) countries between 1980 and 2009, to conclude that urbanization induces energy use. Sardorsky found mixed results while analyzing the impacts of urbanization on energy intensity. He pointed out the dual role of urbanization as a cause of the confounded results, at one hand urbanization increases economic activity to increase

energy consumption, while on the other hand the resultant economies of scale raises the opportunities for energy efficiency. However, the study noted an increase in energy intensity wherever the urbanization coefficients were statistically significant [171]. The results from these national level studies indicate a strong relationship between urbanization and energy consumption. However, the country level information as used in these studies does not reveal much about their urban systems, which may mask out the critical information needed for developing sustainability measures focused on individual countries or regions. Also, the above mentioned studies have only explored the correlation and causal relationships between various aspects of urbanization and energy consumption. The effects of urbanization on energy consumption is an important task that has been much less studied [73]. Several past studies have noted that the sustainability efforts of our present times may be won or lost in the cities [178, 65], urban scale data could help shift the focus from nation to cities and help reach the sustainability goals.

2.3.1   Urban Scaling of energy consumption

Urban scale consumption of resources, generation of socio-economic goods and waste have been compared to a biological metabolic system under the concept known as "urban metabolism". This concept evolved in biological sciences and was later applied in urban science [65]. Kennedy *et al.*, while conducting a meta-analysis of different past studies in urban metabolism, related to water, energy, materials, and nutrients; found that the local climate, socio-economic conditions, urban structure, and urban heat islands plays a crucial role in determining the energy profiles of individual cities [105]. Several studies [171, 65]have pointed out that urban areas may in one hand spur economic growth and industrial activities that could increase the energy consumption, however with increasing urbanization increases the socio-economic as well as the technical output of the city which pave the way for increasing energy efficiency. Thus, the effects of urban dynamics on energy consumption is non-linear and dependent on several local factors. Identification of a simple indicator that effectively summarizes the associated web of complex local interactions is crucial for developing effective and generalizable models. Analogous to the body mass of an organism, which determines it's physiological characteristics, urban population size is widely

accepted as an effective determinant of urban socio-economic characteristics [12]. Urban population size has successfully been used to link urbanization and various socio-economic attributes of urban areas such as urban road network in 425 US cities [172], homicide in Brazilian cities between 1980 - 2009 [3]. Urban productivity in US cities was observed to increase by about 111% with each 100% increase in population [119]. Moreover, analyzing the deviations from the expected values, the study found that highly productive cities offered higher wage and level of employment exceeding the theoretical expectations of wages and jobs from cities their size, whereas cities showing low productivity were found to offer jobs and wages lower than the theoretical expectations. In another study [66], CO_2 emissions were found to scale linearly with urban population size in the US census defined core based statistical areas between 1999-2008, but this study indicated a a mild efficiency in larger urban areas over smaller ones. Bettencourt *et al.* [13] provide the scaling exponents for a large set of urban characteristics and the corresponding population for cities in the US, China, European Union, and Germany. In this study, socio economic indicators were found to exhibit an increasing return to urban scale; number of new patents, total wages, and total electricity consumption, for example, were found to increase by a 127%, 112%, and 107% corresponding to a 100% population growth in the US (2001 and 2002) and Germany (2002). The total housing and household water consumption was found to increase linealyr and at 101% with a doubling of population in the US(2001) and China(2001). Infrastructural parameters, such as gasoline sales and length of electrical cables were found to exhibit a decreasing return to scale, which increased only by 79% and 87% respectively with 100% increase in population.

This population dependent explanation of urban functions provides a simple yet powerful analytical measure to summarize a complex web of demographic, social, and economic interactions at local scales which contributes to a measurable urban characteristic. It has been proven that the exponent of this relationship, between the natural logs of an urban characteristic as the dependent variable and urban population as the independent variable, is scale invariant. This ensures it's applicability to all cities within an urban system with regard to a specific urban property irrespective of their absolute sizes [12, 11]. When applied specifically to model electricity usage, an earlier study [13] using data for Germany and

China in 2002, indicated a 107% increase in total electricity consumption in Germany for 392 cities, and a proportional increase in electricity delivered to households from 377 German cities. The same study also found the household electricity consumption to increase by about 105% corresponding to a 100% increase in population from 295 Chinese cities. Khnert *et al.* [111], found an increase of 101% in usable electricity and a linear increase in electric power delivered to households in Germany in 2002. Electricity consumption in Southern Spain was investigated using data from 134 cities and villages, the authors set up different scaling equations for various consumption sectors such as primary, secondary, tertiary, residential, and administrative consumers to get a range consumption increase from 42% to 121% per population doubling. The study found an overall increase of 106% across all electricity consumer sectors [88]. Barring these studies, not much has been reported on the effects of urbanization on electricity consumption. Mostly due to data constraints, how urbanization affects electricity consumption remains far less studied compared to other energy related issues such as gasoline consumption or emission from energy consumption.

2.3.2   Urban energy statistics and estimation

To explore urban scaling theories, availability of suitable data is a crucial requirement [13, 12], and very few datasets exist that allow for analyzing the implications of the effect of urbanization on energy consumption. Studies that specifically focus on urban scale electricity consumption patterns are severely limited due to lack of reliably measured data. The available datasets are generally not presented at urban scales, and the available datasets are generally presented in aggregated form only at national and regional levels which are not suitable for urban scale analyses. Also, non-standardized data collection and reporting methods impede the available urban datasets from comparative analysis [65]. In the absence of suitable ground data, past researchers have estimated energy consumption using proxy measures. Satellite recorded nighttime light emissions from human settlements have been known to be a useful proxy of human presence and their socio-economic activities [187, 133]. The informational value of nighttime lights data for countries where statistical datasets are sparse or not available was recently highlighted by Chen & Nordhaus [25]. Similarly

Ghosh et al. justified the importance of nighttime lights as an indicator of urban socio-economic activities in places with no available census information [71]. Elvidge *et al.* [50] found a significant correlation between the lit area and economic activity and electricity consumption at national levels in 21 countries. Subsequently, nighttime lights data has been used to estimate GDP at the state level in India, China, Turkey, and the US [188], and to develop spatially disaggregated global and regional level map of economic activity [71]. Amaral *et al.* [4] used nighttime lights to estimate population and energy consumption in Brazilian Amazon. Henderson *et al.* [79] used the growth in nighttime lights to improve economic growth estimates. They found supplemental value in nighttime lights to improve GDP estimates for countries with low quality national accounts data. However, for countries with high-quality statistical data, the nighttime lights offered little usefulness.

By it's design, the nighttime lights data does not capture the energy usage inside human settlements, such as the energy needed for heating and cooling needs as well as for running appliances[65]. Thus, using this data as a proxy for estimating electricity consumption in the urban system of a large and diverse country may mask the spatial heterogeneities arising from the variations in local conditions. For places where socio-economic datasets are accessible, model-based estimation approaches have been used by researchers. Filippini & Pachauri [60] used variables such as prices of electricity and fuel, income and age structure of the population, and size of households, dwelling size, and geographic variability was used to estimate monthly residential electricity consumption in India for multiple seasons; their model explained between 50 - 54% of variance in the data. Price and income elasticities of household electricity consumption were estimated using net income, prices of electricity, natural gas, and oil, climate characteristics, and household cohorts in a pseudo-panel model analysis in Quebec province, Canada [10]. In the United States, climate and weather data, economic variables such as electricity sales, employment, unemployment, gross state product, per capita income of the population, and labor force data was used to study climate sensitivity of electricity consumption in residential and commercial sectors in Florida [142]. Tamayao *et al.* [189] used county-level electricity consumption data for California, climate variables and population data to predict total electricity consumption at US county levels.

These worka highlight the potential for efficiently estimating electricity consumption at urban scales using socio-economic variables.

2.4   Existing knowledge gaps:

On the basis of the ongoing review of the existing body of knowledge, the following three areas were identified as potential areas of knowledge gap. The research objectives of this dissertation is aimed at addressing each of these points. A summary of these objectives and how the following three chapters address them has been presented in section 1.5.

1. Accurate measurement of urban areas is an essential step towards understanding urbanization and associated energy processes. Comparison of high-resolution settlement datasets against their lower and medium resolution counterparts over vastly different geographies, which offer distinct scene characteristics and challenges for remote sensing based data extraction, is needed a comprehensive assessment of their benefits. However, this has not yet been addressed in the present literature. Also, no research has been done assessing the performance of the 8 meter LandScan SL data in comparison with other datasets.

2. While it is evident that texture and object-based image analysis can help characterize urban settlements, no study has tested the applicability of this approach to connect their visible form to their function, such as electricity consumption. In the absence of ground data, especially in vast data-poor regions of the earth, this could provide a high-resolution spatial proxy of energy usage and other socio-economic indicators.

3. Urban electricity consumption is expected to significantly increase in the future, thus, it is imperative to assess the relationship between urbanization and electricity consumption. However, such studies remain incredibly scarce, partly due to the scarcity of suitable data. The key questions that need addressing are (1) In the absence of statistical data, how can a generalizable method be made to estimate urban electricity consumption? (2) How urbanization impacts electricity consumption? And (3) If

urbanization indeed offer the efficiency from returns of scale, what are its limits and implications?

Chapter 3

Global and regional human settlement mapping - a comparative analysis

3.1 Introduction

Urban dwellers made up about 54% of global population in 2014. According to the United Nations, their share is projected to reach 68% by the year 2050 [195]. This unprecedented growth introduces critical challenges for multiple issues such as management of population growth, provisioning of civic services, waste disposal, access to social and health services, supply of natural resources and energy access, and controlling pollution. These challenges necessitate heightened efforts to ensure environmental sustainability, equitable growth and prosperity, and citizens welfare are included in sustainable urban planning for the future. As urban areas keep expanding, identification of new growth and changes to existing areas could point at the hot spots of changing demographic, environmental, and economic scenarios on the ground. Identification of new areas resulting from the peri-urban growth, as well as from infilling of existing urban areas is also important as this is where the most vulnerable members of urban communities are likely to be found. Human settlement maps derived from remotely sensed images offer high spatial detail and can be updated regularly. They are thus better suited for monitoring urbanization than traditional survey or census-based methods [175]. Over the past three decades, satellite data derived coarse and medium resolution datasets have been produced by different organizations. These free and open access global

datasets have made strong scientific contributions in a range of studies related to population distribution modeling [116], environmental impacts [174], urban planning [191], disaster management [61], health [7], and energy studies [219] at regional and global levels. However, the low-resolution datasets fail to identify finer details within the urban landscape, required for current issues in urban dynamics research as outlined above. The insufficiency of low-resolution datasets in accounting for large parts of built-up patches in and around urban areas across the globe and the resulting uncertainties have led to a wide variety of urban and built-up area estimates in reported work [173, 163, 164, 80, 174].

Increased availability of very high-resolution remotely sensed images in recent years, coupled with advancements in high-performance computing resources and efficient image processing algorithms have fostered the development of high-resolution human settlement datasets [26, 203, 153]. While it is expected these high-resolution datasets are more adept at capturing finer details in the urban landscape, comparative evaluation against low-resolution datasets in diverse landscapes could justify their usage in scientific studies and offer new insights into urban dynamics. This chapter presents a comparative evaluation of a set of four distinct human settlement and built-up area datasets. The two high-resolution datasets LandScan SL and GHSL have been analyzed for their mutual consistency; alongside two datasets with coarse and medium resolution MODIS Land Cover and ESA - Climate Change Initiative Land Cover Data. Barring LandScan SL, all three datasets are available as global coverage. Hence, this analysis was conducted using a subset of the data. The study areas, Egypt and Taiwan, were carefully chosen to represent two disparate geographies in terms of demography, economy, and physiography. By analyzing the four datasets at both national and sub-national levels, this chapter aims to identify the comparative performance of these datasets at multiple resolutions. To explore identify inter-dataset correspondence regarding built-up area detection, composite maps were generated at multiple scales, each corresponding to the native spatial resolution of one of these datasets. A quantitative analysis of urban morphology was then performed on this composite maps, using spatial metrics, to draw insight on the congruency of these datasets in capturing the complex settlement patterns and supplement existing knowledge of global human settlement mapping initiatives. The rest of this has been ordered as the description of the datasets used and the preprocessing

methods used, followed by a description of study areas, methodology, results, discussion, conclusion, and lastly the chapter's contribution.

3.2 Datasets

MODIS Land Cover Data: MODIS land cover dataset (also known as MCD12Q1) is a 500 meter data (hereafter MODIS LC) developed by the National Aeronautics and Space Administration (NASA) [68]. MODIS LC is produced using a supervised decision tree classification algorithm on a full year of eight-day MODIS Normalized BRDF-Adjusted Reflectance (NBAR) and MODIS Land Surface Temperature (LST) data composites[185]. The data is presented to the classifier either as monthly composites or annual metrics. The data generation process includes an optimization method termed "boosting" for improving classifier accuracy, which is achieved by systematic variation of training samples [68]. MODIS LC data has been extensively used in urban land cover related studies at global scales [176, 96, 220, 138]. The data carries land cover information as per to five separate land cover classification schemes, one in each of the first five layers. We used the layer depicting the International Geosphere-Biosphere Programme (IGBP) scheme in the 2011 dataset, from which the urban pixels were extracted. MODIS LC dataset may be freely downloaded from the USGS website: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1.

ESA - Climate Change Initiative Land Cover Data: European Space Agency's Climate Change Initiative (CCI) project has developed a global land cover dataset (hereafter CCI-LC) at 300 meter spatial resolution. The datasets were designed to provide a stable, time-series land cover maps, and have been developed for three time periods of 2008-2012, 2003-2007, and 1998-2002, respectively centered on 2010, 2005, and 2000 [82]. In production of CCI-LC, Medium Resolution Imaging Spectrometer (MERIS) full resolution datasets were used as the primary input while MERIS reduced resolution data were used in the absence of full resolution coverage. *Satellite Pour lObservation de la Terre* Vegetation (SPOT-VGT) data was used to fill in gaps in MERIS temporal coverage. These datasets follow the United

Nations' Land Cover Classification Scheme (UN-LCCS) for classifying the land cover classes [82, 106]. We extracted the urban areas based on pixel values from the 2008-2012 dataset, which is freely accessible at https://www.esa-landcover-cci.org/.

Global Human Settlement Layer: The GHSL data is developed by the Joint Research Center at the European Commission, utilizing Landsat time series data collection ranging from 1975 to Landsat-8 images from 2013 and 2014 [157]. GHSL is a global level data which was created using supervised classification algorithms on the Landsat images. An array of training datasets including MODIS 500m Global Urban Extent, Global Land Cover 2000 and GHSL BUREF2010 data, WorldPop project settlement polygons, Open Street Map data, and Geonames derived settlement points were used to refine the results. Finally, automatically generated discriminant rules were applied to the data for information extraction. GHSL data can be freely accessed at http://ghsl.jrc.ec.europa.eu/datasets.php.

BUMIX, a sub-pixel mixture model output at 38.2 meter spatial resolution available within GHSL data suite, for the year 2014 was used in this analysis. In BUMIX data, pixel values are represented as continuous integers in the 0-255 range where 0 represents 0% built-up coverage while a pixel value of 255 denotes 100% built-up coverage within the pixel [157].

LandScan Settlement Layer: Oak Ridge National Laboratory is currently engaged in the refinement of the spatial resolution of the well established LandScan Global (30 arc-second) population distribution dataset to produce a high-resolution (3 arc-second) gridded population dataset known as LandScan HD. One of the many underlying datasets informing the population model is a high-resolution (8m) settlement layer, referred to as LandScan Settlement Layer (LandScan SL). This dataset is derived from the optical and near-infrared bands of very high-resolution images (≤0.5m) processed through a set of supervised pattern recognition algorithms to identify human settlements [26, 153]. For global scale processing, cluster-based parallel computing is utilized which significantly decreases the computational time. Lastly, the binary output (depicting either settlement

or non-settlement) is subjected to a through verification and validation process using semi-automated and manual techniques, which is expanded upon in [207], to resolve commission or omission errors. After the manual review, the dataset is considered to be of reference quality. For this research, settlement data for the year 2014 was used.

3.3 Data processing

The four unique datasets representing human settlements and built-up areas at different native spatial resolution, for the areas of interest, were acquired from their respective sources. A total of five MODIS LC tiles were downloaded (three for Egypt and two for Taiwan) to create mosaics for each country. The CCI-LC dataset was downloaded from the ESA website, and the GHSL data was provided by the European Union's Joint Research Council. The LandScan working group at Oak Ridge National Laboratory provided the LandScan SL data for both countries.

To make the datasets suitable for comparative analysis, they were reprojected to Geographic Lat/Lon projection with WGS1984 datum from their native projection systems. Since the datasets are widely different in their native spatial resolution; to ensure multi-resolution analysis, the exploration of their mutual correspondence was carried out at each dataset's native resolution. Hence, the datasets were resampled to the four different spatial resolutions of 500, 300, 38.2 and 8 meters, using nearest neighbor resampling method. Then the thematic comparisons were conducted at each of these resolutions.

3.3.1 Processing GHSL data:

The pixel values in GHSL data is different from the other three datasets, where a pixel value indicates either built-up or non built-up or other categories, in the sense that GHSL data represents the percentage of built-up areas within a pixel. The percent built-up area coverage per pixel has been scaled between 0 and 255 [157], where the pixel values 0 and 255 represents 0% and 100% built-up area. Therefore, to calculate the actual built-up area, the

pixel values were transformed from percentage to square meters using the following equations 3.1a and 3.1b:

$$P_{ij} = [(DN_{ij} - 0)/255] \tag{3.1a}$$

$$A_{ij} = (Ft_{ij} \times P_{ij}) \tag{3.1b}$$

Where P_{ij} is the proportion of built-up area in pixel ij, the value ranges between 0 and 1 (where 1 and 0 indicates 100% and no presence of built-up areas respectively). DN_{ij} is the value of pixel ij in the GHSL dataset. A_{ij} represents the built-up area in pixel ij, and Ft_{ij} is ground footprint of a GHSL pixel (38.2 X 38.2 sq. meters or 1459.2 sq. meters). The total built-up area (A) for a given area was then estimated using the following equation 3.2:

$$A = \sum A_{ij} \tag{3.2}$$

3.4   Study Area

3.4.1   Egypt:

Egypt is a situated in between the north-east part of Africa and south-west parts of Asia in Northern Africa. It experiences desert climates characterized by hot, dry summers and moderate winters. Its 97 million inhabitants (July 2017 est.) occupy a land area of 995,450 sq. Km with a population density of 90 persons/sq. Km. The country consists of 27 administrative regions known as governorates. Due to the arid physiography, much of these vast lands are uninhabited, and 95% population live within a 20 km zone around the Nile River and its delta region, leaving vast countrysides mostly uninhabited (see figure 3.1). A rapidly growing population, coupled with the short supply of arable land poses challenges for resource management. Two important cities Cairo and Alexandria were inhabited by about 20.1 million and 5.1 million people. Egypt's population grew at 2.45% in 2017 while the country is being urbanized at an estimated rate of 1.86% per year between 2015-2020; about 42.7% of the total national population lived in urban areas in 2018. Economy wise,

Egypt was World's 22nd largest economy with a GDP (on purchasing power parity basis) of $1.20 trillion in 2017, which was growing at a rate of 4.2% annually [28].
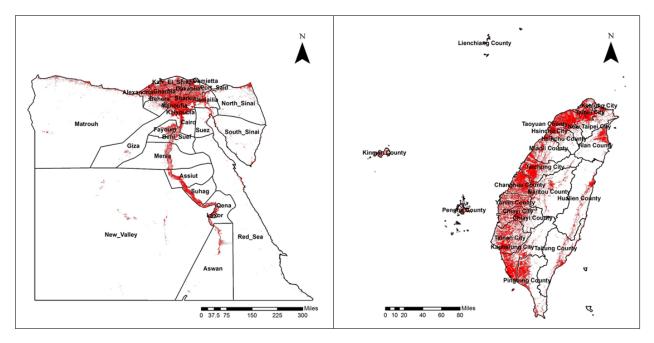
Figure 3.1: Human settlement distribution over Egypt and Taiwan, using LandScan SL data.

3.4.2 Taiwan:

Taiwan is an East Asian state. Compared to Egypt, it has a much higher population density of 648 persons/sq. Km, totaling about 23.5 million residents (July 2017 est.) spread over 32,260 sq. Km. of land area. Taiwan enjoys a tropical marine climate, marked by seasonal southwest monsoonal rainfall from June to August, and persistent cloudiness all year around. The population is centered around the coastlines, and the main concentration occurs along the northern and western coasts (see figure 3.1). The population growth rate was estimated to be 0.17% in 2017, and the urbanization rate between 2015-2020 has been estimated to be 0.8%. Taiwan is comparatively more urbanized than Egypt, approximately 78.2% of the total population was living in urban areas in 2018. Taiwan is divided into 22 administrative regions which include six special municipalities of Kaohsiung, New Taipei, Taichung, Tainan, Taipei, and Taoyuan; three provincial cities and 13 counties. New Taipei city had 4.3 million inhabitants, while the capital Taipei city is home to 2.7 million people. Taipei had $1.19

trillion in GDP (on purchasing power parity basis) in 2017, growing at 2.8% annually, placing it as the 23rd largest in the world[29].

3.5 Methods

Several methods have been used by past researchers for comparison of thematic maps, such as the kappa statistics [161], pairwise comparison [64], error budget approach [32], and quantity and allocation analysis [162] have been used for assessing congruency of land cover datasets. After considering potential advantages and disadvantages of these methods with respect to the present research objective, this problem as approached through quantity and allocation approach as it allows quantification of congruency among the datasets and visual identification of agreements and disagreements. We first quantify the settlement areas as estimated by each dataset at national and sub national levels, followed by a comparison to explore spatial variation in classification at local scales. Secondly, mutual agreement-disagreement maps were developed, at each native resolutions of these datasets to identify the discrete zones of agreement and disagreement and conduct further quantitative analysis to identify trends.

3.5.1 Agreement-disagreement maps

Accurate allocation of pixels to land use classes is a critical aspect of thematic map comparison [162]. The assignment of pixels provides spatial validity and allows for visual comparison of the results. The research objective here is to explore the congruency of pixels assigned as built-up in these datasets, and it's implications. Analysis of spatial agreement and disagreement between datasets was found useful for this purpose [148, 126]. Accordingly, spatial agreement-disagreement maps were created for the study areas and four discrete inter-dataset levels of agreements were identified in the resultant maps, at all four spatial resolutions. The agreement levels are described below: a. No agreement: Only one of the four datasets represent pixel i,j as built-up. b. Low agreement: Two of the four datasets represent pixel i,j as built-up. c. High agreement: Three of the four datasets represent pixel i,j as built-up. d. Full agreement: All datasets represent pixel i,j as built-up.

3.5.2 Landscape indices

Egypt and Taiwan display vast differences in their physiographic, climatic and demographic characteristics, these factors have been found to have a profound impact on the development of human settlement patterns [182]. Therefore, these differences could have resulted in spatially distinct settlement patterns between these two countries. Furthermore. Distinct on-ground settlement patterns, combined with different satellite captured data characteristics could create a several challenges for the datasets based on their input data and processing algorithms. Thus the capabilities of these datasets in accounting for built-up areas over different geographies need to be tested through further analysis. To account for various morphological characteristics of the built-up areas, three landscape indices: Landscape Shape Index (LSI), Mean Patch Fractal Dimension (FRACMN) and Percentage of Likely Adjacencies (PLADJ) were calculated. Each index is briefly described in the following paragraphs. For further background information may be obtained from [129].

- **Landscape shape index:** LSI is a measure of class aggregation or disaggregation of a landscape. The measure is obtained through measuring the ratio between total class edge length or edge density in a landscape to that of a landscape with a standard simple shape such as a square patch, adjusted for the landscape size. For a landscape consisting of a single square shaped patch, the LSI value obtained will be 1. As the landscape becomes more disaggregated, the index value increases infinitely (see equation 3.3).

$$LSI = \frac{e_i}{min(e_i)} \tag{3.3}$$

Where, e_i is the total length of the perimeter or edge of patches involving class i, and $min(e_i)$ is the minimum total perimeter or edge length of class i, both measured in terms of number of cell surfaces.

- **Mean patch fractal dimension:** Fractal dimension measures patch complexity through calculating the perimeter-area relationship, where both numerator and denominator are log transformed. An index value of 1 indicates simple perimeter.

The index value increases with complexity till it reaches 2, which indicates highly convoluted and plane-filling perimeters. The mean index values (FRACMN) were used in this analysis (see equation 3.4).

$$FRAC = \frac{2ln(0.25p_{ij})}{ln(a_{ij})} \qquad (3.4)$$

Where, p_{ij} and a_{ij} are the perimeter and area of the patch ij respectively.

- **Percentage of likely adjacencies:** Contiguity is another critical morphological aspect of a landscape. PLADJ measures patch contiguity by calculating the number of like adjacencies involving one particular class divided by the total number of adjacencies involving the class, multiplied by 100. The index values equal 0 if the class patches are maximally disaggregated with no like adjacencies, and a value of 100 represents a single patch landscape where all the adjacencies are within the same class (see equation 3.5).

$$PLADJ = (\frac{\sum_{i=1}^{m} g_{ii}}{\sum_{i=1}^{m} \sum_{k=1}^{m} g_{ik}}) \times 100 \qquad (3.5)$$

Where, g_{ii} denotes the number of likely adjacencies between pixels of class i and g_{ik} denotes the number of adjacencies between pixels of class i and k, both measured using double-count method.

3.6   Results

3.6.1   Analysis of built-up area estimates:

As a first step, the built-up areas as estimated by each of these datasets were calculated, at their respective original resolution (i.e., before any resampling process) for both national and sub national levels. At the national level, the total built-up area estimates for Egypt ranged from a minimum of 3123.75 sq. Km (GHSL) to a maximum of 8166.53 sq. Km (MODIS LC), indicating a variation of 161%. The country level estimates for Taiwan ranged from

Table 3.1: Built-up areas detected by the datasets at national scale.

Country	Built-up area (in sq. km.)			
	MODIS LC	CCI-LC	GHSL	LandScan SL
Egypt	3132.75	5395.41	8166.53	3867.78
Taiwan	2506.25	4390.92	2373.01	2083.83

2083.83 (LandScan SL) to 4390.92 km (CCI-LC), thereby indicating a variation of 110% in built-up area estimates (see table 3.1).

The estimated built-up areas varied even more severely at the sub national level. For 27 governorates in Egypt, the lowest variation of around 60% was found for Beni Suef, while the estimates of built-up areas varied as much as 900% for South Sinai (see figure 3.2). For the 22 cities, municipalities and counties in Taiwan, the lowest variation in built-up area estimates was obtained for New Taipei City (around 27%), the highest variation was observed for Chiayi County where the estimates varied by about 534% (see figure. 3.3).
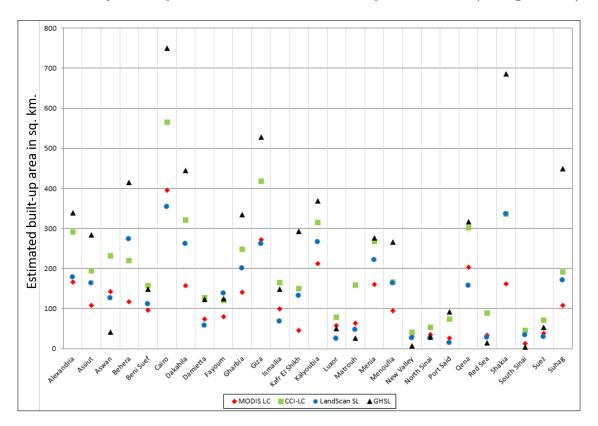
Figure 3.2: Comparison of built-up area estimates at governorate level in Egypt.
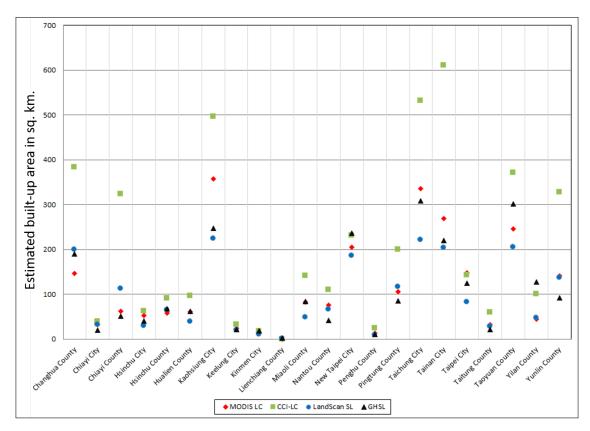
Figure 3.3: Comparison of built-up area estimates at city and county level in Taiwan.

3.6.2 Analysis of spatial congruency among datasets

The agreement-disagreement maps, as described in the methods section, were created for the study areas to assess congruency of the datasets under analysis across four spatial resolutions. Four distinct inter-dataset agreement classes were identified in the resultant maps that help in identification of the agreements and disagreements, and subsequent analysis. Figures 3.4 and 3.5 display the inter dataset agreement in Cairo city, Egypt and Taipei city, Taiwan and their surrounding areas at 8 meter spatial resolution.
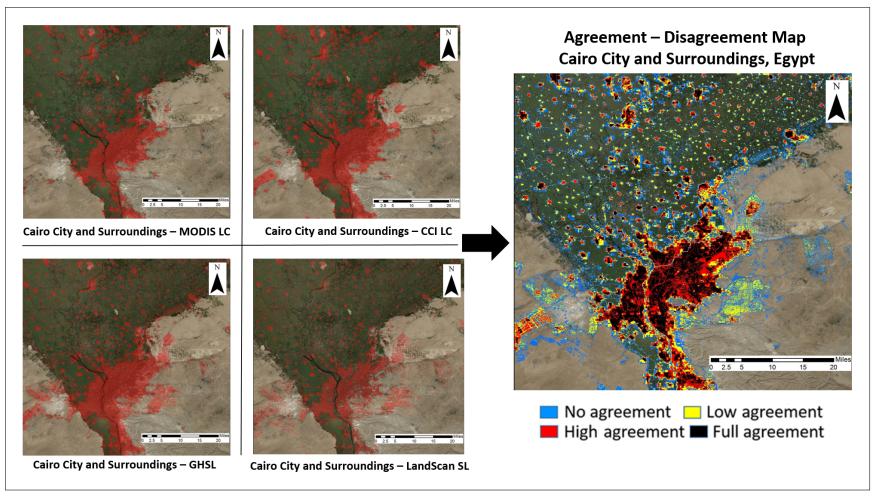
Figure 3.4: Agreement-Disagreement map for Cairo City and surroundings, Egypt.
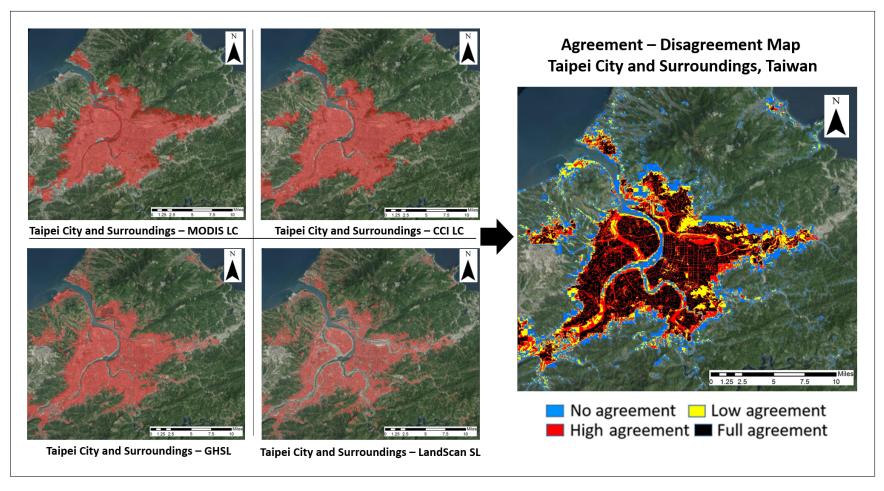
Figure 3.5: Agreement-Disagreement map for Taipei City and Surroundings, Taiwan.

To analyze the spatial characteristics of these discrete patches of agreements and disagreement zones observable in these maps, spatial metrics were used. Spatial metrics have been successfully used by past researchers in studying urban settlement patterns[179, 175]. In this work, three essential morphological characteristics of urban landscape, i.e. complexity, aggregation, and contiguity of built-up patches [81] were calculated and analyzed. By examining these characteristics across the four agreement classes (no agreement, low agreement, high agreement, and full agreement, referred to as CL1, CL2, CL3, and CL4 hereafter) obtained from the agreement-disagreement maps, how differences in these morphological characteristics of built-up areas correspond to congruency among these datasets was explored. The three spatial metrics were calculated for two countries for all built-up patches belonging to CL1, CL2, CL3, and CL4. The interpretation of many spatial metrics is dependent on the size of the landscape. Therefore, fixed landscape size was needed for comparative analysis [129]. Subdividing the countries into smaller square blocks also reduced computational burden, while ensuring uniform landscape size. Hence, the study areas were sub-divided into square blocks of 5km X 5km dimension, a spatial unit that was found sufficiently large to capture urban patterns in the two study areas.

Once the spatial metrics were calculated using Fragstats software, it was necessary to check that the differences in the morphological characteristics across the agreement classes, as represented by these metrics values, were indeed statistically significant. Therefore, analyses of variance (ANOVA) test was conducted for both Taiwan and Egypt for each spatial metric at all resolution levels. The data showed non-homogeneity of variance within the classes, thus a robust ANOVA using Welch's correction was conducted which was found to account for such non homogeneous within class variance [208]. The test results were found to be significant at 95% confidence level for both countries and at all four spatial resolutions. the results of the ANOVA tests have been presented in tables 3.2 - 3.4. For the mean values of these metrics across different inter-dataset agreements, please see tables A.1 through A.4 in appendix A.

The ANOVA results confirms that statistically significant differences do exist in morphological characteristics across the agreement classes at all four spatial scales. This allows for a safe rejection of the null hypothesis that no difference in morphological

Table 3.2: ANOVA results for LSI - Egypt and Taiwan.

Resolution	Egypt	Taiwan
500 m	$F_{[3,2567.3]}$=1026.8, p<.00001	$F_{[3,708.8]}$=31.0, p<.00001
300 m	$F_{[3,2575.7]}$=660.4, p<.00001	$F_{[3,1017.1]}$=40.5, p<.00001
38.2 m	$F_{[3,3826.2]}$=938.9, p<.00001	$F_{[3,1496.4]}$=75.7, p<.00001
8 m	$F_{[3,3440.8]}$=267.0, p<.00001	$F_{[3,1460.3]}$=95.1, p<.00001

Table 3.3: ANOVA results for FRACMN - Egypt and Taiwan.

Resolution	Egypt	Taiwan
500 m	$F_{[3,2274.9]}$=191.2, p<.00001	$F_{[3,662.8]}$=6.8, p=.0002
300 m	$F_{[3,2170.5]}$=32.5, p<.00001	$F_{[3,922.2]}$=9.5, p<.00001
38.2 m	$F_{[3,2735.8]}$=334.7, p<.00001	$F_{[3,1378.7]}$=104.1, p<.00001
8 m	$F_{[3,3177.5]}$=278.4, p<.00001	$F_{[3,1398.1]}$=100.4, p<.00001

characteristics exists across agreement classes. These discrete classes are derived from agreement and disagreement amongst the built-up areas, identified by the four datasets. Therefore, the morphological characteristics associated with the different agreement classes could indicate an association between the urban built-up area morphology and congruency among datasets. Hence, the data was further analyzed to explore the possible effects of built-up area morphology over the inter datasets agreements. The observed agreement classes are discrete in nature, past study have indicated suitability of logistic regression models for their analysis [128]. Thus, four ordered LOGIT regressions were developed, one for each spatial resolution, to be run between the agreement classes (CL1, CL2, CL3, and CL4) as dependent variable and the morphological indices values (LSI, FRACMN, and PLADJ) as independent variables. Equation 3.6 below illustrates the general formulation of the model:

$$\text{CL} = \hat{\beta}_0 + \hat{\beta}_1 \times LSI + \hat{\beta}_2 \times FRACMN + \hat{\beta}_3 \times PLADJ + \varepsilon \qquad (3.6)$$

Table 3.4: ANOVA results for PLADJ - Egypt and Taiwan.

Resolution	Egypt	Taiwan
500 m	$F_{[3,2094.2]}$=310.2, p<.00001	$F_{[3,663.9]}$=19.2, p<.00001
300 m	$F_{[3,2220.9]}$=139.1, p<.00001	$F_{[3,957.4]}$=37.5, p<.00001
38.2 m	$F_{[3,2991.1]}$=718.6, p<.00001	$F_{[3,1485.1]}$=80.4, p<.00001
8 m	$F_{[3,3566.2]}$=626.9, p<.00001	$F_{[3,1436.9]}$=63.9, p<.00001

The regression coefficients for both countries at all four spatial resolutions were found to be significant (see table 3.5 for Egypt and table 3.6 for Taiwan). In terms of overall trend, an increasing disaggregation (LSI) of built-up patches were found to be negatively associated with increasing correspondence among the datasets. The contiguity (PLADJ) of built-up patches was positively associated with inter dataset correspondence. However, the increasing complexity of built-up patches (FRACMN) was found to be strongly associated with inter dataset correspondence. This association could be misleading at the outset and may be misconstrued as an indication of a general association between increasing complexity of built-up patches with increasing agreements among datasets. However, the values of FRACMN metric range between 1 and 2, where values nearing 2 indicate an increase in convoluted, plane filling shapes while values close to 1 indicate simple perimeters such as a square [129]. An increasing FRACMN value here should be interpreted as an indication of increased proximity among built-up of patches with narrow gaps in between. Such conditions are likely to be found in the densely developed urban cores where narrow open areas such as roads, lanes, and parks are intermingled with built-up areas. The PLADJ and FRACMN exhibit negative coefficients for the analysis at 500 meters in Egypt. We interpret this as a likely effect of the coarse pixels, which fails to identify the finer spatial complexities. Also, the settlement layout of Egypt with very dense settlements in and around cities along with vast non-built-up swaths could likely have contributed to the negative association between increasing inter-dataset agreement and PLADJ and FRACMN at 500 meter resolution. For Taiwan, however, no such trend was observed.

Table 3.5: Ordered LOGIT regression results for Egypt.

Variable	At 500 m			at 300 m		
	Reg. Coeff. ($\pm$ SE)	t-value	p-value	Reg. Coeff. ($\pm$ SE)	t-value	p-value
LSI	-0.60 ($\pm$ 0.03)	-20.19	<0.0001	-1.03 ($\pm$ 0.03)	-39.18	<0.0001
FRACMN	-3.65 ($\pm$ 0.87)	-4.18	<0.0001	13.35 ($\pm$ 1.37)	9.72	<0.0001
PLADJ	-0.02 (($\pm$ 0.001))	-14.06	<0.0001	0.03 ($\pm$ 0.001)	14.82	<0.0001
Variable	At 38.2 m			at 8 m		
	Reg. Coeff. ($\pm$ SE)	t-value	p-value	Reg. Coeff. ($\pm$ SE)	t-value	p-value
LSI	-0.21 ($\pm$ 0.005)	-41.11	<0.0001	-0.12 ($\pm$ 0.003)	-37.72	<0.0001
FRACMN	24.87 ($\pm$ 1.27)	19.58	<0.0001	20.19 ($\pm$ 0.95)	21.15	<0.0001
PLADJ	0.04 ($\pm$ 0.001)	32.96	<0.0001	0.06 ($\pm$ 0.001)	37.16	<0.0001

Table 3.6: Ordered LOGIT regression results for Taiwan.

Variable	At 500 m			at 300 m		
	Reg. Coeff. ($\pm$ SE)	t-value	p-value	Reg. Coeff. ($\pm$ SE)	t-value	p-value
LSI	-0.65 ($\pm$ 0.06)	-10.54	<0.0001	-0.37 ($\pm$ 0.03)	-11.07	<0.0001
FRACMN	8.27 ($\pm$ 2.28)	3.62	<0.0001	7.33 ($\pm$ 2.06)	3.56	<0.0001
PLADJ	0.02 ($\pm$ 0.003)	4.47	<0.0001	0.016 ($\pm$ 0.003)	5.45	<0.0001
Variable	At 38.2 m			at 8 m		
	Reg. Coeff. ($\pm$ SE)	t-value	p-value	Reg. Coeff. ($\pm$ SE)	t-value	p-value
LSI	-0.06 ($\pm$ 0.005)	-11.80	<0.0001	-0.05 ($\pm$ 0.004)	-13.51	<0.0001
FRACMN	30.11 ($\pm$ 2.05)	14.71	<0.0001	29.96 ($\pm$ 1.98)	15.15	<0.0001
PLADJ	0.01 ($\pm$ 0.002)	7.08	<0.0001	0.04 ($\pm$ 0.004)	9.82	<0.0001

In spite of Egypt and Taiwan sharing dissimilar physical and demographic characteristics, which could possibly have led to dissimilar built-up patterns, the association between built-up area morphology and inter dataset agreements as observed from regression coefficients remain mostly similar across all four spatial resolutions. This trend indicates that the detection of built-up areas by the individual datasets is fundamentally affected by the morphological characteristics of urban areas irrespective of geographical variances in study areas. Quantification of these effects on inter-dataset congruency could therefore reveal the basic effects of landscape morphology on urban area detection in remote sensing data. As ordered LOGIT coefficients represent the change in a dependent variable in terms of logs of odds ratios, this is often unsuitable for direct interpretation other than their significance and direction of the effect. Under these conditions, representation of these effects as probability is much easier to interpret, which can be done through marginal effects [139]. Marginal effects represent the expected change in a dependent variable from a unit change in a specific independent variable while keeping all other variables constant at their respective means [131]. In this study, in addition to observing the general trend through an ordered LOGIT regression, an assessment of how these morphological characteristics affect the agreement among datasets helped draw more insight. The average marginal effects calculated for Egypt (see table 3.7) show that 1 unit increase in disaggregation (LSI) in a built-up area will lead to an average increase between 0.029 and 0.249 in probability of that area belonging to CL 1 or no agreement zone, while the average decrease in its probability of falling into CL 2 (low agreement), CL3 (high agreement) or CL4 (full agreement) ranges between 0.012 and 0.058; 0.011 and 0.133; 0.006 and 0.059 respectively. A unit increase in shape complexity (FRACMN) leads to an average probability decrease of a built-up patch falling into CL1 between 3.218 and 6.208, while increasing its average probability between 0.735 and 2.761, 1.722 and 2.509, and 0.762 and 0.974 for falling into CL2, CL3 and CL4 respectively, within the resolution range of 8 - 300 meter. Similarly, 1 unit increase in contiguity (PLADJ) leads to a decrease of an average probability between 0.006 and 0.015 for CL1 while increasing its average probability of belonging to CL2, CL3 or CL4 between 0.001 and 0.006, 0.003 and 0.006, and 0.001 and 0.003 within the 8 - 300 meter resolution range. However, it should be noted that at the 500 meter resolution the marginal effects of FRACMN and PLADJ shows

Table 3.7: Marginal Effects of urban morphological characteristics - Egypt.

Variable	At 500 m				At 300 m			
	CL1	CL2	CL3	CL4	CL1	CL2	CL3	CL4
LSI	0.149	-0.058	-0.059	-0.032	0.249	-0.057	-0.133	-0.059
FRACMN	0.911	-0.354	-0.362	-0.196	-3.218	0.735	1.722	0.762
PLADJ	0.005	-0.002	-0.002	-0.001	-0.006	0.001	0.003	0.001
Variable	At 38.2 m				At 8 m			
	CL1	CL2	CL3	CL4	CL1	CL2	CL3	CL4
LSI	0.052	-0.023	-0.021	-0.008	0.029	-0.012	-0.011	-0.006
FRACMN	-6.208	2.761	2.509	0.938	-5.046	2.166	1.906	0.974
PLADJ	-0.010	0.005	0.004	0.002	-0.015	0.006	0.006	0.003

reverse trends than what described above. We consider this as a likely effect of loss of details at coarse resolution.

Similarly, table 3.8 presents the marginal effects for Taiwan which indicates that 1 unit increase in built-up area disaggregation (LSI) will lead to an average probability increase between 0.012 and 0.160 of that area being classified as CL 1 or no agreement zone, while the average decrease in its probability of falling into CL 2 (low agreement), CL3 (high agreement) or CL4 (full agreement) ranges between 0.002 and 0.028; 0.005 and 0.073; 0.005 and 0.059 respectively. An unit increase in shape complexity, indicated by FRACMN, leads to an average decrease in probability of a built-up area to fall into CL1 between 7.224 and 2.031, while increasing its average probability between 0.215 and 1.259, 0.855 and 3.352, and 0.714 and 2.818 for falling into CL2, CL3 and CL4 respectively. Lastly, one unit increase in contiguity (PLADJ) will lead to a decrease of an average probability between 0.004 and 0.009 for CL1 while increasing its average probability of belonging to CL2, CL3 or CL4 between 0.001 and 0.002, 0.002 and 0.004, and 0.001 and 0.003. The PLADJ values at 500 and 300 m resolution showed no affect on a class falling into CL2, however no anomalous pattern like Egypt was observed in Taiwan. Comparing with Egypt, while the absolute numbers change slightly, most likely due to the difference physiographic and demographic factors and their effects on built-up area and settlement characteristics. The overall trends, however, remains the same. This also highlights how basic urban morphological features such as complexity, contiguity or aggregation affect detection of a built-up patch by the datasets and thereby affecting the inter-dataset agreements.

Table 3.8: Marginal Effects of urban morphological characteristics - Taiwan.

Variable	At 500 m				At 300 m			
	CL1	CL2	CL3	CL4	CL1	CL2	CL3	CL4
LSI	0.160	-0.028	-0.073	-0.059	0.092	-0.011	-0.044	-0.037
FRACMN	-2.031	0.355	0.928	0.748	-1.785	0.215	0.855	0.714
PLADJ	-0.004	0.001	0.002	0.001	-0.004	0.000	0.002	0.002
Variable	At 38.2 m				At 8 m			
	CL1	CL2	CL3	CL4	CL1	CL2	CL3	CL4
LSI	0.015	-0.002	-0.007	-0.006	0.012	-0.002	-0.005	-0.005
FRACMN	-7.212	1.042	3.352	2.818	-7.224	1.259	3.218	2.747
PLADJ	-0.003	0.000	0.002	0.001	-0.009	0.002	0.004	0.003

3.7 Discussion

In this work, performance of high-resolution human settlement datasets were assessed against coarse and medium resolution datasets at multiple scales in Egypt and Taiwan. Since the datasets were obtained from different sources, necessary preprocessing were done to make them suitable for comparative analysis. Initial assessment of built-up area estimates highlight that such estimates vary considerably at national scales and exhibits severe variations at sub-national levels. Next, in order to analyze the spatial allocation of built-up areas in these datasets agreement-disagreement maps were developed from these datasets. Using these maps, analysis of built-up area morphological characteristics at different agreement levels was done to highlight the association of basic urban landscape characteristics with the mutual congruency of these datasets.

The general trend amongst the datasets, observed through the analysis in both Egypt and Taiwan, indicates higher inter-dataset agreement over the core urban areas, from where it decreases gradually over fringe and peri-urban areas. The decreasing inter-datasets agreement highlights the inability of some of these datasets to identify the areas away from the core urban areas where the most disaggregated built-up patches may be found. This brings the attention to the spatial resolution of these datasets as a likely reason for such failures in detection. However, the far flung areas from the urban core has been observed to have serious implications for urban sustainability. Peri-urban areas have been identified as the most predominant form of new urban growth, and remote sensing has an important

role to play in capturing this land transformation [177]. These areas, growing rapidly in disjunct fashion, have attracted much attention from researchers, especially in developing countries. For example, many facilities such as manufacturing units, airports, and office spaces were found to have shifted to fringe areas of Sao Paolo, Brazil [31]. Another study found a decrease in population density in core city and along with an increase in the fringe areas of Cairo, Egypt [216]. In this analysis, these essential components of urban areas were observed to have been missed out by the MODIS LC as well as CCI-LC, while being detected by LandScan SL and at most times by GHSL, causing the significant inter-dataset disagreement as observed before. The limitation of low-resolution built-up area datasets and the comparative strength of high-resolution mapping in capturing the most dynamic and rapidly evolving portions urban areas, especially at local scales is starkly visible in these areas.

In terms of the estimation of built-up area extents, past study has argued that even though the coarse and medium resolution images fail to capture the smaller and narrower landscape patches, it does not make a significant difference in the aggregated total area estimates [17]. In this light, the increasing variations in the total built-up area estimates from national to local scales can be explained, as the smaller differences got masked at an aggregated level but become more evident at local scales. Due to larger pixel size of MODIS LC or CCI-LC (500m and 300m), and contingent upon the land cover classification rules applied, it is likely that settlement patches smaller than the ground footprint of these pixel caused the entire pixel to be labelled as either built-up or non built-up, leading to erroneous detection and estimation of built-up area in either case. However, these uncertainties have serious implications for downstream studies, as this may result in misleading results from other studies like hydrological, population, and ecological modeling that use these datasets as main input. In addition to urban extents, identification of fine patterns of urban settlements helps analysis of urban morphology, and its relation to human health, mobility patterns, economic status. However, coarse and medium resolution datasets fail to account for these complex urban patterns emerging out of new urban growths as well as infilling of existing urban areas, creating severe hindrance for their application in urban studies. The new high-resolution human settlement datasets such as GHSL and LandScan SL were found to be able

to account for settlement areas that were unseen through other built-up area and land cover datasets.

Development of urban maps through digitization or crowd-sourced data may provide maps that meet the need of urban dynamics studies, however they comparatively need higher investment of money and time, which could act as a hindrance especially over remote areas. As urban analyses gradually become more focused on regional and local scales, requirements of detection and quantification of complex local-scale urban patterns for understanding urban dynamics will become increasingly vital. The high and very high-resolution image derived human settlement datasets could prove to be an efficient data source to meet such demands. The spatio-temoral urban change detection, which was primarily limited to global or national scale so far could now potentially be conducted at much finer scales, as the high-resolution datasets are produced in regular update cycles.

3.8   Conclusion

With the world becoming increasingly urbanized, analysis of local scale urban dynamics is becoming increasingly important for understanding the inter-linkages between local to regional and global level dynamics. To now, the coarse and medium resolution global and regional level land cover and built-up area datasets have fulfilled the needs of scientific data. However, as they fail to capture the complex, subtle urban patterns at finer scales, their applications become limited for new analyses focused on local areas. As observed from this study, aggregation, complexity, and contiguity of urban areas impact both detections as well as estimates of built-up areas by remote sensing derived datasets. The peri-urban areas, which are less contiguous and highly fragmented, may be missed entirely by low-resolution datasets. On the other hand, urban infilling typically consists of informal settlements and exhibits complex patterns, which again may not accurately be identified and represented by low-resolution datasets. Both of which seriously limits their applicability in urban dynamics analysis. It was also observed that the estimation of urban extents varies increasingly from national to local scale, where the lack of spatial resolution becomes a limiting factor for low resolution datasets. Harnessing the opportunity presented by very high-resolution

satellite imagery and high-performance computing infrastructure, a new set of high-resolution human settlement datasets could address the present limitations in urban remote sensing. Application of such high-resolution data could generate new insights on urban dynamics through spatiotemporal inter-comparison as the high-resolution human settlement datasets are developed for other countries, ultimately satisfying the data requirements of the present and future urban studies.

Chapter 4

Assessment of intra-urban electricity consumption patterns using a data-driven settlement characterization method

As per the most recent estimates, the phenomenal rate of urban growth is set to result in 68% of the global population living in urban areas by the year 2050 [198]. It is very likely that the existing urban infrastructures related to water, energy, and essential services will bear severe stress to accommodate this increasing populace, in near future [127], which are crucial factors in the well-being of modern society. United Nation's Center for Human Settlements considers the access to electricity as a vital indicator of economic growth [200]. It is imperative that the energy systems be planned well ahead in time, which is not only crucial for maintaining the present well-being of urban areas, but also to crucial for ensuring that future urban areas and those in low-income nations gradually get uninterrupted access to electricity. Hence, understanding of energy consumption patterns may not only help understand the present situation but may also assist in projecting future needs and creation of plans to address them [89, 227].

Various factors such as location, climate, topography, and economics govern electricity usage at local levels [132, 225], contributing to wide variations in electricity usage among nations and within geographic regions. In developed countries, much of the local level energy usage information may be captured by intelligent infrastructures such as smart meters capable of recording customer behavior and consumption patterns [76, 6]. The collected information may then be passed on to be analyzed in a central location to support the present, and short-term needs and also prepare the systems to meet the long-term energy requirements [165]. The value of such measurements and associated feedback have been found to be helpful in lowering the household energy wastage [39]. However, for cities in the developing countries, these infrastructure is practically non-existent [83]. At present, energy consumption estimates for most places are based on the Gross Domestic Product (GDP), population growth rates, and calculations of efficiency measures. However, such models may not be effective in capturing the energy dynamics at local levels [202]. GDP values have also been found to be inconsistently measured around the world and thus may not be reliable while trying to estimate future energy needs [59]. The issues surrounding irregular and non-standardized electricity usage surveys only make the task of electricity consumption monitoring much more difficult [59]; and prevents useful comparison among cities, needed for evaluation of efficiency measures. Under current projections, nearly 90% of the global urban expansion in the future will occur in Asia and Africa; while Africa will experience the highest rate of urbanization [180, 73].

In these areas where due to lack of infrastructure and statistics reporting mechanisms, remote sensing based methods could be used in deriving proxies of energy use. The high-resolution optical images are available for almost anywhere on the land surface. Texture analysis of these high-resolution images has been observed to help distinguish among urban settlement types [85, 8, 72]. Different urban settlements display contrasting physical appearance on high-resolution images [145, 72], which could be captured using texture analysis, leading to the development of a functional inventory of settlements. This inventory may then be utilized, on the assumption that, e.g., an industrial building will have very different energy consumption than a residential or commercial building, as a proxy to draw insights on local level socio-economic dynamics. To counter the problem of lack

of electricity statistics, another set of remotely sensed images called the nighttime lights might be used as a spatial proxy for the distribution of electricity usage. The nighttime lights capture visible as well as near-infrared lights emission from human settlements at night under cloud-free conditions. Nighttime lights imaging capabilities provided by Defense Meteorological Satellite Programs Operational Linescan Sensor (DMSP/OLS) have long been exploited by scientists to monitor and map electricity consumption patterns at regional and global scales [209, 210, 4, 23, 114, 224, 214, 215, 20, 78]. The recent introduction of Visible Infrared Radiometer Suite (VIIRS) Day/Night band (DNB) data has added much-needed improvement to existing nighttime imaging capabilities in terms of higher spatial and radiometric resolutions [52]. The initial research findings from the application of VIIRS DNB datasets have indicated higher discerning power than its predecessor [181, 98, 122, 169].

In this chapter, a set of texture measures was applied on the high-resolution optical images for three cities in the developing world, i.e., Ndola, Zambia; Sana'a, Yemen, and Johannesburg, South Africa. This was followed by a factorization based texture segmentation method [217] to classify the settlements into different functional types. Using this human settlement inventory, a correlation between settlement types and associated nighttime lights emission from VIIRS DNB data was drawn, to gather insight into electricity usage patterns. In the following sections, the study areas and datasets have been briefly described, followed by an explanation of the methods applied, and the obtained results. The chapter ends with a discussion on the observed patterns in these three cities and highlights the significance of this work for understanding urban energy dynamics in data-poor regions of the world.

4.1 Data

4.1.1 High-resolution optical images

The high-resolution satellite images for the three cities, with a spatial resolution of 0.3 - 0.5 meter, were obtained from the WorldView 2 and WorldView 3 satellites. The images were observed to encompass a wide variety of human settlement types, from which the functional neighborhoods have been generated. For the city of Sana'a, a collection of images in a strip

was processed. The data was collected by WorldView 2, on December 12, 2016, covering a total area of approximately 1,320 sq. km. Similarly, for the city of Ndola, a collection of images, collected by WorldView 3 on June 24, 2016, arranged as a strip was processed. The images covered an area of approximately 450 sq. km. The amount of geographical area processed for Johannesburg was 11,039 sq km and was spread across four image strips collected from WorldView 2 and WorldView 3 satellites. The acquisition period for those four images strips is from August 2017-December 2017. Across the 3-city study area a total of 6 image strips, further broken down into 32 scenes of approximately 44000-pixel columns by 31000-pixel rows, each scene was between 2-4 GB in size. The images were ordered as 4-band (NIR-Red-Green-Blue) data, and mosaics for each city was created to be used as the input data for texture analysis.

4.1.2 VIIRS DNB nighttime lights data

The 2015 VIIRS DNB annual composite data used in this study was downloaded from the website of the National Geophysical Data Center of National Oceanic and Atmospheric Administration (NGDC-NOAA) (https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html). The VIIRS DNB offer significant improvements over its predecessor DMSP/OLS datasets both in terms of higher spatial resolution (500m vs. 1km) and much higher radiometric resolution (14-bit vs. 6-bit quantization levels). The increased information content may enable minute detection of urban lights at local levels [48]. The VIIRS DNB annual composites are average radiance values of the emitted urban lights in nanoWatts/cm^2/sr. To create an annual composite, all available cloud-free data for a calender year is used to obtain the average values of the pixels. Prior to compositing, anomalies from Aurora, sunlit and moonlit pixels, clouds, lightning, and stray lights have been removed from the data, along with the background values [52]. However, as of now, lights from gas flares and biomass burning are still present in this data. To address this issue, a gas flare mask obtained from NGDC-NOAA was applied to the data, and all such lights associated with gas flares were omitted. After this correction, the data was considered to be free from all known sources of errors.

4.2 Study area

4.2.1 Johannesburg, South Africa:

Johannesburg is the largest city in South Africa (see figure 4.1) and had a population of 9.6 million in 2016. Johannesburg was originally founded in 1886, making it one of the youngest major cities in the world. According to the 2016 World's Cities Report, The United Nations projects Johannesburg to become a Megacity between 2016 and 2030 with am estimated population of 11.5 million by 2030. [197]. Johannesburg has seen many periods of rapid expansion, including after the discovery of gold in the 1890's. Like many cities in the global south, Johannesburg continues to experience the affects of rapid urbanization, frequent migration and crowding, and massive unemployment, among other challenges. The city experiences what can be described as a warm temperate climate, according to the Koppen-Geirger Climate Classification, with dry winters and warm summers. Temperatures are considered mild year round, which is another favorable draw to the constant influx of population [110]. The climate and location are not the only attractive characteristics of Johannesburg. The City of Johannesburg has typically had an economic growth rate higher than the provincial Gauteng and even the national level across all periods between 1997-2016 [102].
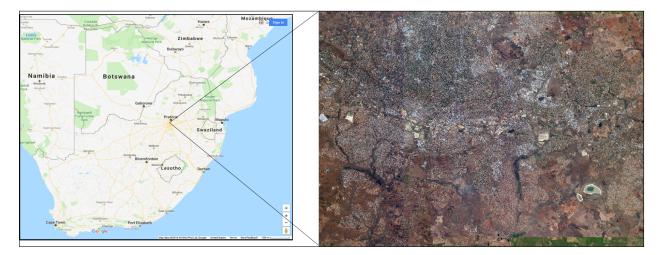
Figure 4.1: Location map and satellite image of Johannesburg, South Africa. (Source: Google Map. Note: Not drawn to scale, for representative purpose only)

4.2.2 Sana'a, Yemen:

Sana'a is the Capital, as well as the largest city in the Republic of Yemen (see figure 4.2). The population of Sana'a was 3,937,500 in 2012. It also carries the distinction of being one of the continuously inhabited oldest cities, as well as one of the highest elevation capital cities in the world. Due to it's long heritage and distinctive architectural patterns, part of the city (Old City of Sana'a) is designated as a UNESCO World Heritage Site. The city experiences mild semi-arid climate which can often be characterized as cold steppe/desert type of climate [110]. Due to the recent conflict in Yemen, many areas, including Sana'a, are experiencing a declining economy which is pushing poverty to abround 80% nationally. The recent influx of population is from Internally Displaced People (IDP) seeking refuge from the conflict. Due to increased political tension, Sana'a is experiencing armed conflict, which currently makes the city a point of constant dynamic population movement [192]
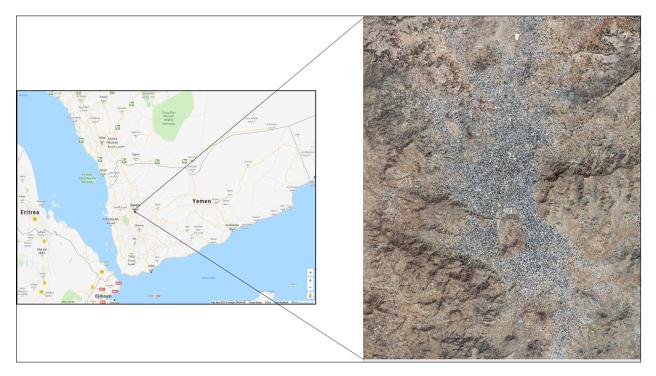
Figure 4.2: Location and satellite image of Sanaa, Yemen. (Source: Google Map. Note: Not drawn to scale, for representative purpose only)

4.2.3   Ndola, Zambia:

Zambia is a landlocked nation located in Southern Africa (see figure 4.3), surrounded on all sides by Democratic Republic of the Congo, Tanzania, Malawi, Mozambique, Zimbabwe, Botswana, Namibia, and Angola. Ndola is the third largest city of Zabia, which is also known as the commercial capital. Ndola experiences a warm climate with a dry winter, and warm summers.[110] As per the 2010 census, Ndola was home to 455,194 people. According to the 2010 Census of Housing and population, Ndola has had a steady growth rate of 1.9 percent between 2000-2010 [22]. It is located in the 'Copperbelt' region, which is well known for copper mining activities. Consequently, Ndola has become the main industrial and commercial hub in the region. The region went through an economic boom during the early 1980's, when multiple industries such as vehicle assembly and clothing manufacturing used operate in the area. However, it went through a lean period thereafter till 2000. Since then, the economy has continued a steady recovery with a 1.8 percent increase in the labor force between 2000-2010, and Ndola has an employment rate of 76.9% [22].


Figure 4.3: Location and satellite image of Ndola, Zambia. (Source: Google Map. Note: Not drawn to scale, for representative purpose only)

4.3 Methods

Two major steps were followed to achieve the research objectives. First, the settlement characterization was done in the study areas. Then, the analysis of the relationships between the settlement types and corresponding nighttime lights emission was carried out to derive insights on electricity consumption patterns. For characterization of human settlements, the texture responses were calculated on for training samples of settlement types in each of the three cities. The samples were collected after careful visual observation of each city so that such samples comprehensively represent the observed settlement types. Once the texture responses were calculated for these training samples, a texture-based segmentation method, as proposed by [217], was applied to the data for delineation of the different settlement types.

The texture measures used in this study as well as the segmentation method has been briefly described in the following subsections.

4.3.1 Texture analysis

Image texture could be characterized by the spatial distributions of grey levels in a given neighborhood of an image. Jain *et.al.,* defined texture as the repeating patterns of the local variations in the intensity of Digital Number (DN) values, which are too fine to be distinguished as separate objects [97]. On high-resolution images, texture analysis has been shown to provide supplementary information on the image properties and improve classification accuracy [1]. In this work, two widely available texture measures, Gabor filter and Laplacian of Gaussian, were used, which are described below:

Gabor filter:

Gabor filter, a linear filter that was originally introduced by Dennis Gabor in 1946, is mostly used for edge detection. Researchers have found successful application of Gabor functions to model the simple cells in the mammalian visual cortex and thus it is thought to be similar to the human visual system in perception. These have widely been used in imagery analysis owing to their spatial locality, orientation selectivity and frequency characteristics [123]. Gabor filters have been widely applied in urban feature extraction from remote sensing

images [170, 226, 205, 184]. Following the work of [62], the Gabor function could be defined as (see equation 4.1):

$$G(x, y | W, \theta, \psi, X, Y) = \exp^{\frac{-[(x-X)^2+(y-Y)^2]}{2\sigma^2}} \times Sin(W(x cos\theta - y sin\theta) + \psi) \qquad (4.1)$$

Where, σ represents the Gaussian width, θ is the filter orientation, W denotes the frequency and ψ indicates the phase shift. X and Y are the center of the filter applied. For an input matrix $L(x, y)$ and a Gabor operator $G(x, y | W, \theta, \psi, X, Y)$, a $G \times L$ spectra could be obtained for different orientations and shifts of the Gabor operator for identifying the texture element, as given below (see equations 4.2a - 4.2c):

$$GL_1(X, Y | W, \theta) = \sum xy G(x, y | W, \theta, 0, X, Y) \times L(x, y) \qquad (4.2a)$$

$$GL_2(X, Y | W, \theta) = \sum xy G(x, y | W, \theta, \frac{\pi}{2}, X, Y) \times L(x, y) \qquad (4.2b)$$

$$S^2(X, Y | W, \theta) = GL_1(X, Y | W, \theta) + GL_2(X, Y | W, \theta) \qquad (4.2c)$$

Where, x and y are indices over the basic matrix elements, ψ denotes the phase shift (i.e. 0 and $\pi/2$) and W is the number of cycles in n pixels, where n is the size of the input pattern. GL_1 and GL_2 are Gabor filter convolutions and textures while S is the locally shift invariant output obtained by using mean of GL_1 and GL_2.

Laplacian of Gaussian:

The Laplacians are the second order spatial derivatives of an image, which helps to identify areas of rapid change such as the edge areas in an image. The input to generating Laplacian is a single gray-scale image, and the process produces another gray-scale image as output. The mathematical formulation of the Laplacian is as follows (see equation 4.3):

$$L(x, y) = \nabla^2 f(x, y) = \frac{\delta^2 f(x, y)}{\delta x^2} + \frac{\delta^2 f(x, y)}{\delta y^2} \qquad (4.3)$$

Here, $L(x, y)$ is the Laplacian of an image where it's DN values are given by $f(x, y)$. In a gray-scale input image, this is calculated by applying discrete convolution kernels. However, the derivative filters are known to be sensitive to image noise. Hence it is required that a smoothing operation is done on the image, which is generally done using Gaussian filters before applying the derivative filters. This way, the process of application of Laplacian filter in an image texture analysis becomes a two-step process and is collectively known as Laplacian of Gaussian or LoG operation. When the Gaussian filter is included for smoothing purposes, a combined Laplacian of Gaussian may be explained by a single equation, under the assumption that the LoG function centered on zero and Gaussian standard deviation is σ (see equation 4.4):

$$LoG(x, y) = \frac{1}{\pi\sigma^4}[1 - \frac{x^2 + y^2}{2\sigma^2}]e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4.4}$$

LoG texture descriptors have found wide acceptance in urban feature extraction from images [84, 108, 218, 33].

4.3.2 Settlement characterization:

In this work, four Gabor filters with orientations at 0, 45, 90, and 135 degrees and a scale value of 2.5 was utilized along with two LoG filters with scale values of 1.2 and 1.8, to capture the textural characteristics of the settlements such as their size, density, and orthogonality of their layout. Satellite images for each city were first visually inspected for distinct settlement patterns. Once the distinct settlement types were identified, individual models for each city were developed using up to 4 training samples per settlement type. First, the texture responses were calculated on the training samples, which is followed by the application of a factorization based texture segmentation approach [217] to segment the data into these settlement types. A settlement area mask was used to keep non-settled areas out of this process, which helped in dimension reduction and computational efficiency. The settlement mask was obtained from the LandScan project team at Oak Ridge National Laboratory. More information on development of the settlement layer may be found in [26, 153]. The output was then checked manually for accuracy of segmentation, and new samples were included

if necessary. This was done repetitively until satisfactory settlement characterization was achieved.

The segmentation method utilizes a specific texture descriptor called local spectral histograms [217], which is calculated based on the local distribution of the filter responses [117]. In this method, the input is represented by an M × N feature matrix, comprising of M dimensional feature vectors computed from N number of pixels. The feature at every pixel is regarded as a linear combination of the representative features, which encodes the criterion for selection of boundaries. Two matrices, containing representative features and their pixel-wise combined weights, are multiplied to obtain the feature matrix. The combined weights indicate the belongingness of the corresponding pixels to different segments. The final segmentation is then achieved through factorizing the feature matrix using singular value decomposition and non-negative matrix factorization. For a detailed description on each of these steps, readers may refer to [217].

Through visual analysis of high-resolution optical images, multiple settlement classes in each of the cities were identified. These three cities are located in significantly different climatic and physiographic settings, and vast differences exist in the socio-cultural and economic conditions of their inhabitants. All such conditions could have distinct effects in the manifestation of the settlement patterns [182]. Distinct patterns were also observed in the settlements of these three cities. For the sake of generalizability and interpretation, five broad settlement types were identified whose interpretation closely aligns for all three cities(see figure 4.4).

	Ndola	Johannesburg	Sanaa
Class 1: High-Density, small-sized urban houses			
Class 2: High to medium density, Medium sized urban buildings.			
Class 3: Medium to low density, medium sized urban and suburban houses.			
Class 4: High-density large sized urban buildings			
Class 5: Very low density rural type developments			

Figure 4.4: Samples showing different settlement classes identified in three cities.

1. Class 1: High-Density, small-sized urban houses. Characterized by non-orthogonal layout.

2. Class 2: High to medium density, medium sized urban buildings. Characterized by orthogonal layout and mixed use commercial and residential. Some suburban developments in case of Johannesburg.

3. Class 3: Medium to low density, medium sized urban and suburban houses. Characterized by orthogonal layout.

4. Class 4: High-density large sized urban buildings. In case of Johannesburg, it includes the downtown area and high-rise buildings. In case of Sana'a, the development is medium to low density, however marked by large urban buildings.

5. Class 5: Very low density rural type developments, includes all other types of developments not included in Classes 1 - 4.

4.3.3 Analysis of correspondence between settlement types and electricity consumption:

As mentioned previously, in the absence of any electricity consumption data for these cities, VIIRS DNB nighttime lights data was used as an indicator for the spatial distribution of electricity usage. However, the 500 meter spatial resolution of VIIRS DNB is much coarser than that of the settlement inventory map used in this study, which is around 8 meter. Hence, a pixel to pixel correspondence was not possible in this study. To examine the relationship between the settlement types and corresponding nighttime lights emission, the presence of different types of settlements within the ground footprint area corresponding to each VIIRS DNB pixels was calculated. In order to calculate the composition of different settlement types, regular grids matching the spatial resolution of VIIRS DNB data covering each city was created. Next the areas of different settlement types within each grid was extracted in square kilometer unit. It was observed that for some parts of the cities, especially in Sana'a and Ndola, VIIRS DNB did not detect any lights. This could be a sign of the lack of electricity access. However, these settlements were most likely inhabited by people

belonging to economically weaker section of the society, hence, these areas should be a part of the analysis. Thus, such pixels were included in the analysis. The implications of this has been mentioned in the discussion section. Additionally, for Ndola, presence of a large flare was observed right on the eastern edge of the city, the affected pixels were removed from analysis using the gas flare mask mentioned earlier. Next, linear regression models were developed to assess the relationship between the settlement types and the corresponding nighttime lights emission, where the VIIRS DNB values were considered as the dependent variable, while the area of the settlement types were included in the model as dependent variable.

The general form of the equation is as follows (see equation 4.5):

$$DN_{VIIRSDNB} = (\sum_{i=1}^{5} \beta_i * Settlement_i) + \varepsilon \tag{4.5}$$

Where, S_i is the area of settlement type i in square kilometer, i denotes the number of different settlement classes. $DN_{VIIRSDNB}$ denotes the pixel values in the VIIRS DNB data, and ε denotes the Gaussian noise. The coefficients obtained from the regression model indicate what is the expected change in the nighttime lights emission in nanoWatts/cm^2/sr, a proxy for the in the electricity usage in the area in the absence of electricity consumption data, corresponding to one square kilometer increase or decrease of the individual settlement types. The data exhibited signs of heteroskedasticity, thus the coefficients and stanrard errors were estimated using a heteroskedisticity consistent method, the White-Huber standard errors. The initial settlement map, classified settlement map, VIIRS DNB nighttime lights observation, and correspondence between lights data and settlements have been graphically provided for each city in figures 4.5 - 4.7.

Subsequently, ANOVA tests were carried out on the regression coefficients, from each of the three, to check whether the light emission associated with each settlement types were statistically significant from each other.
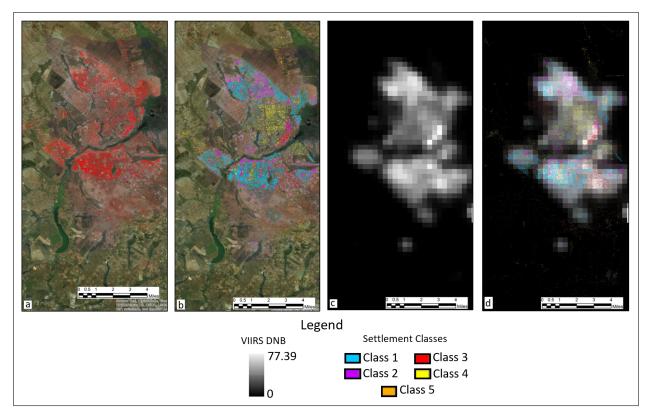
Figure 4.5: From left: Settlement Map, Settlement Classes, Settlement Classes overlaid on VIIRS DNB, and VIIRS DNB for Ndola, Zambia.
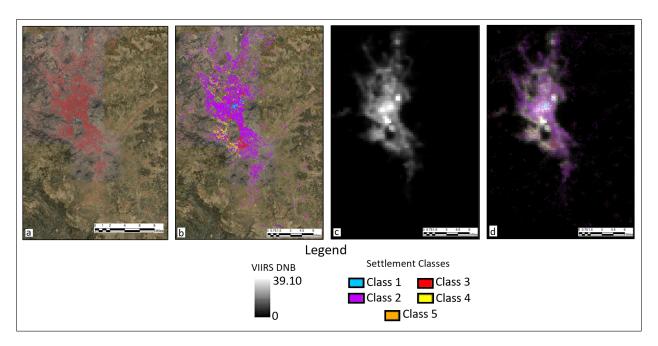
Figure 4.6: From left: Settlement Map, Settlement Classes, Settlement Classes overlaid on VIIRS DNB, and VIIRS DNB for Sana'a, Yemen.
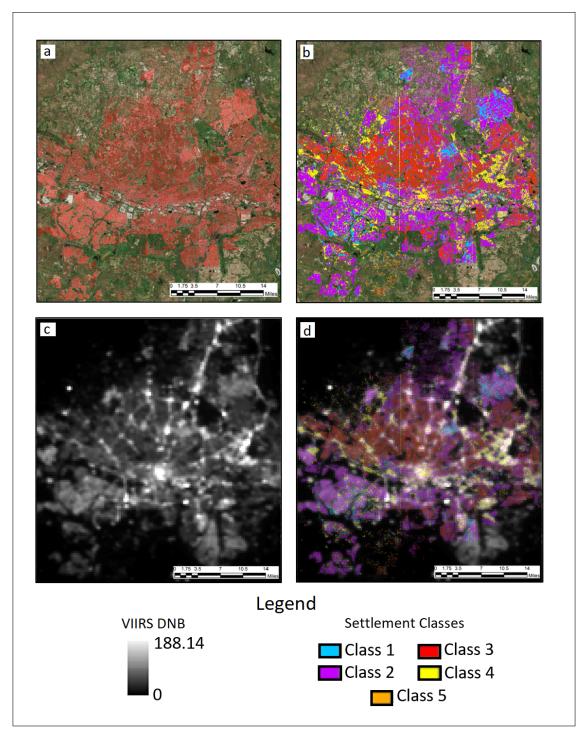
Figure 4.7: Clockwise from top-left: Settlement Map, Settlement Classes, Settlement Classes overlaid on VIIRS DNB, and VIIRS DNB for Johannesburg, South Africa.

4.3.4 Correlating building types to electricity consumption in Johannesburg

As mentioned in the introduction, urban level energy data is almost non-existent in developing worlds. However, linking actual electricity consumption to the building types would undoubtedly establish the applicability of the method presented here. A through search of published high-level reports revealed a 2011 record of 58,839,270 gigajoules (which is equivalent of 16,344,242 Mwh) total electricity consumption in Johannesburg metro area . An attempt has thus been made here to link this consumption to the buildings. However, the temporal mismatch of six years between the energy record and building inventory could not be bridged, and an assumption was made that the spatial distribution of electricity usage remained similar in this time period.

To connect the building types to electricity consumption, the total electricity consumption was spatially disaggregated from city to pixel levels using the nighttime light data, which is a surrogate of electricity consumption. The method of spatially disaggregating electricity consumption using nighttime lights was proposed by [222]. In this method, the sum of all lights in Johannesburg metro area was considered equivalent of the total electcity consumpotion, from there the electricity consumption represented by one unit of sum of light was calculated (see equation 4.6)

$$U_{VIIRSDNB} = Total electricity consumption/Sum of lights \qquad (4.6)$$

The sum of lights in Johannesburg was found to be 313,155. Using equation 4.6, the electricity consumption represented by an unit of sum of light was found to be 187.89 Mwh. Next, a constant raster whose DN value was equivalent to 187.89 was created and multiplied with the VIIRS DNB data. The DN values of the resultant raster now represents the electricity consumption in correspond ng area in Mwh. Again, a linear regression similar to equation 4.5 was developed, where the only exception was in the new equation electricity consumption was used as independent value.

4.4   Results

The results from the regression for Ndola, Sana'a, and Johannesburg indicated a statistically significant correlation between the settlement types and corresponding nighttime lights emission. The regression model for Ndola, Zambia, explained 77% of the variance in the data ($R^2 = 0.77$, $F(5,1634) = 1080$, $p < 0.0001$), while the model for Sana'a, Yemen, explained 71% of the variance ($R^2 = 0.71$, $F(5,5798) = 2862$, $p < 0.0001$). Lastly, the model for Johannesburg, South Africa, explained 45% of the variance in the data ($R^2 = 0.45$, $F(5,14338) = 2372$, $p < 0.0001$).

Table 4.1: Robust coefficients from regression between VIIRS DNB and settlement types in Ndola, Zambia.

Settlement Type	Regression coefficients ($\pm$ SE)	t value	p value
Class 1	1.70 ($\pm$ 0.08)	21.07	< 0.0001
Class 2	2.61 ($\pm$ 0.18)	14.14	< 0.0001
Class 3	3.17 ($\pm$ 0.37)	8.61	< 0.0001
Class 4	2.74 ($\pm$ 0.15)	18.38	< 0.0001
Class 5	0.35 ($\pm$ 0.24)	1.45	0.1465

In Ndola, Zambia, we observed with every square kilometer increase of class 1 type settlement (high-density, small urban houses), the nighttime lights emission increased by 1.70 nanoWatts/cm^2/sr. Similarly a square kilometer increase in high to medium, mid-sized urban buildings that could be used for both commercial or residential use (Class 2) leads to 2.61 nanoWatts/cm^2/sr increase in nighttime lights emission. The increase of per square kilometer area of Class 3 and Class 4 type settlements results in an increase of 3.17 and 2.74 nanoWatts/cm^2/sr respectively. The coefficient for class 5 was found to be extremely small and non-significant in this model. The most plausible reason for this have likely been caused by extremely sparse development in Class 5 for this city, many of which were not identified by VIIRS DNB data (see figure 4.5) . Most nighttime lights emission in Ndola were found to be associated with Class 3 which represents the medium to low-density, medium sized urban and suburban houses (see table 4.1).

For Sana'a, Yemen, most nighttime lights emission was found to be associated with Class 4, showing an increase of 3.86 nanoWatts/cm^2/sr in nighttime lights emission from one

Table 4.2: Robust coefficients from regression between VIIRS DNB and settlement types in Sana'a, Yemen.

Settlement Type	Regression coefficients (± SE)	t value	p value
Class 1	3.60 (± 0.31)	11.79	< 0.0001
Class 2	1.99 (± 0.04)	44.44	< 0.0001
Class 3	2.89 (± 0.33)	8.62	< 0.0001
Class 4	3.86 (± 0.19)	19.91	< 0.0001
Class 5	1.77 (± 0.33)	5.41	< 0.0001

square kilometer increase in Class 4 settlement type. This was closely followed by Class 1, showing 3.60 nanoWatts/cm^2/sr increase in lights emission corresponding to a one square kilometer increase in the area of Class 1 type settlements. Class 3, 2, and 5 were observed as having comparatively less effect on nighttime lights emissions (see table 4.2).

Table 4.3: Robust coefficients from regression between VIIRS DNB and settlement types in Johannesburg, South Africa.

Settlement Type	Regression coefficients (± SE)	t value	p value
Class 1	1.54 (± 0.08)	19.47	< 0.0001
Class 2	2.05 (± 0.03)	70.90	< 0.0001
Class 3	1.50 (± 0.03)	49.51	< 0.0001
Class 4	5.38 (± 0.10)	51.72	< 0.0001
Class 5	1.75 (± 0.32)	5.45	< 0.0001

In Johannesburg, South Africa, Class 4 type settlements, which included the downtown areas and the high-rise buildings, were observed to have the most association with the nighttime lights emission. Results show a 5.30 nanoWatts/cm^2/sr increase in nighttime lights for each square kilometer increase in Class 4 type settlements, while a square kilometer increase in Class 2 type settlement was found to result in an increase of 2.05 nanoWatts/cm^2/sr in nighttime lights emission. Classes 1, 3, and 5 have a lesser and somewhat similar association with nighttime lights (see table 4.3). The analysis of correspondence between electricity consumption and settlement types were done using a linear regression similar to equation 4.5. The results are presented in table 4.4 below:

From table 4.4, the impact of each type of settlement on electricity consumption can be identified. Similar to table 4.3, a square kilometer increase in high-rise, large commercial building stock was seen to cause about 1012 Mwh increase in annual electricity consumption,

Table 4.4: Robust coefficients from regression between 2011 electricity consumption and settlement types in Johannesburg, South Africa.

Settlement Type	Regression coefficients ($\pm$ SE)	t value	p value
Class 1	289.60 ($\pm$ 14.87)	19.47	< 0.0001
Class 2	385.24 ($\pm$ 5.43)	70.90	< 0.0001
Class 3	281.35 ($\pm$ 5.68)	49.51	< 0.0001
Class 4	1011.66 ($\pm$ 19.56)	51.72	< 0.0001
Class 5	328.01 ($\pm$ 60.18)	5.45	< 0.0001

the corresponding electricity consumption increase from class 2 type buildings was 385.24 Mwh.

The ANOVA tests highlighted that the light emission from each type of settlements were indeed statistically significant from each other. Only for Class 5 or the rural type of settlements in Ndola did not show statistical significance. This is most likely an effect of very sparse distribution of this type of settlements in Ndola, from where not much light was captured to render its statistical significance. Results from the ANOVA tests have been provided in tables B.1 - B.3 in appendix B.

4.5 Discussion

The research objective of this study has been to assess the applicability of a data-driven settlement characterization method in understanding local level electricity consumption patterns. Two widely accepted texture descriptors and a texture-based segmentation method proposed by [217] was utilized to develop a functional inventory of urban settlement for Ndola, Zambia; Sana'a, Yemen; and Johannesburg, South Africa. This functional settlement inventory was then used to find a possible correlation with the proxy of electricity consumption. As mentioned earlier, in the absence of measured electricity consumption data, remote sensing based indirect methods such as presented here could help fill the data gap. Hence, the nighttime lights from VIIRS DNB were used here as a surrogate of urban electricity consumption. Since the spatial resolution of the urban settlement raster created in this work and VIIRS DNB are widely dissimilar, the area of different settlement types corresponding to each VIIRS DNB pixel was calculated, next regression

models were developed between the area of different settlement types and corresponding nighttime lights emission, expressed in nanoWatts/cm^2/sr. In all three cities, a common trend was observed to emerge where the Class 4 type settlements (High-density large sized urban buildings; in the case of Johannesburg, it includes the downtown area and high-rise buildings) in Johannesburg, and Sana'a corresponds the most with nighttime lights emissions. In the case of Sana'a, Class 1 (High-Density, small-sized urban houses characterized by non-orthogonal layout) settlements were seen to have the second most association with nighttime lights emission, closely following the Class 4 buildings. In Ndola, the highest association was observed between Class 3 or the medium to low density, medium sized, urban and suburban houses. The rural type, sparsely distributed buildings did not exhibit a statistically significant relationship with the lights; it is assumed that this is a likely effect of their sparse distribution and the fact that many such buildings did not register lights. Using an old statistic in Johannesburg, it was demonstrated that the effect of building stock on electricity consumption can be evaluated. However, even with the nighttime lights unit, the relative interpretations of the effect of the buildings on energy consumption remain the same.

The general trend observed from the three regression results indicate a common correlation between settlement types and corresponding nighttime lights emission, considered a proxy of electricity consumption in this study. This validates the general applicability of this method in understanding local electricity consumption patterns in data sparse regions. The usage of nighttime lights as a spatial indicator of electricity usage further generalizes this approach as this data is freely available for any place on Earth. However, due to the coarse spatial resolution of 500 meters, VIIRS DNB data alone may not be able to address the minute local variation within cities. Characterization of human settlements, as seen in this work, extends the scope of this analysis to a much finer resolution, beyond the generic spatial resolution of VIIRS data. Ratti et al. [167], found useful correlation between urban texture and energy consumption, while assessing the applicability of digital elevation models alongside 3-dimensional urban models . While this work was focused on highly developed cities such as London, United Kingdom; Berlin, Germany; and Toulouse, France, and used much detailed datasets describing urban texture, it nonetheless demonstrated that patterns observed in urban landscapes could be a key to understanding energy consumption dynamics.

This work indicated that urban density and organization patterns have significant influence in determining energy consumption. Comparatively, this work was done using only two remote sensing datasets with an objective to apply these methods where no ground data is available. The fact that results from this study conform to similar patterns, as seen from the past work mentioned above, confirms it's applicability.

As caveats of this study, it first needs to be stated that the nighttime lights data was found not to have captured lights from some parts of these cities, especially in Sana'a, Yemen, and Ndola, Zambia. This is an artifact of the socio-economic conditions in these places as well as the late overpass time of VIIRS which is after midnight local times. Hence, many parts of these cities could have turned their lights off, and the reliability of power grids might also have played a role in non-detection in these parts. Secondly, the coefficients of determination for Johannesburg, South Africa was much weaker than the other two cities. It can be assumed that this is an effect of highly mixed land-use in parts of Johannesburg, resulting in nighttime lights signals that did not always match with the physical appearance of the buildings, e.g., much higher light emission from a building that houses both residential and commercial activities.

This study presents an application of data-driven methods in urban electricity related applications. In the absence of any ground information, which is prevalent in many low-income countries, this method could indicate the differential electricity consumption levels at a local scale. Such information, in conjunction with land-use land-cover information, could provide valuable insights into urban planning. For example, the planners may correlate between existing land use and electricity usage patterns, identify zones that are prime for rolling out efficiency measures or areas that could be developed according to plans meeting future goals.

4.6   Conclusion

This study provides a direction towards using remote sensing data and image processing driven methods to understand local-scale energy consumption patterns, and understanding socio-economic dynamics within cities. Application of these methods over more number

of cities will validate the assumptions this study relied on. The patterns identified in this study may be used in conjunction with other available data such as land use and land cover maps to inform local-level policymaking. Ultimately, helping urban planners make informed decisions in data scarce regions of the world.

Chapter 5

Dynamics of urban electricity consumption in the United States and implications for sustainability

Urban areas consume around three-quarters of the world's primary energy supply and generate almost 70% of global energy consumption related CO_2 emissions [196, 178]. Electricity is the second largest fuel by end-use within the global energy mix. As of the year 2015, electricity accounted for around 18.5% of global energy consumption and occupied an even higher market share of 22.2% in the case of OECD countries [92]. Urban electricity requirements arise from a myriad of activities related to operating and maintaining built-environments, e.g., lighting, space heating and cooling, running appliances in urban households, and operating machines in industries. The urban population currently accounts for 55% of the world's total population; this is expected to reach 68% by the year 2050 [198], expanding urban areas and infrastructures considerably in the due process. Urban electricity consumption can be expected to increase significantly in the future due to unhindered urbanization, up-gradation of existing infrastructures with electrical systems, increasing electrical use in industrial machines, and mass usage of electric vehicles. IEA estimates that 40% of the growth in global energy end-usage by year 2040 will come from increased consumption of electricity, which is presently the fastest growing energy fuel by end-use[93]. Under these circumstances, the energy policymakers face twin challenges of

ensuring uninterrupted future supply of electricity and reducing environmental impacts from electricity generation. The existing electricity generation pathways are known to cause irreversible harm to the environment through emission, water usage, and land-use change. Within the United States, it is estimated that 67% of the generated electricity came from fossil fuel sources in 2016, and US power plants were responsible for the release of 64% and 75% economy-wise release of Sulfur dioxide and acid gases respectively [124].

Several past studies have indicated the dual role of urban areas as both the cause and the solution to global sustainability problems [178, 65, 223]. The urban areas are characterized by high energy demands and diversity of energy usage which presents unique opportunities for implementing energy efficiency measures, energy management, and clean energy practices [21, 73]. Urban infrastructures are constantly expanding to support the inflow of people, which further increases the energy demands. Similarly, urban areas concentrate sets of diversely skilled people within their perimeter, coupled with the free flow of ideas, capital, and materials may also encourage transitional changes to promote efficiency and lower per capita electricity consumptions. These effects depend on the interplay of several socio-economic, geographical, and technological factors. Thus, systematic studies are required to identify the trends at the urban system level. Such studies may serve as a metric used to identify the needs of intervention and analyze the performance of efficiency measures.

However, regardless of the fact that all over the world energy usage is heavily urban-centric, the existing energy statistics mechanisms operate at regional and national scale. Such aggregated datasets mask the variation observed from one urban area to another, and masks any potential for city comparison, energy analysis at the urban system level. Even in the United States, the second largest electricity consumer after China, the urban level electricity data is rare. The US Energy Information Administration (EIA) provides electricity sales statistics at state levels which can be used as an indicator of state-level consumption. Thus, it is critical to develop efficient and generalizable urban level energy data gathering methods, for studying the present dynamics and predicting the future scenarios. In the absence of statistical data, geospatial data based methods may be used to approximate the urban electricity consumption, using pertinent variables such as climate, economy, industrial activities, and population.

In this chapter, an attempt has been made to use the EIA state-level electricity data to downscale the electricity consumption at urban levels, in order to overcome the obstacle of data scarcity. The estimates were then used for the analysis of the effects of urbanization on electricity consumption in the US urban system. The following questions were explored in this study: 1. Can geospatial and ancillary data based models be used to satisfactorily estimate urban electricity consumption in the United States? 2. What is the expected change in electricity consumption in response to urbanization in the United States? 3. If any evidence of returns of scale is found then where are the limits of such energy efficiency? The following sections sequentially describe the datasets used in this study and their sources. Followed by a detailed discussion on the rationale of urban delineation scheme, urban electricity estimation model, and scaling analysis. The results from this study have been presented afterward along with the discussion of the implications, In conclusion, the summary and limitations of this work have been provided.

5.1   Data

The first step in this study was to develop a model to estimate the electricity usage at urban scale. To ensure the general applicability of the model elsewhere, care was taken to include variables that can be easily obtained at other places. As detailed in Chapter 2, urban electricity consumption is shaped by factors such as local climate, economy, industrial activities, and population. Thus, degree days were included as an indicator of local climate, aggregated household income was included to reflect the economic status, and the number of manufacturing industries indicated economic activities. The following describes these datasets and their sources in detail. It also includes descriptions of additional datasets, such as population, extents of built-up areas.

5.1.1   Climate data

The Heating and Cooling degree days (HDD and CDD) are commonly used metrics to estimate the energy requirements to satisfy heating and cooling needs inside buildings, as a response to daily weather across different locations [95]. On a given day, the number of degree

89

days for a place is calculated by comparing the mean temperature to a base temperature, commonly considered as 65°F. For example, a day's a mean temperature of 70°F corresponds to 5 CDDs while a mean temperature of 30°F results in 35 HDDs. NOAA estimates the daily temperatures for each of the 344 climate divisions in the United States by using the records from one to four relevant weather stations. These daily temperature estimates are then used to calculate the number of degree days for respective climate divisions. Using the degree days of the relevant climate divisions, the degree days for the individual states are finally calculated. In this process, the division level records are multiplied by weights proportional to the ratio of the population of the corresponding divisions to that of the state, to reflect the effect if population on energy consumption [146]. Degree days data can be freely downloaded for both station and state levels from NOAA website. The station level data provides the number of CDD and HDDs on a monthly basis for 331 NOAA weather stations situated within the contiguous US, while the state level data is derived from the process mentioned above. Both station and state level datasets for year 2015 were downloaded from NOAA's National Weather Service website (`ftp://ftp.cpc.ncep.noaa.gov/htdocs/products/analysis_monitoring/cdus/degree_days/archives/`).

5.1.2 Electricity data

Retail electricity sales (in Megawatt-hours, Mwh) to the end-consumer segments was considered as a proxy for electricity consumption. 2015 data was downloaded from the EIA website (`https://www.eia.gov/electricity/data/state/`). In this data, sales volumes are recorded across four different consumer segments: residential, commercial, industrial, and transportation. Among these sub-categories, the electricity used for lighting public buildings, streets, and highways, and sales to the public entities have been accounted for in the commercial sector; while agricultural and irrigation related sales are included in the industrial sector. Electricity sold to the transportation sector include electrified rail and urban transit systems. Very few states contain records under this category, and it formed a small fraction of the respective total electricity sales for states where there was any electricity used by the transportation sector. Hence, for the sake of generality, only the first three categories were considered in this study.

California is the only state for which the county level electricity consumption data is available. This data was downloaded from the California Energy Commission website (http://ecdms.energy.ca.gov/elecbycounty.aspx). The data was provided in millions of Kilowatt-hours units. Thus, the necessary conversion factor was applied to the data to make it comparable with the EIA data.

5.1.3   Socio-economic data

The American Community Survey (ACS) is a part of the decennial census program, that is conducted nationwide to identify community-level changes. Various attributes such as income, house values, and age structure are captured in this yearly data. The aggregated household income data for each state and counties was obtained from the ACS database. The 5 year estimates are developed based on the data collected in previous five years proceeding the release year. This version of the data has been indicated to have higher precision than single-year estimates. Thus, the 5 year estimates were chosen for this study and downloaded the data from the US Census website (https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml)

County Business Patterns (CBP) dataset is released annually by the US Census Bureau, containing information on business-related economic parameters such as the number of employees, payrolls, and number of business establishments at multiple geographical scales like states, counties, and Zip codes. Number business establishments are recorded following the North American Industry Classification System (NAICS), information on some establishment types like public administration, government and railroad employees, National Postal Service, and the self-employed have been excluded from this data [18]. 2015 data was obtained from the US Census website: (https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=BP_2015_00A1&prodType=table).

5.1.4   National Land Cover Database

The National Land Cover Database (NLCD) is a time-series land cover dataset, produced at 30-meter spatial resolution by the Multi-Resolution Land Characteristics Consortium at

the United States Geological Survey (MLRC-USGS). The most recent release of this dataset dates to 2011, which was downloaded from the MLRC website (https://www.mrlc.gov/nlcd11_data.php). The developed areas within the contiguous United States were then extracted from this data. NLCD follows a 16 class land cover classification system, which includes forest, shrubs, developed, cultivated, and wetlands, among other land cover classes. The developed class represents impervious areas such as housing, parks and clubs, industrial, and commercial spaces. Based on percent of the impervious area, the developed class has been sub-divided into open spaces, low intensity, medium intensity and high intensity developed areas [87]. Based on the original classification scheme, the pixels were re-classified as either developed or non-developed, by grouping all developed pixels into one class and all other pixels into non-developed class (see figure 5.1).
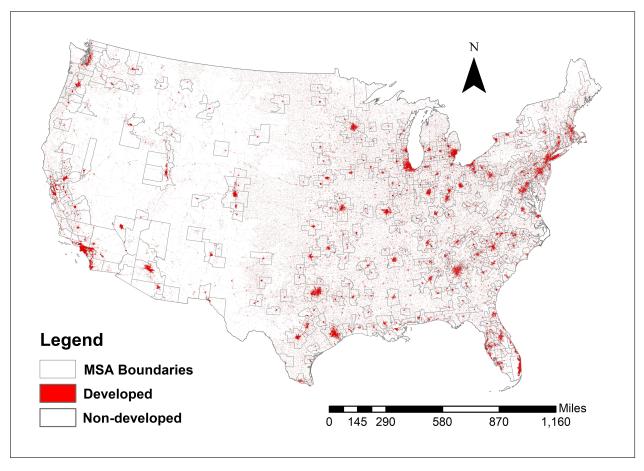
Figure 5.1: Developed areas in the contiguous United States, extracted from 2011 NLCD data.

5.1.5 Administrative boundaries and population data

Administrative boundaries, such as outlines of the contiguous United States; individual states, counties, and MSAs for the year 2015 were downloaded as TIGER (Topologically Integrated Geographic Encoding and Referencing) Shapefiles from the Census Bureau website (https://www.census.gov/geo/maps-data/data/tiger.html). These vector layers also contained attribute information on 2015 population estimates (estimated using 2010 population as the base) and geographic areas of the administrative subdivisions.

5.2 Methods

The research objective in this work required a few key steps be followed in sequential order. The functional definition of urban areas is critical for this analysis, hence an appropriate urban unit was selected first. Next, estimates of county level degree days were made using the NOAA station level degree days data. Finally, the electricity consumption model was developed using state level electricity consumption data as dependent variable and state level degree days, number of manufacturing industries, aggregated household income, and population as independent variables. States are the highest level of spatial granularity at which electricity statistics was available. Thus, downscaling of state level consumptions to the county levels had to be done using the electricity consumption model. Once the county level consumptions were estimated, the estimates were normalized to the corresponding state level totals. Next, the corresponding counties for each urban area were aggregated to obtain the urban electricity consumption estimates. Finally, the scaling analysis was done using the estimates of urban electricity consumption and corresponding census derived population for 2015. The following paragraphs provides a detailed and systematic description of each of these steps.

5.2.1 Delineation of urban areas:

Urban socio-economic dynamics, such as electricity consumption, is characterized by socio-economic interactions between the urban core and its surrounding areas. Changing urban

definitions have been observed to have severe effects on the observed relationship between the size of urban areas and their socio-economic characteristics [11]. Therefore a stable functional definition of urban areas is required for analysis of urban electricity consumption. A functional definition can spatially enclose the network of associated socio-economic interactions, and is more appropriate than administrative boundaries which are often drawn in an arbitrary manner [12, 65]. Moreover, functional definition enables clear demarcation of electricity consumption patterns over space, removing ambiguities over association of spatial patterns with urban entities.

In the United States, the Metropolitan Statistical Areas (MSA) represent a temporally consistent definition of urban areas which is independent of administrative definitions such as states or municipalities. Recent work in urban energetics [65] used multiple sets of urban definitions and observed a wide range of exponents for the relationship between urbanization and consumption of total electricity. The results from [65] further justifies the need of a single, unambiguous, and objective urban definition. Among all urban definitions in the US, MSAs were found to be most aligned with an ideal functional definition of urban areas. MSAs are characterized by the presence of a city with at least 50000 inhabitants or an urbanized area with a minimum of 100000 people. Along with the core urban areas, MSAs also consist of one or more adjacent whole counties that share a high level of socio-economic integration, measured by the patterns of commuting to work, with the urban core. This effectively marks a unified labor market within which the flow of materials, information and people is fully contained. As per the 2010 Census, around 83.7% population in the United States lived in MSAs, with a population density of around 283 persons/mile^2 which is several times higher than the non-urban areas that experience a population density of 10 person/mile^2 [212]. Cumulatively, these MSAs generated around 90% of the national GDP in 2015. All 377 MSAs in the contiguous United States in 2015 were included in this study. The terms urban areas and MSA have been used interchangeably in the rest of this chapter.

5.2.2 Estimation of county degree days:

Yearly number of degree days for each county were estimated using the NOAA station level monthly observations. As climatic phenomena such as temperature, wind, and humidity

are affected by a variety of location specific geographic factors, which may not have been accounted for by the nearest station alone. Thus, it was empirically determined that 6 nearest stations were needed to sufficiently account for the site specific variations in degree days in each county. First, the yearly degree days for each stations was calculated from the monthly data. Six nearest weather stations from each county were then identified. The degree days for each county were then determined through Inverse Distance Weighted (IDW) interpolation, where the effect of each nearest six stations on the estimated CDD and HDD is inversely proportional to the square of their distance from the county centroids. The process is defined in equation 5.1:

$$Z(x) = \frac{\sum w_i z_i}{\sum w_i} \tag{5.1}$$

Where, $Z(x)$ is the estimated degree days at location X, z is observed number of degree days at the weather stations, w is the distance based weight ($w = 1/distance^2$) assigned to each of 6 stations used for estimation, and i is the station index. The estimation results were validated against the actual observations from NOAA weather stations, which indicated an average mean percentage error of 7.2%. The actual degree days count was used for the counties where NOAA weather stations were located. The estimated county level degree days have been presented in figure 5.2.

5.2.3 Estimation of electricity consumption:

Electricity consumption was modeled as a function of climate, industrial activities, economy, and population. The electricity sales for the year 2015 was included as a proxy for consumption. Degree days were used as an indicator of electricity needs in response to local climate, the number of manufacturing establishments was considered a proxy of industrial activities. Aggregate household income was included as an indicator of economic status. The states were split into two categories based on the median aggregated household income, normalized by population to control for the large variations in state population sizes. These groups were included in the model as dummy for economic status. Electricity needs in response to climate is highly dependent on the population, e.g. moderate temperature
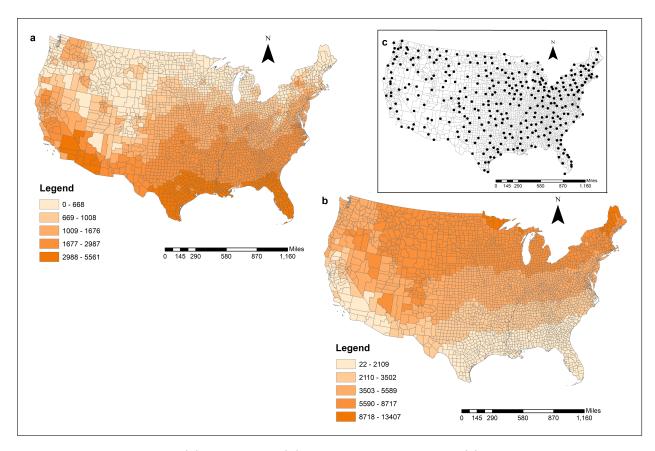
Figure 5.2: Estimated (a) CDDs and (b) HDDs at county level. (c) NOAA weather station locations are shown in the inset.

change in a highly populous area will lead to higher demand than a significant temperature change in a sparsely populated area. Thus, to include the population effect in the model (see equation 5.2), the number of degree days were included as interaction variables with population. In order to check for the main effects in the model, another model with main effects were developed and tested against the model presented here. The Chi Square test between the two models revealed the first model did not fit the data significantly better than the second model. Hence, the reduced model has been used in this study.

$$E = \alpha + \beta_1 \times G + \beta_2 \times (CDD\!:\!POP) + \beta_3 \times (HDD\!:\!POP) + \beta_4 \times M \qquad (5.2)$$

In equation 5.2 above, E indicates state-wise retail electricity sales, G represents the two categories of income, CDD and HDD represent the estimated cooling and heating degree days, POP is the corresponding population in 2015, and M is the number of manufacturing

establishments for each state. The regression was performed with a heteroskedasticity consistent, White-Huber method method, the summary results are presented in table 5.1 below.

Table 5.1: Summary of regression results using heteroskedasticity consistent standard error estimation.

Variable	β	SE(β)	t Value	p Value
Intercept	-1.88e+07	1.30e+07	-1.45	0.153153
Group	1.95e+07	7.16e+06	2.72	0.009282**
I(CDD:POP)	2.30e+03	2.38e+02	9.65	1.958e-12 ***
I(HDD:POP)	2.84e+03	6.22e+02	4.58	3.851e-05 ***
M	3.87e+02	6.85e+02	0.56	0.575181

The coefficient of determination (R^2) was 0.97 ($F(4,44)=391.5$, $p<0.0001$). Compared against the actual data, the model estimates indicated a mean relative error (where, $RelativeError(RE) = (Actual - Estimate)/Actual$) of 0.09 and standard deviation of RE equaling 0.36. The same set of independent variables, as in equation 5.2, were either estimated (degree days) or obtained (aggregated household income, population, number of manufacturing industries) for each county (see section 5.1, for details on the data), on which the model was used to predict county-level electricity consumption.

The initial county estimates (E_c) were aggregated to corresponding state levels (E_s). Normalization factors (N_s) for each state were then calculated, as a ratio between actual (A_s) and estimated consumption (E_s). Finally, the normalized estimates (E_c^*) were obtained by multiplying the initial county estimates by the corresponding normalization factor (See equations 5.3a - 5.3c).

$$E_s = \sum E_c \tag{5.3a}$$

$$N_s = \frac{A_s}{E_s} \tag{5.3b}$$

$$E_c^* = N_s \times E_c \tag{5.3c}$$

Where, s and c are indexes for states and counties respectively. In the next step, the MSAs were matched to the corresponding counties, and the E_c^* for the counties corresponding to each MSA were added up to get the MSA level estimates.

5.2.4 Validation against California data:

For external validation, the electricity consumption estimates were compared against the California data for the same time period. As the California data was also provided at county levels, the counties forming each MSA were added together to calculate actual electricity consumption for California MSAs for the year 2015. A plot of relative errors and log of MSA population has been provided in figure 5.3. The mean RE of 0.12 was obtained across all MSAs in California. The RE for most MSAs lie within ± 0.25, with several of them being tightly knit around the RE = 0 line. REs for eight MSAs fell within $pm(0.25 - 0.5)$. Three MSAs with small population and low consumption have REs the range 0.5 $\geq$RE$\leq$1.0. These were interpreted as an artifact of using a state-level model for county-level estimation, causing it to overestimate at smaller urban areas. However, the model worked reasonably well to capture the moderate to major urban areas in California. Thus, going forward in subsequent analyses.

5.2.5 Scaling of urban electricity consumption:

Urban population size was used as an indicator to assess the electricity consumption change in response to urban growth. Urban population has been widely accepted as an effective determinant of it's socio-economic characteristics [12]. Quantifiable variation in measurable urban characteristics has been observed in response to population change. For example, economic productivity and innovation rates have been observed to increase about 115% with each doubling of population, while the corresponding increase in the length of electric cable was only 90% [14]. The productivity, innovation, and wealth in urban areas rise with the growing population, increasing their attraction, consequently attract more people to urban areas. This sets off a chain event that allows larger urban areas to grow faster and consume resources more rapidly than the smaller ones. Urban scaling refers to a power law theory
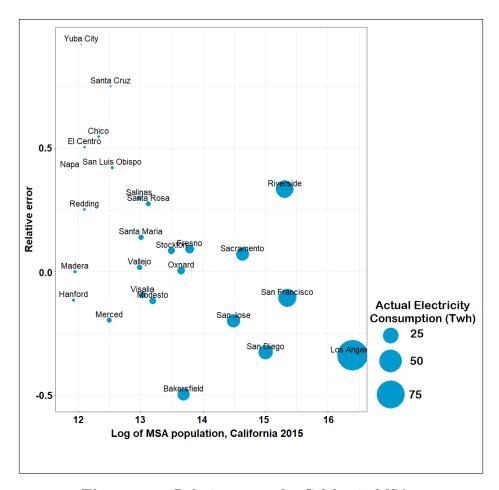
Figure 5.3: Relative errors for California MSAs.

that efficiently captures this functional relationship between urban properties and size, which may be defined as follows (see equation 5.4):

$$Y(t) = Y_0 \times N(t)^{\beta} \tag{5.4}$$

Where $Y(t)$ is the quantity of an urban property, Y_0 is a normalization constant and $N(t)$ is the population size at time t, the scaling exponent is denoted by β which indicates the change in urban property in response to population change. If the per capita quantity at time t is denoted by $y(t) = Y(t)/N$, the per-capita change in the said urban property in response to a fractional population change $(\Delta N/N)$ can be expressed as $\Delta y/y \approx (\beta - 1)\Delta N/N$. [14]. If $\beta = 1$, then $\Delta y/ = 0$, y remains constant, and Y and N exhibit a linear relationship; signifying no change in the per-capita quantity with respect to population growth. Variables related to the individualistic needs have displayed such relationship with population size.

But practically, the scaling exponents of most urban properties differ from this and display considerable variation [14]. Indicators of economic productivity such as number of patents, creative occupations, wages, and gross domestic product have been observed to scale super-linearly ($\beta > 1$), while infrastructural indicators such as road surface, electrical cable lengths, number of gas stations, and volume of gasoline sales exhibit sub-linear scaling ($\beta < 1$) [13]. At any given value of $\Delta N/N$, y is independent of the initial size N and is a function of β. This size invariance of scaling exponent allows for its application over all urban areas within a system, based on their relative differences irrespective of the absolute quantities.

Following equation 5.4, electricity consumption and population size in urban system of the United States was theorized below (see equation 5.5a).

$$E_i = \psi_0 \times N_i^{\beta} \tag{5.5a}$$

Where E_i and N_i are respectively the electricity consumption and population of urban area i, ψ_0 is a normalization constant, and β is the scaling exponent. Linear form of the equation was obtained via log-transforming both sides (see equation 5.5b).

$$log(E_i) = log\psi_0 + \beta \times log(N_i) + \varepsilon \tag{5.5b}$$

In equation 5.5b, β is the elasticity measure of the urban scaling representing the expected percent change in E_i resulting from 1% change in population, and ϵ is the gaussian noise. We determined the exponent using natural logs of estimated electricity and population.

As MSAs are made up of adjacent whole counties, they include all non-developed areas inside the county boundaries. Figure 5.1 highlights the composition of developed and non-developed areas within MSAs. As the human activities related to electricity consumption are almost entirely concentrated within the developed areas, population densities based on geographical areas may provide unreasonable correlation with per capita electricity consumption. Thus, To explore the relationship between per-capita electricity consumption and population density, urban population densities were calculated on the basis of developed area, which was derived from the NLCD data.

5.3 Results

5.3.1 Comparison of electricity consumption among urban areas

The estimates of electricity consumption for 377 metropolitan regions in the US reveal a stark contract between large urban centers against the rest. Figure 5.4 represents a proportional area diagram depicting electricity consumption at each of urban areas, from which it is evident that handful of urban areas consume a major of electricity. The top 10 MSAs together consumed around 27% of the total yearly consumption in 2015, while the same for the bottom 50 MSAs is only around 2.3% (see table 5.2 and 5.3). New York-Newark-Jersey City metropolitan region is by far the largest consumer of electricity, consuming about 6.5% of the total urban consumption in 2015, followed by the Chicago-Naperville-Elgin, and Washington-Arlington-Alexandria metropolitan regions. On the other hand, Great Falls, MT consumed the last amount of electricity, followed by Missoula, MT, and Pocatello, ID.
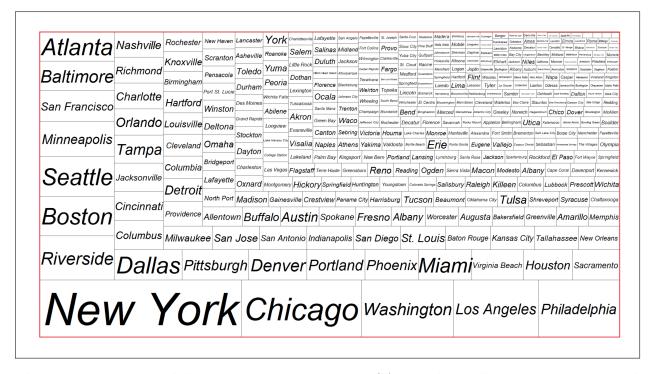
Figure 5.4: Estimated electricity consumption in MSAs. Each small square is proportional to consumption, the rectangle circumscribed in thick red line represents total 2015 urban electricity consumption.

Table 5.2: Ten urban areas consuming most electricity in 2015.

Urban area	Consumption (Mwh)	Population
New York-Newark-Jersey City, NY-NJ-PA	142241597	20118063
Chicago-Naperville-Elgin, IL-IN-WI	84860058	9532569
Washington-Arlington-Alexandria, DC-VA-MD-WV	65037566	6078469
Los Angeles-Long Beach-Anaheim, CA	59938373	13268828
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	57474413	6062303
Riverside-San Bernardino-Ontario, CA	40610356	4475437
Boston-Cambridge-Newton, MA-NH	36577646	4766755
Seattle-Tacoma-Bellevue, WA	36455165	3727097
Minneapolis-St. Paul-Bloomington, MN-WI	32393762	3518252
San Francisco-Oakland-Hayward, CA	28057153	4642227

Table 5.3: Ten urban areas consuming least electricity in 2015.

Urban area	Consumption (Mwh)	Population
Sioux Falls, SD	771264	251889
Las Cruces, NM	759884	213567
Santa Fe, NM	744609	147708
Farmington, NM	736896	118701
Hot Springs, AR	642864	97154
Coeur d'Alene, ID	575655	150364
Rapid City, SD	568878	144059
Pocatello, ID	546068	83911
Missoula, MT	278057	113982
Great Falls, MT	270051	82118

5.3.2 Effect of urbanization on electricity consumption

The scaling of electricity consumption was estimated using equation 5.5b, where E_i is the estimated electricity consumption in each of 377 MSAs, N_i is the corresponding population in 2015 and β is the scaling exponent. The regression was found to be significant at 95% confidence level and the model explained 79% of the variance ($R^2 = 0.79$, $F(1,375)1404$, $p<0.0001$). The scaling exponent indicated an increase of 77% (SE $\pm 2.04\%$, p<0.0001) in electricity consumption corresponding to 100% increase in population. The regression fit is represented in Figure 5.5 (blue line), while the theoretical $\beta = 1$ (black dashed)and $\beta = 7/6$ (red dashed) lines are included in the graph for distinction the observed pattern from the theoretical ones.

Figure 5.5: Scaling of urban electricity consumption and population, MSAs 2015.

5.4 Discussion

In this work, the expected increase in electricity consumption in the United States, in response to urbanization was analyzed. Such studies are severely hampered due to lack of suitable data. To overcome this impediment, an electricity consumption model using a combination of five independent variables, and the 2015 state level electricity sales data as a proxy for consumption was developed. The estimates were validated against the 2015 California data. This external validation showed an average relative error of 12%. The subsequent scaling analysis of electricity consumption with urbanization indicated a 77% increase in electricity consumption corresponding to each doubling of urban population. In the following paragraphs summarize the research findings and their implications for urban sustainability.

Estimation of urban electricity consumption: In the first part of this work, an urban electricity consumption model using four crucial drivers of electricity consumption, i.e. climate, population, industrial activities, economic standards, was developed. The

electricity sales, used as a proxy of consumption, was only reported for the states. Thus, appropriate independent variables, that were also available at the county levels, were required for downscaling the state level consumptions to urban levels. Moreover, it was intended that the analysis approach be sufficiently generalizable for other parts of the world. Which further dictated the variable selection process in this work. Past research in the United States have utilized California data alone for such analysis [65, 189]. The validation from [189] showed an 11% mean error against the state level data. Which is comparable to our results with California urban areas. However, it is should be noted that the state of California is the second most energy efficient state in the nation [46]. Hence, California energy dynamics may stand in stark contrast with some of the other states, and using this state specific data in model development may result in unwanted bias. In this study, state level data and linear regression based model was used to downscale the consumption to urban levels. California statistical data was used for an unbiased external validation of the downscaled estimates, as this is the only statistical data available at this spatial scale. Availability of similar data from other states would have helped in more robust validation. Some estimation errors were observed for the smaller MSAs that returned higher relative errors (see figure 5.4). This is an artifact of the model which was used to downscale the estimates, thus the predictive accuracy decays for new data points further away from the initial data points, as in the case of the smaller MSAs. Similar to this patterns, [189] also indicated narrower error margins for metropolitan counties, indicating the increasing error over less urbanized areas. Therefore, while this model was found to be effective for large urbanized areas, this approach is not applicable for semi and non-urban areas. To ensure any estimation errors for individual data points are are minimized to the fullest possible extend, correction factors at state levels were used to derive normalized county level electricity consumptions (see equation 5.3c). This part of the analysis thus demonstrates that urban level electricity consumption may be estimated using geospatial data modeling approach, and overcome the data scarcity problem.

Scaling of urban electricity consumption: In the second part of this study, the urban electricity consumption estimates were used to study urban scaling of electricity consumption in the United States. The analysis showed a sub-linear scaling of urban

electricity consumption in the United States, which is quite in contrary to the past work suggesting linear [111] or mildly super-linear [88, 13] scaling. The results presented here suggest a 77% increase in urban electricity consumption corresponding to 100% increase in population. From the perspective of economies of agglomeration, instead of return to scale or increasing returns to scale, the observed pattern indicates a economies of scale at 23%. This may be interpreted as an evidence that larger cities indeed consume less electricity per person than the smaller ones. This could be due to their technological advancements, socio-economic conditions, design and infrastructural characteristics, and governing policy. The difference between scaling exponents from past work could also be attributed to the differences in geographies and associated socio-economic processes. Thus, further analysis of these patterns is required to pinpoint the factors behind this observed efficiency. Identification and understanding of urban attributes responsible for such economies of scale may help in developing a global taxonomy of cities, addressing one of the key research areas of sustainable urban science [166]. The evidence of efficiency is nonetheless an encouraging indication for the role of cities in combating the negative consequences of urbanization. However, as a word of caution, it needs to be pointed out that any evidence of efficiencies need to be interpreted with respect to other associated evidences and not be used as a justification for further populating the cities. Urban growth driven by efficiency has theoretically been observed to reach a carrying capacity threshold, beyond which the growth ceases [13]. This is of great implications for sustainable urbanization, urban growth is inevitable but the trajectories of future urbanization can be well understood so as to align the desired sustainability goals with present and future trends.

Bounds of urbanization potential: A plot of urban population density (for detail see methods section) and the estimate per-capita electricity consumption reveals the limits of such economies of scale (see figure 5.6). A loess curve has been fit to the data to assist the reader to follow the trend in data. It may be seen in the plot that maximum efficiency is gained when urban density increases from 200 to 800 persons/sq. km, resulting in capita electricity consumption reduction from about 22 Mwh/year/person to 9 Mwh/year/person. Between 800 and 2000 persons/sq. km the gain is only about 4 Mwh/year/person. Beyond

the 2000 persons/sq. km mark, the efficiency rapidly diminishes. Similar relationship with urban density has been documented for several related areas such as household carbon footprint [100]. High urban density has adverse effects on urban heat islands [147], less access to green space and safety [40]. Given this fundamental nature of urban systems, hypothesized by earlier work [13]. These efficiency thresholds may be different for other urban systems, but it is a fundamental property. Thus, future energy sustainability plans need to be aware of these limitations and aim for the gains within these specific limits.
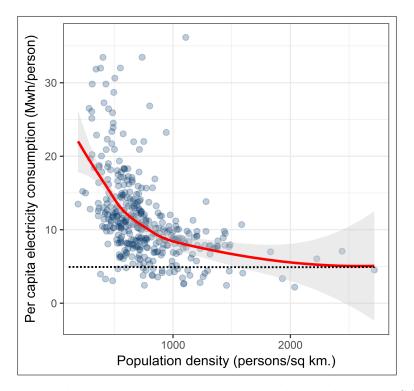
Figure 5.6: Per capita electricity consumption and population density, MSAs in the United States 2015. The black dashed line at $y \approx 5$ indicates the lower limit of per-capita electricity consumption.

To point out the limitations of this work, the weakness of the electricity estimation model over semi-urban and non-urban areas needs to be mentioned first. Due to the lack of urban energy data on urban scales, a linear model to downscale energy consumption at county levels. Linear models have been widely used in such conditions among other common regression methods [193]. However, as the model was calibrated using state level data, smaller counties are much beyond the data range which caused higher estimation error. Similar analysis using actual data, if it becomes available, could reveal specific insights not obtained from

regression predictions.Secondly, this study did not include any other geographic variability other than climate, however other specific geographical variables may be included in future work. Indicators for local energy policies, which can be a strong determinant, was not included in this work.

5.5 Conclusion

This work makes two significant contributions. First, a simple geospatial data driven model to downscale urban electricity consumption from state levels to urban levels was developed for the United States. The validation results showed acceptable results, coparable with past work. This shows an approach to overcome the scarcity of statistical data in urban energy consumption. Second, the expected change in electricity consumption in response to urbanization was explored. The evidences of economies of scale in urban electricity consumption was observed, with an efficiency of around 23%. While this is an encouraging sign of prospects of leveraging urbanization to lower per capita energy consumption, however this should not be considered as a limitless opportunity for efficiency. The subsequent analysis also shows that the efficiency slows down and eventually diminishes at certain population density thresholds in the United States urban system. This work identifies a trend but further work in this area need to be carried out to identify the specific factors leading to the observed energy efficiency. This can lead to development of taxonomy of cities with respect to energy consumption and help sustainable urban science. Also, availability of energy consumption data on urban scales will help in identification of finer trends that are not discernible through estimation models.

Chapter 6

Conclusion

Urbanization is a hallmark of the human civilization in post-industrial revolution era. Urban areas act as the massive socio-economic engines of the society. Given their diverse roles, urban areas can be expected to grow at the present rate in the foreseeable future. However, due to the adverse effects of urbanization on the environment and human well-being, it has become critical to find ways to ensure that the socio-economic growth follows an environment and natural resource-friendly manner. Worldwide, 75% of the total primary energy supply is consumed in urban areas which causes the release of around 70% of the world's energy-related carbon-di-oxide emissions. Electricity is the fastest growing energy fuel by end-usage, and it is expected to contribute to about 40% of the growth in total energy end-usage by 2040. Given the increasing market share of the electricity, ensuring access to reliable and long-lasting electricity supply while minimizing the environmental footprint of electricity generation pathways are critical energy sustainability challenges. Cities are dynamic entities, where the interplay of social, economic, and technical parameters shape the electricity consumption profiles. Most of the past studies have focused on national and regional level correlation analysis between urbanization and electricity consumption, which notes an overall positive relationship between them. However, these studies do not elucidate much detail about the urban systems of these countries. Scholars have argued that the struggle for sustainability will either be won or lost in urban areas. Thus energy-related studies need to be focused on the city scale.

One of the major hindrance to urban level energy analysis is the lack of appropriate data. The energy statistics mechanisms, all around the world, are still nation or region-centric. This practice severely limits the potential of implementation and evaluation of energy efficiency measures, and future capacity building. While a there are a few initiatives starting to shape up, such as the Energy Data initiative in the United States, or the World Bank's Open Energy Data initiative, currently operating in African cities of Accra, Ghana and Nairobi, Kenya. However, the availability of energy data across an urban parameter with uniform detail has not yet been reached. Under these circumstances, this dissertation work has made an effort to explore the critical question of whether geospatial data-driven approaches can fill this data void and answer some of the pressing questions on urban energy sustainability. In order to address this broad research question, three research objectives were identified and answered in this dissertation. The following paragraphs summarize the research questions, and the insights gathered from their exploration.

Urban influence on the global arena is a well-established fact; this necessitates identification and measurements of urban areas for any analysis of their impacts. Satellite image derived urban area maps are far more efficient than the traditional surveys due to high temporal granularity and low costs. Yet, these datasets severely diverge in the way they depict and quantify of urban lands. To understand the effects of urbanization on environmental as well as socio-economic systems, accurate and unambiguous quantification of urban areas at multiple spatial scales is a necessary first step. As the urban area maps derived from high-resolution satellite images offer much more spatial detail than their coarse and low-resolution counterparts, the performance of such high-resolution datasets was analyzed at multiple scales in the third chapter of this dissertation. The analysis, conducted in Egypt and Taiwan at national and regional scales, indicate that the estimated urban extents severely fluctuate amongst these datasets. The variations also display an increasing pattern from national to regional scales, with the regional level variations reaching as high as 900% in Egypt and about 550% in Taiwan. The next step in this analysis was to explore the relationship between the morphological characteristics of the urban landscape and the mutual agreements of these datasets. The agreement-disagreement maps, for these two countries indicated that the agreements amongst these datasets regarding detection of urban lands

quickly diminishes from the urban core to peripheral areas. Landscape morphology analysis highlights that complexity and disaggregation of urban patches contribute to increased inter-dataset disagreements; while increasing contiguity results in higher inter-dataset agreements. Since the high complexity and disaggregation of landscape are more prevalent in peri-urban areas as well as informal settlements within urban infilling zones; these areas may therefore not be identified by all datasets alike. Lack of spatial resolution in the coarse and medium-resolution datasets will, therefore, hinder the accurate detection of these vital areas. These areas may lack electricity access, or consume much more electricity than expected due to the nature of the informal economy. Their non-detection may simply omit one of the most vibrant components of moder urbanization from any analysis. The insights from this analysis firmly highlight the importance of high-resolution urban mapping for energy-related studies. Continued application of these high-resolution datasets could also lead to a convergence in global urban area estimations.

The problem of urban data scarcity is much severe in the developing and under-developed nations. The lack of information severely hinders analysis of the lack of electricity access, and the management and planning of energy systems. Thus, an attempt was made in the fourth chapter to utilize satellite image derived proxy measures to understand electricity consumption patterns in Johannesburg, Sana'a, and Ndola, which represents three cities from the developing nation. As the high-resolution satellite images are capable of recording the size, shape, and orientation of urban structures, this has been exploited by past studies to classify urban structures based on their visual characteristics on satellite images. These capabilities were used to develop a generalizable urban taxonomy, consisting of five classes of settlements. These classes roughly represent high-density, small sized, and non-orthogonally laid out residential structures which are most likely informal; low to medium density residential and commercial buildings; low density urban and suburban houses; large buildings which are most likely of commercial and industrial usage; and lastly the rural houses. This settlement inventory was then linked to nighttime lights emission from these areas. The nighttime lights are a well-known proxy of socio-economic activities including electricity consumption. Regression analyses indicate that in Johannesburg the commercial and industrial buildings consume the most electricity, while in Ndola it is the medium

to low density, medium sized urban buildings that are associated with maximum light emission. However, in the case of Sana'a, it was observed that both the large commercial and administrative buildings as well as small and likely informal settlements that correlate with maximum light emission. This trend is not unexpected given the current political unrest in Yemen that led to the significant internal displacement of people. However, this does indicate that high-resolution image derived metrics can identify minute patterns within urban areas. The results show an association between energy consumption, which is approximated using the nighttime lights as a proxy, and building stocks in these three cities. The planners can use them to promote energy access and efficiency, identify the areas requiring attention in an informed manner. It was also observed that there were some pockets of settlements, especially in Ndola and Sana'a, where nighttime lights data did not detect any light emission. The is a likely indication of lack of access to electricity in those areas, which could again help the planning authorities in focusing their attention and resources more accurately and effectively.

Beyond the immediate issue of the identification of urban areas and patterns of urban electricity consumption, this dissertation work was also aimed at exploring the role of geospatial data-driven modeling to address energy sustainability questions. Thus, the effect of urbanization on electricity consumption in the urban system of the US was addressed in the fifth chapter of this work. The United States is one of the most urbanized nations in the world and the second largest consumer of electricity. To overcome the lack of urban electricity data, a predictive model was developed as a function of local climate, economy, industrial activity, and population. The census defined metropolitan statistical areas were chosen to represent urban areas, which was found to be most consistent with the functional definition of urban areas provided by past studies. The model estimates were found to have an average error of 12% upon validation against the California data, which is the only available county level statistical data in the US. The national electricity data was only available at the state level. This effectively forced the regression model to make predictions in the data range beyond its original configuration, which contributed to some of the errors. The urban population size is a well-known indicator of urbanization, the analysis between urbanization and electricity consumption revealed that in the United States, each doubling of the urban population would

lead to a 77% increase in electricity consumption. This evidence of 23% energy savings is an indication that larger cities could help to reduce the per capita consumption levels. This finding that urbanization could be used to lower energy consumption is vital for urban sustainability. However, the next question that comes with this finding is how large the urban areas can grow? To answer this, the focus was turned to urban population density, a crucial factor in determining urban energy profiles. Subsequent analysis shows that by increasing population density from 200 to 800 persons/sq. km urban areas can lower per capita annual electricity consumption from 22 to 9 MWh. However, between 800 to 200 persons/sq. km., the potential for electricity efficiency is only 4 MWh/person/Year. Beyond the density of 2,000 persons/sq. km., there is almost no efficiency to be gained. This is aligned with the theoretical hypothesis that efficiency driven urban growth eventually comes to a limiting stage [13]. The results highlight the potential of urbanization to alleviate energy sustainability challenges along with the limits of these opportunities, such insights may be used by urban planners and energy policy makers in their energy sustainability efforts.

6.1 Uncertainties

The main focus of this work pertains to utilization of geospatial datasets and modeling to support urban energy related studies. The initial part of this work addresses the issue of mutual consistencies of global and regional human settlement datasets. As these datasets mark the human presence on the face of the Earth, accuracies of such measurements assume paramount importance. There is an urgent need for convergence in these measures, and such goal can only be achieved through multi-scale analysis amongst datasets. While LandScan SL employs a rigorous manual quality control, this dataset can be thought of as reference quality. Thus, the analysis of mutual consistencies in the first part of this study was limited to comparison and contrast among the datasets. However, not all the datasets employ similar quality control measures which can introduce random elements into such analysis. The application of satellite image derived human settlement inventories have been indicated as a potential metric for assessing urban socio-economic dynamics in the past. The work presented in the fourth chapter presents a five class settlement typology for three cities in the

developing world, namely Johannesburg, South Africa; Ndola, Zambia; and Sana'a, Yemen. However, more cities need to be included into such analysis to develop a more generalizable settlement inventory purely based on the data. The small sample size of the rural type settlement in Ndola, Zambia alongwith non-detection of much lights from these buildings have resulted in statistically insignificant coefficients. Such uncertainties can be avoided by including more cities from a country into the analysis. Also, the relationship between nighttime lights and human activities are mediated by cultural and economic factors. It is assumed, however, that the effects of such non-stationarity is minimal within a particular city. Lastly, as the nationwide electricity data was only available at state levels, the regression model in fifth chapter was used to make out-of-sample predictions. This has resulted in some systematic errors where the smaller MSA's were overestimated.

6.2   Limitations

While discussing the limitations of this work, the following points needs to be mentioned:

- Firstly, the LandScan SL is going through a continuous maturation process. Thus, its availability as a global coverage is presently being worked on. Availability of a global coverage will allow testing the potential of this dataset alongside others over more geographies, and at region, nation, to continental scale.

- In the intra-urban electricity consumption pattern analysis, the nighttime lights data was found deficient in capturing lights from some parts of the three cities, especially in Sana'a, Yemen, and Ndola, Zambia. This could either be due to lack of access or due to a combination of late overpass time of VIIRS satellite and socio-economic conditions leading to lights being turned off during data acquisition. The reliability of power grids might also have played a role. The coefficients of determination in the regression developed for Johannesburg was found to be much weaker than the other two. This may be interpreted as is an effect of highly mixed land-use in some areas of Johannesburg city. The mixed land-use may result in nighttime lights signals associated with an economic activity that does not always match with the physical appearance

113

driven taxonomy of the buildings. For example, much higher light emission from a building that houses both residential and commercial activities. These factors caused likely confounded the data and subsequent analysis.

- The limitations in urban electricity estimation model lies in its weakness over semi-urban and non-urban areas. Due to the lack of energy data at urban scales, the linear model was used to downscale energy consumption at urban levels. An Analysis using actual urban statistical data could reveal specific insights not obtained from regression predictions. Also, due to the limited sample size (n=50) which restricts the number of independent variables in the model, no other geographic variable other than climate was included in the study. However, other location specific geographical variables may be included in future work to better incorporate geographical effects on electricity consumption. Policy indicators could also be used to give model more flexibility. Urban level statistical datasets, if they become available, may help inclusion of many more pertinent variables in future research.

6.3 Major contributions

This dissertation work highlights the potential geospatial data, especially high-resolution image derived datasets, and modeling in understanding urban electricity consumption patterns. The major contributions made in this dissertation are:

1. This is work highlights mutual disagreements in the identification of urban areas and their measurements among global and regional datasets, at multiple scales; and illustrates how urban landscape morphology affects the level of disagreements. This insight could be used to choose appropriate urban areas datasets for subsequent analysis by future researchers. The results also establish the strength of high-resolution mapping in urban pattern detection.

2. A method has been presented that highlights the use of complimentary high and low-resolution satellite datasets in understanding electricity consumption patterns in data starved regions. Such methods could be rapidly deployed by urban planners to

develop and visualize an intra-urban energy consumption map, to be used for planning purposes.

3. An urban electricity consumption model has been presented. The subsequent analysis using model estimates reveal that the urban system in the US provides an opportunity for electricity efficiency. The limits within which such efficiency can be realized has also been pointed out. These insights could prove valuable urban energy sustainability

6.4   Future research directions

The future research with regards to the present study will be focused in the following directions:

1. Application-focused comparisons of human settlement datasets need to be carried out. The outcomes can be checked against known data to identify ideal data for specific applications. This may make a strong case for high-resolution data to other domains.

2. Availability of micro-census data can be highly beneficial to develop empirical relationships between settlement types and their socio-economic characteristics. Through analysis of different economies more generalizable empirical relationships can be developed.

3. Urban scaling analysis needs to be done using actual statistical data; strong cases need to be made for collection and publication of such data. Estimates can provide general direction, but subtle variations may get masked due to generalizations and assumptions in the model.

4. Urban socio-economic, technological characteristics, and existing energy policies need to be analyzed against their energy profiles to identify the crucial variables for sustainability and to develop specific sustainability plans.

Bibliography

[1] Aguilar, M. A., Fernández, A., Aguilar, F. J., Bianconi, F., and Lorca, A. G. (2016). Classification of urban areas from geoeye-1 imagery through texture features based on histograms of equivalent patterns. *European Journal of Remote Sensing*, 49(1):93–120. 70

[2] Albert, A., Strano, E., Kaur, J., and Gonzalez, M. (2018). Modeling urbanization patterns with generative adversarial networks. *arXiv preprint arXiv:1801.02710.* 24

[3] Alves, L. G., Ribeiro, H. V., and Mendes, R. S. (2013). Scaling laws in the dynamics of crime growth rate. *Physica A: Statistical Mechanics and its Applications*, 392(11):2672–2679. 34

[4] Amaral, S., Câmara, G., Monteiro, A. M. V., Quintanilha, J. A., and Elvidge, C. D. (2005). Estimating population and energy consumption in brazilian amazonia using dmsp night-time satellite data. *Computers, Environment and Urban Systems*, 29(2):179–195. 26, 36, 65

[5] Anderson, B., Lin, S., Newing, A., Bahaj, A., and James, P. (2017a). Electricity consumption and household characteristics: Implications for census-taking in a smart metered future. *Computers, Environment and Urban Systems*, 63:58–67. 8

[6] Anderson, B., Lin, S., Newing, A., Bahaj, A., and James, P. (2017b). Electricity consumption and household characteristics: Implications for census-taking in a smart metered future. *Computers, Environment and Urban Systems*, 63:58–67. 64

[7] Bailey, P. E., Keyes, E. B., Parker, C., Abdullah, M., Kebede, H., and Freedman, L. (2011). Using a gis to model interventions to strengthen the emergency referral system for maternal and newborn health in ethiopia. *International Journal of Gynecology & Obstetrics*, 115(3):300–309. 40

[8] Banzhaf, E. and Hofer, R. (2008). Monitoring urban structure types as spatial indicators with cir aerial photographs for a more effective urban environmental management. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1(2):129–138. 9, 30, 64

[9] Bazilian, M., Rogner, H., Howells, M., Hermann, S., Arent, D., Gielen, D., Steduto, P., Mueller, A., Komor, P., Tol, R. S., et al. (2011). Considering the energy, water and food nexus: Towards an integrated modelling approach. *Energy Policy*, 39(12):7896–7906. 2

[10] Bernard, J.-T., Bolduc, D., and Yameogo, N.-D. (2011). A pseudo-panel data model of household electricity demand. *Resource and Energy Economics*, 33(1):315–325. 36

[11] Bettencourt, L., Lobo, J., and Youn, H. (2013). The hypothesis of urban scaling: formalization, implications and challenges. *arXiv preprint arXiv:1301.5919.* 34, 94

[12] Bettencourt, L. M. (2013). The origins of scaling in cities. *science*, 340(6139):1438–1441. 34, 35, 94, 98

[13] Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences*, 104(17):7301–7306. 34, 35, 100, 105, 106, 112

[14] Bettencourt, L. M. A., Lobo, J., Strumsky, D., and West, G. B. (2010). Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLOS ONE*, 5(11):1–9. 98, 99, 100

[15] Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1):2–16. 20

[16] Boyd, D. S. and Foody, G. M. (2011). An overview of recent remote sensing and gis based research in ecological informatics. *Ecological Informatics*, 6(1):25–36. 23

[17] Boyle, S. A., Kennedy, C. M., Torres, J., Colman, K., Pérez-Estigarribia, P. E., and Noé, U. (2014). High-resolution satellite imagery is an important yet underutilized resource in conservation biology. *PloS one*, 9(1). 60

[18] Bureau, U. S. C. (2017). County business patterns. https://www.census.gov/programs-surveys/cbp/about.html. Accessed: 2017-11-12. 91

[19] Burney, N. A. (1995). Socioeconomic development and electricity consumption a cross-country analysis using the random coefficient method. *Energy Economics*, 17(3):185–195. 31

[20] Cao, X., Wang, J., Chen, J., and Shi, F. (2014). Spatialization of electricity consumption of china using saturation-corrected dmsp-ols data. *International Journal of Applied Earth Observation and Geoinformation*, 28:193–200. 65

[21] Carreón, J. R. and Worrell, E. (2018). Urban energy systems within the transition to sustainable development. a research agenda for urban metabolism. *Resources, Conservation and Recycling*, 132:258–266. 4, 88

[22] Central Statistical Office of Zambia (2014). Accessed: 2018-06-18. 69

[23] Chand, T. K., Badarinath, K., Elvidge, C., and Tuttle, B. (2009). Spatial characterization of electrical power consumption patterns over india using temporal dmsp-ols night-time satellite data. *International Journal of Remote Sensing*, 30(3):647–661. 26, 65

[24] Channan, S., Collins, K., and Emanuel, W. (2014). Global mosaics of the standard modis land cover type data. 21

[25] Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594. 35

[26] Cheriyadat, A., Bright, E., Potere, D., and Bhaduri, B. (2007a). Mapping of settlements in high-resolution satellite imagery using high performance computing. *GeoJournal*, 69(1-2):119–129. 20, 29, 40, 42, 72

[27] Cheriyadat, A., Bright, E., Potere, D., and Bhaduri, B. (2007b). Mapping of settlements in high-resolution satellite imagery using high performance computing. *GeoJournal*, 69(1):119–129. 23

[28] CIA (2018a). World factbook: Egypt. https://www.cia.gov/library/publications/the-world-factbook/geos/eg.html. Accessed: 2018-09-06. 45

[29] CIA (2018b). World factbook: Taiwan. https://www.cia.gov/library/publications/the-world-factbook/geos/tw.html. Accessed: 2018-09-06. 46

[30] CIESIN, IFPRI, and CIAT (2011). Global rural-urban mapping project, version 1 (grumpv1): Urban extents grid. 21

[31] Cohen, B. (2004). Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts. *World development*, 32(1):23–51. 60

[32] Congalton, R. G., Gu, J., Yadav, K., Thenkabail, P., and Ozdogan, M. (2014). Global land cover mapping: A review and uncertainty analysis. *Remote Sensing*, 6(12):12070–12093. 46

[33] Cooner, A. J., Shao, Y., and Campbell, J. B. (2016). Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 haiti earthquake. *Remote Sensing*, 8(10):868. 72

[34] Corbane, C., Pesaresi, M., Politis, P., Syrris, V., Florczyk, A. J., Soille, P., Maffenini, L., Burger, A., Vasilev, V., Rodriguez, D., Sabo, F., Dijkstra, L., and Kemper, T. (2017). Big earth data analytics on sentinel-1 and landsat imagery in support to global human settlements mapping. *Big Earth Data*, 1(1-2):118–144. 23

[35] Creutzig, F., Baiocchi, G., Bierkandt, R., Pichler, P.-P., and Seto, K. C. (2015). Global typology of urban energy use and potentials for an urbanization mitigation wedge. *Proceedings of the National Academy of Sciences*, 112(20):6283–6288. 3, 4

[36] D., S. and Bourdeau, P. (1995). Europe's environment. the dobris assessment. 3

[37] Dahl, C. and Erdogan, M. (1994). Oil demand in the developing world: lessons from the 1980s applied to the 1990s. *The Energy Journal*, pages 69–85. 30

[38] Danko, D. M. (1992). The digital chart of the world project. *Photogrammetric engineering and remote sensing*, 58(8):1125–1128. 22

[39] Darby, S. et al. (2006). The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*, 486(2006):26. 64

[40] Dempsey, N., Brown, C., and Bramley, G. (2012). The key to sustainable urban development in uk cities? the influence of density on social sustainability. *Progress in Planning*, 77(3):89–141. 106

[41] Dobbs, R., Smit, S., Remes, J., Manyika, J., Roxburgh, C., and Restrepo, A. (2011). Urban world: Mapping the economic power of cities. Technical report, McKinsey Global Institute. 2

[42] Dobson, J., Bright, E., Coleman, P., and Bhaduri, B. (2003). Landscan: a global population database for estimating population at risk. In Messev, V., editor, *Remotely Sense Cities*, pages 267–280. CRC Press. 21

[43] Dodman, D. (2009). Blaming cities for climate change? an analysis of urban greenhouse gas emissions inventories. *Environment and Urbanization*, 21(1):185–201. 5

[44] Du, X., Leinenkugel, P., Guo, H., and Kuenzer, C. (2017). Open earth observation data for measuring anthropogenic development in coastal zones at continental scales. In *AGU Fall Meeting Abstracts*. 24

[45] Dugmore, K., Furness, P., Leventhal, B., and Moy, C. (2011). Information collected by commercial companies: what might be of value to ons. *International Journal of Market Research*, 53(5):619–650. 20

[46] Eldridge, M., Neubauer, M., York, D., Vaidyanathan, S., Chittum, A., and Nadel, S. (2008). The 2008 state energy efficiency scorecard. 104

[47] Ellis, E. C. and Ramankutty, N. (2008). Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, 6(8):439–447. 20

[48] Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., and Ghosh, T. (2017). Viirs night-time lights. *International Journal of Remote Sensing*, 38(21):5860–5879. 66

[49] Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., and Davis, C. W. (1997a). Relation between satellite observed visible-near infrared emissions, population,

economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6):1373–1379. 26

[50] Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., and Davis, C. W. (1997b). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6):1373–1379. 36

[51] Elvidge, C. D., Baugh, K. E., Sutton, P. C., Bhaduri, B., Tuttle, B. T., Ghosh, T., Ziskin, D., and Erwin, E. H. (2011). Who's in the darksatellite based estimates of electrification rates. *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*, 250:211–224. 26, 27

[52] Elvidge, C. D., Baugh, K. E., Zhizhin, M., and Hsu, F.-C. (2013). Why viirs data are superior to dmsp for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network*, 35:62–69. 26, 27, 65, 66

[53] Elvidge, C. D., Cinzano, P., Pettit, D. R., Arvesen, J., Sutton, P., Small, C., Nemani, R., Longcore, T., Rich, C., Safran, J., Weeks, J., and Ebener, S. (2007). The nightsat mission concept. *International Journal of Remote Sensing*, 28(12):2645–2670. 26

[54] Elvidge, C. D., Ziskin, D., Baugh, K. E., Tuttle, B. T., Ghosh, T., Pack, D. W., Erwin, E. H., and Zhizhin, M. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, 2(3):595–622. 27

[55] Erkin, Z. and Tsudik, G. (2012). Private computation of spatial and temporal power consumption with smart meters. In *Applied Cryptography and Network Security*, pages 561–577. Springer Berlin Heidelberg. 8

[56] Esch, T., Taubenböck, H., Roth, A., Heldens, W., Felbier, A., Thiel, M., Schmidt, M., Müller, A., and Dech, S. (2012). Tandem-x mission - new perspectives for the inventory and monitoring of global settlement patterns. *Journal of Applied Remote Sensing*, 6(1):061702–1. 23

[57] European Environment Agency (2017). Corine land cover 2017. `https://land.copernicus.eu/user-corner/publications/clc-flyer/at_download/file`. Accessed: 2018-08-29. 21

[58] Fang, S., Gertner, G., Wang, G., and Anderson, A. (2006). The impact of misclassification in land use maps in the prediction of landscape dynamics. *Landscape ecology*, 21(2):233–242. 23

[59] Feige, E. L. and Urban, I. (2008). Measuring underground (unobserved, non-observed, unrecorded) economies in transition countries: can we trust gdp? *Journal of Comparative Economics*, 36(2):287–306. 64

[60] Filippini, M. and Pachauri, S. (2004). Elasticities of electricity demand in urban indian households. *Energy policy*, 32(3):429–436. 36

[61] Fleiss, M., Kienberger, S., Aubrecht, C., Kidd, R., and Zeil, P. (2011). Mapping the 2010 pakistan floods and its impact on human life: A post-disaster assessment of socioeconomic indicators. *Geoinformation for Disaster Management (GI4DM), Antalya, Turkey, CD-ROM*. 40

[62] Fogel, I. and Sagi, D. (1989). Gabor filters as texture discriminator. *Biological cybernetics*, 61(2):103–113. 71

[63] Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., et al. (2005). Global consequences of land use. *science*, 309(5734):570–574. 20

[64] Foody, G. M. (2004). Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, 70(5):627–633. 46

[65] Fragkias, M., Lobo, J., and Seto, K. C. (2016). A comparison of nighttime lights data for urban energy research: Insights from scaling analysis in the us system of cities. *Environment and Planning B: Planning and Design*, page 0265813516658477. 5, 8, 30, 33, 35, 36, 88, 94, 104

[66] Fragkias, M., Lobo, J., Strumsky, D., and Seto, K. C. (2013). Does size matter? scaling of co2 emissions and us urban areas. *PLoS One*, 8(6):e64727. 34

[67] Freire, S., Kemper, T., Pesaresi, M., Florczyk, A., and Syrris, V. (2015). Combining ghsl and gpw to improve global population mapping. In *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, pages 2541–2543. IEEE. 24

[68] Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X. (2010). Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114(1):168–182. 21, 41

[69] Gahegan, M. (2009). visual exploration and explanation in geography analysis with light. In *Geographic data mining and knowledge discovery*, pages 291–324. CRC Press. 11

[70] Gamba, P., Dell'Acqua, F., Lisini, G., and Trianni, G. (2007). Improved vhr urban area mapping exploiting object boundaries. *IEEE Transactions on Geoscience and Remote Sensing*, 45(8):2676–2682. 29

[71] Ghosh, T., L Powell, R., D Elvidge, C., E Baugh, K., C Sutton, P., and Anderson, S. (2010). Shedding light on the global distribution of economic activity. *The Open Geography Journal*, 3(1). 36

[72] Graesser, J., Cheriyadat, A., Vatsavai, R. R., Chandola, V., Long, J., and Bright, E. (2012). Image based characterization of formal and informal neighborhoods in an urban landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4):1164–1176. 9, 28, 29, 30, 64

[73] Grubler, A., Bai, X., Buettner, T., Dhakal, S., Fisk, D. J., Ichinose, T., Keirstead, J. E., Sammmer, G., Satterthwaite, D., Schulz, N. B., et al. (2012). Chapter 18 - urban energy systems. In *Global Energy Assessment - Toward a Sustainable Future*, pages 1307–1400. Cambridge University Press, Cambridge, UK and New York, NY, USA and the International Institute for Applied Systems Analysis, Laxenburg, Austria. 1, 2, 4, 5, 6, 7, 8, 33, 64, 88

[74] Güneralp, B. and Seto, K. C. (2008). Environmental impacts of urban growth from an integrated dynamic perspective: A case study of shenzhen, south china. *Global Environmental Change*, 18(4):720–735. 2

[75] Halicioglu, F. (2007). Residential electricity demand dynamics in turkey. *Energy economics*, 29(2):199–210. 32

[76] Hancke, G. P., Hancke Jr, G. P., et al. (2012). The role of advanced sensing in smart cities. *Sensors*, 13(1):393–425. 64

[77] Hay, S., Noor, A., Nelson, A., and Tatem, A. (2005). The accuracy of human population maps for public health application. *Tropical Medicine & International Health*, 10(10):1073–1086. 23

[78] He, C., Ma, Q., Liu, Z., and Zhang, Q. (2014). Modeling the spatiotemporal dynamics of electric power consumption in mainland china using saturation-corrected dmsp/ols nighttime stable light data. *International Journal of Digital Earth*, 7(12):993–1014. 65

[79] Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, 102(2):994–1028. 36

[80] Herold, M. (2009). Some recommendations for global efforts in urban monitoring and assessment from remote sensing. In Gamba, H. and Herold, M., editors, *Global mapping of human settlements: Experiences, Data Sets, and Prospects*, pages 11–23. CRC Press. 21, 22, 40

[81] Herold, M., Couclelis, H., and Clarke, K. C. (2005). The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment and Urban Systems*, 29(4):369 – 399. 53

[82] Herold, M., Groenestijn, A., Kooistra, L., Kalogirou, V., and Arino, O. (2011). Land cover cci user requirements document rev. 2. http://www.esa-landcover-cci.org/?q=webfm_send/46. Accessed: 2016-08-22. 21, 41, 42

[83] Hilbert, M. (2013). Big data for development: From information-to knowledge societies (ssrn scholarly paper no. id 2205145). *Social Science Research Network, Rochester, NY.* 8, 64

[84] Hinz, S. and Baumgartner, A. (2003). Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1-2):83–98. 72

[85] Hofmann, P., Strobl, J., Blaschke, T., and Kux, H. (2008). Detecting informal settlements from quickbird data in rio de janeiro using an object based approach. In *Object-based image analysis*, pages 531–553. Springer. 29, 64

[86] Holtedahl, P. and Joutz, F. L. (2004). Residential electricity demand in taiwan. *Energy economics*, 26(2):201–224. 32

[87] Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., and Megown, K. (2015). Completion of the 2011 national land cover database for the conterminous united states–representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, 81(5):345–354. 92

[88] Horta-Bernús, R., Rosas-Casals, M., and Valverde, S. (2010). Discerning electricity consumption patterns from urban allometric scaling. In *Complexity in Engineering, 2010. COMPENG'10.*, pages 49–51. IEEE. 35, 105

[89] Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., and Modi, V. (2012). Spatial distribution of urban building energy consumption by end use. *Energy and Buildings*, 45:141–151. 8, 63

[90] Hussey, K. and Pittock, J. (2012). The energy–water nexus: Managing the links between energy and water for a sustainable future. *Ecology and Society*, 17(1). 7

[91] International Energy Agency (2017). Energy access outlook 2017 - from poverty to prosperity. 5

[92] International energy Agency (2017a). Key world energy statistics. `https://www.iea.org/publications/freepublications/publication/KeyWorld2017.pdf`. Accessed: 2018-08-18. 3, 4, 87

[93] International energy Agency (2017b). World energy outlook 2017 - executive summary. `https://www.iea.org/Textbase/npsum/weo2017SUM.pdf`. Accessed: 2018-08-18. 87

[94] International Finance Corporation - The World Bank Group (2014). Ifc in the power sector. `https://www.ifc.org/wps/wcm/connect/b82f23fc-6285-43ef-9cd5-938f725fe604/PowerInfographic.pdf?MOD=AJPERES`. Accessed: 2018-08-22. 6

[95] Isaac, M. and van Vuuren, D. P. (2009). Modeling global residential sector energy demand for heating and air conditioning in the context of climate change. *Energy Policy*, 37(2):507 – 521. 89

[96] Jacobson, M. Z. and Ten Hoeve, J. E. (2012). Effects of urban surfaces and white roofs on global and regional climate. *Journal of climate*, 25(3):1028–1044. 41

[97] Jain, R., Kasturi, R., and Schunck, B. G. (1995). *Machine vision*, volume 5. McGraw-Hill New York. 70

[98] Jaturapitpornchai, R., Kasetkasem, T., Kumazawa, I., Rakwatin, P., and Chanwimaluang, T. (2015). A level-based method for urban mapping using npp-viirs nighttime light data. In *Information and Communication Technology for Embedded Systems (IC-ICTES), 2015 6th International Conference of*, pages 1–6. IEEE. 28, 65

[99] Jiang, L., Deng, X., and Seto, K. C. (2013). The impact of urban expansion on agricultural land use intensity in china. *Land use policy*, 35:33–39. 2

[100] Jones, C. and Kammen, D. M. (2014). Spatial distribution of us household carbon footprints reveals suburbanization undermines greenhouse gas benefits of urban population density. *Environmental science & technology*, 48(2):895–902. 106

[101] Jones, D. W. (1991). How urbanization affects energy-use in developing countries. *Energy policy*, 19(7):621–630. 31

[102] Karuaihe, S. (2013). The state of the economy: City of johannesburg. 67

[103] Kaufmann, R. K., Seto, K. C., Schneider, A., Liu, Z., Zhou, L., and Wang, W. (2007). Climate response to rapid urban growth: evidence of a human-induced precipitation deficit. *Journal of Climate*, 20(10):2299–2306. 2

[104] Keirstead, J., Jennings, M., and Sivakumar, A. (2012). A review of urban energy system models: Approaches, challenges and opportunities. *Renewable and Sustainable Energy Reviews*, 16(6):3847–3866. 8

[105] Kennedy, C., Cuddihy, J., and Engel-Yan, J. (2007). The changing metabolism of cities. *Journal of industrial ecology*, 11(2):43–59. 33

[106] Kirches, G., Brockman, C., Boettchar, M., Peters, M., Bontemps, S., Lamarche, C., Schlerf, M., Santoro, M., and Defourny, P. (2014). Land cover cci product user guide version 2. http://www.esa-landcover-cci.org/?q=webfm_send/84. Accessed: 2016-04-20. 42

[107] Kit, O. and Lüdeke, M. (2013). Automated detection of slum area change in hyderabad, india using multitemporal satellite imagery. *ISPRS journal of photogrammetry and remote sensing*, 83:130–137. 20

[108] Kit, O., Lüdeke, M., and Reckien, D. (2012). Texture-based identification of urban slums in hyderabad, india using remote sensing data. *Applied Geography*, 32(2):660–667. 72

[109] Klotz, M., Kemper, T., Geiß, C., Esch, T., and Taubenböck, H. (2016). How good is the map? a multi-scale cross-comparison framework for global settlement layers: Evidence from central europe. *Remote Sensing of Environment*, 178:191–212. 22, 23, 24

[110] Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World map of the köppen-geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263. 67, 68, 69

[111] Kühnert, C., Helbing, D., and West, G. B. (2006). Scaling laws in urban supply networks. *Physica A: Statistical Mechanics and its Applications*, 363(1):96–103. 35, 105

[112] Lariviere, I. and Lafrance, G. (1999). Modelling the electricity consumption of cities: effect of urban density. *Energy economics*, 21(1):53–66. 31

[113] Lenzen, M., Wier, M., Cohen, C., Hayami, H., Pachauri, S., and Schaeffer, R. (2006). A comparative multivariate analysis of household energy requirements in australia, brazil, denmark, india and japan. *Energy*, 31(2-3):181–207. 31

[114] Letu, H., Hara, M., Yagi, H., Naoki, K., Tana, G., Nishio, F., and Shuhei, O. (2010). Estimating energy consumption from night-time dmps/ols imagery after correcting for saturation effects. *International Journal of Remote Sensing*, 31(16):4443–4458. 27, 65

[115] Liddle, B. and Lung, S. (2014). Might electricity consumption cause urbanization instead? evidence from heterogeneous panel long-run causality tests. *Global Environmental Change*, 24:42–51. 32

[116] Linard, C., Gilbert, M., Snow, R. W., Noor, A. M., and Tatem, A. J. (2012). Population distribution, settlement patterns and accessibility across africa in 2010. *PloS one*, 7(2). 40

[117] Liu, X. and Wang, D. (2002). A spectral histogram model for texton modeling and texture discrimination. *Vision Research*, 42(23):2617–2634. 73

[118] Liu, Y. (2009). Exploring the relationship between urbanization and energy consumption in china using ardl (autoregressive distributed lag) and fdm (factor decomposition model). *Energy*, 34(11):1846–1854. 32

[119] Lobo, J., Bettencourt, L. M., Strumsky, D., and West, G. B. (2013). Urban scaling and the production function for cities. *PLoS One*, 8(3):e58407. 34

[120] Lofman, D., Petersen, M., and Bower, A. (2002). Water, energy and environment nexus: The california experience. *International Journal of Water Resources Development*, 18(1):73–85. 2

[121] Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W. (2000). Development of a global land cover characteristics database and igbp discover from 1 km avhrr data. *International Journal of Remote Sensing*, 21(6-7):1303–1330. 21

[122] Ma, T., Zhou, C., Pei, T., Haynie, S., and Fan, J. (2014). Responses of suomi-npp viirs-derived nighttime lights to socioeconomic activity in chinas cities. *Remote Sensing Letters*, 5(2):165–174. 28, 65

[123] Marmol, U. (2011). Use of gabor filters for texture classification of airborne images and lidar data. *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, 22. 70

[124] Massetti, E., Brown, M. A., Lapsa, M., Sharma, I., Bradbury, J., Cunliff, C., and Li, Y. (2017). Environmental quality and the u.s.power sector: Air quality, water quality, land use and environmental justice. Technical report, Oak Ridge National laboratory. 7, 88

[125] Mathur, O. (2013). Urban poverty in asia: Study prepared for the asian development bank. https://www.adb.org/sites/default/files/project-document/81002/urban-poverty-asia.pdf. Accessed: 2018-08-18. 2

[126] McCallum, I., Obersteiner, M., Nilsson, S., and Shvidenko, A. (2006). A spatial comparison of four satellite derived 1km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation*, 8(4):246–255. 46

[127] McCormick, K., Anderberg, S., Coenen, L., and Neij, L. (2013). Advancing sustainable urban transformation. *Journal of Cleaner Production*, 50:1–11. 2, 63

[128] McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, pages 109–142. 54

[129] McGarigal, K. (2013). Fragstats 4.2 help. *University of Massachusetts, Amherst, MA*. 47, 53, 55

[130] McGranahan, G., Balk, D., and Anderson, B. (2007). The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and urbanization*, 19(1):17–37. 21

[131] McGuirk, M. A. and Porell, F. W. (1984). Spatial patterns of hospital utilization: the impact of distance and time. *Inquiry*, pages 84–95. 57

[132] McLoughlin, F., Duffy, A., and Conlon, M. (2012). Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An irish case study. *Energy and Buildings*, 48:240–248. 64

[133] Mellander, C., Lobo, J., Stolarick, K., and Matheson, Z. (2015). Night-time light data: A good proxy measure for economic activity? *PloS one*, 10(10):e0139779. 35

[134] Miller, H. J. (2010). The data avalanche is here. shouldnt we be digging? *Journal of Regional Science*, 50(1):181–201. 11

[135] Miller, H. J. and Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, 80(4):449–461. 11, 12

[136] Miller, R. B. and Small, C. (2003). Cities from space: potential applications of remote sensing in urban environmental research and policy. *Environmental Science & Policy*, 6(2):129–137. 21

[137] Mishra, V., Smyth, R., and Sharma, S. (2009). The energy-gdp nexus: evidence from a panel of pacific island countries. *Resource and Energy Economics*, 31(3):210–220. 32

[138] Miyazaki, H., Shibasaki, R., and Nagai, M. (2016). An automated method for time-series human settlement mapping using landsat data and existing land cover maps. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pages 1784–1787. IEEE. 41

[139] Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1):67–82. 57

[140] Moore, M., Gould, P., and Keary, B. S. (2003). Global urbanization and impact on health. *International journal of hygiene and environmental health*, 206(4-5):269–278. 2

[141] Morton, A., Nagle, N., Piburn, J., Stewart, R. N., and McManamay, R. (2017). A hybrid dasymetric and machine learning approach to high-resolution residential electricity consumption modeling. In *Advances in Geocomputation*, pages 47–58. Springer. 9

[142] Mukherjee, S. and Nateghi, R. (2017). Climate sensitivity of end-use electricity consumption in the built environment: An application to the state of florida, united states. *Energy*, 128:688–700. 36

[143] Nair, S., George, B., Malano, H. M., Arora, M., and Nawarathna, B. (2014). Water–energy–greenhouse gas nexus of urban water systems: Review of concepts, state-of-art and methods. *Resources, Conservation and Recycling*, 89:1–10. 2

[144] Netzband, M. and Jürgens, C. (2010). l. In *Remote sensing of urban and suburban areas*, pages 1–9. Springer. 2

[145] Niebergall, S., Loew, A., and Mauser, W. (2008). Integrative assessment of informal settlements using vhr remote sensing datathe delhi case study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1(3):193–205. 9, 28, 30, 64

[146] NOAA (2005). Degree days explanation. http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/cdus/degree_days/ddayexp.shtml. Accessed: 2017-12-06. 90

[147] Omer, A. M. (2008). Energy, environment and sustainable development. *Renewable and sustainable energy reviews*, 12(9):2265–2300. 2, 106

[148] Orenstein, D. E., Bradley, B. A., Albert, J., Mustard, J. F., and Hamburg, S. P. (2011). How much is built? quantifying and interpreting patterns of built space from different data sources. *International Journal of Remote Sensing*, 32(9):2621–2644. 46

[149] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66. 28

[150] Pachauri, S. and Jiang, L. (2008). The household energy transition in india and china. *Energy policy*, 36(11):4022–4035. 31

[151] Parikh, J. and Shukla, V. (1995). Urbanization, energy use and greenhouse effects in economic development: Results from a cross-national study of developing countries. *Global Environmental Change*, 5(2):87–103. 31

[152] Patino, J. E. and Duque, J. C. (2013). A review of regional science applications of satellite remote sensing in urban settings. *Computers, Environment and Urban Systems*, 37:1–17. 30

[153] Patlolla, D. R., Bright, E. A., Weaver, J. E., and Cheriyadat, A. M. (2012). Accelerating satellite image based large-scale settlement detection with gpu. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, pages 43–51. ACM. 20, 40, 42, 72

[154] Payne, J. E. (2010). Survey of the international evidence on the causal relationship between energy consumption and growth. *Journal of Economic Studies*, 37(1):53–95. 4

[155] Perdigao, V. and Annoni, A. (1997). Technical and methodological guide for update corine land cover data base. 21

[156] Peres, C. A., Barlow, J., and Laurance, W. F. (2006). Detecting anthropogenic disturbance in tropical forests. *Trends in Ecology & Evolution*, 21(5):227–229. 23

[157] Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A. J., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., and Syrris, V. (2016a). Operating procedure for the production of the global human settlement layer from landsat data of the epochs 1975, 1990, 2000, and 2014. publications office of the european union. http://publications.jrc.ec.europa.eu/repository/handle/111111111/40182. 23, 42, 43

[158] Pesaresi, M. and Freire, S. (2014). Buref ? producing a global reference layer of built-up by integrating population and remote sensing data - background paper prepared for the global assessment report on disaster risk reduction. http://publications.jrc.ec.europa.eu/repository/handle/JRC90459. Accessed: 2016-04-20. 21

[159] Pesaresi, M., Gerhardinger, A., and Kayitakire, F. (2008). A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE Journal of selected topics in applied earth observations and remote sensing*, 1(3):180–192. 23, 29

[160] Pesaresi, M., Syrris, V., and Julea, A. (2016b). A new method for earth observation data analytics based on symbolic machine learning. *Remote Sensing*, 8(5):399. 23

[161] Pontius, R. G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric engineering and remote sensing*, 66(8):1011–1016. 46

[162] Pontius Jr, R. G. and Millones, M. (2011). Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15):4407–4429. 46

[163] Potere, D. and Schneider, A. (2007). A critical look at representations of urban areas in global maps. *GeoJournal*, 69(1-2):55–80. 20, 22, 40

[164] Potere, D., Schneider, A., Angel, S., and Civco, D. L. (2009). Mapping urban areas on a global scale: which of the eight maps now available is more accurate? *International Journal of Remote Sensing*, 30(24):6531–6558. 20, 21, 22, 40

[165] Rahman, A., Liu, X., and Kong, F. (2014). A survey on geographic load balancing based data center power management in the smart grid environment. *IEEE Communications Surveys & Tutorials*, 16(1):214–233. 64

[166] Ramaswami, A., Bettencourt, L., Clarens, A., Das, S., Irwin, E., Pataki, D., Pincetl, S., Seto, K., and Waddell, P. (2018). Sustainable urban systems: Articulating a long-term convergence research agenda. a report from the nsf advisory committee for environmental research and education, prepared by the sustainable urban systems subcommittee. Technical report, Advisory Committee For Environmental Research and Education - National Science Foundation. 1, 2, 20, 105

[167] Ratti, C., Baker, N., and Steemers, K. (2005). Energy consumption and urban texture. *Energy and buildings*, 37(7):762–776. 84

[168] Rindfuss, R. R. and Stern, P. C. (1998). Linking remote sensing and social science: The need and the challenges. *People and pixels: Linking remote sensing and social science*, pages 1–27. 20

[169] Román, M. O. and Stokes, E. C. (2015). Holidays in lights: Tracking cultural patterns in demand for energy services. *Earth's future*, 3(6):182–205. 28, 65

[170] Ruiz, L., Fdez-Sarría, A., and Recio, J. (2004). Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study. In *20th ISPRS Congress*, volume 35, part B, pages 1109–1114. 71

[171] Sadorsky, P. (2013). Do urbanization and industrialization affect energy intensity in developing countries? *Energy Economics*, 37:52–59. 33

[172] Samaniego, H. and Moses, M. E. (2008). Cities as organisms: Allometric scaling of urban road networks. *Journal of Transport and Land use*, 1(1):21–39. 34

[173] Schneider, A., Friedl, M. A., McIver, D. K., and Woodcock, C. E. (2003). Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, 69(12):1377–1386. 22, 40

[174] Schneider, A., Friedl, M. A., and Potere, D. (2010). Mapping global urban areas using modis 500-m data: New methods and datasets based on ?urban ecoregions? *Remote Sensing of Environment*, 114(8):1733–1746. 22, 40

[175] Schneider, A. and Woodcock, C. E. (2008). Compact, dispersed, fragmented, extensive? a comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. *Urban Studies*, 45(3):659–692. 20, 39, 53

[176] Schwarz, N., Lautenbach, S., and Seppelt, R. (2011). Exploring indicators for quantifying surface urban heat islands of european cities with modis land surface temperatures. *Remote Sensing of Environment*, 115(12):3175–3186. 41

[177] Seto, K. (2009). Global urban issues–a primer. In Gamba, H. and Herold, M., editors, *Global mapping of human settlements: Experiences, Data Sets, and Prospects*, pages 3–9. CRC Press. 20, 21, 60

[178] Seto, K. C., Dhakal, S., Bigio, A., Blanco, H., Delgado, G. C., Dewar, D., Huang, L., Inaba, A., Kansal, A., Lwasa, S., et al. (2014). Human settlements, infrastructure and spatial planning. In *Climate Change 2014: Mitigation of Climate Change. IPCC Working Group III Contribution to AR5.* Cambridge University Press. 1, 30, 33, 87, 88

[179] Seto, K. C. and Fragkias, M. (2005). Quantifying spatiotemporal patterns of urban land-use change in four cities of china with time series landscape metrics. *Landscape Ecology*, 20(7):871–888. 53

[180] Seto, K. C., Güneralp, B., and Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*, 109(40):16083–16088. 2, 64

[181] Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., Chen, L., and Wu, J. (2014). Evaluating the ability of npp-viirs nighttime light data to estimate the gross domestic product and the electric power consumption of china at multiple scales: A comparison with dmsp-ols data. *Remote Sensing*, 6(2):1705–1724. 27, 65

[182] Small, C. (2011). The human habitat. In *Human Population*, pages 27–46. Springer. 47, 73

[183] Smith, K. R. (1993). Fuel combustion, air pollution exposure, and health: the situation in developing countries. *Annual Review of Energy and the Environment*, 18(1):529–566. 7

[184] Stan, M., Popescu, A., Datcu, M., and Stoichescu, D. A. (2015). An assessment of feature extraction methods for sentinel-1 images on urban areas. In *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, pages 369–372. IEEE. 71

[185] Strahler, A., Muchoney, D., Borak, J., Friedl, M., Gopal, S., Lambin, E., and Moody, A. (1999). Modis land cover product algorithm theoretical basis document (atbd) version 5.0. http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf. 41

[186] Su, W., Li, J., Chen, Y., Liu, Z., Zhang, J., Low, T. M., Suppiah, I., and Hashim, S. A. M. (2008). Textural and local spatial statistics for the object-oriented classification

of urban areas using high resolution imagery. *International journal of remote sensing*, 29(11):3105–3117. 29

[187] Sutton, P., Roberts, D., Elvidge, C., and Baugh, K. (2001). Census from heaven: An estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing*, 22(16):3061–3076. 35

[188] Sutton, P. C., Elvidge, C. D., Ghosh, T., et al. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, 8(S07):5–21. 36

[189] Tamayao, M., Blackhurst, M., and Matthews, H. (2014). Do us metropolitan core counties have lower scope 1 and 2 co2 emissions than less urbanized counties? *Environmental Research Letters*, 9(10):104011. 36, 104

[190] Taubenböck, H. and Kraff, N. (2014). The physical face of slums: a structural comparison of slums in mumbai, india, based on remotely sensed data. *Journal of Housing and the Built Environment*, 29(1):15–38. 30

[191] Taubenbock, H., Wurm, M., Setiadi, N., Gebert, N., Roth, A., Strunz, G., Birkmann, J., and Dech, S. (2009). Integrating remote sensing and social science. In *2009 Joint Urban Remote Sensing Event*, pages 1–7. 40

[192] The World Bank (2018). The world bank in yemen. 68

[193] Trzaska, S. and Schnarr, E. (2014). A review of downscaling methods for climate change projections. *United States Agency for International Development by Tetra Tech ARD*, pages 1–42. 106

[194] UN-Habitat (1997). Regional development planning and management of urbanization - experiences from developing countries. http://sbd.iuav.it/sbda/mostraindici.php?&EW_D=NEW&EW_T=TF&EW_P=LS_EW&EW=118390&EW_INV=AA_CIA_000013913&. Accessed: 2018-08-20. 3

[195] United Nation (2014). World urbanization prospects: The 2014 revision, highlights. department of economic and social affairs. *Population Division, United Nations.* 39

[196] United Nations (2012). Un-habitat - energy. https://unhabitat.org/urban-themes/energy/. Accessed: 2018-07-24. 1, 87

[197] United Nations (2016). The world's cities in 2016. 67

[198] United Nations - Department of Economic and Social Affairs (2018). 2018 revision of world urbanization prospects. 1, 63, 87

[199] United Nations - Sustainable Development Goals (2018). Affordable and clean energy. https://www.un.org/sustainabledevelopment/energy/. Accessed: 2018-08-18. 5, 6

[200] United Nations Centre for Human Settlements (1984). Energy requirements and utilization in rural and urban lowincome settlements. 63

[201] Vatsavai, R. R., Bhaduri, B., and Graesser, J. (2013). Complex settlement pattern extraction with multi-instance learning. In *Urban Remote Sensing Event (JURSE), 2013 Joint*, pages 246–249. IEEE. 23

[202] Victor, N. and Nichols, C. (2014). Census from heaven: An estimate of global electricity demand'if everyone lived like in oecd'. *United States Association for Energy Economics - Research Paper Series*, 14-192. 64

[203] Vijayaraj, V., Bright, E. A., and Bhaduri, B. L. (2007). High resolution urban feature extraction for global population mapping using high performance computing. In *2007 IEEE International Geoscience and Remote Sensing Symposium*, pages 278–281. 23, 40

[204] Vijayaraj, V., Cheriyadat, A. M., Sallee, P., Colder, B., Vatsavai, R. R., Bright, E. A., and Bhaduri, B. L. (2008). . In *Applied Imagery Pattern Recognition Workshop, 2008. AIPR'08. 37th IEEE*, pages 1–8. IEEE. 29

[205] Wang, A., Liu, P., and Xie, C. (2016a). Urban land use classification from high-resolution sar images based on multi-scale markov random field. In *Geoinformatics, 2016 24th International Conference on*, pages 1–4. IEEE. 71

[206] Wang, Y., Chen, L., and Kubota, J. (2016b). The relationship between urbanization, energy use and carbon emissions: evidence from a panel of association of southeast asian nations (asean) countries. *Journal of Cleaner Production*, 112:1368–1374. 32

[207] Weber, E. M., Seaman, V. Y., Stewart, R. N., Bird, T. J., Tatem, A. J., McKee, J. J., Bhaduri, B. L., Moehl, J. J., and Reith, A. E. (2018). Census-independent population mapping in northern nigeria. *Remote Sensing of Environment*, 204:786 – 798. 43

[208] Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3/4):330–336. 53

[209] Welch, R. (1980). Monitoring urban population and energy utilization patterns from satellite data. *Remote sensing of Environment*, 9(1):1–9. 25, 65

[210] Welch, R. and Zupko, S. (1980). Urbanized area energy-utilization patterns from dmsp data. *Photogrammetric Engineering and Remote Sensing*, 46(2):201–207. 26, 65

[211] Weng, Q. (2009). Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4):335–344. 21

[212] Wilson, S. G. (2012). *Patterns of metropolitan and micropolitan population change: 2000 to 2010*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau. 94

[213] World Energy Council (2018). Average electricity consumption per electrified household. https://wec-indicators.enerdata.net/household-electricity-use.html. Accessed: 2018-08-22. 6

[214] Xie, Y. and Weng, Q. (2016a). Detecting urban-scale dynamics of electricity consumption at chinese cities using time-series dmsp-ols (defense meteorological satellite program-operational linescan system) nighttime light imageries. *Energy*, 100:177–189. 65

[215] Xie, Y. and Weng, Q. (2016b). World energy consumption pattern as revealed by dmsp-ols nighttime light imagery. *GIScience & Remote Sensing*, 53(2):265–282. 26, 65

[216] Yin, Z.-Y., Stewart, D. J., Bullard, S., and MacLachlan, J. T. (2005). Changes in urban built-up surface and population distribution patterns during 1986–1999: A case study of cairo, egypt. *Computers, Environment and Urban Systems*, 29(5):595–616. 60

[217] Yuan, J., Wang, D., and Cheriyadat, A. M. (2015). Factorization-based texture segmentation. *IEEE Transactions on Image Processing*, 24(11):3488–3497. 30, 65, 70, 72, 73, 83

[218] Yuan, J., Wang, D., and Li, R. (2014). Remote sensing image segmentation by combining spectral and texture features. *IEEE transactions on geoscience and remote sensing*, 52(1):16–24. 72

[219] Zhang, Q. and Seto, K. C. (2011). Mapping urbanization dynamics at regional and global scales using multi-temporal dmsp/ols nighttime light data. *Remote Sensing of Environment*, 115(9):2320 – 2329. 40

[220] Zhang, X., Li, P., and Hu, H. (2016). Regional urban extent extraction using multi-sensor data and decision rules. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pages 1778–1781. IEEE. 41

[221] Zhao, N., Ghosh, T., and Samson, E. L. (2012a). Mapping spatio-temporal changes of chinese electric power consumption using night-time imagery. *International journal of remote sensing*, 33(20):6304–6320. 27

[222] Zhao, N., Ghosh, T., and Samson, E. L. (2012b). Mapping spatio-temporal changes of chinese electric power consumption using night-time imagery. *International Journal of Remote Sensing*, 33(20):6304–6320. 80

[223] Zhao, P. and Zhang, M. (2018). The impact of urbanisation on energy consumption: A 30-year review in china. *Urban climate*, 24:940–953. 30, 88

[224] Zhao, S., Da, L., Tang, Z., Fang, H., Song, K., and Fang, J. (2006). Ecological consequences of rapid urban expansion: Shanghai, china. *Frontiers in Ecology and the Environment*, 4(7):341–346. 65

[225] Zheng, M., Meinrenken, C. J., and Lackner, K. S. (2014). Agent-based model for electricity consumption and storage to evaluate economic viability of tariff arbitrage for residential sector demand response. *Applied Energy*, 126:297–306. 64

[226] Zhong, P. and Wang, R. (2007). A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3978–3988. 71

[227] Zhu, D., Tao, S., Wang, R., Shen, H., Huang, Y., Shen, G., Wang, B., Li, W., Zhang, Y., Chen, H., et al. (2013). Temporal and spatial trends of residential energy consumption and air pollutant emissions in china. *Applied energy*, 106:17–24. 63

Appendices

A   Additional tables for chapter 3

Table A.1: Group means of calculated spatial metrics across agreement classes at 500 meter resolution. (Standard errors are given in the parenthesis.)

Class	Egypt			Taiwan		
	LSI	FRACMN	PLADJ	LSI	FRACMN	PLADJ
CL1	2.419 (0.01)	1.046 (0.00)	41.198 (0.29)	2.350 (0.03)	1.025 (0.00)	21.909 (0.50)
CL2	2.410 (0.02)	1.034 (0.00)	32.316 (0.35)	2.206 (0.03)	1.023 (0.00)	20.084 (0.58)
CL3	1.795 (0.02)	1.027 (0.00)	26.415 (0.42)	2.083 (0.04)	1.024 (0.00)	20.239 (0.74)
CL4	1.467 (0.02)	1.029 (0.00)	35.806 (0.90)	1.848 (0.05)	1.033 (0.00)	30.447 (1.29)

Table A.2: Group means of calculated spatial metrics across agreement classes at 300 meter resolution. (Standard errors are given in the parenthesis.)

Class	Egypt			Taiwan		
	LSI	FRACMN	PLADJ	LSI	FRACMN	PLADJ
CL1	3.012 (0.02)	1.022 (0.00)	22.799 (0.27)	3.044 (0.04)	1.027 (0.00)	25.585 (0.52)
CL2	2.470 (0.02)	1.020 (0.00)	19.930 (0.30)	2.980 (0.05)	1.025 (0.00)	23.130 (0.53)
CL3	1.975 (0.02)	1.026 (0.00)	27.325 (0.41)	2.746 (0.06)	1.025 (0.00)	22.050 (0.58)
CL4	1.636 (0.03)	1.027 (0.00)	33.326 (0.73)	2.303 (0.06)	1.035 (0.03)	35.151 (1.16)

Table A.3: Group means of calculated spatial metrics across agreement classes at 38.2 meter resolution. (Standard errors are given in the parenthesis.)

Class	Egypt			Taiwan		
	LSI	FRACMN	PLADJ	LSI	FRACMN	PLADJ
CL1	9.258 (0.09)	1.036 (0.00)	50.541 (0.31)	8.986 (0.15)	1.035 (0.00)	52.160 (0.60)
CL2	7.192 (0.08)	1.043 (0.00)	60.917 (0.38)	8.769 (0.21)	1.042 (0.00)	52.631 (0.53)
CL3	5.279 (0.10)	1.052 (0.00)	70.968 (0.43)	9.548 (0.28)	1.049 (0.00)	57.092 (0.52)
CL4	3.409 (0.07)	1.054 (0.00)	74.361 (0.62)	5.765 (0.19)	1.049 (0.00)	66.899 (0.87)

Table A.4: Group means of calculated spatial metrics across agreement classes at 8 meter resolution. (Standard errors are given in the parenthesis.)

Class	Egypt			Taiwan		
	LSI	FRACMN	PLADJ	LSI	FRACMN	PLADJ
CL1	11.959 (0.13)	1.059 (0.00)	72.873 (0.28)	12.932 (0.23)	1.061 (0.00)	83.099 (0.32)
CL2	10.494 (0.13)	1.069 (0.00)	80.906 (0.30)	12.291 (0.33)	1.071 (0.00)	84.046 (0.34)
CL3	9.045 (0.19)	1.076 (0.00)	87.517 (0.29)	13.899 (0.45)	1.075 (0.00)	86.512 (0.32)
CL4	5.963 (0.17)	1.075 (0.00)	89.256 (0.36)	7.535 (0.27)	1.071 (0.03)	90.141 (0.45)

B Additional tables for chapter 4

Table B.1: ANOVA table for the regression coefficients from the Ndola model.

Settlement Class	df	Sum sq.	Mean Sq	F Value	Pr(>)F
1	1	26999.6	26999.6	2642.7132	<0.00001∗∗∗
2	1	12931.5	12931.5	1265.7304	<0.00001∗∗∗
3	1	7867.6	7867.6	770.0794	<0.00001∗∗∗
4	1	7376.0	7376.0	721.9655	<0.00001∗∗∗
5	1	15.6	15.6	1.5291	0.2164

Table B.2: ANOVA table for the regression coefficients from the Sana'a model.

Settlement Class	df	Sum sq.	Mean Sq	F Value	Pr(>)F
1	1	7632	7632	1789.06	<0.00001∗∗∗
2	1	44405	44405	10408.71	<0.00001∗∗∗
3	1	2424	2424	568.21	<0.00001∗∗∗
4	1	6347	6347	1487.78	<0.00001∗∗∗
5	1	245	245	57.40	<0.00001∗∗∗

Table B.3: ANOVA table for the regression coefficients from the Johannesburg model.

Settlement Class	df	Sum sq.	Mean Sq	F Value	Pr(>)F
1	1	61801	61801	432.445	<0.00001∗∗∗
2	1	473771	473771	3315.1812	<0.00001∗∗∗
3	1	87448	87448	611.910	<0.00001∗∗∗
4	1	1065948	1065948	7458.898	<0.00001∗∗∗
5	1	6021	6021	42.131	<0.00001∗∗∗

Vita

Pranab Kanti Roy Chowdhury was born and raised in Kolkata (formerly known as Calcutta), India. Pranab completed middle and high schooling at Barrackpore Government High School. He obtained a Bachelor of Science degree from the erstwhile Presidency College University of Calcutta, Kolkata (now known as Presidency University) in Geography (Honors), which was followed by a Master of Science degree in Geography from the same institution. Later, he obtained a Master of Technology degree from Indian Institute of Remote Sensing, the capacity building arm of the Indian Space Research Organization, located at Dehradun, India. Here, his specialized in the application of remote sensing and geographical information science in urban and regional planning. Subsequently, he worked as a Scientist in Government of India for the next three years. In order to pursue his Ph.D., Pranab joined the Bredesen Center for Interdisciplinary Research and Graduate Education at the University of Tennessee Knoxville in Fall 2014. His dissertation research was carried out at Oak Ridge National Laboratory, in collaboration with the Geographical Information Science and Technology Group and the Urban Dynamics Institute. His present focus areas are energy geography, urban science, land-use land-cover change, and urban sustainability. He was awarded a Chancellor's Citation for Extraordinary Professional Promise at the University of Tennessee in May 2018. He received his Ph.D. in December 2018.