The application of optical satellite imagery and census data for urban population estimation: A case study for Ahmedabad, India

Zur Erlangung des akademischen Grades einer

DOKTORIN DER NATURWISSENSCHAFTEN

von der Fakultät für

Bauingenieur-, Geo- und Umweltwissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Geol. Eike – Marie Nolte

aus Hildesheim

Tag der mündlichen Prüfung: 15.07.2010

Referent: Prof. Dr. Friedemann Wenzel

Korreferent: Prof. Dr. Stefan Hinz

Karlsruhe 2010

Wissen ist der einzige Rohstoff, der sich durch Gebrauch vermehrt.

(Jürgen Mlynek, 2007)

Abstract

In the last decades of the 20^{th} century, the rapid growth of the world population has led to dynamic and complex urbanization. By 2050, it is predicted that 6 billion people – by then about 70% of the world population – will reside in urban areas (UN, 2008). However, this urban growth is geographically disproportionate. The rapid and often unplanned growth of urban areas in lesser developed and newly industrialized countries, with annual growth rates between 3% and 4%, directly contrasts the stagnating urban growth within industrialized countries. India is an useful example of a nation where the phenomenon of urbanization is rapid, spatially varied, and exceptionally dynamic.

By 2050, India's population is expected to overrun China's population (1,4 billion) with a total population of 1,6 billion and an urban population of 0,9 billion (UN, 2008). The considerably high annual urban growth rates are expected to remain constant within the coming decades. As a consequence, many cities (today already more than 43 cities in India have more than 1 million inhabitants) will become megacities in the near future. One example for the dynamic and uncontrolled development of urban areas is the city of Ahmedabad in northwest India.

India has an exceptional census history. Since 1872, the population is enumerated every ten years by the official Indian Census Bureau. Thus, the availability of population data for Indian cities is restricted to these decennial Census data. But as most Indian cities face a dynamic population growth, a population count on a decennial basis is not sufficient to display the population development. To estimate the intercensal population for these dynamic cities, modern technology such as satellite imagery plays an important role as it provides information for large areas and can be used for identifying and classifying for example land cover like urban extent. In addition to the large area coverage with a spatial resolution depending on the imaging technology used, observations are feasible repeatedly, depending on the revisit time of the satellite, so that changes of land cover can be monitored.

In this thesis, the question of how the population in large cities can be spatially modelled is addressed. A methodology is developed to spatially model the population of the Ahmedabad Municipal Cooperation (AMC) area using satellite images and census data. The developed method allows for modeling its spatial distribution for different times of day on city, district and building level. The developed models enable the generation of population data with different spatial resolutions and different information detail at each level, with the amount of required input data increasing with the information detail and resolution of the resulting population data. This way, the developed methodology allows for generating population information for large cities without depending on detailed survey information.

At all spatial levels, the population estimation models incorporate GIS-based data and therefore an administrative map for the AMC area is essential, for example for mapping the population distribution on city and district level and for spatially analyzing the statistical data. Therefore, a methodology was developed for generating an administrative map using GoogleEarth images and an analog paper map. The resulting map yield an accuracy of 93,49%. At city level (tier 1), three models were developed to estimate the population within the AMC area and within the urban extent of the AMC area. The extent of the urban areas was extracted from moderate optical satellite images (Landsat 5 TM) and very high resolution optical satellite images (Quickbird). The city population for 2008 was projected to be 3.152.108 people, with a population density of 23.091 people/km^2 in the AMC area (model I). For the urban population (model II), the population density was estimated 56.791 people/km^2 for the Quickbird image and 34.446 people/km^2 for the Landsat image. The density difference is due to the fact that the urban extent extracted from Landsat is larger than the urban extent extracted from Quickbird because of mixed pixels. Model III allows for population estimation and for modelling the spatial distribution of the population for different times of day. In this study, two occupancy–based approaches were tested: A binary approach which distinguishes only residential and non-residential occupancy (Coburn & Spence, 2002) and an approach provided by HAZUS using 4 different occupancy categories (FEMA, 2008).

Tier 2 operates on district level and provides 3 different models to estimate and model the population. The first model (model IV) provides the district population as an aggregated value for the administrative boundary. The second model (model V) calculates the population in the urban areas within each district. The results from model IV and V revealed that the district population varies considerably in the AMC area. The districts with the highest population are located in the south-eastern periphery of the AMC area. However, the districts with the highest population density are located in the central part of the AMC area due to the much smaller administrative area. This shows that the assumption of a constant population density which proved to be valid on city level cannot be readily transferred to district level. At tier 3 - building level, two models were developed to estimate the population and model the population distribution within the districts. Model VII estimates the population on building block level, the second model VIII operates on single building level. Because of limited data availability, only the building block model was tested for the city of Ahmedabad.

This dissertation thesis addresses the question how optical satellite imagery and census data can be used to develop a tiered modelling method for urban population of large cities. The developed methodology is summarised in the final chapter, where the achievements as well as scope for improving the currently proposed methodology is described.

Key words: population modelling, remote sensing, census, India, megacities, GIS, disaster management

Kurzfassung

Anwendungsmöglichkeiten optischer Satellitenbilder und Zensusdaten zur Bevölkerungsmodellierung am Beispiel der indischen Stadt Ahmedabad

Unsere Welt wird immer stärker durch Städte geprägt. Der Anteil der Menschen, die in Städten leben wächst ständig. Lebten 1950 nur 29,1% der Weltbevölkerung in Städten, sind es 2010 mehr als die Hälfte (UN, 2008). In absoluten Zahlen hat sich die Weltbevölkerung seit 1950 deutlich mehr als verdoppelt. 1950 lebten 2.535.093 Menschen auf der Welt, 2010 sind es fast sieben Milliarden (UN, 2008). Aber dieses Wachstum ist global ungleich verteilt. Das explosionsartige Wachstum der urbanen Räume in Entwicklungs- und Schwellenländern mit jährlichen Wachstumsraten von 4,0% bzw. 3,0% steht einer regelrechten Stagnation des urbanen Wachstums in den Industrieländern gegenüber.

In Indien ist das Phänomen der Urbanisierung besonders gut zu beobachten. Zwischen 2007 und 2025 ist ein Zuwachs der urbanen Bevölkerung um 197 Millionen Menschen zu erwarten (UN, 2009). Es wird davon ausgegangen, dass die urbanen Wachstumsraten in Indien in den nächsten Jahrzehnten unverändert hoch bleiben werden. Viele der heutigen Städte, mehr als 45 der indischen Städte zählen schon jetzt mehr als 1 Million Einwohner, werden also in naher Zukunft zu Megastädten werden. Die Stadt Ahmedabad in Nordwestindien ist ein gutes Beispiel einer Millionenstadt mit sehr komplexen Bebauung und stark variierender Bevölkerungsdichte, die einer hohen Dynamik mit schnellen strukturellen und sozioökonomischen Wandel unterworfen ist.

Die Bevölkerungsschätzungen für indische Städte sind meist auf die Daten des indischen Zensus, der alle zehn Jahre durchgeführt wird, beschränkt. Vor dem Hintergrund der hohen Bevölkerungsdynamik indischer Städte sind diese Daten nicht ausreichend, um die Bevölkerungsentwicklung für die Jahre zwischen zwei Bevölkerungszählungen mit hinreichender Genauigkeit abzubilden. Für die Erfassung der räumlichen Verteilung der Bevölkerung in diesen Zeiträumen, spielen Fernerkundungsdaten als Datenquelle für flächenhafte, großräumige Information eine wichtige Rolle, da traditionelle Techniken zur Datenerfassung an ihre Grenzen stoßen. Anhand von Fernerkundungsdaten können Informationen über die Landnutzung, die Bebauung und die Ausdehnung des Stadtgebietes gewonnen werden. In Abhängigkeit von der Wiederkehrperiode des Satelliten ist es außerdem möglich, Veränderung der Landnutzung, des Inventars sowie anderer Parameter in regelmäßigen Zeitintervallen zu erfassen.

Die zentrale Fragestellung dieser Dissertation ist daher, wie die räumliche Bevölkerungsverteilung in indischen Großstädten mit Hilfe optischer Satellitendaten, geokodierten Informationen und statistischen Daten modelliert werden kann. Hierzu wurde eine hierarchische Methode entwickelt, die es ermöglicht für unterschiedliche administrative Ebenen Bevölkerungsdaten zu generieren. Die Hierarchie der entwickelten Methode umfasst drei räumliche, administrative Ebenen: Stadt-, Distrikt- und Gebäudeebene. Für jede Ebene wurden verschiedene Modelle zur Modellierung der Bevölkerungsverteilung entwickelt. So ist es möglich, je nach Datenverfügbarkeit Datensätze mit unterschiedlicher Detailgenauigkeit zu entwickeln. Für den

Fall, dass für das gewählte Jahr weder administrative Grenzen noch Bevölkerungsdaten zur Verfügung stehen, wurde eine Methode entwickelt, um eine geokodierte Karte der administrativen Grenzen unter Verwendung von GoogleEarth und einer analogen Karte zu erstellen. Um für das Zieljahr den grundlegenden Bevölkerungsdatensatz zu berechnen, wurden unterschiedliche Projektionsmethoden vorgestellt. Im Fallbeispiel wurde die *vital statistic procedure* verwendet, um ausgehend von der indischen Bevölkerungszählung im Jahr 2001 die Bevölkerung der Stadt Ahmedabad im Jahr 2006 abzuschätzen. Dieser Wert wurde anschließend dazu verwendet, existierenden Bevölkerungszahlen für Ahmedabad zu verifizieren. Die Übereinstimmung der projizierten Bevölkerung betrug 98,4%. Daher wurden die existierenden Bevölkerungsdaten, die von der *Ahmedabad Municipality Cooperation* erhoben wurden, als ausreichend genau betrachtet.

Auf Stadtebene wurden 3 Modelle entwickelt (Modell I – III), die die zeitliche und räumliche Verteilung der Stadtbevölkerung, der urbane Bevölkerung und der Bevölkerung nach der Gebäudenutzung modellieren. Model I liegt die Annahme zu Grunde, dass die Bevölkerung innerhalb der administrativen Stadtgrenze homogen verteilt ist. Für das Jahr 2008 wurde für Ahmedabad eine Bevölkerungsdichte von 23.091 Einwohner/km^2 errechnet. Model II ermöglicht es, die urbane Bevölkerung für Ahmedabad zu berechnen und ihre Verteilung zu modellieren. In einem ersten Schritt wurden die bebauten Anteile des Stadtgebiets mit Hilfe von Satellitenbildern (Quickbird und Landsat TM5) extrahiert. Hierzu wurden NDVI Grenzwerte verwendet. Ein direkter Vergleich der Ergebnisse zwischen Quickbird (55,50 km^2) und Landsat TM 5 (91,50 km^2) ergab, dass aufgrund der Mischpixelproblematik die bebaute Fläche aus Landsat TM 5 Bildern deutlich über der Fläche aus Quickbird lag. Die urbane Bevölkerung wurde unter Verwendung beider Ergebnisse berechnet. Die Bevölkerungsdichte für Quickbird im Jahr 2008 beträgt 56.791 Einwohner/km^2 und 34.446 Einwohner/km^2 für Landsat 5 TM. Model III basiert auf der Annahme, dass die Verteilung der Bevölkerung maßgeblich von der Flächen bzw. Gebäudenutzung abhängt. In einem ersten Schritt wurden statistische Daten untersucht, um die Zahl der Beschäftigten in bestimmen Sektoren zu ermitteln. Es wurden folgende Sektoren unterschieden: Industrie, Gewerbe, Dienstleistung und öffentliche Einrichtungen. Bei einer Gesamtbevölkerung von 4.005.084 im Jahr 2008, gingen 1.447.364 aus unterschiedlichen Gründen wie zum Beispiel Arbeitslosigkeit oder Alter keiner Beschäftigung nach. Insgesamt waren offiziell 796.747 Menschen als erwerbstätig registriert. Damit liegt die offizielle Erwerbstätigkeitsquote in Ahmedabad bei 19,64%. Allerdings werden 44 % der Gesamtbevölkerung nicht von der Statistik erfasst. Es ist davon auszugehen, dass ein Großteil der nicht erfassten Bevölkerung im informellen Sektor arbeitet. Daher ist die berechnete Anzahl der Beschäftigten in den einzelnen Sektoren als Mindestschätzung anzusehen. Um die Verteilung der Bevölkerung für verschiedene Tageszeiten zu berechnen wurden zwei Methoden getestet. Zu einem die Gebäudenutzungskurven von Coburn & Spence (2002) und zum anderen die Gebäudenutzungsfunktionen, die im Rahmen des HAZUS Programms zur Anwendung kommen. Die beiden Ansätze unterscheiden sich in zwei Punkten. Erstens ermöglichen Coburn & Spence (2002) die Berechnung der Dynamik der Bevölkerungsverteilung im Tagesverlauf, während HAZUS nur Funktionen für drei ausgewählte Zeitpunkte verwendet (02:00 Uhr, 14:00 Uhr und 18:00 Uhr). Zweitens unterscheiden Coburn & Spence (2002) nur zwei Nutzungskategorien (Wohngebäude und Nicht-Wohngebäude), während in HAZUS bis zu 7 Nutzungskategorien unterschieden werden können. Ein direkter Vergleich der beiden Methoden hat gezeigt, dass die unterschiedlichen Ergebnisse vor allem darauf zurückzuführen sind, dass unterschiedliche Bevölkerungsanteile für die jeweilige Nutzungskategorie angenommen werden. Zum Beispiel nimmt HAZUS für 02:00 Uhr morgens an, das sich 99 % der Bevölkerung in Wohngebäuden aufhalten, Coburn & Spence (2002) hingegen nur 76 %. Für die Distriktebene wurden ebenfalls drei Modelle (Modell IV – VI) entwickelt, die es ermöglichen die

Gesamtdistriktbevölkerung, die urbane Distriktbevölkerung und die Bevölkerungsverteilung nach der Gebäudenutzung zu erfassen. Die Bevölkerungsverteilung nach Model IV und Model V zeigte, dass die peripheren Distrikte zwar eine höhere Gesamtbevölkerung haben, aber aufgrund der großen Distriktfläche die niedrigste Bevölkerungsdichte besitzen. Model VI ermöglicht es, die Bevölkerung für verschiedene Nutzungskategorien der Gebäude abzuschätzen. Geringe Veränderungen in der Bevölkerungsdichte im Tagesverlauf eines Distrikts können beispielsweise in einem hohen Anteil an Slums, in denen Menschen gleichzeitig arbeiten und wohnen, begründet liegen. Da keine statistischen Daten über die Beschäftigungsverhältnisse in den einzelnen Distrikten vorlagen, wurden die Annahmen von Stadtebene auf Distriktebene übertragen. Für die Gebäudeebene (Ebene 3) wurden zwei Modelle entwickelt, um die Bevölkerung für Wohnblöcke (Model VII) und für einzelne Wohngebäude (Modell VIII) abzuschätzen. Aufgrund fehlender Daten konnte Model VIII nur konzeptionell entwickelt und nicht für Ahmedabad getestet werden. Zwei Konzepte wurden für Model VIII vorgestellt. Eine Bottom-up Methode mit der basierend auf Befragungen repräsentativer Bevölkerungsgruppen, die Bevölkerung für die restlichen Gebäude modelliert wird. Und eine Top-Down Methode mit der Bevölkerungsschätzungen für höhere administrative Raumeinheiten für einzelne Gebäude disaggregiert werden.

Im Mittelpunkt der entwickelten Methode steht die Notwendigkeit die Verteilung der Bevölkerung in großen Städten modellieren zu können ohne von der Verfügbarkeit detailierter Informationen abhängig zu sein. Die Identifikation der inhärenten Daten- sowie der Modellungenauigkeiten bildetet einen zweiten Schwerpunkt. So ist es möglich, Bevölkerungs-informationen mit ausgewiesener Genauigkeit zu generieren. Im Hinblick auf den zentralen Aspekt dieser Dissertation „Anwendungsmöglichkeiten optischer Satellitenbilder und Zensusdaten zur Bevölkerungsmodellierung am Beispiel der indischen Stadt Ahmedabad“ lässt sich resümieren, dass die Kombination aus Satellitenbildern und Zensusdaten die Modellierung der zeitlichen und räumlichen Verteilung ermöglicht. Das Potential der Methode liegt vor allem in der Flexibilität der entwickelten Modelle, die zeitliche und räumliche Modellierung der Bevölkerungsverteilung für unterschiedliche administrative Ebenen in Abhängigkeit der Daten-verfügbarkeit mit unterschiedlicher Genauigkeit ermöglicht.

Danksagung / Acknowledgments

Ich danke Prof. Dr. Friedemann Wenzel für die Betreuung und dafür, dass er mir die Möglichkeit zu dieser interessanten Arbeit gegeben hat. Die engagierte Betreuung, die stets konstruktive Kritik, aber auch das entgegengebrachte Vertrauen haben maßgeblich zum Gelingen dieser Arbeit beigetragen. Ebenso bedanke ich mich bei Prof. Dr. Stefan Hinz vom Institut für Photogrammetrie und Fernerkundung für die anregenden Diskussionen, für die Unterstützung sowie für die Übernahme des Zweitgutachtens.

I would like to thank all colleagues from ImageCat Inc. for their support and motivation. Especially, I would like to thank Beverley Adams and Ron Eguchi, who have been major participants in this research and shaped the way in which this research was conducted.

Bedanken möchte ich mich auch bei den Kollegen vom Deutschen Zentrum für Luft- und Raumfahrt, insbesondere bei Hannes Taubenböck für die stets anregenden, wissenschaftlichen Diskussionen.

Diese Arbeit entstand am Center for Disaster Management and Risk Reduction Technology (CEDIM). Meinen Kollegen von CEDIM danke ich für die fortwährende Unterstützung sowie für die interessante Zusammenarbeit. Mein besonderer Dank gilt Bijan Khazai, der mir stets eine große Hilfe war.

Ich danke meinen Bürogefährtinnen Nina und Tanja für ihre Unterstützung in allen Phasen der Promotion, für die hilfreichen Diskussionen und zahlreichen Anregungen. Ich danke auch Juliane Lüke, Rebecca Harrington, Julia Kubanek und James Daniell für ihre Unterstützung und ihre Solidarität in schwierigen Projektphasen.

Schließlich möchte ich mich bei allen Mitarbeiterinnen und Mitarbeitern des Geophysikalischen Instituts für die gute Zusammenarbeit und netten Kaffeepausen bedanken. Besonderer Dank gilt Petra Knopf, die sich stets engagiert darum gekümmert hat, technische Schwierigkeiten aus dem Weg zu räumen. Auch Gaby Bartmann, Claudia Payne und insbesondere Monika Hebben möchte ich für Ihre fortwährende Unterstützung in Verwaltungsfragen danken.

Zu guter Letzt möchte ich mich bei Marco bedanken, für seine Geduld und liebevolle Unterstützung während der Promotion. Meiner Familie möchte ich für ihre moralische Unterstützung in schwierigen Phasen der Arbeit danken. Auch meinen Freundinnen Leila El Gasmi, Sabine Stabenow, Maren Rapp und Kerstin Petrick gebührt herzlicher Dank für ihre jahrelang moralische Unterstützung, die fortwährende Motivation und ihre Freundschaft.

Table of Contents

Introduction

The 21st century is the century of the city. Of the Earth's 6,8 billion human inhabitants, about 3 billion live in urban settlements (UNCHS, 2008). The world population is projected to reach 7 billion early in 2012 and surpass 9 billion people by 2050 (UN, 2009). The cities and towns of the less developed regions are expected to absorb the population growth expected over the next four decades while at the same time drawing in some of the rural population. According to the UN Centre for Human Settlements (UNCHS, 2008), 3 million people per week migrate to cities of the developing world. In India, the phenomenon of urbanization is exceptionally dynamic. Between 2007 and 2025, the urban areas in India are projected to gain 195 million inhabitants. In 2050, India will have the second largest urban population of 0,9 billion, with China in first place with 1,0 billion (UN, 2009).

Population information is publicly available for many countries, but the quality and up-to-dateness of the dataset is sometimes questionable, especially if no metadata are provided. In rapidly growing urban areas, the maintenance of up-to-date population data sets is very time- and cost - intensive. This is especially relevant in the case of India where very large cities, rapid urban development, and very complex urban structures demand new methodologies for population data generation.

In this context, remote sensing, especially optical satellite images, plays a significant role as an independent source of information. Space-borne remote sensing provides detailed geospatial information at different resolutions and over time periods that could not be obtained before. In combination with a geographic information system (GIS), remotely sensed information can be used to generate valuable input information for urban population estimation. The required input information includes inventory data including built-up area and information on the socio-economic structure of the city. While physical measures related to buildings can be directly be calculated from the geometry visible on the space born imagery, socioeconomic parameters need to be inferred from a combination of physical parameters and secondary information.

Population information is employed in a range of different applications. In the field of earthquake risk modelling, population data are used to estimate the probable level of casualties. Casualty estimation is an important step to reduce the impact of future earthquakes. The impact of earthquakes on large-scale urban areas has recently been demonstrated by the 2010 Haiti earthquake (magnitude 7,0) which resulted in approximately 92.000 to 300.000 fatalities. The urban earthquake risk of today's cities and tomorrow's megacities result from the combination of high probability of a large magnitude earthquake, vulnerable structures and a high population density. With more than half of the world's largest cities lying preferentially in seismically active regions and more and more people moving to, and living in urban areas of high risk, the severity and the extent of earthquake-related urban disasters will increase in the future. Another field of application is development aid in which population is an important variable that if carefully assessed and analysed can help to target interventions to reduce poverty and improve living conditions in developing countries.

In this dissertation thesis, a tiered method for population estimation and modelling is developed which integrates information extracted from satellite image and secondary information. This method aims to provide information on the population and its distribution on three spatial tiers: (1) City level, (2) district level, and (3) building level. At each tier, different models are provided for which processing time, input data requirement and associated costs increase with information detail. The rapidly growing urban agglomeration of Ahmedabad located in the state of Gujarat in northwest India is selected as the test site for the methodology development. The city of Ahmedabad is an example of a dynamic metropolis with a population of approximately 3,5 million in 2001 and an annual population growth rate of 2,4%.

With the developed methodology, Tier 1 operates on city level and three different population estimation models are provided. Model I: City Population includes two methods to estimate population and population density. With model I, the population is supposed to be uniformly distributed in the city area. This generalizing assumption is refined with model II, which considers the extent of the built-up area of the city which is extracted from satellite images. Model III considers the different occupancy types in the city. With model III, two procedures for occupancy based population estimation and modelling are tested for the city of Ahmedabad. Tier 2 operates on district level and three models are developed to generate population data for individual districts. At tier 3, models are provided to generate population data on building-block and building level.

1.1 Problem statement

An exhaustive per-dwelling enumeration acquired through fieldwork is the accepted golden standard for counting people. However, a detailed building survey is a time- and cost-intensive task and is therefore in many countries only conducted every 5 to 10 years. This results in time gaps for which no population information is available. Especially in very dynamic urban areas with rapid population growth, population data need to be updated more frequently in order to provide accurate information. A consistent methodology is needed to generate urban population data for inter-census time periods which allows for estimating and modelling the population without relying on detailed building surveys.

1.2 Goals of the study

The overall goal for this dissertation research is to **develop a tiered estimation method for urban population that combines optical satellite imagery and census data to provide population information for large cities**. The tiered method is developed for the Ahmedabad case study.

In order to address this overarching goal, the following specific goals are identified:

1. **Identify the most common spatial levels population data are applied on.**

2. **Evaluate the accuracy of available data for the city of Ahmedabad and generate inter-census population input data set for the year of satellite image acquisition 2008 (target year).**

3. **Analyze and evaluate existing applications of satellite imagery for population estimation and modeling.**

4. **Develop population estimation models for each tier, for which processing time, data requirement and cost increase with population information detail.**

1.3 Limitation of the study

The method only allows for generating population information on city, district and building level. Other spatial scales for example topology – related units relevant for different applications could be implemented.

While this research is focused on population estimation and modeling for Indian cities and employs Indian census data, the implemented models have certain data requirement which have to be met in order to be applicable for a study site.

1.4 Organization of the study

This study is organized by objective topic. Chapter 2 presents the selected study site, the city of Ahmedabad in northwest India. In section 2.3, the underlying data for the development of the population models and data generation methodologies for inter-census population information and a geocoded, administrative map are presented. In section 3.1, existing techniques for population estimation and projection are reviewed. In section 3.2, different applications of population data are presented. Section 3.3 gives a detailed review on sources of population data focusing on the Indian census. In section 3.4, previous studies using remote sensing for population estimations are reviewed. Chapter 4 presents the population estimation model I – III for generation population estimates and population maps which display the population distribution on city level (tier 1). In chapter 5, model IV – VI operating at tier 2 (district level) are developed to generate population information on district level. In Chapter 6, population estimation models which operate on building block and building level are developed. In chapter 7, the findings of this study are discussed and an outlook for future research tasks is given.

The study framework demonstrates the motivation behind each of these chapters and sections. This framework serves as a road map for this study, identifying the aims and objectives, techniques for achieving them and key outputs.

General aim	To develop a tiered estimation method for urban population that combines optical satellite imagery and census data to provide population information for large cities		
Objective	**1**. Introduce study site and underlying data for this study, generate inter-census, and administrative data sets	**2**. Examine existing literature on techniques for population estimation and projection, previous studies on remote sensing and population estimation	**3**. Identify applications of population data and most commonly used spatial levels
How objective relates to aim	Provides location of the study area and overview on the available data and newly generated data sets	Provides theoretical basis for work on this subject	Provides the spatial tiers the population estimation models are developed for
Approach	Evaluate the data quality and data availability, generate inter-census and administrative data sets	Investigate previous work to provide a starting point for population estimation method development	Analyse existing applications with respect to the employed spatial units of analysis
Data required	Information on population (census, statistics), satellite imagery	Previously published research on population estimation, studies on remote sensing and population estimation	Information on existing applications in different fields
Sources of data	Online databases, reports from different institutions, GoogleEarth, Digital Globe	Literature review of previously published research	Handbooks, literature review of previously published research
Objective	**4**. Develop population estimation models at tier 1 – city level	**5**. Develop population estimation model at tier 2 – district level	**6**. Develop population estimation model at tier 3 – building level
How objective relates to aim	Provides population estimation models with different complexity, accuracy and data requirement on city level	Provides population estimation models with different complexity, accuracy and data requirement on district level	Provides population estimation models with different complexity, accuracy and data requirement on building level
Approach	Develop population estimation models, starting with the simplest possible case with the least data requirement	Develop population estimation models, starting with the simplest possible case with the least data requirement	Develop population estimation models, starting with the simplest possible case with the least data requirement
Data required	Population data for 2008, admin. boundary for the city of Ahmedabad, satellite images	Population data for 2008, admin. boundary for district of Ahmedabad, satellite images	Population data for 2008, admin. boundary for district of Ahmedabad, satellite images
Sources of data	Digital Globe, objective 1 – data generation	Digital Globe, objective 1 – data generation, objective 4 - results	Digital Globe, objective 1 – data generation, objective 5 - results

2 Study site and data

The aim of this chapter is twofold: firstly to introduce the study site, and secondly to provide an overview of the underlying data used in the study including the generation of new data sets (objective 1).

2.1 City of Ahmedabad

The city of Ahmedabad is chosen as the study site location in northwestern India. It is located in the state of Gujarat, approx. 450 km north of Mumbai. The location of today's Ahmedabad is first documented from prehistoric times (12.000 to 5.000 BC) when it was known as Shwabhramati. The next historical records are of much younger age. In the 11^{th} century, Ahmedabad was known as Ashaval, after its founder Ashaval, a Bhil King. At the end of the 11^{th} century, Ashaval was defeated by the Solakian King Karandev I, who established the city of Karnavati. The Solanki rule over Ahmedabad lasted until the 13^{th} century when Karnavati was conquered by the Sultanate of Dehli (Gillion, 1968). In 1411, the Ahmed Shah I, son of the Muzaffarid Sultan Zafar Khan, Karnavati became the capital of Gujarat and was renamed after Ahmed Shah I into Ahmedabad. Ahmedabad was ruled by the Muzaffarid dynasty until 1575 when Gujarat was conquered by the Mugal emperor Akbar. The Mughal rule lasted until 1650, when Ahmedabad was conquered by the Marathas. In 1818, Ahmedabad was invaded as part of the British conquest of India (Gillion, 1968). The Indian independence movement has strong relations to Ahmedabad. In 1930, Mahatma Ghandi initiated the Salt march from Ahmedabad. Following the partition of India in 1947, the city faced communal violence between Hindus and Muslims. After the State of Bombay was divided into Maharashtra and Gujarat in 1960, Ahmedabad became the capital of the state of Gujarat. Unlike Bombay, Calutta, Madras, and Kanpur, Ahmedabad is not a creation of the British, but a city which, while remaining true to itself, successfully adapted to the new industrial age, carrying over commercial and industrial skills and patterns of traditional social organization. In no great city of India can the continuity of past and present be seen as clearly as in Ahmedabad (Gillion, 1968).

Due to its geologic setting, Ahmedabad is prone to earthquakes and lies within the moderate damage risk zone III which is specified in the Indian Seismic Code 1893 (Jain, 2002). This zone corresponds to the MSK VII (very strong) intensity. There are historical records of earthquakes in the state of Gujarat from 1819 to 2001. In 2001, a M7.9 earthquake struck the Kutch region in Gujarat, affecting more than 20 districts. The far-reaching ground motions also affected Ahmedabad, approximately 150 km east of the blind trust fold on which the earthquake occurred (Eidinger, 2001). In Ahmedabad, a peak ground acceleration of 0,.11 g was recorded (EERI, 2001). The anomalously high PGA values given an approximate distance to the epicentre of 150 km can be explained by the local, geological setting of Ahmedabad. The city is located in the Khambat graben, which contains several kilometres of Tertiary and Quaternary sediments. In addition, the near subsurface beneath the city consists of thick, alluvial sediment deposited by the Sabarmati River. Liquefaction of the thick alluvial deposits along the Sabarmati River lead to the collapse of mostly modern, middle to high-rise buildings (Saito, 2004). An analysis of historical seismicity in the Kutch region shows a recurrence of approximately 200 years for large magnitude events such as the 1891 Kachchh and 2001 Kutch earthquake.

Besides being prone to earthquakes hazard, Ahmedabad also faced other natural and man-made hazards in the past e.g. extensive flooding in 1868, 1875 and 1927; famine in 1819; 1918 – 1919 plagues and influenza pandemic (Gillion, 1968).

Figure 1: Example of a collapsed building in Ahmedabad after the 2001 Kutch earthquake (Geotechnical Earthquake Engineering Server, 2001).

2.2 Administration of Ahmedabad

There are two local main bodies in Ahmedabad – Ahmedabad Municipal Cooperation (AMC) and Ahmedabad Urban Development Authority (AUDA). The Ahmedabad Municipal Cooperation (AMC) was constituted under the Bombay Provincial Act in 1949 and is the elected urban local government (Bhatt, 2003). The AMC comprises an area of 190 km^2 and is subdivided into 43 municipal election districts across 5 zones (central, east, west, north and south). The limits of AMC were extended three time in the last two decades: towards the east in 1987 (including 92,65 km^2) and twice towards the west in 2006. The Ahmedabad Urban Development Authority (AUDA) was constituted under the Gujarat Town Planning and Urban Development Act, 1976 in 1978 to regulate and monitor the development in the periphery of the cooperation limits and the adjoining 300 villages and 9 municipalities.

2.3 Underlying data

This section presents an overview of the different datasets used in this study. Section 2.3.1 provides an overview on the employed satellite images and the technical characteristics. A short overview of previous studies on Ahmedabad used as sources of information is presented in section 2.3.2; an extensive list is presented in Appendix II. Section 2.3.3 and section 2.3.4 describe the generation of datasets necessary for the population estimation which are not readily available. In section 2.3.3, an inter-census population dataset is generated for 2008; in section 2.3.4 a geocoded, administrative dataset of the AMC area is developed.

2.3.1 Satellite imagery

The identification of small scale parameters with sufficient accuracy is feasible using satellite images with an appropriate spatial resolution which depends on the size of the parameter to be extracted. For individual residential buildings, the spatial resolution of the imagery should be less than or about 1 m in order to be able to extract the building outline with sufficient accuracy. As the introduction of processing and cost efficiency is highly desirable at all tiers of modelling, testing different types of satellite images becomes mandatory. Appendix I lists all currently available satellites and some future satellites.

Based on this list and the regional availability of satellite images for the study site, two image types are selected for testing. Quickbird is selected to represent the very-high resolution satellite image type and Landsat 5 is selected as an example for moderate resolution images. Quickbird is one of the highest resolution, commercial sensors available at present, with a spatial resolution of 2,4 m for the multispectral bands and a panchromatic band of 60 cm. Landsat 5 is the fifth satellite of the Landsat program which started in 1972 with the launch of Landsat 1. Landsat 5 has 7 spectral bands and a spatial resolution of 30 m, except for 120 m for the thermal infrared band (see Table 1 for more technical details). The Landsat images are available online. The largest selection of Landsat images can be downloaded from the Global Land Cover Facility webpage (Global Land Cover Facility , 2009) or from the US Geological Survey (US Geological Survey, 2009). In contrast to the Landsat 5 images which are freely available, the Quickbird satellite is owned by Digital Globe and images have to be purchased from regional resellers. To give an idea of the associated costs: the GAF AG charges 16,18 € for a single user license for 1 km^2 of Quickbird bundle image (multispectral, panchromatic, orthorectified) (GAF AG, 2008). The minimum extent is 25 km^2. For this study, a Quickbird image scene covering 190 km^2 is needed. However, the Quickbird image scene with sufficient quality in terms of cloud cover (0%) is only available for 37 out of 43 districts. Therefore, a 136,51km^2 image scene is purchased from GAF at a price of 2200 €. In contrast, the Landsat image scene is freely available. In this study, the estimation of the urban population is limited to the 37 districts. However to simplify matters, this area is referred to as the city of Ahmedabad or the AMC area. In section 4.2.1, the built-up area of the city of Ahmedabad is extracted from Quickbird and from Landsat 5 TM imagery.

For the accuracy assessment of the results, no ground truth reference information is available because a survey to collect information on the spectral characteristics of the study site was not feasible. Instead, a set of reference pixels is generated and classified using the panchromatic Quickbird image. It is important to note that this kind of validation data set is biased by the subjective interpretation of the image. Nevertheless, the generation of sample data set follows the same statistical principles as a real sample data collection in the field.

Table 1: Quickbird and Landsat 5 TM characteristics (Digital Globe, 2007; US Geological Survey, 2009).

Sensor	Quickbird	Landsat 5 TM
Launch Date	October 18, 2001	March 1, 1984
Orbit Altitude	450 km	705 km
Revisit Time	1 – 3,5 days	16 days
Radiometric Resolution	11 bits	8 bits
Acquisition date	31.05.2008	23.05.2009
Acquisition time	06:06:56	n.n.
Spatial Resolution	Pan: 61 cm (nadir) to 72 cm (25° off-nadir) MS: 2,44m (nadir) to 2,88m (25° off-nadir)	MS: 30 m (nadir) TIR: (120 m)
Spectral Resolution	Blue: 450 – 520 nm	Blue: 450 – 520 nm
	Green: 520 – 600 nm	Green: 520 – 600 nm
	Red: 630 – 690 nm	Red: 630 – 690 nm
	NIR: 760 – 900 nm	NIR: 760 – 900 nm
		MIR: 1550 – 1750 nm MIR: 2080 – 2350 nm TIR: 1040 – 1250 nm

2.3.2 Previous studies on Ahmedabad

An overview on Ahmedabad's urban history is provided by Gillion (1968). The history of Ahmedabad displays the long way of transforming from a traditional centre of trade and industry into a modern industrial city. Gillion (1968) emphasises the importance of the period of Ahmedabad's industrialization and point out the consequences from the decline of the mill industry for the urban development. Rice (1958) gives a detailed analysis of the experiment of introducing of a production system, so called socio-technical system in some of Ahmedabad's textile mills. This analysis gives valuable insight to the social conditions of industrial Ahmedabad. Focusing on the social consequences of Ahmedabad's industrialization, (Hamesse, 1983) performed a detailed population analysis. This analysis includes information on location and size of slums, population density in different districts, housing and dwelling. Information on the interrelation between different areas within the city and between the city and its surrounding rural areas is provided by Hamesse (1983). Hamesse (1983) analyzed the structure of different parts of the city, providing information on infrastructure, occupancy and residential pattern. There are also studies available on perception of the population of their living environment. Desai (1985) conducted field survey in the walled city of Ahmedabad to obtain information on health, housing and environmental quality of living. A more recent study Pandya et al. (2001) provides an overview over Ahmedabad's culture including details on architecture and building types from different epochs. Most of the available information on Ahmedabad is not up-to-date. However, it proves useful for analyzing trends, calculating averages or to obtain a general idea of Ahmedabad's urban structure. A comprehensive list of the previous studies on Ahmedabad used in this study is presented in Appendix II.

2.3.3 Generation of inter-census population data for 2008

The introduction of satellite images as an information source requires that the target year for the population estimation is set to the image acquisition year. Therefore, all modelling processes in this study have the target year 2008. This leads to the need to project the underlying population input data from last Census in 2001 to 2008. Projected population data for Ahmedabad are available from the Statistics Department of the AMC. Since the quality of this population projected in unknown, the *vital rate procedure* is used in this section to generate validation population counts to verify the quality of available population data.

Vital statistics are very often available for countries, in some cases for cities, and occasionally for districts and sub-districts. In this study, most recent records on birth or death rate are available for 2006, so the validation of the AMC population projection is conducted for the year 2006 instead of the target year 2008. The premise of this method is that the observed change in birth or death rates between most recent census and the estimation year for a reference area will occur to roughly the same degree for the selected study area. As the reference area the state of Gujarat is selected and the study area is Ahmedabad. Table 2 lists all required data sets for the population projection using the *vital rate procedure*. A detailed overview of the input data and equations utilized is expanded in Appendix III. In order to ensure that the selected reference and study area fulfil the underlying premise of similar population change, the death and birth rates for Gujarat (urban and total) and Ahmedabad (total) are displayed in Figure 2 and Figure 3. For the death rates, it becomes obvious that there is general declining trend from 1991 to 2006 although some fluctuations are observed in 2002. For the birth rates, the general declining trend is more evident than for the death rates. The birth rate for Ahmedabad in 2006 was the lowest since 1991. Since the underlying assumption of the *vital rate procedure* is fulfilled by both measures showing similar trends for Gujarat and Ahmedabad, this procedure is applied to estimate the total population for Ahmedabad in 2006. Ahmedabad is an urban area, so it is assumed that the population estimation based on the urban birth and death rates of Gujarat are closer to reality than the estimates which also include rural areas. Using the birth rates of Ahmedabad and the urban birth rate of Gujarat, the population of Ahmedabad in 2006 is estimated to be 4.042.413.

Table 2: Datasets required for using the *vital rate procedure* for population estimation after Rives and Serow (1984).

Required dataset	Available dataset	Source
Death rate of study area	Death rate of Ahmedabad 2001 and 2006	Statistical Department of AMC
Birth rate of study area	Birth rate of Ahmedabad 2001 and 2006	Statistical Department of AMC
Population of study area	Population of Ahmedabad 2001 and 2006	Statistical Department of AMC
Death rate of reference area	Death rate of Gujarat 2001 and 2006	Directorate of Economics and Statistics, Government of Gujarat
Birth rate of reference area	Birth rate of Gujarat 2001 and 2006	Directorate of Economics and Statistics, Government of Gujarat
Population of reference area	Population of Gujarat 2001 and 2006	Directorate of Economics and Statistics, Government of Gujarat

From the population estimates based on the death rates of Ahmedabad and the urban Gujarat death rate, the population of Ahmedabad in 2006 is estimated to be 3.874.434. In a next step, the results are averaged and then compared to population projection data. In 2006, the averaged urban population in Ahmedabad is estimated to be **3.958.423**. The Statistics Department of Ahmedabad Municipal Cooperation (AMC) projected the population of Ahmedabad in 2006 to be **3.894.430**. Table 3 displays the AGR applied by the Statistics Department of AMC to project the population for Ahmedabad based on the census 2001. The average urban population of 3.958.423 for 2006 calculated in this section using *the vital rate procedure* is very close to projection by the AMC, with a deviation of 1,6 % i.e. 63.984 people . These results show that the population projected for 2006 by the AMC can be verified by the very similar results from population projection for 2006 using the *vital rate procedure*. From this, it can be concluded that the AGR applied for the projection (see Table 3) are sufficiently accurate for estimating Ahmedabad's population.

In this study, birth and death rate are not available for 2008; therefore a population projection using vital rate procedure is not feasible for 2008. However, considering the minimal deviation of the projected population counts from both methods for 2006, it is assumed that the 2008 from Statistics Department of AMC is sufficiently accurate to be used as input data in this study. The projected population for 2008 provided by the AMC is **4.055.085**. It has to be noted that the spatial extent considered in the population projection has to match the image extent. Therefore, 6 districts of the AMC area have to be excluded which are not covered by the satellite imagery. The total population of the remaining 37 districts is calculated to be **3.152.108**. This estimate serves as the basic input population for model I to III (city level) in chapter 4.

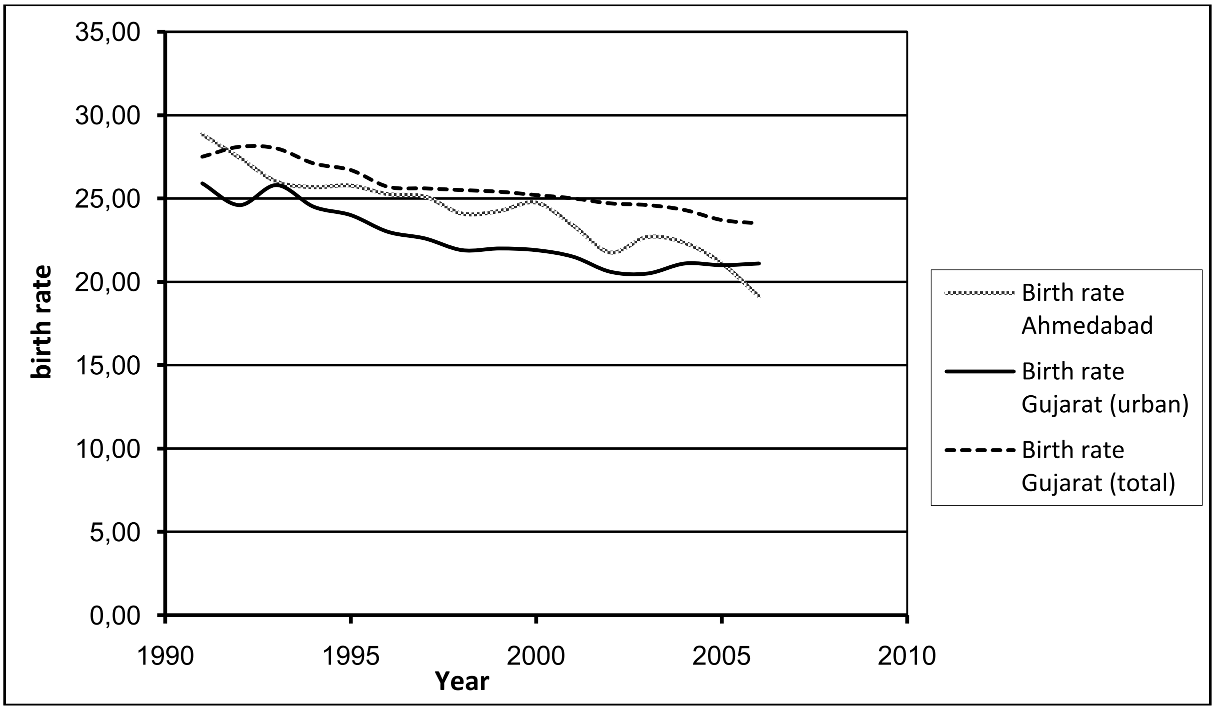


Figure 2: Birth rates of Ahmedabad and Gujarat 1991 – 2006. The birth rates of Ahmedabad and Gujarat display the same declining trend.

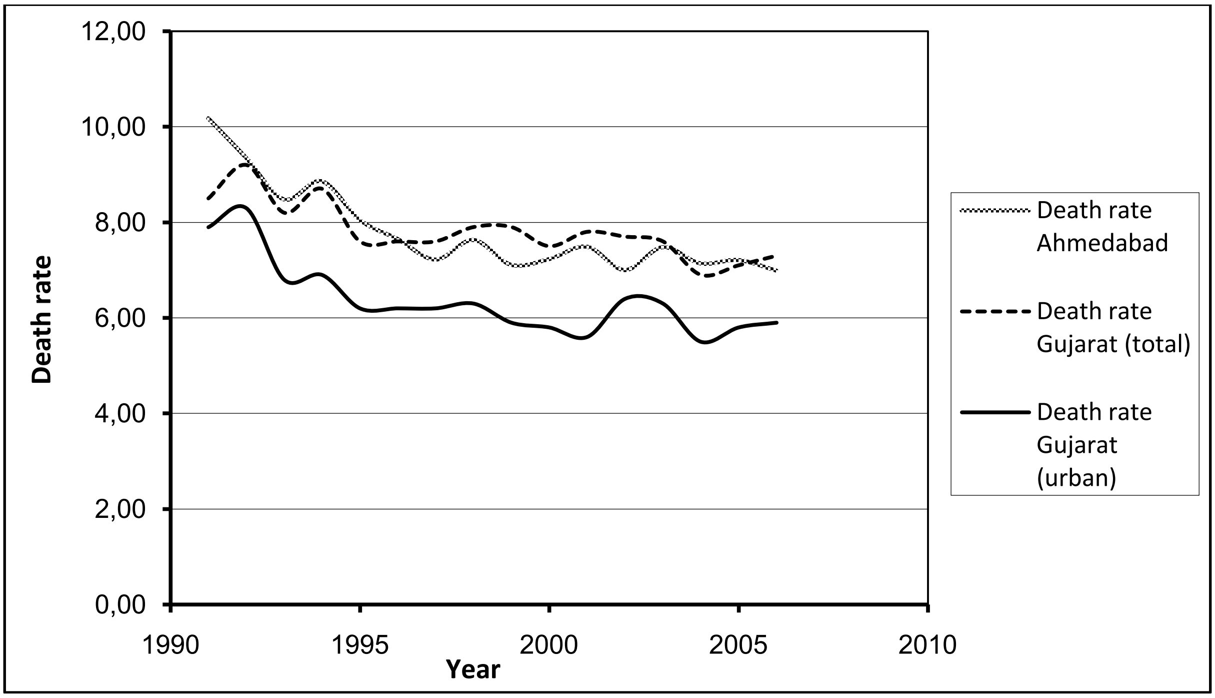


Figure 3: Death rates of Ahmedabad and Gujarat 1991 - 2006. The death rates of Ahmedabad and Gujarat show a generally declining trend.

Table 3: Annual growth rates for Ahmedabad based on Census 2001 population count (Statistics Department of Ahmedabad Municipal Cooperation (AMC).

Year	Annual growth rate (AGR) in %
2000 – 2001	2,03
2001 – 2002	2,05
2002 – 2003	2,04
2003 – 2004	2,04
2004 – 2005	2,04
2005 – 2006	2,04
2006 – 2007	2,04
2007 – 2008	2,04

2.3.4 Generation of geocoded, administrative map for Ahmedabad

In this study, the generation of the population distribution maps is GIS-based. Therefore, the administrative boundary dataset has to be digital and geocoded in order to be suitable for the analysis of the spatial population distribution. For the AMC area, there are a limited number of sources for digital, geocoded dataset on city scale. Table 4 provides an overview about the data available from different sources and the associated costs. The prices listed in Table 4 reveal that digital city maps are disproportionately expensive, in particular as the up-to-dateness and accuracy of the products is unknown in advance.

In the following, a methodology for generating a geocoded, administrative map for Ahmedabad including city and district boundaries combining an analogue paper map purchased from CE Info Systems (P) Ltd and information available from Google Earth is presented. The accuracy of the generated data set is validated using district area information available from Census 2001. Based on an administrative paper map for Ahmedabad, the district boundaries are manually digitized from Google Earth images in combination with the digital street layer available from Google Earth. The kml file is then exported to a shape format, imported into ArcGIS and for each district the area is calculated. The validation reveals that the deviation between the overall area from Census 2001 and from the generated district map is 8,35 km^2 which corresponds to a deviation of 6,51% (see Table 5). On district level, the deviation range from an underestimation of 0,89 km^2 to an overestimation of 1,03 km^2 (see Figure 4). The Sabarmati district (no. 15) shows an exceptional high overestimation by 7,28 km^2 because of the airport area, which is not included by the AMC and difficult to exclude when digitizing from satellite imagery since additional map features like street are not available due to security reasons.

Table 4: Acquisition costs for geocoded, administrative maps for the AMC areas (CE Info Systems (P) Ltd, 2008; Biond Software Technologies, 2008; Risk Management Solutions, 2009).

Area covered	URL	Scale	Format	Costs €
Ahmedabad City ca. 190 km^2	www.mapmyindia.com	1:67.500	Paper map	100
Ahmedabad City ca. 190 km^2	www.biondsoftware.com	1:1	Geocoded data	1000
Ahmedabad City ca. 190 km^2	http://www.rms.com/	500 m grid	Grid	1000

Table 5: Administrative map and total area generated in this study, excluding 6 districts not covered by the satellite image acquired for 2008.

Administrative Map	Source	Area (km^2)
AMC district	Census 2001	128,16
Generated district map	This study	136,51

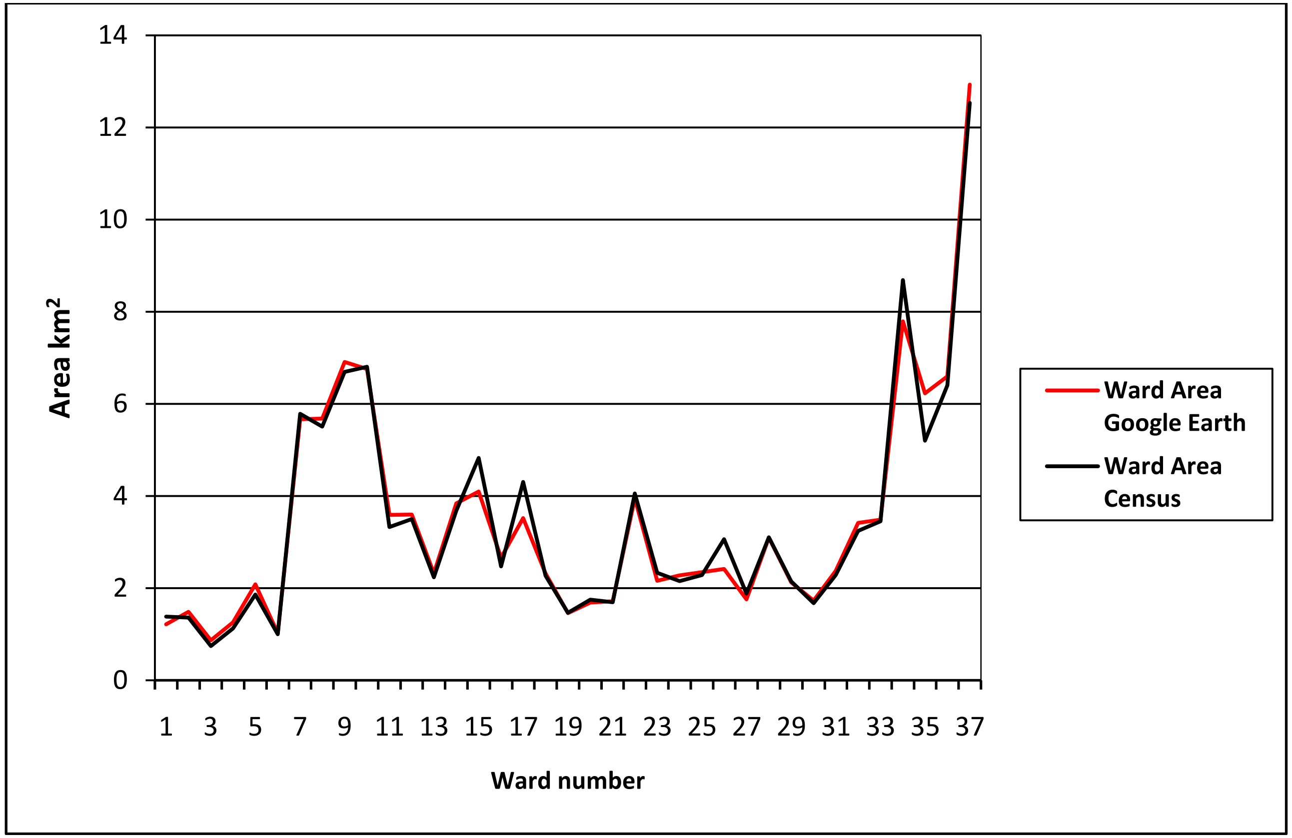


Figure 4: Comparison of district areas for the AMC area. The highest deviation is observed for district No. 15, in which the airport area is included by the AMC but is difficult to exclude manually. The high deviation are also observed for districts which do not have a large street network and distinctive landscape features which can be used as reference points for digitization.

3 Literature review

The literature relevant to this study comprises the following main subject areas: techniques for population estimation and projection, applications of population data, and the role of remote sensing in population estimation (objective 2). The purpose of this chapter is to investigate existing population estimation and projection methodologies and previous applications of remotely sensed data to this field to identify existing methodological gaps which apply to this study. The spatial levels of different applications for population data are identified and serve as a basis for the methodology development in chapter 4 to 6.

3.1 Review of techniques for population estimation and projection

Despite the long history of population estimates - the first historical evidence dates back to Babylonian times 3800 BC (Statistics Canada, 2009) - only a handful of methodologies are documented in the literature. To avoid confusion, it is important to emphasize the main difference between population estimation and population projection. Population estimation is used for the present and the past, whereas population projections are used to guesstimate future population size (Hardin et al., 2007).

In the field of demography, population is generally defined as "the collection of persons alive at a specified point in time who meets certain criteria" (Preston et al., 2008). In general, population has three facets: (1) absolute size, (2) distribution, and (3) density. These three facets related describe a population at specific point in time. However, information on the present population is only of limited help to make prediction about the population in the future. Therefore, population growth - the change in population size – is a focus of demographic research. There are two immediate causes of population growth: People are born into the world, increasing the population and people die, decreasing the population. If the so called "social mobility" is considered, the net migration also influences the population growth (Bryan, 2004). The emerging megacities are an illustrative example of how net migration substantially influences population growth. The very rapid population growth in these areas is not due to extremely high birth rates or extraordinary good health of the city's inhabitants but due to the large number of rural migrants. Net migration can be defined as the imbalance between people moving in and people moving out of a specified area for which the population growth is calculated. These three influencing factors are included in the basic population growth equation (Bryan, 2004) (see Equation 1).

$$P_2 - P_1 = B_{(1,2)} - D_{(1,2)} + M_{(1,2)}$$ **Equation 1**

Where:

P_1 = earlier population count
P_2 = later population count
$B_{(1,2)}$ = number of births in the specified time interval (1 - 2)
$D_{(1,2)}$ = number of deaths in the specified time interval (1 - 2)
$M_{(1,2)}$ = net migration in the specified time interval (1 - 2)

Population projections are an important demographic technique to show the future development of a population based on a set of assumption regarding the future course of fertility, mortality, and migration. The quality of projection strongly relies on whether the underlying assumptions are correctly made and whether the prediction corresponds to the subsequent events (Preston et al., 2008). A number of population projection methodologies exist, ranging from very simple approaches to very complex calculations. A simple technique which allows projecting the total population for a time T when the total population is know at a time 0 by aggregating fertility, mortality and mobility to a single measure of demographic change – the average annual growth rate (AGR) (Rives & Serow, 1984). This technique is commonly referred to as *mathematical extrapolation*. The AGR is calculated for a reference period, which is normally the time period between the two most recent censuses. In case the average annual growth rate (AGR) is unknown, the simplest projection is to assume the population size will remain constant in the considered time period. As the annual growth rates is most often a few percent or less, this assumption provide a fair approximation for short time periods. However, for urban areas with very high population growth this assumption cannot be made.

Another technique for population projection which has modest data requirements is the *vital rate procedure*. Vital statistics are very often available for countries, in some cases for cities, and occasionally for districts and sub-districts. In section 2.3.3 of this study, the *vital rate procedure* is used to estimate the urban population of Ahmedabad in 2006. Another approach for projecting the total population is the *housing unit procedure*. The *housing unit procedure* can be used to project the population of cities, census tracts, zip codes and other spatial areas. It is based on change in the housing stock of an area from the base date to the projection date. The total population is subdivided into household population and group quarters population like hospitals, military barracks, or dormitories. Additionally, the specification of the vacancy rates and average household size is required. Once specified, the housing unit change is multiplied by the occupancy rate and average household size to estimate population change. As a refinement, separate estimates can be constructed by housing types like single – family dwelling, 2 to 4 units, 5+ units. Recent examples for applying the housing unit method can be found in Smith and Cody (1999). Compared to the *mathematical extrapolation* and the *vital rate procedure*, there are more potential problems with source, types, and accuracy of data in the application of housing units. Many demographers believe the *housing unit method* is an inferior estimation method with a district bias and severe data limitations (Morrison, 1971; Brockway & Wurdock, 1981). In contrast, Smith and Mandell (1984) point out that the poor estimates produced by the housing unit method in some cases are results of the specific data and techniques not of the method itself. Naturally, the three components (occupancy rate, number of persons per household, and group-quarters population) are never known exactly. Many techniques and data sources can be used to estimate each component, which gives rise to the great diversity that exists among population estimates by the housing unit model. Apart from the techniques discussed in this chapter, procedures for estimation and projection of population with more substantial data requirements have been developed. An overview of more advanced population projection methods can be found in Preston et al. (2008).

For example, the *ratio-correlation method (R-CORR)* is based on the assumption that the ratio of the share of the total population represented by a given subarea for adjacent time periods (typically the most two recent census) is a function of that subarea's share of several other symptomatic variables which are likely to mirror change in population. Within a defined geographic region, a changing ratio is assumed to be function of changing population ratio (Plane & Rogerson, 1994). Ideally, these variables should temporally co-vary with population change in a predicable fashion (Hardin et al., 2007). For example, the number of residential building permits should be positively correlated to an increasing population in the considered administrative unit. Other symptomatic variables commonly used include school enrolment, vital events, or tax returns. The regressions coefficients are estimated across all subunits included and are assumed to remain constant over the period since the most recent census (Rives & Serow, 1984). A detailed discussion on data requirements and computational aspects can be found in Rosenberg (1968) and Pursell (1970). The *component method* was initially developed by the US Census Bureau to produce individual estimates for the different components of population change. An enhanced version the *component method II* is a type of component method that results in a population estimate through the use of data on the three major components of population: birth, deaths, and net migration (Raymondo, 1992). With this method, the base population is updated by taking into account birth and death in the estimate areas between the census data and the estimate date, and by estimating the amount of net migration in the areas. The greatest difficulty with this approach is correctly specifying this migration factor (Hardin et al., 2007). To estimate the net migration, other administrative data like school enrolment are employed.

Beside aforementioned methods which mostly employ vital statistics, a number of population estimation techniques using *urban occupancy categories* exist. In 1985, the Applied Technology Council (ATC) published the report "ATC – 13 Earthquake Damage Evaluation Data for California" and proposed an inventory methodology that includes a technique for estimating the number of occupants during daytime (3:00 pm) and night time (3:00 am) (ATC, 1985). In this approach, the building square footage is multiplied by an occupancy factor to calculate the number of occupants per 1000 Square Feet. This factor is calculated from US census data for different occupancy categories. The prerequisite for this method to be applicable is extensive information on e.g. employment, residential population, and industrial facilities, and an inventory data set with information on the single building square footage. Another approach is presented as part of the casualty estimation module in the risk assessment methodology Hazards U.S. Multi-Hazard (HAZUS-MH) developed by the US Federal Emergency Management Agency (FEMA, 2008). Relationships between occupancy categories and population distribution are provided for three times of day an earthquake might strike: night time scenario at 2:00 am, day time scenario at 2:00 pm, and commute time scenario 5:00 pm. The night time scenario is expected to generate the highest casualties for the population at home. Using the provided relationships, the population in different occupancies can be calculated for the three time of day. There are two multiplies associated with each relationship: (1) The fraction of a population component present in an occupancy, and (2) the fraction of the population component that is in- and outdoors. The default relationships are based on the US census tract level data. The data include population information distributed into six basic occupancy groups: residential, commercial, educational, industrial, hotel, and commuting population. A more simplistic approach for occupancy based population estimation is presented by (Coburn & Spence, 2002) as part of the lethality ratio procedure, which defines the ratio of the number of people killed to the number of occupants present in collapsed buildings. The lethality ratio is calculated for different building types. Using occupancy curves, the occupancy of residential and non-residential buildings is estimated as percentage of the total population. For different times of day, default occupancies of buildings by rural, agricultural population and default occupancies of residential and non-residential buildings by urban population are provided. In this study, the occupancy based population methods developed within HAZUS – MH and the occupancy curves provided by Coburn and Spence (2002) are employed in the population estimation models on city (see section 0) and district level (see section 5.3).

3.2 Review of applications of population data

In this section, the different fields in which population data are used on various different spatial levels are reviewed (objective 3). In the field of development aid, population information is employed to calculate a number of indices. For example, the human development index is calculated by combining indicators such as life expectancy, educational attainment and income. This information is obtained from population surveys. Population data are also of importance in city planning because the need for various local utilities for housing, schools, shopping, and recreational facilities is related to population growth. Unreliable population data lead to a lack of balance between demand and supply of these facilities and services. This also applies to the planning of lifelines for example drinking water and waste water supply. Population growth and urbanization are already challenging water management systems in many parts of the world.

In the field of disaster management, population data are used in all phases of the so called disaster management cycle. In the post-event response and recovery phase, population data are used determine the number of people missing and to allocate the resources needed in affected regions. This applies not only to earthquakes but also to other kinds of natural hazards. Population rise increases the pressure on the population to live in flood prone areas and therefore, increases the potentially affected population. During the last decade, floods have affected more than 1,5 billion people worldwide, which equates to 75% of all people affected by disasters (European Space Agency, 2006). For emergency response, disaster management agencies, and policy makers also need reliable population data to be able to determine what level of resources will be needed in order to effectively respond to future floods. In the pre-event phase, earthquake risk models are used to calculate future earthquake scenarios (see section 3.2.1). In order to determine the expected affected population, earthquake risk models include an assessment of the probable levels of human casualties. In risk management, population data also serve as a basis to assess the vulnerability and resilience, and coping capacity of the population.

3.2.1 Casualty estimation and population data

The primary impact of concern with regard to earthquakes is life safety. For earthquake risk models and loss estimation studies to be useful for earthquake protection, they need to include an assessment of the probable levels of human casualties (Coburn & Spence, 2002). The goal of this section is to identify the most common operational levels of earthquake risk models and the integrated casualty estimation modules (see Table 8) and to explore their population data requirements.

Various earthquake risk modelling software packages which have been developed in the past incorporate casualty estimation modules. The early attempts of casualty estimation relied on judgment, construction data, aggregated population and historical earthquake statistics (Seligson & Shoaf 2003). In 1970, the National Oceanic and Atmospheric Administration (NOAA) initiated two studies to estimate the regional impact for large earthquakes in the San Francisco and Los Angeles areas (NOAA, 1972 & 1973). The studies tabulate casualties per 100.000 people for historic earthquake events (1886 Charleston, South California to 1971 San Francisco, California). The number of predicted deaths is estimated by applying selected historical death rates to the assumed population in various types of structures. For example using the 1971 San Francisco death ratio, night time deaths in wood-framed structures from a M8.3 earthquake on the San Andreas Fault is estimated 12 / 100.000. From these studies, the conclusion was drawn that serious injuries could be determined from death estimates using a ratio of 4:1, while minor injuries could be determined using a ratio 30:1.

In 1985, the Applied Technology Council (ATC) published the "ATC - 13 *Earthquake Damage Evaluation Data for California"* report (ATC, 1985). Funded by the Federal Emergency Management Agency (FEMA), the ATC-13 report was developed to provide expert-opinion earthquake damage and loss methodology to estimate local, regional and national economic impacts of a major California earthquake. The report documents provide weighted damage factor statistics for 78 classes of structures. This factor enables the user to estimate the mean damage combing the Modified Mercalli Intensity and the facility class. It should be noted that the damage functions are intended to present classes of structures in California, rather than individual buildings. To estimate casualties, ATC -13 provides injury and death rates related to the building's level of damage or damage state (see Table 6). The rates can be used to estimate casualties when the level of damage and the occupancy level of the building are known. In case the exact occupancy level is not know, ATC-13 provides an algorithm to estimate approximate daytime or night time occupancy load. It is important to note that, due to limited data availability, these estimates consider only two 03:00 am and 03:00 pm. It is assumed that at 03:00 am the greatest portion of population would be at home in bed and at 03:00 pm the greatest portion would be away from home. In contrast, the earlier NOAA study (NOAA, 1972 & 1973) concludes that in dense metropolitan areas the number of deaths and injuries will be greater at peak rush hours, between 04:00 pm and 06:00 pm than at the two times generally used for night and day time population i.e. 03:00 am and 03:00 pm.

In the 1990s, the technology advances facilitated the development of automated loss-estimation tools. The FEMA and the National Institute of Building Science (NIBS) developed the risk assessment software HAZUS (NIBS/FEMA, 1999). HAZUS is a standardized, US-applicable earthquake and multi-hazard loss estimation system. The earthquake model of the latest version HAZUS-MH produces quantitative estimates of losses to buildings and infrastructure, estimates of casualties, displaced households and other population impact measures. In HAZUS, the damage is expressed in term of the probability of a building being in any of four state of damage (slight, moderate, extensive, and complete damage). To categorize injuries, four severity levels are used (see Table 7) (NIBS / FEMA, 2002). In addition, indoor and outdoor casualty rates are tabulated at each injury severity level by building type and damage state.

Besides HAZUS, there are many other automated risk or/and loss-estimation tools (for a comprehensive overview see Table 8). For example, the Early Post-Earthquake Damage Assessment Tool (EPEDAT) was designed by the California Office of Emergency Services (OES) to produce regional damage and casualty estimates for emergency response, especially for southern California.

Table 6: Injury and death rate related to building damage states (ATC, 1985).

Nr.	Damage state	Range	Minor injuries	Serious injuries	Dead
1	None	0	0	0	0
2	Slight	0 – 1	3 / 100.000	1 / 250.000	1 / 1000.000
3	Light	1 – 10	3 / 10.000	1 / 25.000	1 / 100.000
4	Moderate	10 – 30	3 / 1000	1 / 2.500	1 / 10.000
5	Heavy	30 – 60	3 / 100	1 / 250	1 / 1000
6	Major	60 – 100	3 / 10	1 / 25	1 / 100
7	Destroyed	> 100	2 / 5	2 / 5	1 / 5

Table 7: Severity levels for casualties and description of the types of injuries for each level as used in HAZUS (NIBS / FEMA, 2002).

Severity Level	Remarks
1	Injuries required medical aid that could be administered by paraprofessionals. Injuries of lesser severity that could be self treated are not estimated by HAZUS.
2	Injuries requiring a greater degree of medical care and use of medical technology, but not expected to progress to a life threatening status.
3	Injuries that impose an immediate life threatening condition if not treated adequately.
4	Instantaneously killed or mortally injured.

The casualty models in EPEDAT are based on existing models (ATC, 1985; Whiteman et al., 1974) and day / night time population distribution. These casualty models estimate the death and injury rates for any building within a given damage state. Peek-Asa et al. (2000) and Shoaf et al. (1998) developed relationships to estimate the total injury caseload for a moderate-sized earthquake in California from Modified Mercalli Scale alone. The application of these scenario-based models is restricted to earthquake similar to the 1994 Northridge earthquake which was used as a case study.

In 2006, the USGS introduced an automated system to rapidly assess the number of people, cities, and regions exposed to severe shaking by an earthquake worldwide (Wald et al., 2006). This system called PAGER (Prompt Assessment of Global Earthquakes for Response) incorporates a tool to rapidly evaluate the potential of casualties. The casualty estimates are based on the fraction of people residing within each building type at the time of the earthquake. PAGER utilizes the population exposure as derived from the Oak Ridge National Laboratory's LandScan 2005 database as the basic input for estimating the likely number of people exposed to variable ground shaking. Another example is the tool EXTREMUM which uses inventory data extracted from population and building information in 89 regions with 2800 administrative units of the Russian Federation (Shakhramanjyan et al., 2000). A version of EXTREMUM called QUAKELOSS is currently being used by World Agency of Planetary Monitoring and Earthquake Risk Reduction (WAPMERR). It is used to provide near-real time (within 2 hours) estimates of building damage and casualties. A scenario-based building loss and casualty estimation model developed by Bogazici University (Erdik & Aydinoglu 2002; Erdik et al. 2003; Erdik & Fahjan 2006) for estimating a worst case scenario of earthquake on main Marmara fault near Istanbul (Turkey) utilizes population information from the State Statistical Institute.

Meeting state-off-the-art internet application technologies, INLET (Internet-based Loss Estimation Tool) incorporates a risk model to provide online estimates for building damage, transportation impacts, and casualties. INLET was developed by ImageCat. Inc as part of the RESCUE project (RESponding to Crises and Unexpected Events) in 2006. Another project related to loss estimation, LESSLOSS was launched by European Centers of Excellence in earthquake and geotechnical engineering. For casualty estimation, the revised model by Spence (2007) is implemented. In original model by Coburn and Spence (2002), the number of deaths is calculated considering the average number of people in each collapsed building, percentage of occupants indoors at time of shacking, expected trapped occupants, mortality at collapse and mortality post-collapse. This methodology was limited to completely damaged buildings. The revised model allow for distinguishing different building types and damages states.

3.2.2 Findings

The presented review provides examples for different applications of population data. The focus is on the use of population data in earthquake risk modelling and loss estimation. In section 3.2.1, the spatial levels on which population data are used are identified for 23 earthquake modelling software packages and tools. The spatial levels commonly used and the types of population data employed were identified (see Table 8). The spatial levels existing risk modelling tools operate on can be subdivided into two categories: (1) Administrative boundaries such as city, district or sub-district, or (2) urban topology-related such as region, homogeneous zones or neighbourhoods. For the risk models to generate realistic casualty estimations, information on population needs to be provided to the identified spatial levels. However to date, it is common practice to increase the resolution of engineering parameters such as building types by conducting surveys. However, the resolution of population data is increased based on very generalizing assumptions because detailed information are not available. Population and occupancy – related questions are seldom included in the surveys because this information is more critical to obtain due to privacy reasons. In order to be able to generate population data independently of the availability of detailed survey information, the prerequisite of the methodology development in chapter 4 and 6 is that the basic input data for the population models are available at the relevant spatial levels. In general, this kind of information such as vital rate is available for administrative units rather than for urban-typology related units. Therefore, the methodology developed in this study is based on a 3-tiered approach in order to generate population information on different administrative levels. Tier 1 operates on **city level** and is the largest scale applied in this methodology. Tier 2 operates on **district level** and tier 3 on **building level**.

Table 8: Overview of Risk Modelling and Loss Estimation software and the application of population data, sorted by region (modified after Daniell, 2009).

Software / tool	Applicability	Owner	Population data	Spatial level
CATS	Global	DTRI, FEMA	Yes	Country, region, city district,
DEBLA	Global	EUCENTER	Yes	City, district
EmerGeo	Global	EmerGEO	Yes	District
Extremum	Global	Extreme Situations Research Center Ltd.	Yes	Multiple
Quakeloss	Global	WAPMERR	Yes	Settlement Level
OPENRISK	Global	AGORA, USGS, OpenSHA	Yes	Multiple
OSRE	Global	Kyoto University, AGORA	Yes	Multiple
PAGER	Global	USGS, FEMA	Yes	Multiple
QLARM	Global	WAPMERR	Yes	City, district
ELER	Europe	JRA-3, NERIES	Yes	City, district
EQSIM	Europe	KIT	Yes	City, district
SES 2002 & ESCENARIS	Europe	DGPC, Spain	Yes	Country, region, city district,
SIGE	Europe	OSSN, Italy	Yes	Country, region, city district,
StrucLoss	Europe	Gebze IT, Turkey	Yes	City, district
CAPRA	Central America	EIRD		
EPEDAT	North America	EQE International, California OES	Yes	City, district
EQRM	Australia / Asia	Geoscience Australia	Yes	City, district
HAZ – Taiwan	Asia	National Science Council	Yes	Country, region, city district,
HAZUS - MH	North America	FEMA, NIBS	Yes	Country, region, city district,
InLET	North America	ImageCat Inc.	Yes	City, district
MAEViz	North America	University Illinois	Yes	District
REDARS	North America	MCEER, FHWA	Yes	City, regional, district
RiskScape	Australia	NIWA, GNS	Yes	City, regional, district Day and night time

3.3 Review of sources for population data

To determine to what extent existing population data match the data requirements identified in section 3.2, a review of the available population data is presented in this section. Population data are available from a number of sources. The primary source for population information in this study is the Indian census and is therefore addressed in an individual section (see section 3.3.1). In the first part of this section, a short overview of the existing databases for population data and information on human settlements is given (see Table 9). The second part of this section focuses on geocoded population datasets which are important for spatial applications.

The International Data Base (IDB), created by the US Census Bureau's International Program Centre (IPC), is a source of demographic and socio-economic statistics for 227 countries and areas from 1950 to date and additionally includes population projections till 2050. The Demographic Yearbooks of the UN Statistics Division (UNSD) constitutes another source. The population data are based on country-wise data from national authorities since 1948, through a questionnaire dispatched annually over 230 national statistical offices (UNSD, 2010). In addition, the UNSD also provides population data on capital cities and cities with more than 100.000 inhabitants. Apart from the Demographic Yearbooks, the United Nations offer a large number of population data sources (see Table 9). For example, the United Nations Human Settlement Program offers the online database UrbanInfo.

Complementing the statistical data records, a number of institutions and organization have developed geocoded population datasets in the past years. An overview of the available datasets is given by Table 10 and Table 11. In 1992, the Digital Chart of the World (DCW) was developed by the Environmental Systems Research Institute Inc. (ESRI) founded by the US Defence Mapping Agency (DMA). The original DCW is a global, vector base map. The primary data were the Operational Navigation Chart (ONC) Series and information provided by other national military mapping authorities. In 1997, the National Imagery Mapping Agency (NIMA) published an updated version of the DCW database, called Vector Smart Map (VMap0). The populated places layer might be the most extensive dataset on towns and cities with about 5.000.000 features. On the advance of new satellites, the National Oceanic and Atmospheric Administration (NOAA) developed Nighttime Lights datasets from the United States Air Force Defense Meteorological Satellite Program (DMSP) - Operational Linescan System (OLS) taken for a 6 month time period between 1994 and 1995. Using different types of light, four different datasets have been created: human settlements like cities or towns, industrial sites, gas flares, fire and heavily lit boats (Imhoff et al., 1997). A new series of annual global OLS Nighttime lights was produced for the 1992 to 2003 time period (Lee et al., 2004). For details on population estimation using DMSP-OLS data see Sutton et al. (1997) in section 3.4.2.

The first major attempt to generate a consistent global georeferenced population dataset was made in 1995 by the National Center for Geographic Analysis (NCGIA) when the Gridded Population of the World (GPW) project was initiated (Tobler et al., 1995). Since then the GPW data sets were frequently updated in 2000 and in 2004 by Center for International Earth Science Network (CIESIN) (Deichmann et al., 2001; Balk & Yetmann, 2004). The first GPW datasets were solely based on administrative boundaries and population datasets. The number of administrative units considered increased from the second version (GPW2) with 128.000 to latest version with 375.000 units (GPW3). In addition, the population data were matched to the United Nations population estimates. The advantage of the GPW datasets is that is relies on a very simple area-weighted scheme for reallocation and on high quality census and administrative data

(Balk & Yetmann, 2004). A limitation is its coarse resolution of 2,5 arc-minutes, which corresponds to a resolution of approx. 5 km at the equator. In 1998, the Oak Ridge Laboratories developed LandScan in order to overcome the limitations of GWP. In contrast to GWP, LandScan utilizes different sources including DCW, Global Land Cover Characterization (GLCC), night time and high-resolution satellite images plus aerial photographs. In the 2001 version, major improvements were achieved by implementing second order administrative boundaries outside the US. In 2002, the Moderate Resolution Imaging Spectroradiometer (MODIS) land cover database was implemented as in input data base. The LandScan dataset has a resolution of approx. 1 km at the equator and the most recent census data are used to calculate the population distribution. The population counts are calculated by an interpolation method that assesses the relative likelihood of population occurrence in cells on the basis of road proximity, slope, land cover, and night time lights. One advantage of LandScan is its increased resolution. One disadvantage is that LandScan does not distinguish between rural and urban population.

The goal of the Global Rural Urban Mapping Project (GRUMP) is to overcome this problem by developing a model for redistributing population within administrative units by combining data from several sources (Balk et al., 2004). The GRUMP database includes global point datasets on cities and towns attributed with its population and the data source, an urban extent dataset which globally covers the extent of human settlements based on night time light datasets, DCW populated places and others (Imhoff et al., 1997). The population distribution was calculated using a mass-conserving algorithm called GRUMPe that reallocates the people into urban areas. The main advantages of GRUMP are: (1) The use of census data as population input data, (2) the identification of urban areas is not solely based on nightlight images, and (3) the population distribution is more realistic than in GWP or LandScan. Despite all the progress in generate global population data sets, GRUMP still faces some limitations like overestimation the urban extent in order to capture small settlements, problems to identify a globally valid threshold for nightlight and other parameters.

Beside these globally available datasets, other georeferenced datasets exist for different regions. The United Nations Environment Program (UNEP) and partner institutions compiled dataset for Africa, Asia and Latin America according to the GPW methodology but extended it to redistribute the population (for details see Deichmann (1996), Hyman et al. (2000), and Nelson (2004).

Table 9: Overview of databases and products related to population and human settlements provided by various institutions.

Institution / organization	Database / product	Online source
Central Intelligence Agency (CIA)	CIA World Factbook	https://www.cia.gov/library/publications/the-world-factbook
Earthquake Engineering Research Institute (EERI); International Association for Earthquake Engineering (IAEE)	World Housing Encyclopaedia (WHE) Housing Report, tutorials, confined Masonry Network	www.world-housing.net
Earthquake Engineering Research Institute (EERI); International Association for Earthquake Engineering (IAEE); US Geological Survey (USGS)	WHE Pager data base	www.world-housing.net
European Commission	eurostat	http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/
Ministry for Extraordinary Situations (Russia)	EMERCOM	www.fas.org
Organisation for economic co-operation and development	OECD database population statistics	www.oecd.org
US Census Bureau	US Census	www.census.gov
US Geological Survey	PAGER	www.earthquake.usgs.gov/earthquakes/pager/
United Nations	Statistics from the UN Economic Commission for Europe (UNECE)	www.unece.org/stats
	Gender and special population groups data (UNECE PAU)	www.unece.org/stats
	Department of Economic and Social Affairs - UN Population Division	www.un.org/esa/population
	UN Human Settlements Programme – UrbanInfo	www.unhabitat.org www.devinfo.info/urbaninfo
	Cyberschoolbus	www.cyberschoolbus.un.org
World Agency of Planetary Monitoring and Earthquake Risk Reduction (WAPMERR)	WAPMERR database	http://www.wapmerr.org/

Table 10: Chronological list of geocoded population dataset available on different spatial scales developed by different institutions.

Datasets	Year	Source	Remarks
Digital Chart of the World (DCW)	1992	US Defence Mapping Agency's (DMA) Operational Navigation Chart (ONC) **Further Reading**: (Danko, 1992); (Tveite & Langaas, 1995)	Computerised, global maps from georeferenced datasets for settlements, county boundaries and other layers (total 16 layers)
Vector Smart Map (VMap)	1997 2000	National Imagery Mapping Agency (NIMA)	Point dataset for millions of human settlements with administrative attributes
Nighttime Lights	1997	National Oceanic and Atmospheric Administration (NOAA) **Further Reading:** (Imhoff et al., 1997); (Sutton et al., 1997)	Different light types from Nighttime Lights Datasets are used to distinguish human settlements from other light sources.
National Land Cover Data	1992 2001	US Geological Survey (USGS) **Further Reading:** (Vogelmann et al., 2001)	The 1992 project was based on Landsat 5 TM data from leaf-on and leaf-off seasons and additional data from Digital Terrain Elevation Data (USGS), statistical US databases.
Global Land Cover Characteristics (GLCC)	1997	United States Geological Survey (USGS) University of Nebraska-Lincoln Joint Research Centre (JRC)	The data are based on advanced very-high resolution radiometer (AVHRR) data and the built-up areas were taken from NOAA 1994 /1995
Moderate Resolution Imaging Spectro-radiometer (MODIS)	1999	National Aeronautics and Space Administration (NASA)	17 land cover types including urban areas
GLCC 2000	2000	Joint Research Centre (JRC) and others	The data are based SPOT-Vegetation data at 1km resolution and the built-up areas were taken from NOAA 1994 /1995
Land Cover Classification System (LCCS)	2000	Food and Agriculture Organization (FAO) United Nation Environment Programme (UNEP)	The built-up areas are taken from DCW.

Table 11: Chronological list of globally available, geocoded population dataset available on different spatial scales developed by different institutions.

Dataset	Year	Source	Remarks
Gridded Population of the World (GPW)	1995 2000 2004	National Center for Geographic Information Analysis (NCGIA) Center for International Earth Science Network (CIENSIN) **Further Reading**: (Deichmann et al., 2001)	375.000 administrative boundaries considered in GWP3 Population data matched to United Nations estimates
LandScan Global Population Database	1998 2000	Oak Ridge National Laboratories (ORNL) **Further Reading**: (Dobson et al., 2000)	Population distribution calculated based on most recent census
Global Rural Urban Mapping Project (GRUMP)	2004	Center for International Earth Science Network (CIENSIN) International Food Policy Research Institute (IPFRI) World Bank	Considers the extent of urban areas in the population distribution

3.3.1 Census - the main source of population information

The primary conventional sources of population data are national censuses and related vital statistics, commonly collected by field survey, interviews, and / or mail response. According to Newell (1988), a census produces a record of individuals at a particular instant in time. It yields a static, cross-sectional snapshot of the population stock, its size and structure as it was on census night. No international standard for the time interval at which census should be conducted exist so far. The United Nations recommend at least a decennial interval and census years ending with 0 or 1 (UNSD, 2010). At the moment, even highly developed countries complete a population census only every five or ten years. The American decennial census mandated by the U.S. constitution is an example. The United Kingdom also conducts enumerations only periodically every 10 years. The main reason for these time gaps is that a census is a complex undertaking. It requires significant human, technological, and fiscal resources to plan and execute. In general, census taking is much less frequent in most developing nations and some parts of the world – mainly on the African mainland - have yet to complete a population count or record of population distribution.

In comparison, India has an exceptional census history. The earliest historical evidence for population census in India can be found in the Rigveda, an ancient collection of Sanskrit hymns, which date back to roughly between 1750 - 1200 BC (Doniger, 1994). Another evidence for an early census in India comes from the Arthashastra, a ancient manual of statecraft attributed to Kautilya, prime minister during the Maurya Empire between 321 to 185 BC (Scharfe, 1993). The first modern Indian Census was conducted in 1872 and is still repeated decennially today. The Census Act passed in 1948 forms the basis for the census procedure in India after the Independence in 1947. The procedure of the India Census usually consisted of two phases: (1) *House listing Operation* and (2) *Population Enumeration*. In phase 1, a building inventory is generated including data on housing conditions, amenities and assets. In phase 2, all persons who are present in the household during the entire period of enumeration or who are known to be usual residents of the households are enumerated in a household.

The conduction of the last Indian census in 2001 was disrupted by a number of difficulties. Certain areas of the country were snowbound and inaccessible for the first enumeration periods of February 9th and 28th and the second, revision enumeration. Therefore, the enumeration was already conducted in January 2001 for Jammu and Kashmir. The flash flood in August 2000 delayed some enumerations which were carried in May 2001 instead. The Bhuj earthquake struck the State of Gujarat just two weeks before the beginning of the first population enumeration. In some heavily affected district, the enumeration was postponed. The large area to be covered – 28 states plus 7 union territories - and the large population of 1.028.610.328 in 2001 make the undertaking of the Indian census exceptionally challenging. In the 2001 census, about 2 million enumerators and supervisors were involved covering 593 districts, 5463 sub-districts, 5161 towns and 638.588 villages. The processing of the 202 million schedules with 1028 million records took 10 months time, starting in October 2002. To illustrate the dimension of the Indian census, the housing listing phase of the 15th Indian Census 2011 has already started in April 2010 and will be finalized in September 2010. The population enumeration will be conducted between 9th and 28th of February 2011. This shows that the completion of the Indian Census takes approximately 4 years, keeping in mind that the final version of the Census 2001 was published in 2003, two years after the enumeration. Census data provide information on various subjects including population, economy, finance, literacy, sex ratio and others. On country and state level the data a publicly available from the official website of the Office of Register General and Census Commissioner of

India (www.censusindia.gov.in). Detailed information about the definition and terminology used in the Indian census, data collection and processing are available on the official census website as well.

In the second part of this section, the uncertainties and errors associated with population counts are discussed, focusing on the India Census because the Indian Census data from 2001 are used in this study. In general, all census population counts suffer from three kinds of errors: (1) errors of coverage, (2) errors of content, and (3) errors of estimation (Yaukey et al., 2007). Sources for errors of coverage are for example counting some people more than once or, more likely, not at all. Even when enumeration is almost complete on the average, particular types of people like marginal segments of the population are more likely to be overlooked. For example, in 1990 US Census, the Census Bureau found that 5,7% of the black population but only 1,3 % of the white population were omitted from the census (Robinson et al., 1993). Errors of content arise when information about the people being enumerated is either misrecorded by the interviewer or misreported by the respondents. Misrecording can be due to ignorance, a desire to maintain personal privacy, or the anguish of possible consequence such as reduction of governmental welfare. Estimation errors are related to the population estimates calculated based on the data collected. In order to insure that the data recorded are sufficiently accurate a consistent validation of the enumeration is vital. The validation possibilities include post-enumeration surveys, data validation, and comparison with secondary sources. In the 2001 Indian Census, a Task Force on Quality Assurance was appointed to supervise the compliance of quality standards developed by the Office of the Registrar General & Census Commissioner. The quality assurance included the evaluation of internal consistency, comparison with similar data from past surveys and validation with data from external sources. In addition, local knowledge and perception is employed to understand and verify regional trends in population distribution. Another focus of the quality assurance is on complete coverage and geographical linkage of each enumeration block.

Other sources of error are related to the nature of the census. It is important to note that yearly demographic data are sometimes reported for the calendar years, sometimes reported for other annual periods, such as from July 1th to July 1th of each year, and may even be reported for other periods, such as five-year intervals. In addition, it is important to remember that midyear population always refers to the population midway between the beginning and end of the period for which the data are reported. As the decennial census provides information on the population size and its composition only at the "census moment", information on events (birth, death etc.) is not provided. To complement the census, in many western countries a registration system for population events has been developed. In many developing countries registration is either weak or almost entirely lacking. In India, however, the Statistical Offices provide annual vital statistics.

Census accuracy also varies considerably from country to country. The amount of error in the census of the best developed countries ranged from 2 to 5 %. In other parts of the world previously published population data had 77 to 208 % undercounts (UNSD, 2010). Moreover, the lag time between a census count the publication is considerable, a year or more being not uncommon in the best cases. For the Indian census 2001, the census moment was 00.00 hours of March 1st; the final results were released in April 2003. In addition, in certain areas enumeration have been undertaken at different points of time and with different reference dates and no adjustments have been made to the enumerated population so as to bring all of them to the common reference date of March 1st 2001. Beside the accuracy issue, another problem with census undertaking is related to privacy and confidentiality and the related mistrust of the people being enumerated. For example, Sweeney (2000) has demonstrated that anonymous datasets can often be readily re-identified. Re-identification is the process of linking anonymous data to the actual identity of an individual. Using 1990 US census data, Sweeney (2000) proofed that individuals with infrequently occurring demographic values can be re-identified in putatively anonymous datasets.

3.3.2 Findings

In this chapter, a comprehensive overview of population estimation procedures, population data generation and sources for population data was given. The review of existing techniques for population projection and estimation presented in section 3.1 revealed that the accuracy of the population data strongly depends on the data requirements of the existing techniques. However, it is important to consider that the required accuracy of the population data always depends on the application they are generated for. In some cases, the data accuracy might be sufficient for a coarse analysis but not adequate for a more detailed analysis. For example, to assess the assets at risk on city level not every building needs to be considered, but a sufficient accuracy can be achieved by using the built-up area as a generalizing parameter instead. For this reason, a tiered approach to develop population data is proposed in this study, in which for different administrative levels, population data can be generated with different degree of detail. Most of the existing methods presented in section 3.1 cannot be applied to the city of Ahmedabad because a detailed survey was beyond the scope of this study. Therefore, simplifying procedures for estimating the population are developed which rely only on generally available census and other statistical data on city and district level.

3.4 Review of remote sensing and population estimation

Due to the cost, effort, frequency and boundary designation problems associated with census (see section 3.3.1), the utility of remote sensing for population estimates has been continuously explored since the 1950s (see Table 14 for a summary of the studies reviewed in this section). Remotely sensed images of various spatial and spectral resolutions have been employed for population estimation on different spatial scales. In this section, a range of applications from fine to coarse spatial image resolution is presented. From a methodological standpoint, the previous studies are assigned to categories which are introduced at the beginning of this section. It has to be pointed out that in the following review the term semi-automated is used for image analysis procedure which involved some degree of user interaction, whereas visual inspection only relies on the image interpreter abilities, and automated image analysis refers to image analysis algorithms which do not require user interventions apart from starting the program.

Different categorization schemes for population estimation using remote sensing are documented in the literature (see Table 12). In each, different, ambiguous sorting criteria are applied. Lo (1986) specified the spatial unit as the sorting criteria for three out of four categories and defined the last category in terms of analytical processing, called “automated digital image analysis” which implies that the analysis in first three categories is manually conducted. This also applies to the scheme introduced by Hardin et al. in 2007. While the first category refers to a statistical methodology, the others refer to spatial units. However, category two and three indicate so called “real world” spatial units whereas the forth category refers to the spatial resolution of the imagery. A more consistent schema is introduced by Jensen & Cowen (1999) which is based exclusively on spatial units. Here, a new scheme to categorize the previous studies on population and remote sensing is developed, based on the information relevant to this study. The following sorting criteria are introduced: type of imagery, methodological approach, and spatial unit of analysis (see Table 13).

Table 12: Overview of documented categorization schemes for population estimation using remote sensing.

No.	Hardin et al. (2007)	Lo (1986)	Jensen & Cowen (1999)
1	Allometric population growth models	Counts of dwelling units	Counts of individual dwelling units
2	Dwelling unit type as surrogate for family size - dwelling Identification	Measurement of areas of urbanization	Measures of urban extent
3	Land type zone as surrogate for family size	Measurement of areas of different land-use	Land use / land cover classification
4	Pixel-based method	Automated digital image analysis	

Table 13: Criteria utilized in this study for literature review scheme and identified subcategories.

No.	Criteria	Subcategories
1	Image type	Ikonos, Quickbird, Aerial Photographs, Landsat, ETM +
2	Methodological approach	Dwelling unit identification (d), allometric population models (a), landtype zone (l), pixel based (p)
3	Spatial unit of analysis	Local test site / city / country

3.4.1 Fine resolution – from aerial photographs to VHR satellite imagery

The increasing availability of remotely sensed imagery has sparked new interest in using very-high and high resolution images for population estimation. Before very-high resolution imagery was available from space, the highest spatial resolution was achieved in aerial photographs. The idea of using aerial photographs for population estimation goes back to Green (1956) and Green & Monier (1957) who postulated that the social structure of a city could be determined through analysis of aerial photography. Green (1956) suggested that the identification of dwelling types is the first step to the use of air photography for demographic application. In this pioneering research, Green (1956) examined 17 residential neighbourhoods in Birmingham (UK) based on stereo air photography (1:8.000). In this study, Green (1956) developed a photographic key for housing identification, including roof structure and form, overall shape and size of the building, situation of the building i.e. location of the building with respect to the street etc., vehicle accommodation, pedestrian accommodation, shape and size of yards, courts etc. Today, Green's study is recognized as foundation of the *dwelling unit identification* approach. Other studies akin to Green (1956) were conducted by Hadfield (1963) and Binsell (1967).

At this point one should be aware that the term "dwelling" does not necessarily refer to single buildings but to a "housing unit someone is living in". This means a number of dwellings could be co-located within the same structure which has been remotely extracted as a single building. With the terminology not explicitly defined, confusion arises from studies in which dwelling density has been calculated from images which do not allow for single building extraction. Watkins (1984) focused his research on aforementioned problem of correctly counting the number of dwelling units in a multi-unit structure. He developed a photographic key which included guidelines for differentiating between residential and non-residential building and instructions for discriminating between converted single-family structures and archetypical apartment structures. This key was applied to three test sites in Boulder (Colorado, USA). From a 1980 aerial image, a total of 2545 buildings were identified. Compared to the census data, the error rate ranged from 1,64% to 4,91%. Following the previous studies, Collins and El-Beik (1971) used dwelling identification methods based on aerial photographs (1:10.000) to estimate the population of Leeds (UK). All houses within the study area were classified into one of three categories: Semidetached, terraced and back-to-back dwelling types. From 1961 UK census, Collins and El-Beik (1971) derived multiplication factors linking dwelling type to inhabitant numbers. Only half of the enumeration districts were used to derive these factors whereas the other half was saved for validation purposes. Compared to 1961 census data, the average error for terraced houses was + 0,87%, back-to-back houses +0,32%, and semidetached dwelling – 6,4%. The largest error among individual enumeration districts were underestimates among semi-detached home with unexpectedly large families. According to Collins and El-Beik (1971), the accuracy of this approach depends primarily on two variables: the identification of the housing types and the representation of the target areas of the same dwelling category by the calibrated multiplier. Throughout the literature, only a single study is documented in which *dwelling unit identification* is applied to an entire city. Lo (1986) estimated the population in 93 traffic zones in Athens, Georgia (USA) from aerial photography. He used a simplified residential structure scheme that included only a few structural types. Comparison of the estimates with 1980 census data revealed an average population count underestimate of 1,7% per traffic zone. To date, this study by Lo (1986) represents still the state-of-the-art in the *dwelling unit identification* to intra-urban population estimation.

However, the application of aerial photography is not limited to dwelling unit identification. Taragi et al. (1994) conducted a study in Saharanpur City (Uttar Pradesh, India) to analyze the feasibility of population estimation methods for built-up areas in Indian cities using aerial photos of 1:10:000 scale. Since the counting of dwellings from aerial photograph was not feasible, a survey was conducted to determine the population of a test site in the centre of the city. The population density was calculated using the area which was determined from aerial photographs. The results were compared to the census data, an error of 5 % was determined. Taragi et al. (1994) point out that the accuracy depends on the interpretation ability to identify homogeneous built-up areas on air-photos. Another study using orthophotographs to identify built-up areas for population estimation was conducted by Wu et al. in 2006. Wu et al. (2006) conclude from a study using orthophotographs with a 0,61 m spatial resolution that the fine resolution allowed to detect detailed variation between human structures. In this study, Wu et al. (2006) applied pixel-based spectral variation statistics on 3 bands, high-resolution orthophotographs. Texture statistics measure the degree on spectral variation between pixels and indicate the degree of landscape heterogeneity. From a methodological perspective, this study can be categorized as a *pixel-based*. For example, residential areas with small houses and small distances between houses represent a more heterogeneous landscape and usually have high values for image texture statistics. It can be concluded that with very-high resolution images, the major problems arise from the count of dwelling units, not from the identification and count of residential structures.

Space-based products that do possess sufficient resolution, such as Ikonos and Quickbird, are relatively new (1999 and 2000 - 2001, respectively). The new generation of high-resolution sensors meet the criterion of very-high spatial resolution, with a spatial resolution ranging between 1.80 m – 0.6 m for pansharpened images. Despite technical progress in the field of feature extraction and the fact that the high-resolution satellite imagery has the same capability as a high-resolution aerial photograph to be used to count buildings either by visual interpretation or semi-automated analysis, only a few authors link the single building footprint to population estimation. Instead a number of examples are documented on the application of very-high resolution satellite imagery to distinguish zones of similar buildings. Taubenböck and Roth (2006) applied the concept of urban homogeneous zones on IKONOS imagery for sample test sites in Istanbul (Turkey). The number of houses was calculated by combining the remotely sensed built-up areas for each zone and the average number of inhabitants acquired from a field survey.

Souza et al. (2003) evaluated the utility of IKONOS images for population estimates for Sao Jose dos Campus (Brazil) using GIS technologies. They identified intra-urban areas with similar residential occupancy features - so called homogeneous zones. They applied official census sectors within these zones and counted the number of houses for each sector. Despite the high accuracy, this manual technique was tested for sample areas only. Liu et al. (2006) explores the possible correlation between the population density and textures in IKONOS images for a study site in Santa Barbara County (USA). This study is based on the assumption that neighbourhoods with similar housing characteristics tend to have similar population densities. Liu et al. (2006) concludes that despite the strong signature of the housing characteristics on the built-up texture of the image, texture alone is unlikely to be sufficient, and that additional interpretation keys have to be explored to achieve satisfactory accuracy. Another study on image texture of very-high resolution imagery was conducted by Almeida et al. (2007) using Quickbird imagery to investigate the correlation between census population density and image texture. In this study, an object-oriented approach was employed using a street network layer for initial image segmentation into streets and blocks; in the classification 9 building classes were distinguished. The classification yield a Kappa Index of 78 %. Almeida et al. (2007) agree with Liu (2006) that the correlation is not strong enough to predict residential population.

3.4.2 Moderate resolution – from Landsat to SPOT

The moderate scale of resolution proves problematic for estimating population size by methods like dwelling unit identification because individual structures are "hidden" in the pixel. Despite the technical fact that the delineation of single structures from moderate resolution images is not feasible with the exception of very large buildings in the dimension of the pixel size, studies on the application of such imagery for population estimation on building level are documented in the literature. However, the majority of studies focus on *land type zone*, *pixel-based* approaches and *allometric models*. Nordbeck (1965) was the first to discuss the relationship between area and population in terms of the law of allometrics. He concluded that built-up area (A) of a settlement should be proportional to the population (P) raised to some power (b): $A = aP^b$. Lo and Welch (1977) used *allometric models* in combination with Landsat imagery to obtain quantitative population estimates for Chinese cities with 500.000 to 2.500.000 people. In this study, the city boundaries were delineated from Landsat imagery. The authors introduced two methodologies. The multiplication method involves the detection and classification of dwelling units or land use classes. The population per dwelling unit or spatial unit of land use is determined from census. Estimates are generated by multiplication of the population factor with the number of dwelling unit/acre per class to give the total population (Lo & Welch, 1997). The second method involves the image measurement of variables such as built-up area, transportation links between urban areas, which are used in a multiple regression model to predict population. Kraus et al. (1974) were some of the first researches to advance the *land type surrogate* approach for population estimation. Using moderate resolution aerial photographs, they estimated the population for four cities in California (USA). The interpretation schema was limited to four land use types. In order to obtain characteristic population densities for the residential land use categories, information from 1970 U.S. Census was used. For 3 of 4 cities, the population was underestimated by an average of 7,2%. Langford et al. (1991) modelled the 1981 population of 49 districts in four districts of northern Leicestershire (UK) using land type surrogate from Landsat-1 image.

Using an automated image processing method, the image was classified to generate land type zones. An overlay analysis using the classified image and the Census district boundaries was conducted which permitted the area of each type to be tailed for each of the 49 districts. A simple correlation analysis revealed that the total district population was highly correlated with the ordinary residential class.

In 2002, Harvey proposed a *pixel-based* approach using Landsat TM images for two Australian test sites. Using a pixel-based classification, the image was classified into residential and non-residential areas. Reference population data were employed by uniformly distributing the population of each zone across the residential pixels. An expectation-maximization algorithm was used to iteratively delineate the pixel population on spectral indicators and re-estimate the pixel population. Similar large-scale correlations have been reported between population and indicators from later orbital sensors. Sutton et al. (1997) modelled population density with night-time satellite imagery and GIS. He presented two approaches for urban population estimation using night-time satellite imagery. In the first approach, an aggregate estimation of total city population was conducted using DMSP OLS imagery. Here, adjacent pixels are clustered, the cluster's area is measured and then the cluster is overlaid by the census population. For countries without census data, relationships for known cities are used. The second, disaggregated or intra urban approach allocates the aggregated estimates to pixels within the urban cluster using linear relationships between light intensity and population density.

In addition to optical images of various resolutions, there has been research on the application of radar imagery for population estimation. Ali (1997) explored the potential of SAR imagery (15-30 m spatial resolution) for population estimation using Almaz SAR Imagery for Riyadh City (Saudi Arabia). This study is based on the assumption that the population is somehow related to the areas occupied by residential buildings. Using the census population (1986) from Riyadh District Authority for seven out of ten regions, the SAR imagery and topographic maps to distinguish residential from non-residential areas, a regression analysis was conducted. The results showed an overestimation of 20,5% compared to the census data when transferring the relation between area and population to the remaining 3 regions. Henderson and Xia (1997) present a comprehensive overview on the application of SAR measurements for human settlement detection, population estimation and related fields.

Instead of using moderate satellite images in single application, Al-Grani (1996) presents an Urban Geographic Information System (UGIS) which integrates satellite images to estimate the population of the Uraija area (Riyadh City, Saudi Arabia) using the sub-model PEUGIS (Population Estimation using UGIS system). The study site was selected using SPOT imagery. The information for the sample areas were collected in the field: total number of dwelling units, total population, and family size (person per dwelling units). The plots were obtained from plot maps from Riyadh Municipality. The population estimation module is based on a *dwelling unit* approach and uses aerial photographs to delineate single structures. Alternatively, Landsat and SPOT imagery were combined with other data to estimate population size. Various studies (Rashed et al., 2001; Weeks et al., 2004; Weeks, 2005) found that good results could be obtained in Egypt using imagery that combined higher-resolution (in this case 5-meter panchromatic Indian Remote Sensing imagery) with moderate-resolution 24-meter Indian Remote Sensing multispectral imagery. From this merged dataset, the studies were able to distinguish detailed differences among neighbourhoods throughout the greater Cairo area. More detailed data within the dense core of Cairo were obtained by merging 0.6-meter panchromatic and a 2.4-meter multispectral Quickbird images.

3.4.3 Review of remote sensing and building inventory

From the review of previous studies on population estimation using remote sensing in section 3.4.1 and 3.4.2, it becomes obvious that identification of residential areas or at a higher resolution building types play a key role. To date, the extraction of building inventory parameters from remote sensing imagery has been focused on high to very high resolution imagery. Herold et al. (2003) stress the potential of very high resolution sensors for accurate and detailed mapping of urban environments. In this section, an overview will be given of the various theoretical approaches and data types which have been employed for the extraction of geometric building parameters. The recent literature shows that the extraction of building parameters from high to very high resolution imagery is focused on four areas:

- The delineation of building outlines (2D) and shapes (3D) from very high resolution satellite imagery
- The combination of satellite imagery and GIS data for building parameter extraction
- The combination of satellite imagery with elevation information such as SRTM
- The up-date or development of building inventory databases

The application of very-high-resolution satellite imagery for delineating two-dimensional building outlines is widely documented, especially with regard to segmentation and classification procedures. Lee et al. (2003) applied segmentation and classification techniques separately on panchromatic and multi-spectral imagery to identify potential buildings. They point out the difficulty of extracting buildings that have a low contrast to their surroundings. Extensive studies on their geometric characteristics within IKONOS imagery can be found in Baltsavias et al. (2001), Fraser et al. (2001), Son and Dowman (2001). For building extraction based on spectral and shape characteristics, a multi-resolution segmentation procedure in Definiens software was used by Tian et al. (2003) for a study site in Beijing (China). Again, problems were encountered due to spectral similarity between the building and its surroundings.

A range of approaches have been suggested for automated building footprint extraction. For example, (Haverkamp, 2004) used an edge-based procedure to extract building footprints from IKONOS panchromatic imagery. Sohn and Dowman (2001) have proposed a method of extracting polygons from Ikonos images by analyzing the edges of buildings. A software package for built-up area recognition, so called BREC, was developed by Paolo Gamba at the University of Pavia in collaboration with ImageCat. This software is able to automatically analyze, extract and rectangularize building outlines, plus derive approximate heights from remotely sensed images. Another approach is documented by Liu and Prinet (2005). They applied probabilistic models for extracting buildings from high-resolution panchromatic images in dense urban areas. Probability functions are computed for regions extracted from a previous segmentation. An object-oriented procedure for a study site in Brazil using Quickbird imagery is suggested by Centeno and Miqueles (2004). They implemented the methodology of Haala and Brenner (1999) in an object-oriented classification procedure for a study site in Brazil using Quickbird imagery. They derived 6 land-cover classes including trees, grass, roads, yards, roofs and bare soil. Overall, a large range of implementable techniques exist for building outline delineation, which methodologically, is a promising starting point for the rapid and efficient development of inventory measures from remote sensing sources.

In addition to building outlines as a two-dimensional parameter, techniques have been developed to extract three-dimensional shape information from satellite and aerial images. Haala and Brenner (1999) used a detailed digital terrain model based on laser scanning data to define regions of interest (ROI). Combing these ROIs with building outlines, singles buildings are delineated. This approach is limited to flat, gable and hip roofs. This ROI approach is adapted by Ameri (2000) to automatically identify building roofs. Using the methods of least squares flat and roof areas are identified and combined to three-dimensional roof polygons which serve as a basis for a rough building model. The development of TOBAGO (Topology Builder for the Automated Generation of Objects from Pointclouds) and ARUBA (Automatic Reconstruction of Sub-Urban Buildings from Aerial Images) by Henriccson et al. (1996) form the basis for the so called Cyber City Modeller. Building roofs are measured as 3D point clouds. For different buildings types, roof models are calculated and stored in an object catalogue. In case a new point clouds is measured, it is compared to the models stored in the catalogue. The texture of roofs and walls are extracted by projection of the models into the aerial photos. The 3D models developed by CyberCity can be implemented into GoogleEarth. The software package ObeX is developed by Gülch et al. (1999). With ObeX in a first step, a building is selected from aerial photographs and its building outline is defined by explicit points set by the user. For example, in case of a gable roof, three points need to be set: 1 ground point and 2 gable points. Krauß et al. (2007) presented an approach for developing a simple 3D urban model using two high-resolution stereo Ikonons images for test sites in Athens (Greece) and Munich (Germany). The resulting city model does not include texture information.

From this above review, it can be concluded that high resolution of the satellite imagery allows the identification of structures on the ground. However, in terms of their direct applicability for efficient inventory development, few if any studies consider a broad geographic extent. Instead, in the studies described above, very-high-resolution images were applied on study site scale only. The time and cost efficiency of applying per-building methodologies on an entire city has to be questioned in case of the dynamic and fast growing Indian megacities. In addition to the large areas to be covered, problems arise from missing or weak contrast between buildings and their surroundings. This is in particular problematic in dense urban areas for example the historical inner city of Ahmedabad. To encounter this problem, the integration of GIS data has proven in past studies to achieve valuable improvements. The combination of satellite imagery and GIS-based building outlines was proposed by Marangoz et al. (2006). They extracted buildings from panchromatic IKONOS images for a study site in Turkey. They applied an object-oriented approach with a GIS supported segmentation and a classification based on spectral, textural and shape features, with an overall accuracy of 86.11%. Similar results are reported from Fraser et al. (2000) for visual interpretation of IKONOS images. They concluded that approximately 15% of buildings could not be extracted from the images. An example for manual intervention to encounter accuracy problems is shown by Shan and Lee (2002). They developed an approach for post-processing building segments in order to obtain a more realistic shape. The approach combines distance-based and curvature-based generalization to simplify, refine and shape the delineated polygons. Post-processing such as manual adjustments is useful to improve the results for building outline generation but is very time consuming.

Another problem arises from the discrimination between buildings and other elevated objects such as trees. Therefore elevation information was utilized in a number of studies, solely or in combination with satellite imagery. Gamba et al. (2002) used SRTM data to identify buildings. To estimate the underlying terrain, they created a lower-resolution 'bare earth' DEM and subtracted it from the original SRTM layer. The remaining areas with sufficient height were considered as buildings. A similar approach has been used by Brunn & Weidner (1997) to distinguish between vegetation and buildings. Guo and Yasuoka (2002) used Ikonos imagery combined with height information from DMS for building extraction. The increasing geometric resolution of airborne SAR systems offers the opportunity to use these technologies for data generation in urban areas (Balz & Haala, 2005). Height information derived from laser scanner was applied for building detection (Maas and Vosselman, 1999; Haala, 1994). Haala and Brenner (1999) applied a pixel-based classification combining multi-spectral information and geometric information from laser scanner.

In the light of the advances of building extraction utilizing a range of remotely sensed data sources and the variety of techniques and procedures covering different geographic scales, the question of the practical utility of the outputs arises. In the past years, a few studies were carried out using outputs from satellite imagery analysis to develop and up-date building and infrastructure inventories. The main focus is on up-dating and improving pre-existing data, rather than producing the original dataset. Miura et al. (2006) developed a procedure to update GIS building inventory data using remote sensing data for Metro Manila (Phillipines). The building extraction is based on the Canny method (Canny 1986, edge-detection algorithm). The number of storeys was estimated from the building's shadow length. Dutta (2004) presents a methodology to develop an up-to-date building inventory using remote sensing and existing databases for Bangkok (Thailand). This methodology involves an object-oriented image classification. It achieved an overall accuracy 85% for several building types: reinforced and non-reinforced concrete buildings depending on the roof materials and roof types. In a study conducted by Chung et al. (2004), the potential of very-high-resolution satellite imagery to update building inventory in HAZUS was investigated. The results for square footage estimates were comparable to HAZUS default square footage and a more accurate representation of building heights was achieved.

3.4.4 Findings

In this section, the key findings of the literature review regarding the application of remote sensing to population estimation are presented. With numerous population estimation techniques already developed and population databases being available, the question of what are the benefits of employing satellite images for population estimation arises. It has to be pointed out that the following statements are based on the assumption of satellite image being available. However, in some regions of the world and for some specific dates this might not be the case. The most obvious benefit is that existing population data can be refined by using more detailed land use information extracted from satellite image. In case land use information was already considered in the generation of the population data, satellite images can be used to assess the quality of the data. In case no information on the year the population estimate is provided for is given, the data can at least be assigned to a pre- or post-image acquisition time period. Another important added value becomes obvious if one imagines a case of very limited GIS data availability. Basic GIS layers such as administrative boundaries can be generated using satellite image with sufficient accuracy for city and district level analysis (see section 2.3.4 this study). In the context of population estimation, GIS and remote sensing are two complementary data sources, which can, if combined, achieve a higher degree of data quality and reliability. The best combination of remote sensing and GIS strongly depends on the scale of analysis. In essence, the relatively inexpensive, widely available, moderate- and low-resolution imagery has proven to be successful in coarse, large-scale population estimation with considerable accuracy. The task of detailed population estimation, however, generally requires higher resolution imagery.

In order to be able to quantify the overall accuracy of the resulting population data, it is important to identify the sources of error associated with the procedure used. From the literature review in this section, different sources of error are identified. One commonly encountered error is that misclassification of impervious surfaces as buildings due to the spectral similarity. In this study, this problem is encountered with slums and surrounding bare ground. Due to the very similar spectral characteristics, it is difficult to semi-automated extracted slums as a residential land use class. In addition, as the huts in a slum are very closely built, distinguishing individual huts is not feasible. Another source of confusion arises from missing standards for the terminology used in urban remote sensing. For example, the term dwelling does not necessarily refer to single buildings but to “a housing unit someone is living in”. This means a number of dwellings could be co-located within the same structure which has been remotely extracted as a single building. With the terminology not explicitly defined, studies in which dwelling density has been calculated from moderate resolution images are questionable and raise wrong expectations of the potential of satellite images. From a technical perspective, the delineation of built-up areas and impervious surfaces is feasible at resolution of between 80m (Landsat 4 and 5) and 30m for Thematic Mapper and Landsat 7, but the extraction of dwelling units as single building is not. This means that in order to calculated building-level parameters such as building density from built-up area information, secondary information needs to be employed. For example, the delineated impervious surface – in case impervious surfaces such as road are excluded – can be used to calculate the average building density by dividing the imperious surface by the average building area.

In this study, the identification of single structures and the number of dwellings within the structure is not feasible due to the very complex structure of the residential areas. The only locations where the identification of single structures might be feasible are high income areas in which single structures are surrounded by a plot which is covered by some sort of vegetation. Another problem is that in many cases the geographical extent of the previous studies is limited to test sites and thus, the applied methodologies are often only designed for small-scale studies. In this study, the task to estimate the population of Ahmedabad with approx. 4 Million people and a city extent of 190km^2 is very challenging. This leads to the need introduce a certain degree of generalization into the building level analysis.

Table 14: Literature review for remote sensing and population estimation using the categorization scheme developed in this study (D = Dwelling unit identification, A = Allometric population models, L = Land type zone, P = pixel based).

Study	Image data	Method	Study area	Remarks
Al-Garni (1996)	Spot, Aerial photographs	D	Riyadh City (Saudi Arabia)	Using the Population Estimation Urban Geographic Information System (PEUGIS) Al-Garni (1996) estimates the population using field survey data and plot maps, attribute delineation from aerial photographs. SPOT imagery is used for study site selection.
Ali (1997)	SAR	L D	Riyadh City (Saudi Arabia)	The application of SAR for population estimation was tested for 10 districts in Riyadh City. For these districts the population was known from statistical data. Using the distinct reflection of the building corners, the residential areas in the districts were identified and their areas were calculated using planimetric techniques. A regression analysis was performed for 7 districts and the resulting regression equation was applied to estimate the population for the remaining 3 districts. The relative error was approx. 21%.
Almeida et al. (2007)	Quickbird	D	Sao Jose Dos Campo (Brazil)	In this study, an object-oriented image analysis was conducted to identify homogenous residential areas in terms of density of occupation and building standards.
Collins & El Beik (1971)	Aerial Photographs	D	Leeds (UK)	Collins & El-Beik (1971) used the dwelling identification method to estimate the population of the city of Leeds based on aerial photographs. In a first step they applied eight overarching land use categories to classify the image. Each category was then subdivided into more detailed subcategories. For example, the residential land use category was subdivided into 6 building types.
Green (1956)	Aerial Photographs	D	Birmingham (USA)	For the city of Birmingham (Alabama), Green (1956) developed a methodology to identify and classify residential structures from aerial photographs.
Green & Monier (1957)			Rochester (USA)	The methodology developed for Birmingham by Green (1956) was test for the city of Rochester (New York) to evaluate the adaptability of the interpretation key for a different geographic location. This study showed that residential structures can be identified on US census tract level with an average error of 2%, in contrast to 9% for Birmingham. In addition, several indices representing combinations of various building types were computed to establish a ratio for residential, commercial and industrial structures. These were used to derive a socio-economic status classification and ranking. A comparison with the socio-economic classifications from the census data shows a positive correlation.
Hadfield (1963)	Aerial Photographs	L D	Fox River Valley (USA)	Hadfield (1963) conducted a traffic study of the Fox River Valley Region (Illinois) to test the use of aerial photography as a principal source for land use and population data.

Study	Image data	Method	Study area	Remarks
Harvey (2002)	Landsat TM	P	Australia	A pixel-based and a zonal-based model for population estimates are compared. Harvey (2002) uses a pixel-based approach to classify the image into residential and non-residential areas.
Kraus et al. (1974)	Aerial Photographs	L	California (USA)	In this study, the built-up area in the test cities was classified into four land use types including three residential categories. The area of each residential category was calculated and located on the census block map 1970. To calculate the spatial population density random block sample within each land use category were selected to determine population density within these categories.
Liu et al. (2006)	Ikonos	L P	Santa Barbara (USA)	In this study, the correlation between population density and image texture is explored. The spatial unit for the analysis was census block with homogeneous land use.
Lo (1986)	Aerial Photographs	D	Athens (USA)	Lo (1986) used a simplified residential structure schema to estimate the population in 93 traffic zones from aerial photographs. The schema included three categories (1) small single family structures, (2) large single family structures and multifamily structures. For each category resident counts were estimated.
Lo & Welch (1977)	Landsat	L D	China	Lo & Welch (1997) applied allometric models to Landsat image to quantitatively estimate the population for Chinese cites.
Lo & Mesev (2003)	Landsat	A		Lo (2003) attempted to develop a statistical correlation between urban census tract population and the area within each tract, classified as residential from satellite imagery.
Pozzi and Small (2002)	Landsat			Pozzi and Small (2002) attempted to establish a reliable relationship between vegetation and population density.
Rindfuss et al. (2003)				Rindfuss (2003) points out that there is no reason why an association at the administrative level would be necessarily the same as that found at the household level.
Sutton et al. (1997)	DMSPOLS		USA, Korea, Japan, Global	Sutton et al. (1997) found a significant log/log correlation between the light cluster area and the metropolitan estimation in the US and cities with similar level of economic development. This relation tends to decay with smaller populations.
Souza et al. (2003)	Ikonos 2	L D	Sao Jose de Campo (Brazil)	In this study, Ikonos2 data were used to estimate the urban population in inter-census periods in Brazil. The concept of homogeneous zone was applied to identify intra-urban areas with similar residential occupation features. Combining average household occupancy from census and remotely identified housing units, the population was estimated.

Study	Image data	Method	Study area	Remarks
Taubenböck & Roth (2006)	Ikonos	L D	Istanbul (Turkey)	In this study Taubenböck & Roth (2006) conducted a land cover classification to delineate 5 thematic classes. Using the parameters from the classification results they derived homogeneous zone with the two test sites. Using the housing unit method they calculated the number of inhabitants in each zone by using an average of occupants per house obtained by a sample survey.
Taragi et al. (1994)	Aerial photographs (1:10.000)	D L	Saharanpur City (India)	Taragi et al. (1994) tested population estimation methods for built-up areas in Saharanpur City using aerial photographs. In this study a survey to enumerate the dwelling units was conducted and the results combined with built-up areas delineate from the aerial photographs.
Watkins (1984)	Arial Photographs		Boulder (USA)	Watkins (1984) focused on the problem of correctly counting dwelling units in multi-unit structures. In his study he developed a photographic key to differentiate multi and single family dwellings using features such as roof division, outside fire escapes etc. He applied this key to a test site in Boulder (Colorado) using panchromatic imagery. The derived dwelling count was compared to Census data.
Wikantika et al. (2005)	Quickbird	L	Java (Indonesia)	Wikantika et al. (2005) identified different land use types for 6 villages on Java to redistribute existing population data and to recalculate the population density.
Wu & Murray (2004)	Landsat 7 ETM + Orthophotos	L P	Columbus Metropolitan Area (USA)	In their study Wu & Murray (2004) estimated the impervious surface in residential areas using a spectral mixture analysis on ETM+ images. In addition a maximum likelihood classification was used to delineate residential pixels. The population for the residential areas was calculated using a co-krigging interpolation technique.
Wu et al. (2007)	Landsat 7 ETM+	L P	Dayton (USA)	Wu et al. (2007) tested two regression approaches for urban population estimation with remote sensing information. Zonal and pixel-based models were applied to Landsat and ETM images for a test site in Dayton, Ohio.
Wu et al. (2006)	Orthophotos		Austin (USA)	Wu et al. (2006) develop a dasymetric mapping method for remodelling census population. In this method the population density is modelled from texture statistics from remote sensing images with the same land use class.

4 Tier 1 - City level

At tier 1, the population is estimated and its spatial distribution is modelled on city level (objective 4). The proposed approach considers three population estimation models, for which processing time, data requirement and cost increase with information detail. Model I: city population uses an area overlay methodology based on the simplified assumption that population is uniformly distributed across the city (section 4.1). In contrast to this very simple case, model II: urban population employs satellite images to extract the urban area (oppose to non-inhabited regions) in which population is supposed to be uniformly distributed (section 4.2). As this assumption is still a simplification, model III: occupancy based population differentiates occupancy categories within urban areas so that population can be estimated for each of the categories (section 0). It has to be pointed out the model I and model II estimate city and urban population at night time, whereas model III allows for modelling the spatial distribution of the population based on the occupancy for different times of the day.

4.1 Model I: City population

This approach is based on the generalized assumption that the population is equally distributed within the selected spatial unit. In this study, this unit is the administrative area of the Ahmedabad Municipal Cooperation (AMC). With model I, two possible ways exists to calculate the city population. One is simply assigning a population count value to the spatial unit the population was counted for. In case, the extent of the considered spatial does not match the population count areas, the second method - a simple area weighting approach - can be used. The simple area weighting approach is based on geometric intersection of a source and a target zone (Langford, 2006). In Figure 5, an illustrative example is shown for the AMC area. In the left figure, a random target zone lies completely within the source zone and occupies about 20% of its area. Thus, the estimated target zone population is one fifth of the source zone population. In the right figure, the extent of the target zone equals the extent of the source zone. Therefore, the population of the target zone is assumed to be equal to the population of the source zone.

In this study, the target and the source zone are identical and therefore the distinction between source and target zone becomes obsolete and the first approach can be used. For the reduced AMC area, the population is projected to be 3.152.108 for 2008 (see section 2.3.3). The reduced AMC area excludes the districts no covered by the satellite images employed later in this study. The population density is calculated for the reduced AMC areas using this projected population and the reduced AMC administrative area (136,51km^2) calculated in section 2.3.4. The calculated population density for the AMC area is 23.091 people per km^2. The advantages of this approach is that no further secondary information is required. In case the target and the source zone have the same extent, a spatial reference is not obligatory. In case the target zone lies within the source zone, the implementation of model I is straightforward. The only data requirement is a geocoded vector file of the boundaries of the target and source zone. Using ESRI ArcGIS as an example, the first step is to import the boundary shape file. Then the areas of the polygons are calculated and the area of the target zone is calculated in percentage of the area of the source zone. Finally, the population value for the target zone is estimated by using the field calculator function of the attribute table of the vector file.

The area percentage of the target zone with respect to the source zone is used to calculate the population in the target zone as a fraction of the population in the source zone.

A limitation of model I is that no information on the spatial distribution of the population within the selected, geographic region can be obtained. However, model I allow for generating a population map on city level with very basic data requirement – a population value and a geocoded, administrative data set. In case, the administrative data set is not available, an open source, data generation methodology is developed in section 2.3.4 of this study. In case the population value is not available for the required year, a procedure for population projection using vital rates from Indian census is presented in section 2.3.3. If the population map is not required, the population density can be calculated without the employment of some kind of GIS technology.

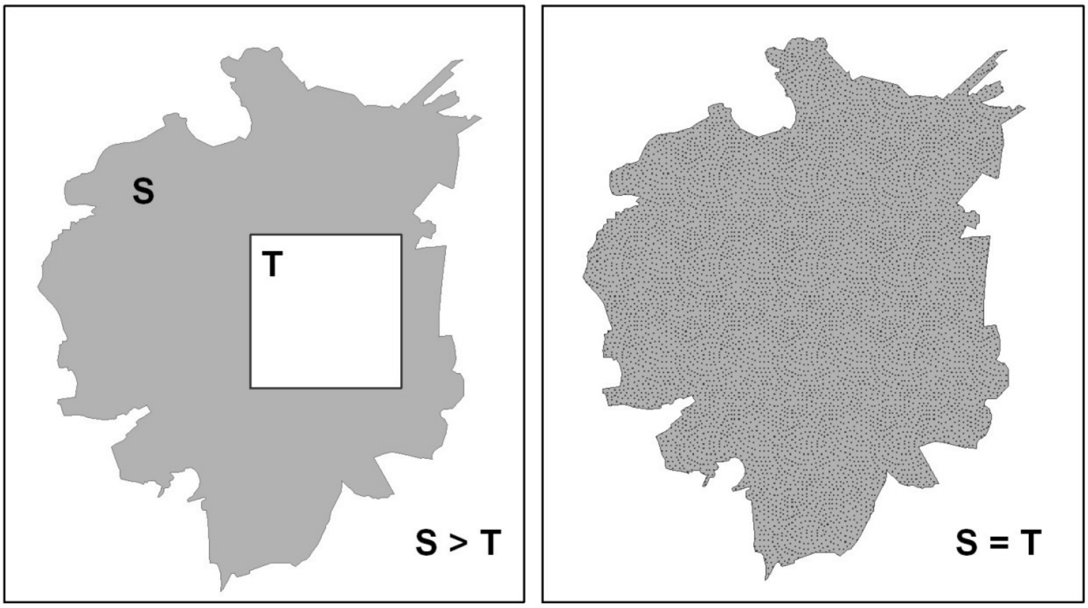


Figure 5: Model I: City population. (Left) Simple area weighting: The target zone lies completely in the source zone within the source zone. The proportional coverage of the target is assumed to correspond to the population fraction of the source area living there. (Right) The target and the source zone are identical and therefore the population of the target zone is assumed to the same as in the source zone.

4.2 Model II: Urban population

The purpose of including secondary information in model II is to obtain a more detailed picture of the city population distribution. Based on the assumption that people reside in the built-up area, information of the urban, built-up extent is employed. In this study, the urban areas are delineated from satellite images. The integration of spatially referenced information in model II requires the use of a GIS environment for the population estimation.

The first step for urban population estimation using model II is the creation of an urban area mask which includes all built-up areas. In the following section, the procedure to create an urban areas mask from Quickbird and Landsat 5 TM satellite images is described. The analysis of accuracy and the associated costs between urban mask creation using very-high resolution

Quickbird images and moderate resolution Landsat 5 TM image is working towards introducing efficiency in the population estimation method.

4.2.1 Creation of the built-up area mask

For the creation of the built-up mask, a commonly used feature to distinguish vegetation from non-vegetation, the Normalized Difference Vegetation Index (NDVI) is explored. Due to the fact that living plants reflect more in the near infrared and absorb more in the read than non living surfaces, the degree of photosynthetically active vegetation within each pixel can be estimated using the NDVI. Based on the assumption that built-up area pixels include significantly less vegetation than non-built up areas, the NDVI can be used to distinguish built up and non built-up areas. Two bands are needed to calculate this index: one containing reflectance values for the visible red (630 – 740 nm) spectrum and the second containing reflectance values for the near infrared (780 – 1400 nm) portion of the spectrum. The NDVI is the quotient of the difference and sum of these two datasets (Mather, 2004). The NDVI values vary in relation to the absorption of red light by plant chlorophyll and the reflection of infrared radiation by water-filled leaf cells. NDVI values range between -1 and 1, with values of 0,5 for dense vegetation and values < 0 indicating no vegetation.

$$NDVI = \frac{NIR - R}{NIR + R}$$

Equation 2

Where:

NIR = reflectance value for the near infrared band
R = reflectance value for the red band.

Due to the very small pixel size (0,61m), the Quickbird image scene has a very large data volume and needs to be divided into two dataset to compute the NDVI. It is computed separately for both image scenes and then merged back together. Due to the larger pixel size (30m) and resulting smaller data volume, this intermediate step is not necessary for the Landsat 5 image. After the a NDVI image is computed from both images, a threshold value for discriminating between built-up areas and non built-up areas is determined by analysing the urban areas in the original images and the corresponding NDVI values of these pixels. Table 15 lists the determined thresholds for the Quickbird and Landsat 5 TM image.

The accuracy of the built-up area classification is assessed using an error matrix. An error matrix is a table that lists the correct classified pixels against the reference class of a pixel. It allows assessing the accuracy of the image classification (Congalton & Green, 1998). In this study, no real ground truth reference information is available (see section 2.3.1). Therefore, a set of reference pixels is selected from the image scene and classified as built-up or non built-up. To assure the statistical significance of the reference pixel, the method by which the reference pixels are collected needs to be carefully chosen. The critical aspects are the size of the reference dataset and the sampling method (Jensen, 1996). There are a number of methods for sampling reference data. The most common are random, stratified random, systematic, and cluster sampling (Taylor, 1977). With random sampling, each sample pixel has an equal probability of being selected.

A disadvantage of random sampling is that abundant classes are over-represented in the sample, since an abundant class covers more area and has thus a higher probability of containing more sample sites. A number of methods for determining the appropriate sample size are documented in the literature. One of the most popular methods is based on binomial theory (for details see Van Genderen et al., 1978, Fitzpatrick-Lins 1981, Goodchild et al., 1993). In this study, the sample size is determined using a formula presented by (Saito, 2008).

$$N = \frac{Z^2(p)(q)}{E^2} \qquad \textbf{Equation 3}$$

Where:

N = total number of pixels to be used for sampling
p = expected classification accuracy,
q = 1 – p
E = allowable error
Z = 2 from the standard normal deviate of 1,96 for the 95% two tailed confidence level.

Table 15: NDVI thresholds for discriminating built-up and non built-up areas in the AMC area using Quickbird and Landsat 5.

Satellite image	Spatial resolution	NDVI threshold	
		Min	Max
Quickbird (pansharpend)	0,61 m	- 0,01	0,05
Landsat 7	30,0 m	- 0,35 - 0,30	- 0,25 - 0,16

From Equation 3, it becomes obvious that the appropriate number of reference pixels strongly depends on the permissible level of error and the desirable level of confidence. In general, the required sample size decreases with the decreasing level confidence required on the classification. Given that land cover classification is necessarily interpretive, in this study relaxed requirements are used in terms of acceptable levels of error as well as confidence levels. A sample size that assures a confidence level of 95% with an acceptable sample error of 2,5% and an expected classification accuracy of 85% is used. From Equation 3, the required minimum sample size is 1024. Although the sample size can be determined using this method, the number of required reference pixel for each class remains unknown. According to Foody (2002), a sample size can be regarded as appropriate if all classes are adequately represented. In this study, the sample size of each class is selected considering the frequency of the class in the image. From visual inspection is becomes obvious that urban areas are the abundant land cover class. Therefore the sample size for urban areas is 816 and the sample size for non-urban areas is 656 for the Quickbird image with a total of 1472 reference pixels. The accuracy assessment for the Landsat image is based on 843 reference pixels (30 m). The class to which each of the reference pixels belongs to is assessed by visually checking the panchromatic image. Because the reference pixels are derived using visual interpretation, an inherent uncertainty within reference data is unavoidable.

Table 16 displays the error matrix for the binary image classification of the pansharpened Quickbird image. The diagonal cells show the number of pixels that are correctly classified, whereas the off-diagonal pixels indicate the number of pixels where the reference pixels (visually interpreted) contradict the pixel classification. For the Quickbird image, the overall accuracy which refers to the percentage of correctly classified pixel is 81% (see Equation 4). For the Landsat image, the overall accuracy was calculated to be 78,45%. To assess the classification accuracy for each class, it is possible to calculate the producer's and consumer's accuracy errors using the error matrix. The producer's accuracy shows what percentage of a particular class is correctly classified with regard to the reference pixel (see Equation 5). The user's accuracy shows what percentage of a particular class *n* correctly classified with regard to the total number of pixels assigned to this class (see Equation 6). Table 17 shows a comparison of the classification results from Landsat and Quickbird image. The comparison shows that the percentage of correctly classified pixels is very similar for the Quickbird and Landsat image. A close inspection of the accuracy table shows a very-high user's accuracy for the built – up area which means that approx. 90 % of the pixels were correctly assigned to the built - up class. From the standpoint of the objective of this image analysis i.e. the extraction of urban areas and the generalizing nature for model II, this accuracy is sufficient. This result is comparable with previous studies in which Landsat images were used to extract built - up areas. For example, Wu and Murray (2007) identified impervious surface from Landsat Image for Columbus (Ohio) with a user accuracy of 91%. However, the similar classification accuracy does not provide any information about the consistence of the spatial extent of the extracted built-up area.

$$Overall\ accuracy = \frac{Numer\ of\ correctly\ classified\ pixels}{Totel\ number\ of\ classifications}$$ **Equation 4**

$$Producer's\ accuracy = \frac{Numer\ of\ correctly\ classified\ pixels\ for\ class\ n}{Totel\ number\ of\ reference\ pixels\ for\ class\ n}$$ **Equation 5**

$$User's\ accuracy = \frac{Numer\ of\ correctly\ classified\ pixels\ for\ class\ n}{Totel\ number\ of\ classifications\ for\ class\ n}$$ **Equation 6**

Table 16: Error matrix for the land cover classification of the Quickbird image calculated using 1472 reference pixels for the two land cover classes (built-up/non built-up).

Classification data	**Reference data**		
	Non-built up	Built-up	Row total
Non - built up	596	217	813
Built - up	60	599	659
Column total	656	816	1472

Table 17: Results for the accuracy assessment for the binary (built-up/non built-up) image classification for Quickbird and Landsat.

	Quickbird		Landsat	
Overall accuracy	81,00 %		78,45 %	
Land cover class	**Built-up**	**Non built-up**	**Built-up**	**Non built-up**
User's accuracy	90,90 %	73,31 %	91,40 %	55,71 %
Producer's accuracy	73,41 %	90,85 %	76,36 %	80,54 %

From visual inspection it appears that the built-up area delineated from the Landsat image is much larger than the built-up area from Quickbird. In order to verify this initial impression, the fraction of urban area for each district is calculated for both datasets. Figure 6 displays the urban area in km^2 extracted from the images for each district. The general trend can be observed that the urban area extent extracted from Landsat is larger than the urban area extracted from Quickbird (see Figure 7). This can be explained by the fact that urban areas in Ahmedabad are too complex and heterogeneous to be captured in detail by a moderate resolution image like Landsat with $900m^2$ pixel size. This problem is commonly referred to as the "mixed pixel" problem. With the applied binary classification, each pixel is can only be assigned to one class. This implies that the land cover exactly matches with the pixel boundaries. The study area is largely dominated by built-up land cover, the generalization of the built-up areas can be assumed to be the dominant phenomena. Consequently, a larger built-up area extent is extracted from the very high resolution image. With very-high resolution pansharpened Quickbird images, the fine grained detail of the built-up areas can be resolved. As less non built-up areas are misclassified as built-up areas, the total extent of the extracted built-up area is smaller. In total, a built-up area of $55{,}50km^2$ is extracted from Quickbird and $91{,}50km^2$ from Landsat.

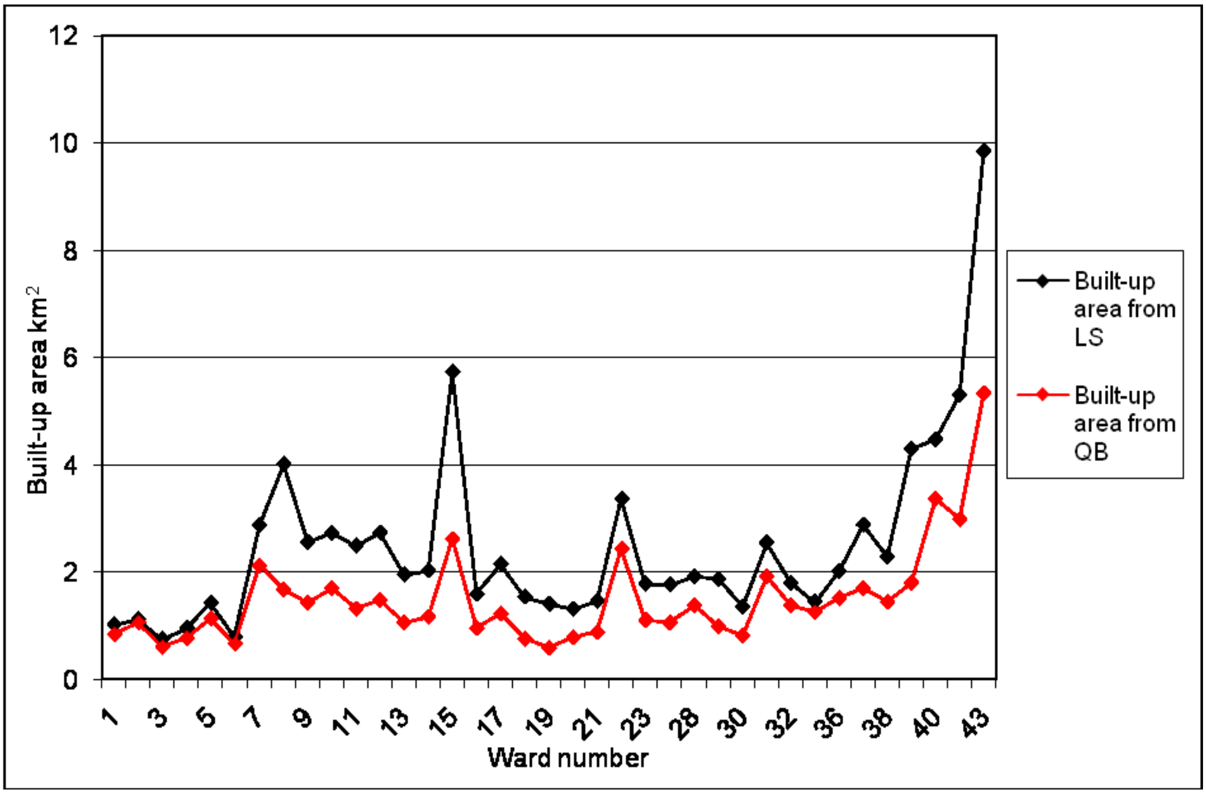
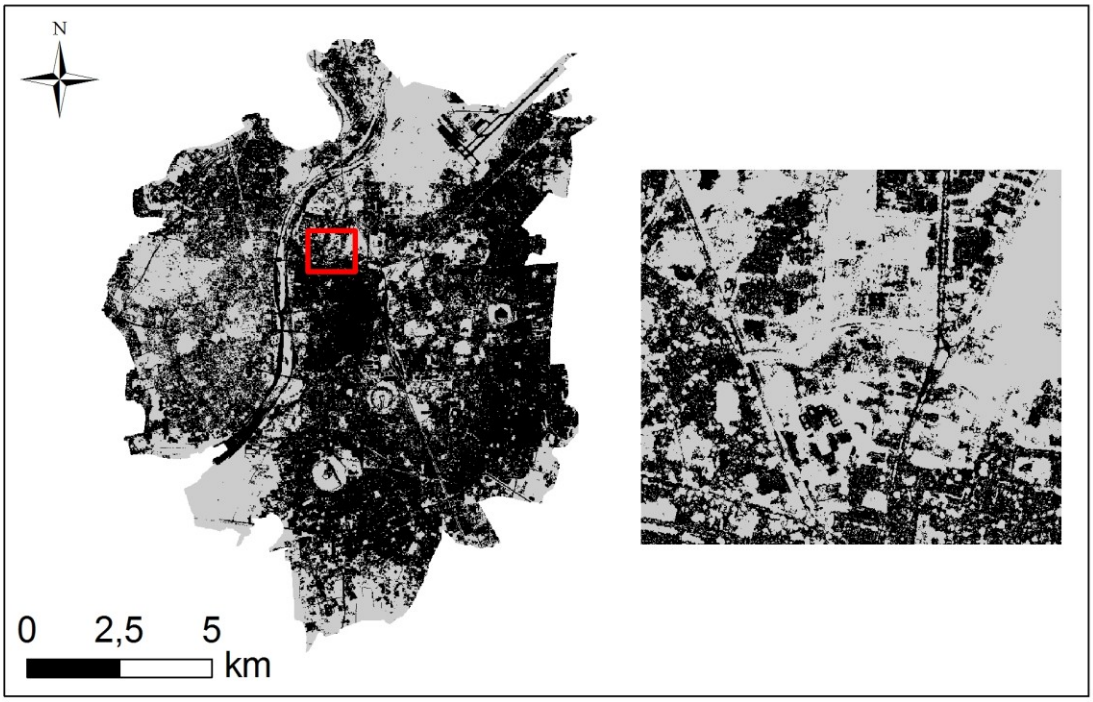


Figure 6: Built-up area for each district in the AMC areas extracted from Quickbird image (red line) and Landsat image (black line).



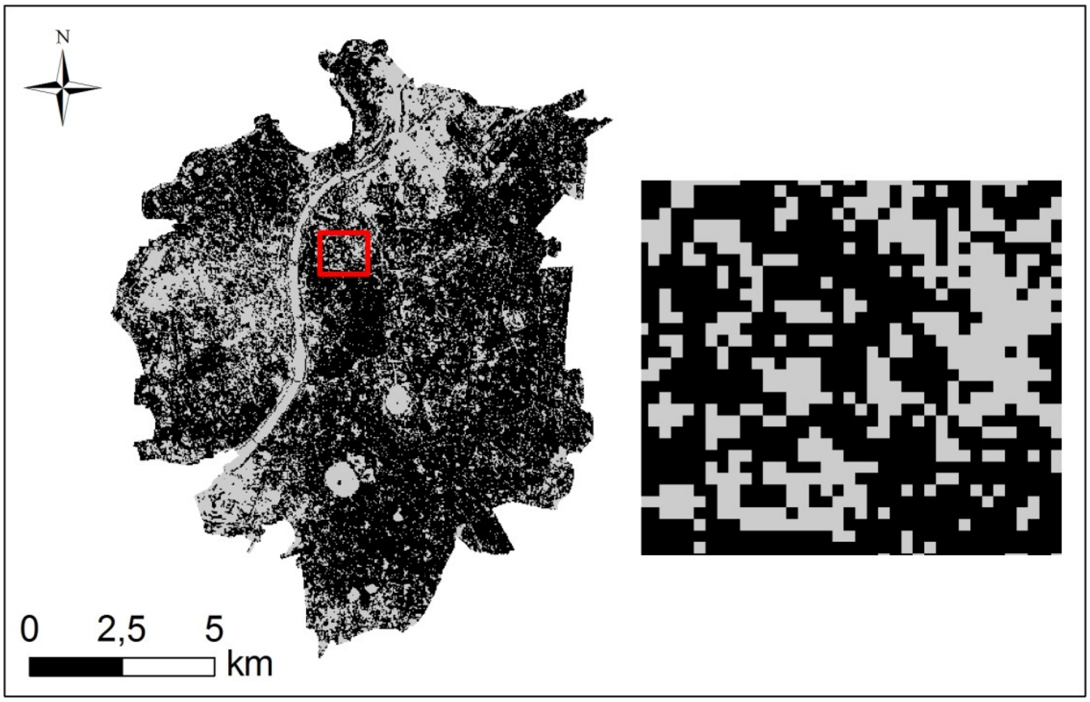


Figure 7: Built-up areas extracted from satellite images for the AMC area: (Upper figure) Quickbird image and (lower figure) Landsat image.

4.2.2 Urban population estimation

Going back to the concept of area weighting (see section 4.1), the extracted built-up area is now the target zone and the administrative area remains the source zone for which the population is known for 2008. The population density is calculated by dividing the population projected for 2008 3.152.108 by the built-up area extracted from satellite images. Table 18 displays the population density and the corresponding built-up area. As the built-up area extracted from Quickbird image is much smaller than from Landsat, the resulting population density is proportionally larger.

Table 18: Population density calculated for the reduced AMC area using the built-up extent extracted from Quickbird and Landsat image.

Parameter	Quickbird	Landsat	Simple areal weighting
Area (km²)	55,50	91,50	136,51
Population Density (people/km²)	56.791	34.446	23.091

4.2.3 Findings

In this section, the findings of the results obtained with model I and model II in section 4.1 and 4.2 are presented. For the purpose of evaluating the quality of the resulting population densities, a validation data set for the total population for the reduced AMC area is calculated. In a first step, the city level population densities (see Table 19) are applied on district level and multiplied with the extracted built-up area for each district. In a second step, the district-wise population values are summed up to calculated the total city population (see Table 19 and Figure 8). It is interesting to observe that despite the large difference of the population densities and the very different fractions of built-up area for the individual districts, the total population for the AMC area is almost identical. The calculated population shows an underestimation of – 20 people for Quickbird and + 23 people for Landsat compared to the projected AMC area population. This proves that population densities derived on city level using satellite images achieve the same degree of accuracy as the official census count. This means in case a population count from a known source and satellite image are available, a reasonable urban population density can be calculated. The population estimation with model I and II is based on the assumption that the population density is constant in the AMC area. This assumption proved to be valid on city level. However, it is questionable whether the assumption of a constant AMC population density holds true for the built-up areas within individual districts. In order to evaluate this assumption, the population for district *i* is calculated based on the constant city population density (see Table 19) and the built-up area or administrative area of district *i* (see Equation 7). Displaying the population and fraction of built-up area for each district, Figure 9 and Figure 10 show that assuming a constant population density for all districts, results in a district population which increases with increasing built-up fraction of the district. To verify to what degree this assumption is valid, a population density validation data set on district level is generated by dividing the district population projected based on district population data from the Statistical Department of Ahmedabad (not census) by the urban district area derived from Quickbird and Landsat images. The comparison reveals under- and overestimation of the district population. Figure 10 displays the district population estimated using the built-up area from Quickbird and the district population estimated using a district-specific population growth rate.

It is interesting to observe that for 20 out of 37 districts the population is overestimated by the constant population density (QB). With Landsat, for 15 out of 37 districts the population is overestimated by the constant population density (LS).

Table 19: Population density, area and total population calculated using administrative AMC area, built-up area derived from Quickbird and Landsat.

Parameter	Quickbird	Landsat	Simple areal weighting
Area (km^2)	55,50	91,50	136,51
Population density	56.791	34.446	23.091
Total population	3.152.128	3.152.085	3.152.108

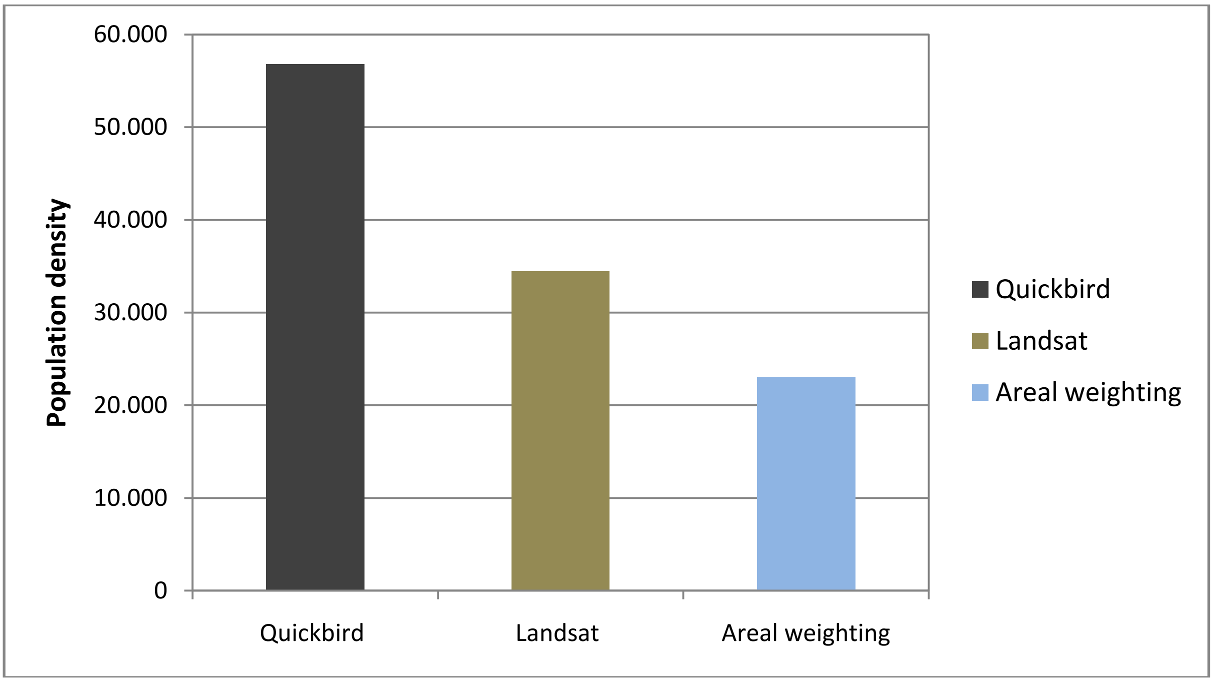


Figure 8: Night time population densities estimated for Ahmedabad at city level (tier 1) using different estimation models. The population density increases with decreasing area extent. For the areal weighting procedure the density is lowest because the entire administrative area of the Ahmedabad Municipal Cooperation is considered.

$$P_{i,built-up} = d_{AMC} * A_{i,built-up} \qquad \textbf{Equation 7}$$

Where

$P_{i,built-up}$ = the population of built-up area in district *i*

d_{AMC} = constant population density on city level

$A_{i,built,up}$ = built-up area of district *i*

For an overestimation of the population density, two reasons are identified. First, the extracted built-up area is very heterogeneously distributed throughout the districts. So if a district has more built-up area then it would have if the built-up area was equally distributed which is assumed for model II, the population density is smaller than the constant city population density. Therefore, the constant city population overestimates the district population density. The underestimation can be explained the same way. If the built-up fraction of a district is smaller than assumed by model II, the population density for this district is larger and therefore, underestimated by the city level density. The over- and underestimations prove that constant population density on city level cannot be directly transferred to district level. The linear correlation which is displayed in Figure 9 and as a black line in Figure 10 using a constant population density does not exist on district level as no linear relation between the fraction of built-up area and the resulting population density is observed from the data. In order to investigate whether a non-linear relation exists between the involved parameters, an analysis of the relations between population (census and Statistical Department AMC), and fraction of built-up areas with respect to the population density deviation is conducted. However, for the AMC area no district level correlation is observed.

From the above discussion of population estimated using model I and model II, it can be concluded that population densities calculated for built-up areas on city level cannot be directly transferred to spatial subunits (districts). To achieve reasonable accuracy results for population estimation on district level, the individual population density in each district or in the built-up area within the district needs to be considered. However, on city level, the comparison of the city population estimation using model I and model II proves on city level population densities derived using satellite images achieve the same degree of accuracy as the official census count.

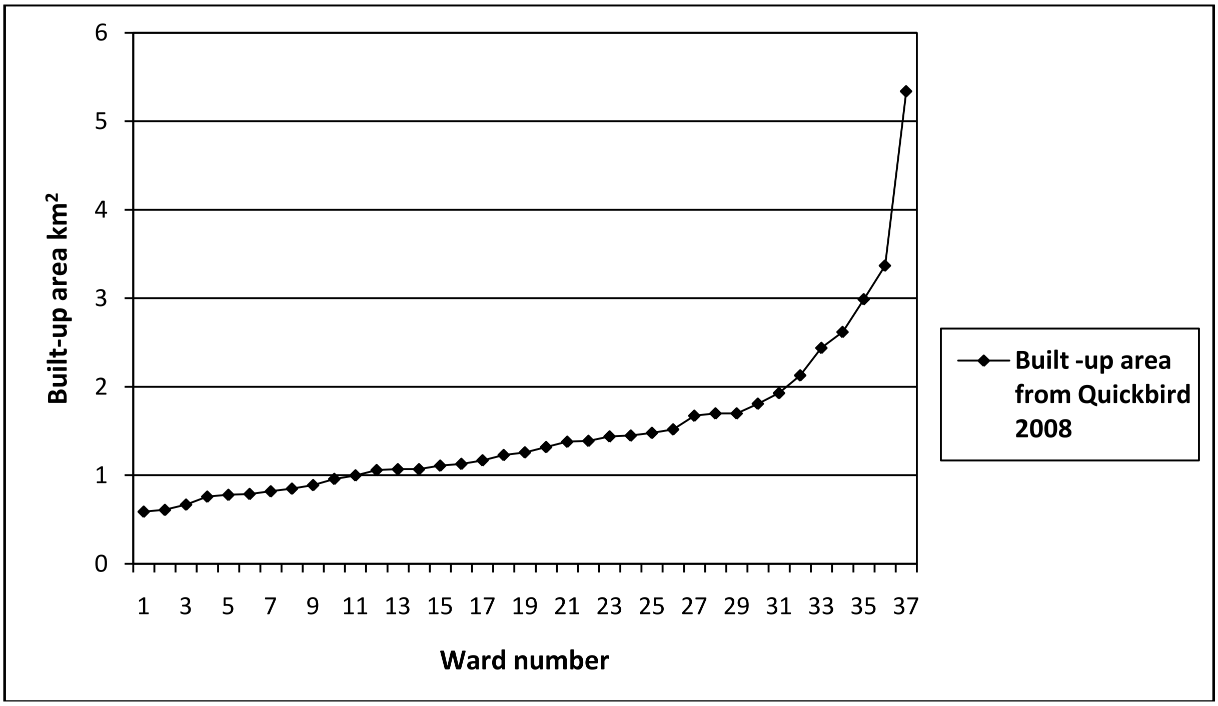


Figure 9: Built-up area extracted from Quickbird image in km² for each district, sorted by increasing built-up district area (km²).

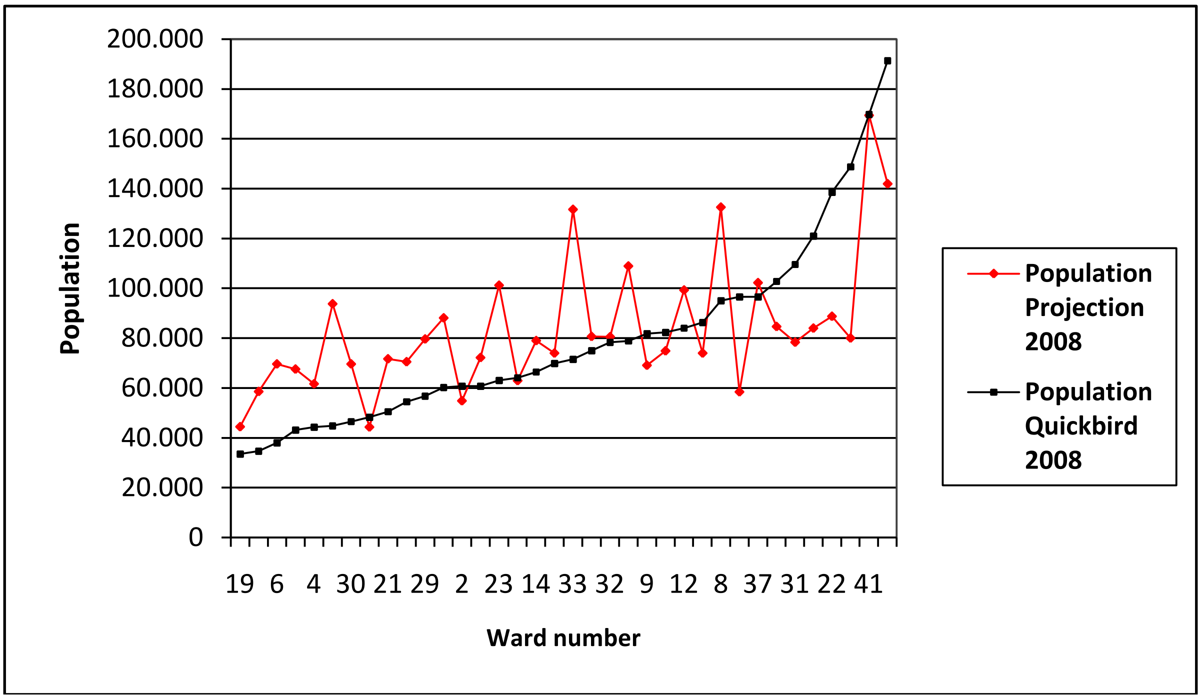


Figure 10: (Black line) Population for each district in the AMC area calculated using a constant population density value, sorted by increasing urban population. (Red line) Population projected for 2008 using a district specific growth rate.

4.3 Model III: Occupancy based population estimation

To overcome the simplification of an equally distributed population in the built-up area and to introduce the distinction of day and night time population, model III considers occupancy categories which allow for disaggregating a single population count according to occupancy relationships. For model III, a methodology is developed to infer occupancy categories on city level for Ahmedabad to estimate the population for different times of the day, at least day and night time population. A detailed overview of the existing occupancy based population estimation techniques is presented in section 3.1. The HAZUS approach is based on US census tract level, a spatial unit which is not available from Indian census (FEMA, 2008). However, some of the occupancy categories used by HAZUS match the occupancy categories identified for Ahmedabad and the occupancy function can be used. The approach by Coburn and Spence (2002) is based on occupancy curves which distinguish two occupancy categories – residential and non-residential. In the following, both methods are tested for the city of Ahmedabad.

4.3.1 Occupancy extraction from satellite imagery

As a first step, occupancy categories to be used in this study are defined from explanatory data analysis. The occupancy categories are characterized by different population density at different times of day and by different building typology, which is important for the occupancy extraction from satellite image and for the generation of the occupancy map. The following 6 occupancy categories are identified:

- **Slums and chawls (low income areas)**
- **Commercial**
- **Pol (traditional housing, walled city)**
- **Service (governmental)**
- **Industrial**
- **Residential (middle / high income)**

From a visual image inspection, the complexity of Ahmedabad's urban typology becomes obvious (see Figure 11). Considering Figure 11e, the *commercial* area incorporates many long, narrow and separated buildings with mainly greyish rooftops. The commercial areas are mostly located near to major roads and have extensive car parks. The *industrial* areas (Figure 11c) are characterized by a high density of warehouses and manufacturing style buildings. Due to the metallic rooftops, the overall dominating colour is grey. There is little or no vegetation, despite single trees. Figure 11f shows the complexity of a *pol* structure. This grouping of houses is typical in the historical, walled city centre of Ahmedabad. A pol incorporates elementary, rectangular units evolved linearly along the street. The *residential* area (Figure 11a) does show many variations in building type and style. Semi-detached houses and terraced houses are very typical. The modern, mostly white painted houses contrast with the traditional houses and the shanties and are therefore easy to identify. Beside the colour, the regular arrangement of the houses is a striking feature of the residential areas. The service areas e.g. Ahmedabad University incorporates a variety of large, flat-roof buildings and a high percentage of recreational areas (see Figure 11b). Figure 11d shows a *slum* in Ahmedabad. The houses are mostly masonry houses with tine or tile roof. The streets are irregular and unpaved. Strictly speaking the categories 1, 3 and 6 are residential areas, but because of their very different structure they were classified separately. In a next step, the feasibility of automatically extracting the identified occupancy categories from satellite image on city level is explored. The aim is to created "new" images from the original image data in order to increase the amount of information between the feature that can be interpreted and used to distinguish objects of interest, in this case occupancy categories. Due to the large study area and the resulting large amount of image data, it is essential to introduce efficiencies into the processing sequence wherever possible. Therefore, the focus is on the least processing intensive image enhancement techniques. According to Lillesand et al. (2008), there are no simple rules for producing the single "best" image for a particular application. This is why testing and evaluating different image enhancement techniques is essential in determining the most suitable technique or combination of techniques. In this study, radiometric and geometric image enhancement techniques are tested as well as spectral band algebra (see Appendix IV for the image enhancement details). The results of the tests reveal that from an image analysis perspective the spectral characteristics of the occupancy categories are not distinct enough to separate the occupancy categories. However, the enhanced image can be useful to assist the manual digitization of the occupancy categories as the enhanced image display useful details the original image do not display. In this study, an occupancy data set on city level is manually digitized from the panchromatic and pansharpened Quickbird images. Due to the moderate resolution of the Landsat image, the digitization of the occupancy categories is not feasible.

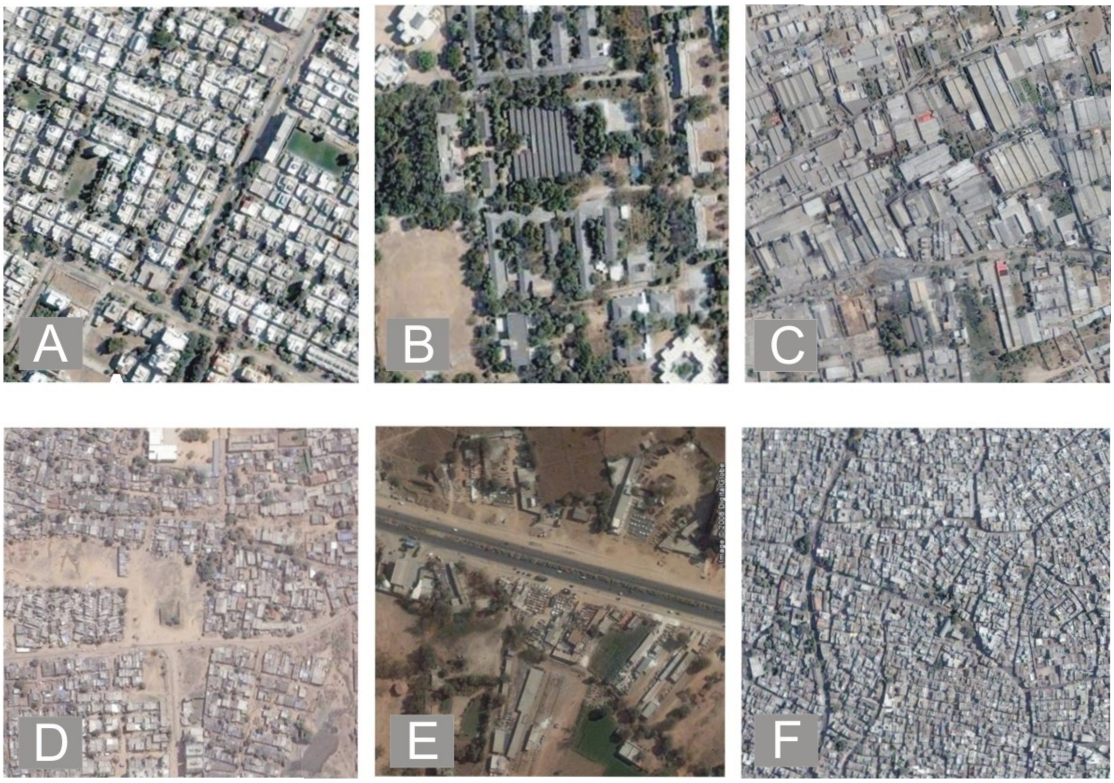


Figure 11: Occupancy categories in Ahmedabad with different characterizing urban typology and building types.

4.3.2 Occupational pattern of Ahmedabad

In this section, the underlying information required for the occupancy based population estimation is developed. Information on residential and non-residential population is sufficient for the occupancy based population modelling for different times of day following the binary approach by Coburn and Spence (2002). In order to be able to apply the HAZUS approach however, the population share for additional occupancy categories needs to be calculated. For this, information on the occupational pattern of Ahmedabad is analysed in the following sections.

Residential Areas

In this section, the share of population residing in each of the three residential occupancy categories is estimated based on statistical data from various sources. Three residential categories are considered: (1) Slums and chawls as low income areas, (2) residential areas – middle and high income areas, and (3) pol – historical residential structures.

Slums and chawls in Ahmedabad

Two types of low income housing can be found in Ahmedabad: Slums and chawls. A slum can be defined as a group of buildings or an area characterized by overcrowding, deterioration and unsanitary conditions. The density of population is very high due to huddling together of houses in marginal areas of the city such as industrial areas, river fronts and low lying areas (Rao, 1990). In Ahmedabad, the development of the slums is strongly related its long history of trading and business. The economy of Ahmedabad has undertaken several considerable transformations, from an ancient, traditional trading centre to a centre of modern textile industry, known as the "Manchester of India". However, the declines of the textile industry in the 1970s to 1980s lead to massive unemployment. The lack of diversification in the industrial base which followed the decline of the textile industry has been mainly responsible for an increase of casual and informal employment (Kundu & Mahadevia, 2002). The related absence of employment contracts, health insurance and unions to represent the worker's interests increases the number of people which could not afford proper housing. As a consequence, the number of people forced to live in slums and to work in the informal sector increased.

The second type of low income housing is also strongly related to the economic history of Ahmedabad. Chawls were originally built in the mill premises for workers, whereas the slums are mostly illegal settlements in the city (Bhatt, 2003). Chawls are planned single room housing units, laid out in rows. Due to the Rent Control Act 1960, the rents for chawls remained constant over time. Because of the very low rent level (around 20 US $ a month), the owner's interest in maintaining the housing units was rather marginal. In consequence, many of the chawls deteriorated in quality (Metha & Metha, 1992). During the textile crisis in the late 1980s, many of the textile mills were closed, the chawls, however, remained. Physically, chawls are mostly multi-storied buildings. A usual tenement consists of a one room accommodation with common circulation passage and toilet facilities. The dwelling unit size varies between 15 to 25 m^2. Chawls can be found at 1383 locations in Ahmedabad, about 1 million people living there (AMC, AUDA, CEPT, 2002; Bhatt 2003).

The share of slum and chawl population in Ahmedabad can only be roughly estimated for 2008. A chronological and consistent series of information on slum population does not exist. Nevertheless, the trend of increasing slum population can be traced throughout the last three decades (Figure 12). The earliest information on slum population is available from a census carried out by the Operation Research Group (ORG) in 1972 (Bhatt, 2003). According to this census, in 1971, 17,1% of the population lived in slums. Baring in mind the much lower total population at that time, about 270.000 people lived in slums in 1971. By 1981, the percentage of people living in slums had increased by 4,3% to 21,4% (Core Consultants, 1983). A decade later, the slum population had risen to 25,6% of the total population (Metha & Metha, 1992). However, a study published by the AMC states that only 16% of the population lived in slum in 1991, referring to 456.000 people (AMC, AUDA, CEPT, 2002). This estimate is about 38% less than the figure of 736.438 people published by Metha and Metha (1989). Recent data on the number of slums in Ahmedabad is available from a slum survey conducted by SEWA and SAATH in 2000 (Joshi, 2002). According to this survey, 710 slums exist in Ahmedabad housing 883.541 people. Considering a total population of 3.520.085 people as per census 2001, 25,1% of Ahmedabad population lived in slums in the year 2000. A survey conducted by the AMC in 2001 displayed similar figures estimating the slum population to be 25,8% of the total population, referring to 906.000 people (AMC, 2003). Information from the most recent Indian census 2001 on slum population constitutes only 12,5% of Ahmedabad population i.e. 439.843 people lived in slums. This discrepancy can be explained by the fact that the census only included those slums which encompass more than 60 hutments or house more than 300 people respectively. Considering the continuously increasing slum population and the unchanged number of slums, it can be concluded that the population density in slums exhibits the same increasing trend as the absolute slum population.

The information available on chawl population is even scarcer than on slum population. In 1990, chawls could be found at 1409 locations in Ahmedabad housing 40 % of the population (Bhatt, 2003). According to the 1991 Indian census, 2.876.710 people lived in Ahmedabad. Therefore, it can be estimated that about 1.150.000 people lived in chawls in 1991. According to the latest survey conducted by the AMC in December 2001, there are 958 chawls with about 1 million people living there (AMC, 2003; Bhatt, 2003).

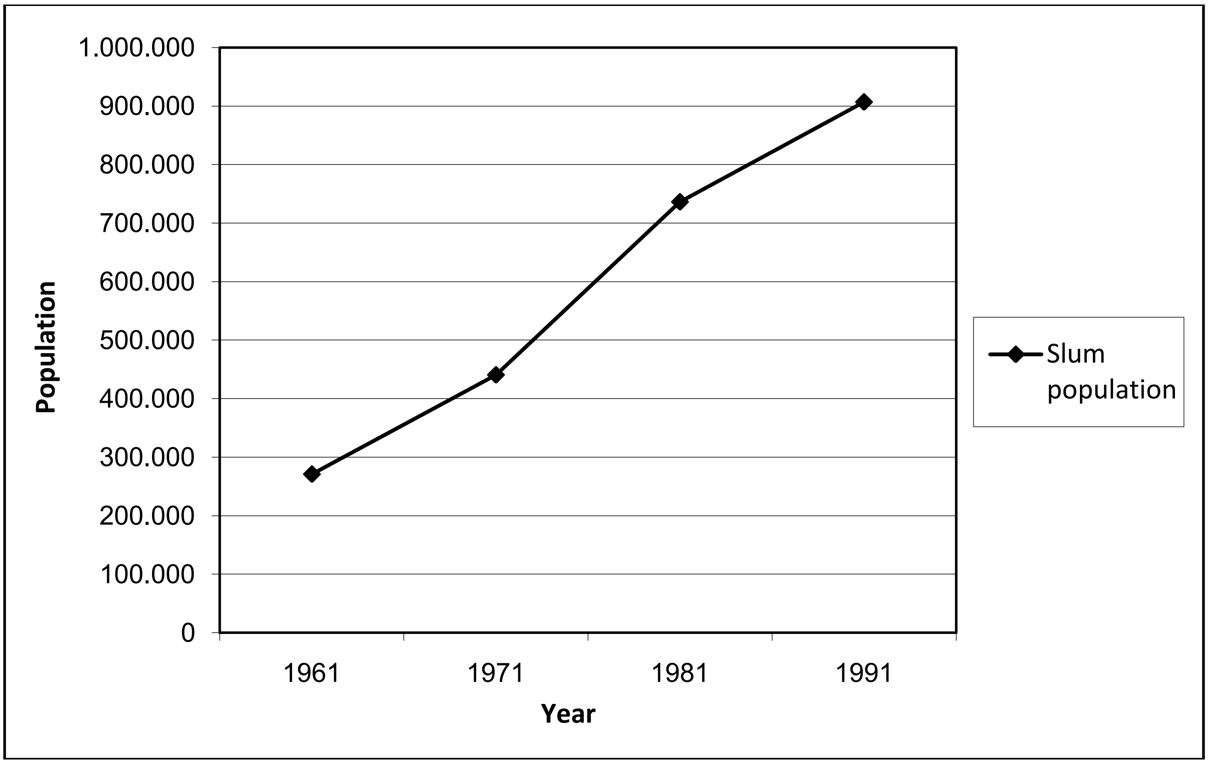


Figure 12: Development of slum Population in Ahmedabad. Data Source: Year 1971 (Bhatt, 2003), year 1981 (Core consultants, 1983), year 1991 (Bhatt, 2003), and year 2001 (AMC, 2003).

Table 20: Chronological list of the slum population estimates between 1971 and 2001 by different institutions and slum population expressed in percentage of the total population.

Year	Slum Population	Slum Population (%)	Author / Institution
1971	270.000	17,10	ORG
1981	440.781	21,40	Core Consultants
1991	736.438	25,60	Metha & Metha
1991	456.000	16,00	AMC
2001	883.541	25,10	SEWA , SAATH
2001	906.000	25,77	AMC
2001	439.843	12,50	Indian Census

Estimation of slum and chawl population in Ahmedabad

In this study, the chawl and slums are combined to a single occupancy category – low income areas. In a first step, the slum and chawl population is estimated separately and then summed up to the total low income area population. To estimate the slum population, the existing data are analyzed to determine an average annual growth rate. Due to the inconsistency of the available data, the average AGR differ significantly depending on which data are used. Based on the estimates excluding the Indian census constrained by the number of hutments, the determined average AGR range between 2,94% and 3,05%. Including the constrained, Indian census data, the annual growth rate is significantly lower 1,64%. Since the unconstrained estimations have the same order of magnitude, an average growth rate of 3,0% is assumed to be realistic.

In order to project the slum population for 2008, the average slum population of 895.334 people for 2001 is calculated based on the unconstrained surveys by SEWA and AMC. From this, the slum population in 2008 is estimated to be 1.083.354. As no detailed information on chawl population exists, the approximate figure of 1 million people estimated by AMC (2003) is used. This results to a total population of **2.083.354** people living in low income areas.

Pols in Ahmedabad

Pols are traditional residential areas located within fortified, historical part of Ahmedabad (also called Gamtal) (Gillion, 1968). As part of Ahmedabad's vernacular architecture, pols are characterized by densely packed clusters of continuous rows of houses, flanked around meandering streets. Each pol has an entrance gate on the main road. Sometimes a large open space, known as "chowk", is seen at the center of the pol and is surrounded with houses in a circular fashion (Desai, 1985). A pol may be housing one or more clusters of residential units belonging to homogeneous groups with their own amenities. A pol is comprised of three to four stories high, narrow but deep houses on either side of the winding main street. Traditionally, member of the same caste or occupational group resided in a pol which was self-governed by the elders. In modern terms, a pol is basically a housing community enclosed on all sides with one or two restricted for a group of up to 500 houses (Pandya et al., 2001).

Estimation of pol population in Ahmedabad

The population of the walled city constantly increased from 1921 to 1971. As new residential areas established in the western part of the city and the commercialization of the walled city increased, people started to move from the walled city (Bhatt, 2003). This led to a residential population decline from 480.735 people in 1971 to 372.622 in 2001 (- 22,50%) (Census of India, 2007). A study conducted by CEPT and GIDB in 2005 estimates an residential population decrease of 21,74% in the walled city between 1981 and 2001. The underlying population figures are very close to the Census estimations. This allows for assuming a relative high level of accuracy. The study also projected the population of the walled city for 2011 and 2010 based on a annual growth rate of – 1,13%. Using this growth rate, the residential population of the walled city for the image acquisition year 2008 is estimated to be **342.381**.

Middle and high-income areas in Ahmedabad

It can be assumed that the population not living in slums, chawls or pols reside in middle to high income residential areas. The number of people residing in middle and high income areas in Ahmedabad is calculated by subtracting the share of population in low income areas and pols from the total city population for 2008.

Table 21: Population in different residential occupancy types for Ahmedabad in 2008.

Occupancy	Year	Population
Slums and Chawls	2008	2.083.354
Middle and high income areas	2008	1.629.349
Pol	2008	342.381

Urban economy in Ahmedabad

Ahmedabad is the industrial centre of Gujarat and was formerly known as “the Manchester of India”. With the first modern textile mill opening in 1861, Ahmedabad has witnessed a flourishing textile industry until the early 20th century. From 1886 to 1930 the number of textile mills increased from 1 to 72. The Great Depression had a severe impact on British India. Between 1925 and 1938 several textile mills in Ahmedabad were closed. Despite two major communal riots between 1941 and 1946, by 1950 125.000 people were employed in the textile industry (Metha & Metha, 1992). After independence in 1947, the socialist reforms between 1950 and 1975 lead to low annual economic growth rate. The growth rate stagnated between 2% and 4 %. The Indian economist Raj Krishna referred to this phenomenon as the “Hindu rate of Growth” (Williamson & Zagha, 2002). This stagnation results in a number of textile mills closing, frequent commercial riots, deterred additional investment in the city. Between 1970 and 1980 Ahmedabad was referred to as the “dying city”. Between 1985 and 1986 the uneconomical mills were nationalized and placed under the Gujarat State Textile Cooperation and the number of mills decreased from 85 to 58. By 1994, only 23 textile mills remained in production. Chawdhury (1996) estimated that around 50.000 people lost their jobs between 1986 and 1994. Although Ahmedabad continued to be dominated by its textile industry, the economic liberalization in the 1990s and 2000s lead to a sectoral shift. New industrial branches such as chemical and engineering goods evolved, leading to a diversification in small and medium industries emerged. However, as south Gujarat witnessed a rapid growth of chemical and petrochemical industries, in Ahmedabad the share of industries significantly declined (AMC, 2003). At the same time, the tertiary sector including business and commerce, transportation and communication, construction and other services evolved. The new policies facilitated the opening for international trade and investment, deregulation, initiation of privatization, tax reforms and inflation control measures. This shift towards the tertiary sectors is reflected in the occupational pattern. In the following section, employment data are analyzed to form a basis for determining the share of population employed in the industrial, commercial, and service sector in 2008.

Employment in Ahmedabad

For the year 2008, no reliable information on the overall employment rate in Ahmedabad is available from the literature. Therefore census data and other secondary, statistical data are analysed to estimate the total share of working population in the AMC area. It is important to note that the concept of the economic census has changed throughout the last three census periods. In the 1961 and 1971 census a person was treated as a worker if he spends his time mainly in work, at least a day of regular work a week. Since 1981, three categories of workers have been distinguished in the census: workers, marginal workers and non workers. Table 22 displays the number of main workers, marginal workers and non-workers as enumerated by the census 1991 and 2001. Since the number of workers and the number of non-workers equally increased between 1991 and 2001, the overall employment rate remained more or less constant, with 29,25% in 1991 and 31,95% in 2001 (see Table 23). According to the 2001 census, in 2001 there were 1.071.011 main workers, 53.497 marginal workers and 2.395.577 non-workers in Ahmedabad, thus the total number of workers was 1.121.995.

Table 22: Number of main workers, marginal workers and non-workers in Ahmedabad as per census 1999 and 2001.

Census	Total Population	Main workers	Marginal workers	Non-workers
Census 1991	2.876.710	831.459	10.109	2.035.142
Census 2001	3.520.085	1.071.011	53.497	2.395.577

Table 23: Total working population and employment rate in Ahmedabad as per census 1999 and 2001.

Census	Total working population	Employment rate in %
Census 1991	841.568	29,25
Census 2001	1.124.508	31,95

Industrial, commercial and service sector

Time series data on the number of registered factories in Ahmedabad show a continuous increase of the number of registered factory. The decrease in the average of workers employed per factory mirrors the trend to smaller factories. In 2003, 4.859 factories were registered in Ahmedabad. Information on the number of workers daily employed in registered factories is only available until 2006. Based on the annual growth rate from 2001 to 2006, the average AGR was calculated to be 5,4%. Based on the assumption that the average increase is also valid for 2007 and 2008, the number of daily employed workers in factories for 2008 is estimated to be 231.156. Beside the regular employed people, a large share of Ahmedabad's population is employed as daily-wage labours. There are about 16 Kadiya nakas (KN) where people assemble every morning waiting for potential employers. Six KN are located on the eastern side of the Sabarmati river whereas the remaining 10 are located in the western part of the city. The number of people employed as day labours range between 12.000 and 15.000 people. In addition to the number of people employed in factories, employment areas like mining and quarrying are categorized as well as industrial. According to the Regional Employment Exchange Office (2007), in 2006 247.926 people were employed in the industrial sector. In contrast to the number of people employed in registered factories, the number of people employed in the industrial sector shows a decreasing trend as a consequence of the emerging tertiary sector. Based on the information from the Regional Employment Exchange Office (2007), in 2008 **239.081** people were estimated to be employed in the industrial sector. The information available on employment in the commercial and service sector are less detailed than the information on the industrial sector. In 2007, 437.840 people were employed in the commercial sector, including establishments and shops (Shop & Establishment Department, AMC 2007). According to the Regional Employment Exchange

Office (2007) in 2006, 107.800 people were employed in community services (social / personal). This includes, among others, education and transportation services. Assuming a constant AGR, in 2008 **109.326** people are estimated to have been working in community services. Table 24 lists the number of people employed in the industrial, commercial and service sector calculated based the results of the above analysis.

Table 24: Number of people employed in different sectors in 2008.

Sector	People employed in 2008
Industry	239.081
Commerce	437.840
Service	109.326
Daily-wage labours	13.000
Total employment	**796.747**

Informal sector

The only study available on the informal sector in Ahmedabad was conducted by Breman (2002). The following information is taken from this study titled "An Informalised Labour System, End of Labour Market Dualism". In this study, Breman (2002) points out that in the early-1970s the informal sector was estimated to account for around half of all work in the urban economy and by the end of the 20th century it had grown to around three-quarters and four-fifths. Breman (2002) defines informal sector work as work on one's own account which generates income but is not regulated by an explicit employment contract and enjoys no protection. This includes people who work in the street, in homes, small-scale enterprises, power loom workshops etc. In this study, the number of people employed in the informal sector can only be indirectly estimated. Assuming that remaining population which is not regularly employed or non-working such as children less than 6 years, is employed in the informal sector. From this, the percentage of the total population working in the informal sector is 44,65%. This is less than other studies estimated for Delhi and Mumbai, where the informal sector is estimated to account for 66,7% (Delhi) and for 68 % (Mumbai) of the total employment (Srivastava, R., 2005). Including the informal sector, the overall employment rate for Ahmedabad in 2008 was 64,30%. Not including the informal sector, the overall employment rate is 19,64%.

It is important to note that quality of the figure estimated in this section varies. The estimate for the informal sector is based on the generalizing assumption that the share of working population not officially registered works in the informal sector. However, to validate this estimate is very difficult because the informal sector is not considered in the official census employment statistics. Another source of uncertainty is that the information on sectoral employment is rather scarce and not up-to-date. However, to provide an insight into the occupancy based population distribution in Ahmedabad this kind of information is sufficiently accurate. The estimation of the share of non-working population which is too young to work can be assumed to be rather accurate as birth rates are a reliable parameter to base this estimate on. The non-working population which is retired is very difficult to estimate because is no generally valid retirement age in India. The retirement age for public, governmental employees has recently been raised from 60 to 62 years. This however does not automatically apply for other employment sectors. In addition with a life expectancy of 62,5 years for male and 64,6 years for the female population, the share of population in retirement can be expected to rather small.

Unemployment and non-working population

The number of unemployed people can only be estimated from the results of the above analysis and the information available on employment seekers in the AMC area. In 2005, a total of 86.096 employment seekers were registered (Regional Employment Exchange Office, 2007). In 2006, 83.004 people were registered as unemployed and 3092 physically handicapped employment seekers were registered. So in total 86.096 people were registered as unemployed in 2006.

In addition to the unemployed population, the non-working part of Ahmedabad's population also includes the number of children < 6 and children attending primary and secondary schools. The number of school children and children under 6 years can be estimated based on the number of live birth, infant mortality rate and children survival. Between 2001 and 2006, the infant mortality rate remained quite constant with an annual average of 2,76%. The average of live births was 83.612. Based on this, the number of children younger than 6 was estimated to be 571.856 for 2008. In 2003, a total of 564.378 children attended private and municipal primary schools. In 2003 the number of students attending secondary and high education remained constant was 225.034. As no further, more recent information is available and a consistent projection is not feasible because of missing growth rates, a total of 1.361.268 student and children are considered in the non-working population.

Table 25: Non-working population for Ahmedabad including children less than 6 years, school children, students and registered unemployed people.

Non-working population	Number of children / students
Children under 6 years (2008)	571.856
School children (2003)	564.378
Students (2003)	225.034
Unemployed (2006)	86.096
Total	**1.447.364**

4.3.3 Occupancy based population estimation over the course of day

In this section, the occupancy based population distribution is modelled for Ahmedabad in 2008 over the course of day. The proposed method is based on the occupancy curves provided by Coburn and Spence (2002). The occupancy curves are binary, only distinguishing urban non-residential and urban residential population. For each time of day, the occupancy of residential and non-residential buildings is provided as percentage of the total population (see Table 26 and Figure 13). In addition, the residential category is adjusted to the local condition in Ahmedabad by subdividing it into three residential categories (low income areas, middle and high income areas, and pols) (see Figure 14). Based on the generalizing assumption, that the population is equally distributed within each occupancy category, the population and the average population density is calculated for each category (see Table 27 and Table 28). Then, using the polygon area information, the population is calculated for the individual occupancy polygons. From this, GIS-based maps displaying the occupancy based population distribution for different times of day are generated. Table 27 displays the population distribution maps for 0:00am, 6:00am, 12:00am and 6:00pm.

Table 26: Typical occupancy of residential and non-residential buildings by urban population after Coburn & Spence (2002).

Time of day	Occupancy of residential buildings (%)	Occupancy of non-residential buildings (%)
00:00 am	78	13
06:00 am	60	16
12:00 am	40	85
06:00 pm	61	52

Table 27: Number of people in residential and non-residential occupancy in Ahmedabad for selected times of day calculated using the typical occupancies after Coburn and Spence (2002).

Time of day	Number of people in residential occupancy	Number of people in non- residential occupancy
00:00 am	3.123.966	520.661
06:00 am	2.403.050	640.813
12:00 am	1.602.034	3.444.372
06:00 pm	2.443.101	2.082.644

Table 28: Population density in residential and non-residential occupancy in Ahmedabad for selected times of day calculated using the population estimated (Table 27) and the occpancy area.

Time of day	Population density (p/km^2) in residential occupancy	Population density (p/km^2) in non- residential occupancy
00:00 am	48.209	20.228
06:00 am	37.084	24.896
12:00 am	24.723	133.814
06:00 pm	37.702	80.911

The accuracy of the population modelling strongly depends on the quality of the occupancy polygons since the polygon area forms the basis for the population estimation. Due to the very high complexity of Ahmedabad's urban structure, uncertainty arises from the generalization of using just two occupancy categories (1) residential and (2) non-residential. In Ahmedabad residential areas cover low income areas, pols and high income areas. Slums can only be digitized as single polygon, resulting in a large areas and thus, in a large population. On the contrary, middle and high income areas can be digitized in more detail as structures can be easier identified. The area of the single polygons is smaller, and therefore the resulting population is smaller. This is way according to the population distribution map in Figure 15 much less people live in the north-western part than in the other parts as very few slums are located in this part of Ahmedabad. Looking at the north-western part of Ahmedabad in more detail, it becomes obvious that the population distribution is more complex than displayed in the city map in Figure 15. This complexity is difficult to capture on city level, but is considered on district level in section 5.3 of this study.

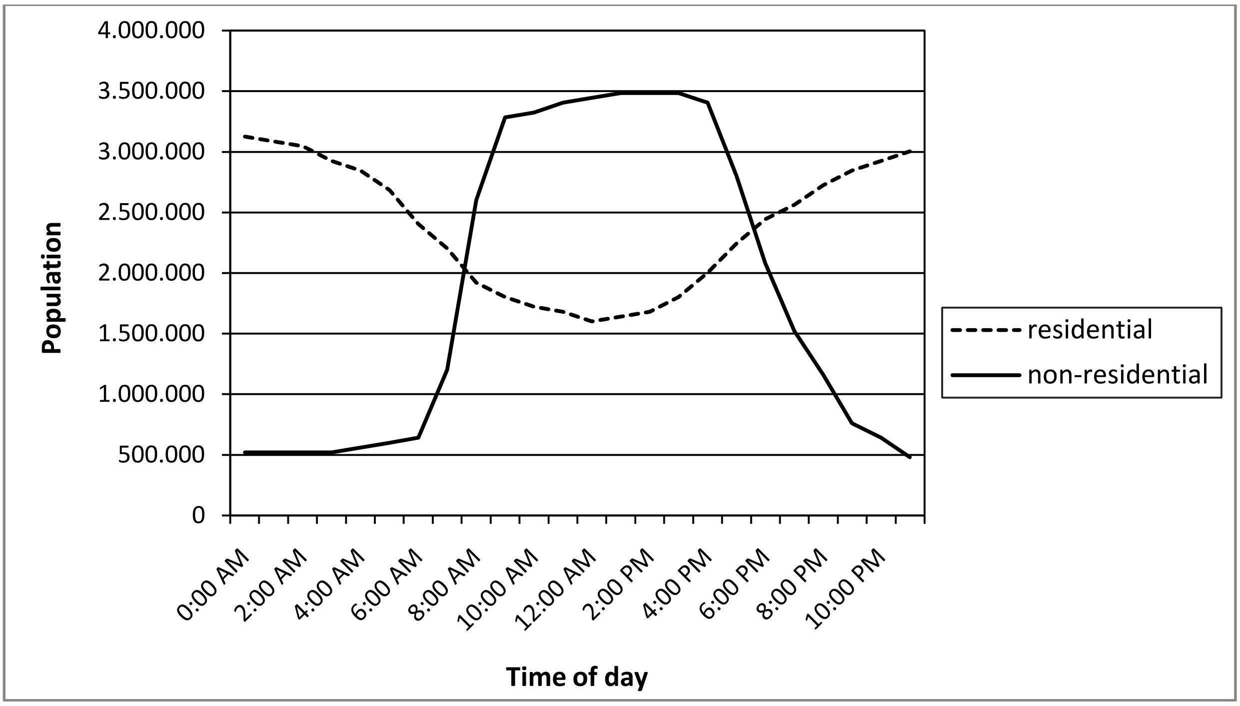


Figure 13: Number of people in residential and non-residential occupancy over the course of day calculated using the occupancy curves by Coburn & Spence (2002).

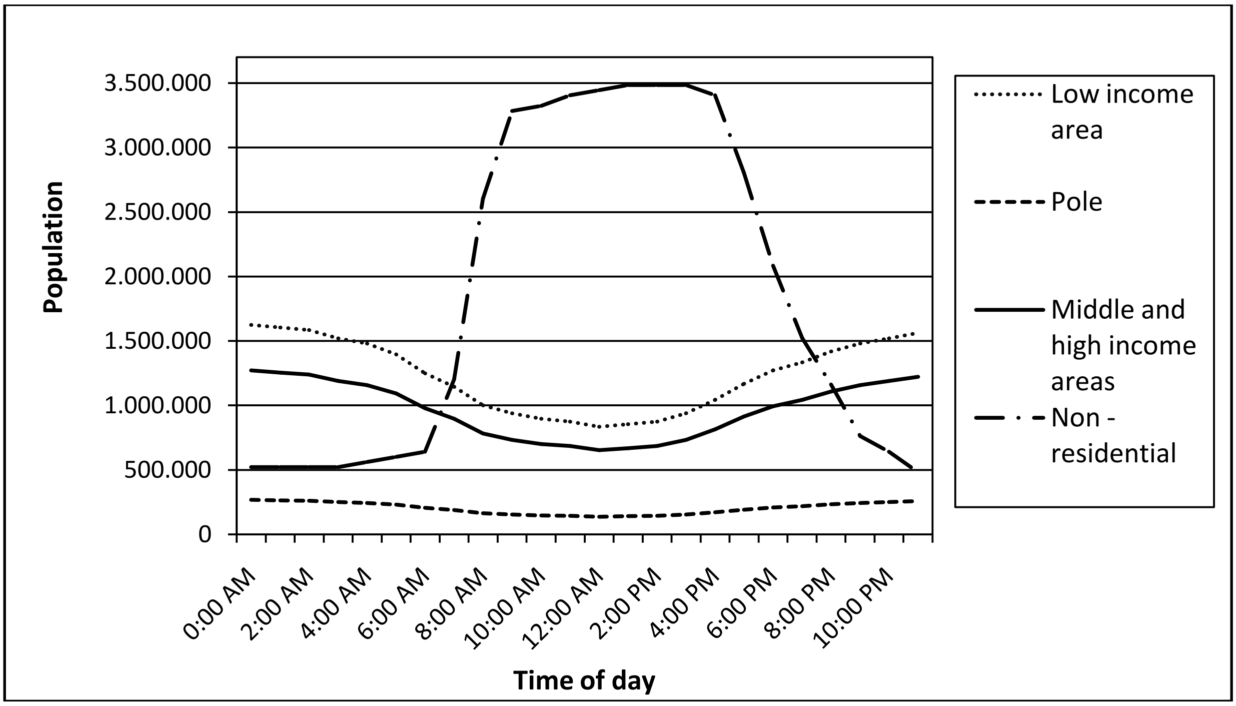


Figure 14: Numer of people in non-residential and residential occupancy over the course of day adjusted to Ahemdabad by subdividing the residential occupancy into thee subcategories.

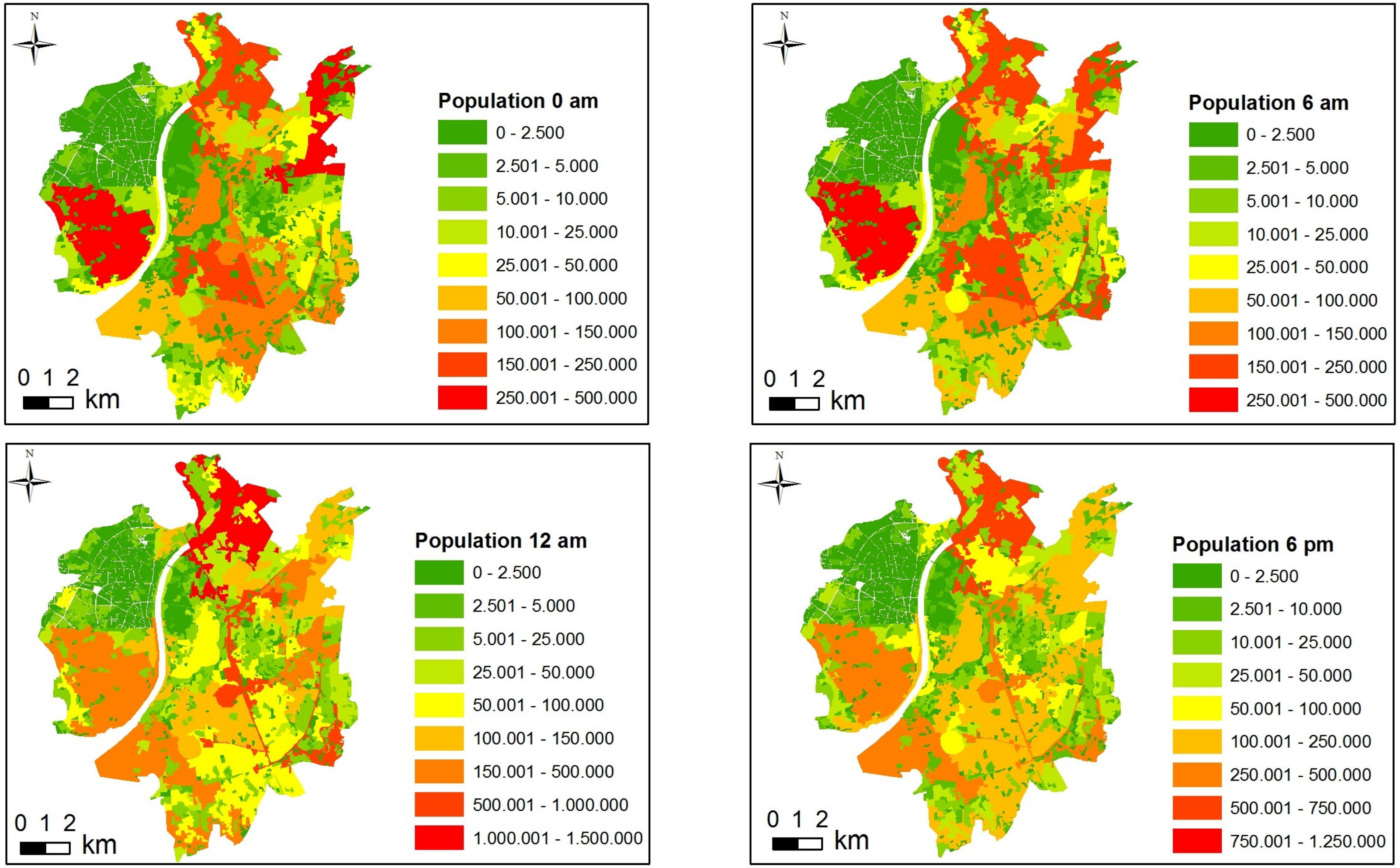


Figure 15: Occupancy based population distribution maps for four different times of day (0:00 am, 6:00 am, 12:00 am, and 6:00 pm) after Coburn and Spence (2002)

.

4.3.4 Occupancy based population estimation for three times of day

In this section, the occupancy relations provided in HAZUS are applied on city level for Ahmedabad. Due to limited data availability, only four of the 6 available occupancy categories can be applied. No information is available on the commuting population and the number of people employed in hotels. For each of the occupancy categories default relationships are provided for three times of day 2:00am, 2:00pm, and 5:00pm (see Table 29). Using the population estimates for different employment categories from section 4.3.2, the relationships are applied for Ahmedabad (see Table 30 and Table 31). To generate the population distribution map, the geocoded, occupancy data set is used. For each of the three times of day, a population distribution map is generated. As educational occupancy cannot be visually identified, this category is excluded from the map. In a first step, the population density is calculated for the occupancy categories based on the occupancy area. In a second step, for each occupancy polygon the population is estimated based on the area of the polygon. The results are displayed in population distribution maps for three times of day (Figure 16).

Table 29: Default relationships for estimating population distribution provided by HAZUS on US census tract level.

Occupancy	2:00 AM	2:00 PM	5:00 PM
Residential	0,99 (NRES)	0,75 (DRES)	0,5 (NRES)
Commercial	0,02 (COMW)	0,98 (COMW) + 0,2 (DRES)	0,98 (COMW) + 0,1 (DRES)
Educational		0,8 (GRADE) + 0,8 (COLLEGE)	0,5 (COLLEGE)
Industrial	0,1(INDW)	0,8 (INDW)	0,5 (INDW)

Where

NRES is the night time residential population
DRES is the day time residential population
COMW is the number of people employed in the commercial sector
INDW is the number of people employed in the industrial sector
GRADE is the number of children in school
COLLEGE is the number of student on college or university campuses

Table 30: Number of people in different occupancy categories for three times of day for Ahmedabad calculated using the HAZUS relationships.

Occupancy	2:00 AM	2:00 PM	5:00 PM
Residential	3.965.033	3.003.813	2.002.542
Commercial	8.757	429.083	429.083
Service	0	631.530	112.517
Industrial	23.908	191.265	119.541

Table 31: Population density per km² for different occupancy categories at three times of the day according to HAZUS (1999).

Occupancy	2:00 AM	2 :00 PM	5:00 PM
residential	61.151	46.327	30.884
commercial	41.501	203.357	203.357
service	0	179.872	32.047
industrial	1.189	9.511	5.944

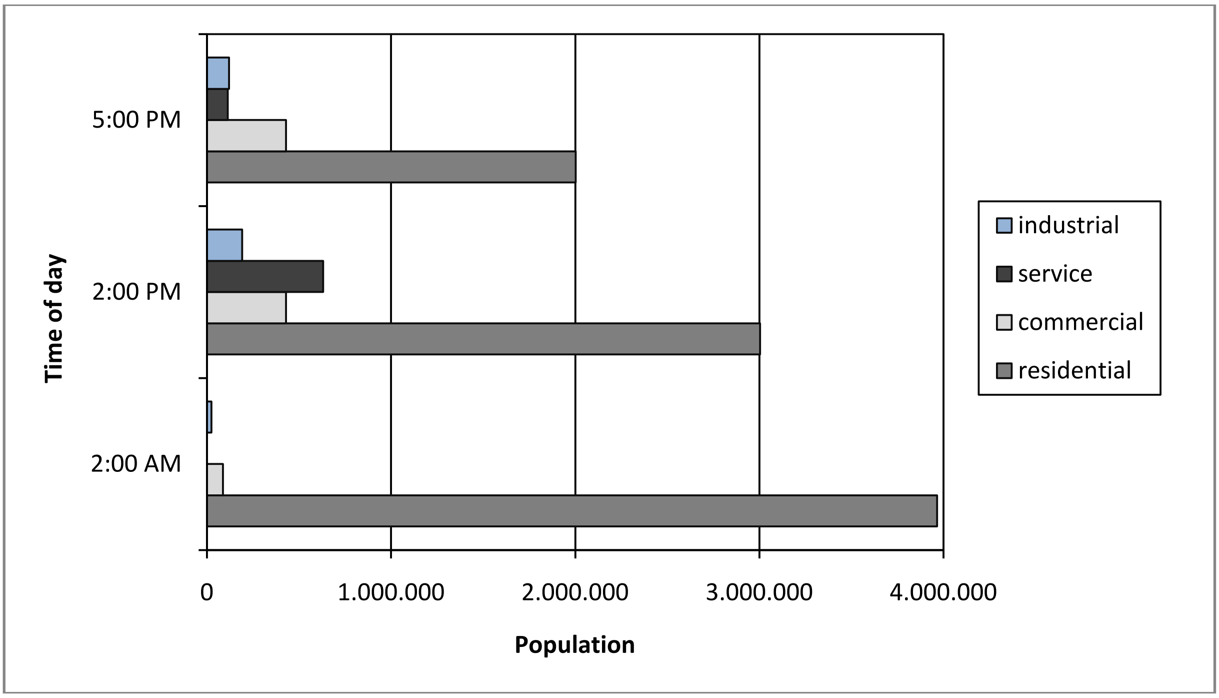


Figure 16: Number of people in different occupancy categories for three different times of the day calculated using the HAZUS relationships.

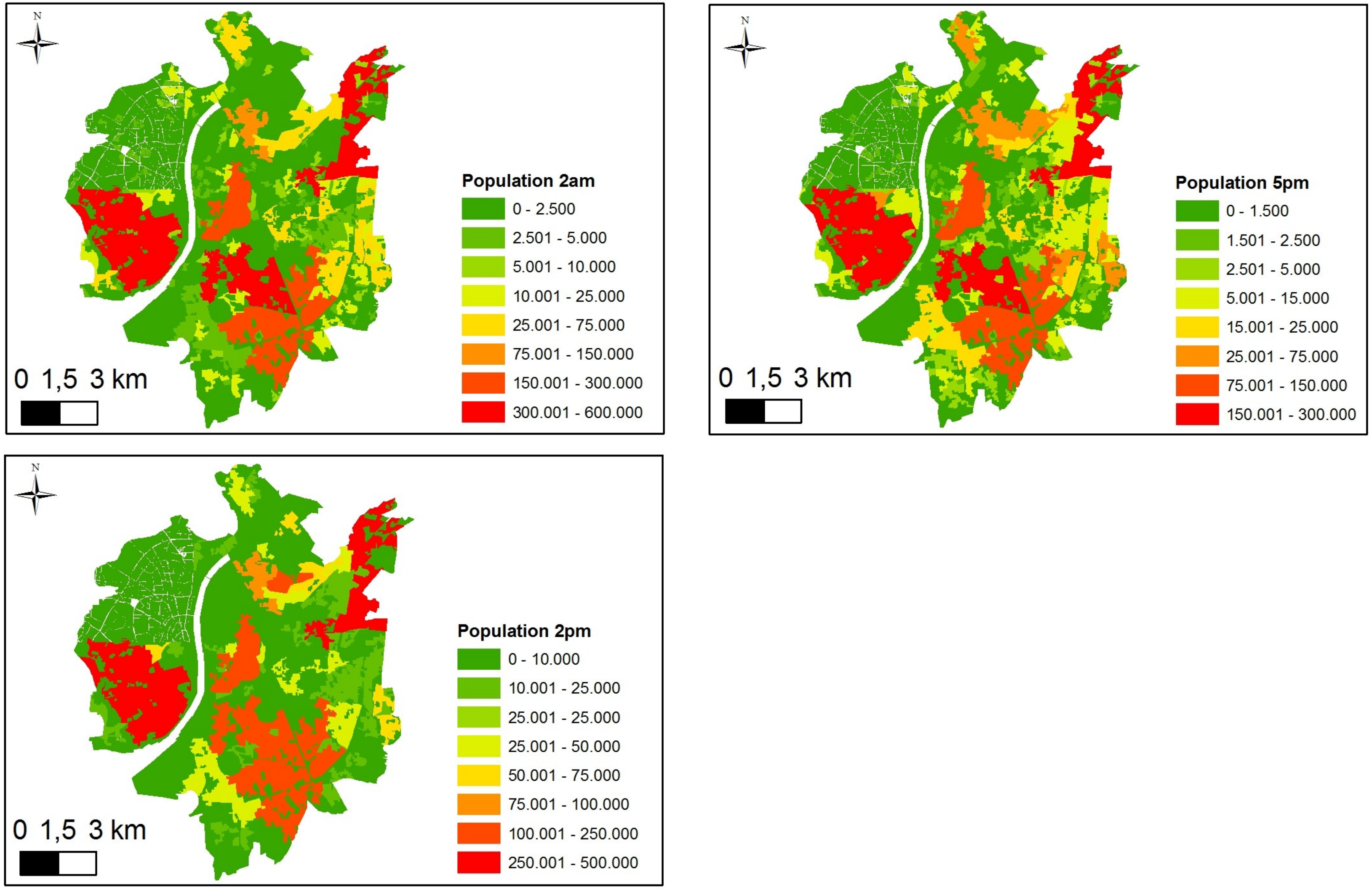


Figure 17: Occupancy based population distribution maps for three different times of the day (2:00 am, 5:00 pm, 2:00 pm) using the HAZUS occupancy relations.

4.3.5 Findings

In this section, the findings of the occupancy based population estimation are discussed. From an analysis of the occupational pattern of Ahmedabad in section 4.3.2, the number of people working in three different employment sectors is calculated and their spatial distribution is modelled using the HAZUS relationships for three times of day. Assuming that the relationships are generally valid for Ahmedabad, the analysis of the population distribution in section 4.3.4 reveals, that with 3.965.033 people the population in residential occupancy is nearly two time larger for the night time scenario (2:00am) than for the commute time scenario (5:00pm) when only 2.002.543 people are in residential occupancy. This is the largest fluctuation in population in all occupancy categories. The smallest fluctuation is observed for industrial occupancy.

However, it is important to note that it is very likely that the number of people employed in the different sectors is underestimated since only regularly employed people can be considered due to data availability. It is concluded that the employment estimates in a certain sector calculated in section 4.3.2 have to be considered as a minimum estimate. The limited availability of employment data is a one source of error associated with the application of the HAZUS relationships to Ahmedabad. Especially, reliable information on the large informal employment sector is very hard to obtain. In this study, the percentage of the total population employed in the informal sector is estimated to be 44, 65%, whereas 19,64% of the total population is estimated to be regularly employed. So information on population distribution in different occupancies is not available for almost half of Ahmedabad total population. In addition, this means that the occupancy for the residential category can be assumed to be overestimated. With no information on the employment of almost half the population, a reliable adjustment of the default relationships to the local conditions is not feasible.

The second approach is based on the occupancy curves for residential and non-residential buildings by urban population developed by Coburn and Spence (2002). The curves were initially developed for calculating the occupancy on building level for different times of day. In this study, the curves are applied on city level with two occupancy categories residential and non-residential to estimate the percentage of the total population in these two occupancies for different times of day. For the modelling of the population distribution, the city area is binary divided into residential and non-residential areas based on the manually digitized occupancy data set. The estimation using the occupancy curves revealed that more than 3 Million people are in residential occupancy at 0:00am. The least number of people in residential occupancy are calculated for 12:00 am. This means that more than 2.8 million people commute from residential to non-residential occupancy between 6:00am and 12:00am. In the non-residential areas, this leads to a very high population density of 133.814 people/km^2 between 6:00am and 12:00am.

Compared to the occupancy relationships provided by HAZUS, the occupancy curves by Coburn and Spence (2002) are generalizing. The binary division of residential and non-residential occupancy does not allow for including secondary information on the non-residential occupancy i.e. employment or non-working population. The HAZUS relationships offer different occupancies which can be used. This allows for adjusting the occupancies to be considered to the study areas, although the adjustment of the relationships strongly depends on the data availability. Another important difference is that the HAZUS methodology is limited to three times of day whereas the occupancy curves by Coburn and Spence can be applied continuously over the course of day. The three time scenarios within HAZUS are predefined by the casualty module. In the HAZUS handbook, it is stated that these scenarios are expected to generate the highest casualties for the population at home (2:00am), the population at work/school (2:00pm) and the population during rush hour (5:00pm). From this it can be concluded that with HAZUS the highest number of casualties is expected to be directly linked to the highest number of people. With Coburn and Spence (2002), the number of casualties is related to the extent of building damage and not to the highest population density. Their fine grained casualty estimation module operates on building level and uses a set of parameters to estimate the proportions of people rescued and trapped and the injury distribution among them. Very detailed information is employed, so called "M – parameters" including number of people per building, number of storey and building type. So, it is confusing that only two categories are provided for the estimation of the building occupancy.

So far, no methodology of estimating and modeling occupancy had been developed for city level application. The results of the city level modeling in this study show that the tested methodologies allow for generating occupancy estimates and population distribution maps on city level. In order to evaluate whether the two methods produce comparable results for a selected area, the calculated population is compared for selected times of day. For the night time scenario at 2:00am, the HAZUS method estimates a population density of 61.189 people/km^2 for residential occupancy and about 4.332 people/km^2 for non-residential occupancy. Using the occupancy curve by Coburn and Spence (2002), the population density is estimated to be 46.972 people / km^2 in residential and 20.226 people/km^2 in non-residential occupancy. The deviation is due to the different percentage of people in a specific occupancy. For 2:00am, the HAZUS method estimates 99% of the residential population to be in residential occupancy, whereas Coburn and Spence (2002) method estimates 76% to be in residential occupancy. For 5:00pm, the HAZUS method estimates 30.903 people/km^2 in residential occupancy and 25.694 people/km^2 in non-residential occupancy. Coburn and Spence (2002) method estimates 34.611 people/km^2 in residential occupancy and 108.911 people/km^2 in non-residential occupancy. Although the estimates for residential occupancy are similar for 5:00pm – commuting time scenario, the estimates for non-residential occupancy show large difference. The small estimate of 25.694 people/km^2 for non-residential occupancy is due to the underestimation of number of people employed in different sectors. Compared to the population estimates and population maps generated by model I and II, the occupancy based estimation and modeling provides more details of the population distribution at different times of day, whereas model I and II only provide night time, residential population estimates.

5 Tier 2 – District level

At tier 2, the population is estimated on district level (objective 5). The proposed approach considers three population estimation models, for which processing time, data requirement and cost increase with information detail. Model IV provides population information for individual districts based on the assumption of a uniform population distribution in the district. Model V employs satellite images to extract the built-up area of each district in which the population is supposed to be uniformly distributed. Model VI uses occupancy categories to model the population distribution within individual districts.

5.1 Model IV: District population estimation

In the following section, the underlying district population data set is generated by combining different statistical population datasets. As a first step, the decadal change in population for each district between census 1991 and census 2001 is analysed. In total, there are 43 districts within the AMC area. In 7 districts the population decreased from 1991 to 2001, whereas the population the 36 remaining districts increased (Census of India, 2007). Thus, the underlying premise of a constant annual growth (AGR) for Ahmedabad in model I and II is not valid on district level. If the AGR of 2,05% (2002) and 2,04% (2002 - 2007) is applied to each district over 7 years (2001 – 2008), this would result in an overall population increase of 14,29% in each district and the total population of all 43 districts for 2008 would be 4.023.105. Surprisingly, despite the fact that the constant AGR for each district does not reflect the districtwise variations in population change, the aggregated population shows a deviation of less than 1% from the city population projected by AMC (see Table 32).

To generate a district population data set, the district-wise population for 2008 is calculated using a district-specific AGR determined from population data (1991 – 2001) provided by the AMC. The resulting aggregated total population of the AMC area is even closer to the projection by AMC. Again, it is surprising to find that the difference for the two population estimations is only 0,29% (see Table 32) and that the influence of the population variations on district level has no major influence on the overall city population estimate for Ahmedabad.

The district level population map generated by model IV provides two kinds of population information: (1) Population estimation for 2008 and (2) population density based on the population estimation and the district area (see section 2.3.4 for the generation of the district map). The calculated population estimates and population density values for each district are listed in Appendix V.

Table 32: Projected population data for 2008 using a district-specific AGR and population projected by AMC using a constant city AGR (see Table 3).

Technique	Population 2008 aggregated from district level	Population 2008 projected by AMC using AGR	Deviation from population 2008 projected by AMC in %
Constant AGR	4.023.105	4.055.087	0,79
District specific AGR	4.034.785	4.055.087	0,29

The resulting population and population density maps are displayed in Figure 18 and Figure 19. From the population map is becomes obvious that the most people live in the southern districts of the AMC area. The least people are estimated for the district within the historical centre of Ahmedabad. However, looking at the population density map it can be observed that southern districts have a smaller population density as the central districts. This is due to the relatively small area of the central districts compared to the number of people living there. In contrast, the southern districts are fairly large and thus, less densely populated. This makes clear that the assumption of a uniform population distribution on district level is too generalizing to obtain a realistic picture of the population distribution within each individual district.

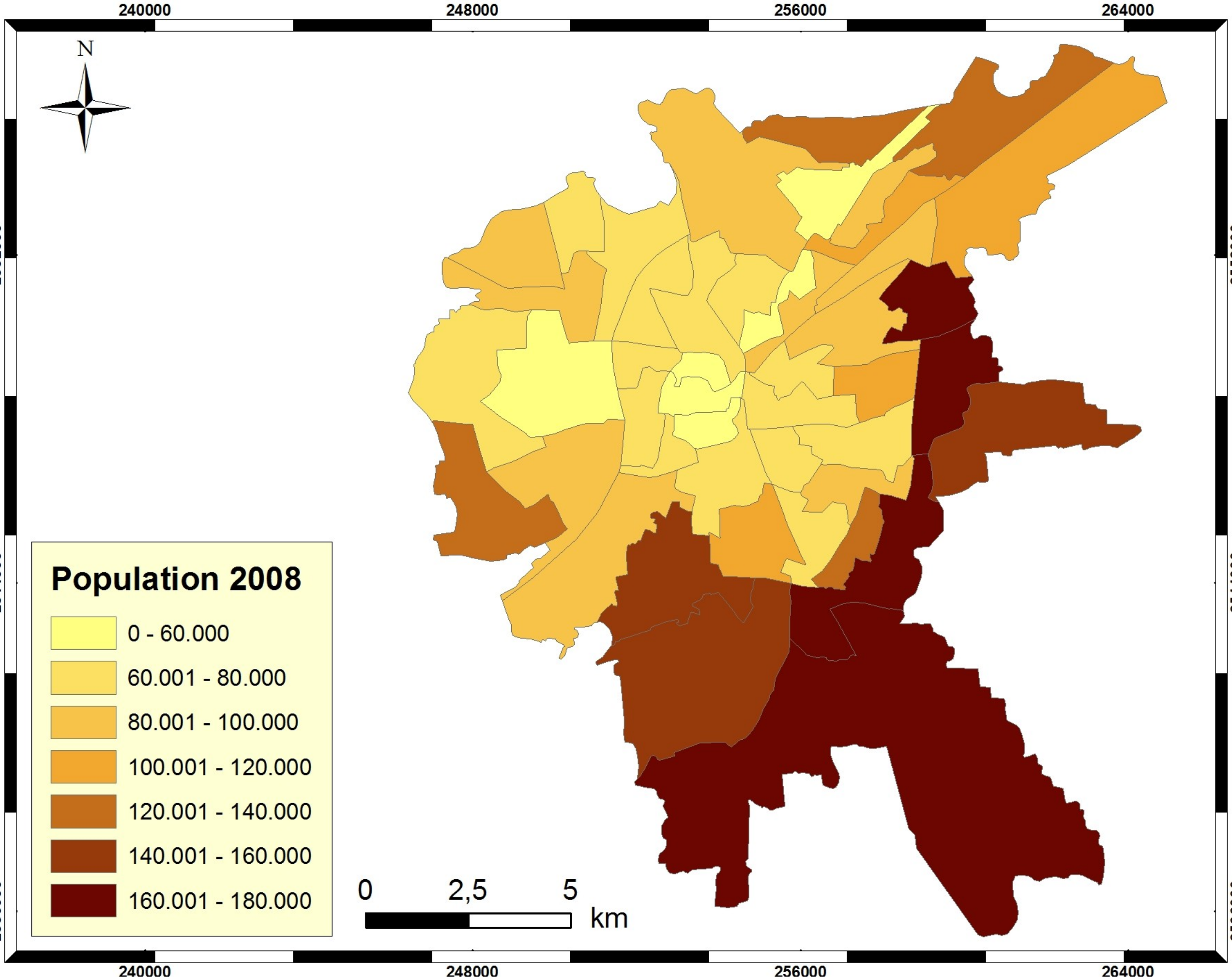


Figure 18: District level population map for 2008. The population estimate for each district is calculated based on a district specific AGR.

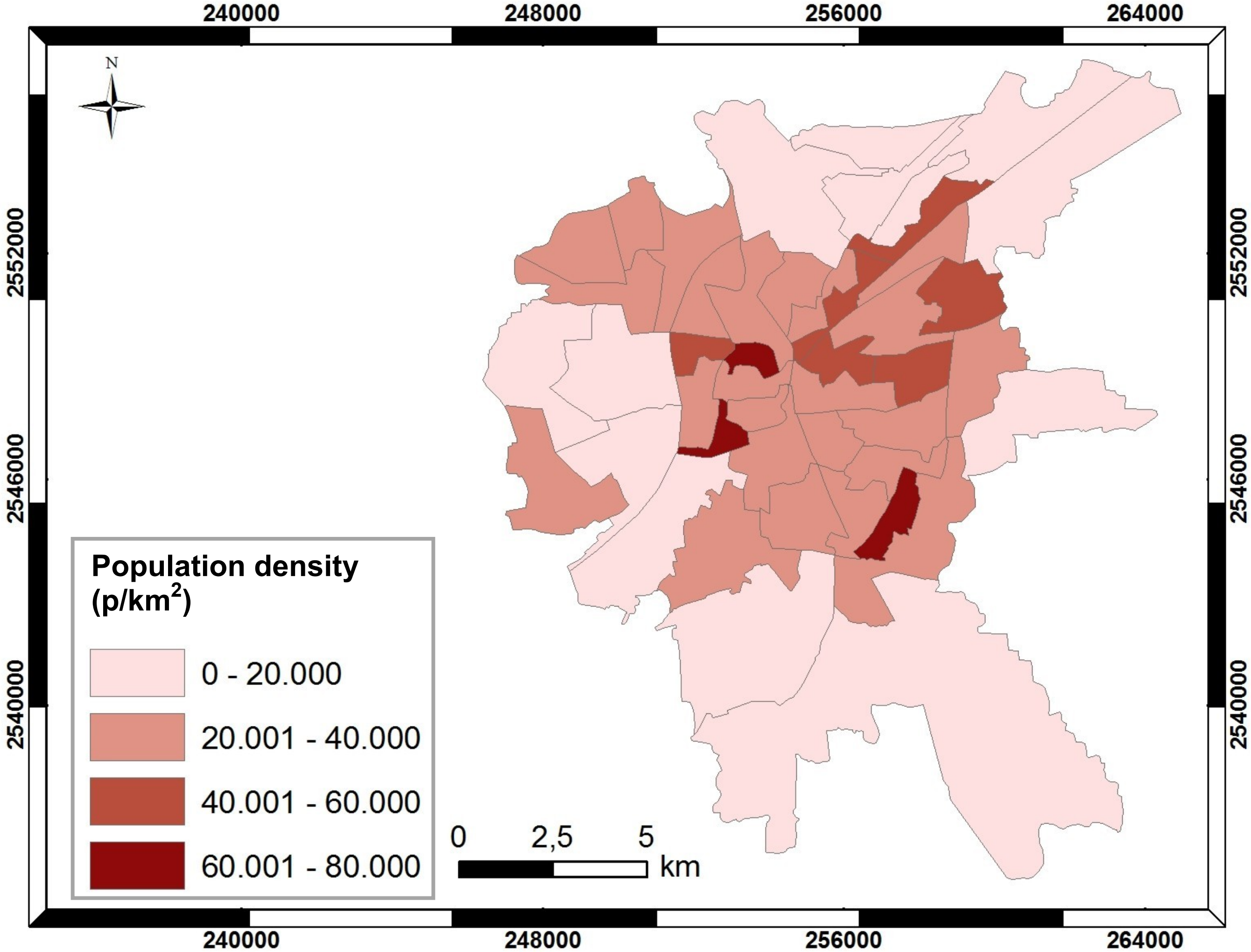


Figure 19: District level population density map. The population density is calculated based on the population estimates for 2008 using a district specific AGR and the district area determined from the administrative district map.

5.2 Model V: Urban district population estimation

In this section, urban population and the urban population density in estimated for each district. Based on the assumption that people reside in the built-up parts of the districts, information on the built-up extent in each district is integrated in model V. The built-up areas are extracted from Quickbird and Landsat 5 TM satellite images (see section 4.2.1) and therefore, two different population density maps are generated. The spatial extent of the maps is limited to the extent of the image and therefore, only 37 out of 43 districts are considered. The integration of spatially referenced information in model V is only feasible if the analysis is conducted using GIS software. This way information on the spatial distribution of the population in each district can be provided. In a first step, the built-up percentage of each district is calculated using the administrative district map and the built-up areas extracted from satellite images. Using the population estimate for each district calculated with model IV in section 5.1 using a district specific AGR, the population density within the built-up areas is calculated. Figure 20 and Figure 21 display the population density maps based on the built-up area extracted from Quickbird and Landsat 5 TM images. The higher population density calculated using the built-up areas can be explained by the fact that the built-up area extracted from Quickbird image is smaller than the areas extracted from Landsat (for details see section 4.2.1). For the discussion of the results see section 5.4 this chapter.

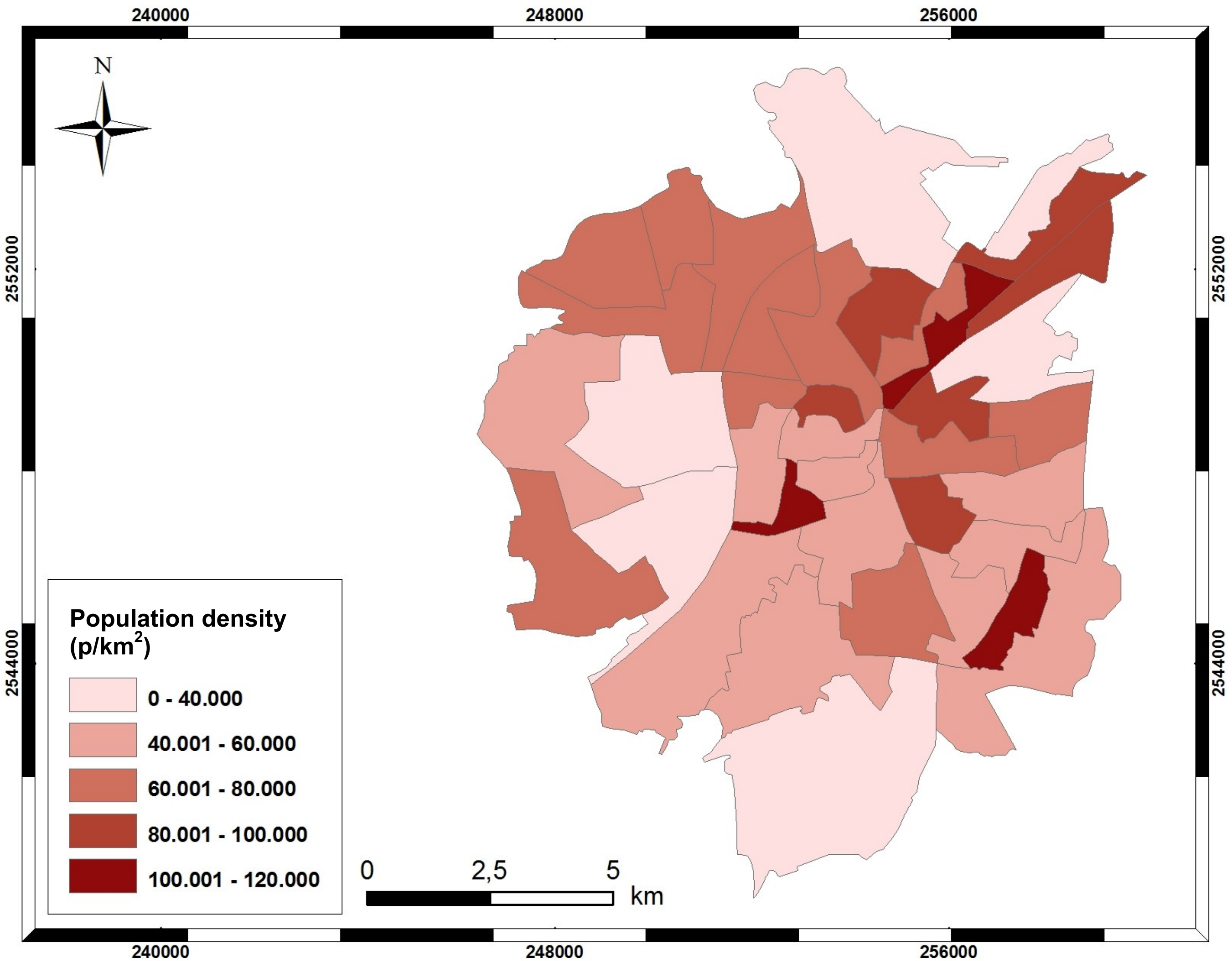


Figure 20: Population density map (people/km²) generated based on the built-up area in each district extracted from Quickbird image.

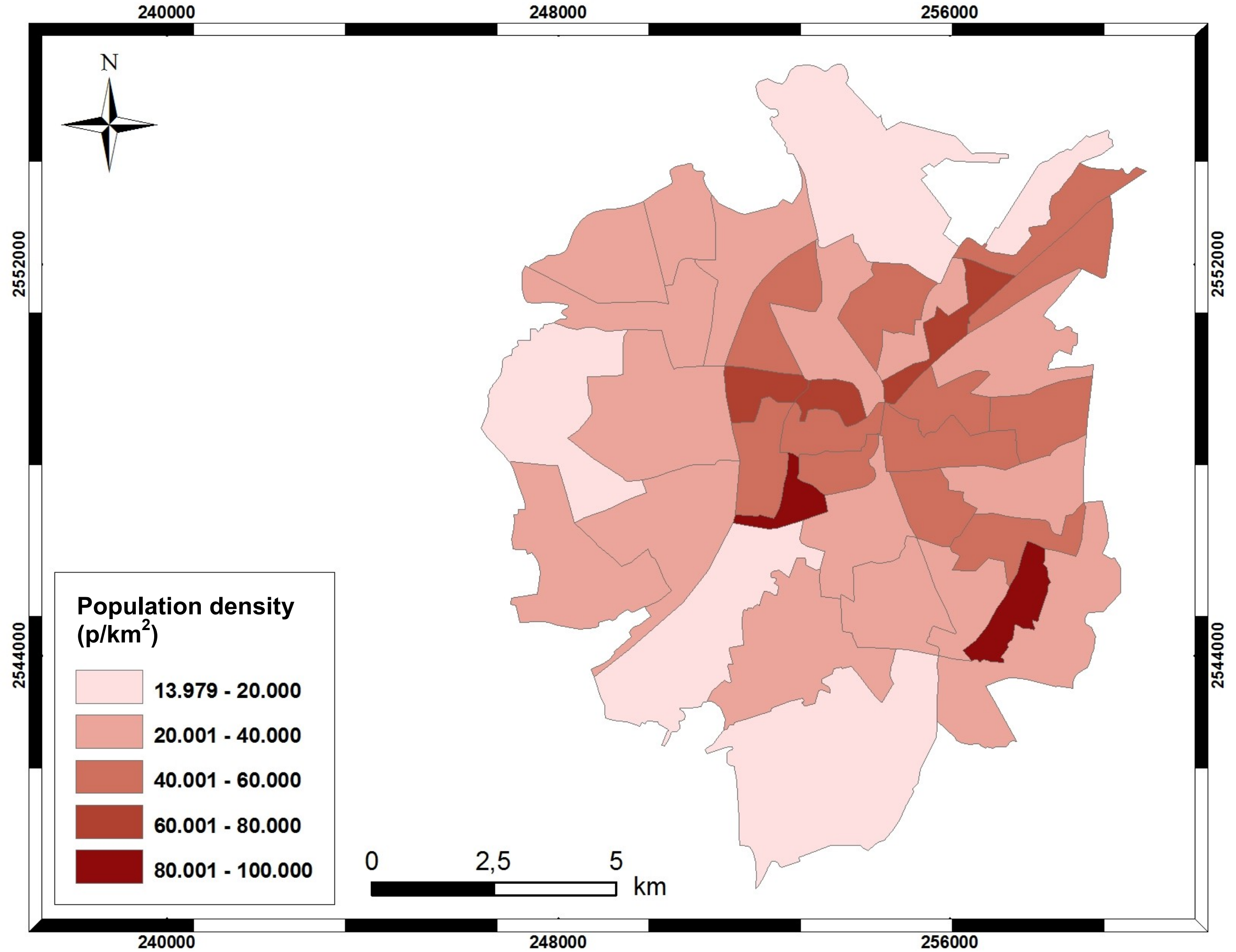


Figure 21: Population density map (people/km^2) generated based on the built-up area in each district extracted from Landsat 5 TM image.

5.3 Model VI: Occupancy based district population estimation

In this section, the district population is modelled based on the different occupancies identified in section 4.3.2. With model VI, it is assumed that the population density is the same in the occupancy categories throughout the districts. This simplification is necessary since employment data are not available on district level and thus, occupational pattern for each individual district cannot be developed. In a first step, the fraction of different occupancies in each district is determined by overlaying the occupancy data set with the district map.

The population densities calculated using the HAZUS relationships on city level are applied to calculate the population for each occupancy category for each district. Then the population is divided by the district area to calculate the occupancy specific population density for the district. The results are displayed in population distribution maps for the three HAZUS scenarios (02:00am, 05:00pm, and 02:00pm). As an example, Figure 22 displays the population distribution on district level for residential occupancy for the three time scenarios. This kind of map can be produced for each considered occupancy category.

The occupancy curves by Coburn and Spence (2002) are applied on district level for four selected times of day (0:00am, 6:00am, 12:00am, and 6:00pm). The population densities calculated using the occupancy curves with model III are applied to each district to estimate the population in each occupancy category. In the next step, the occupancy specific population density is calculated for each district using the district area. Figure 23 displays the population density maps for the four time scenarios (00:00am, 06:00pm, 12:00am, and 06:00pm).

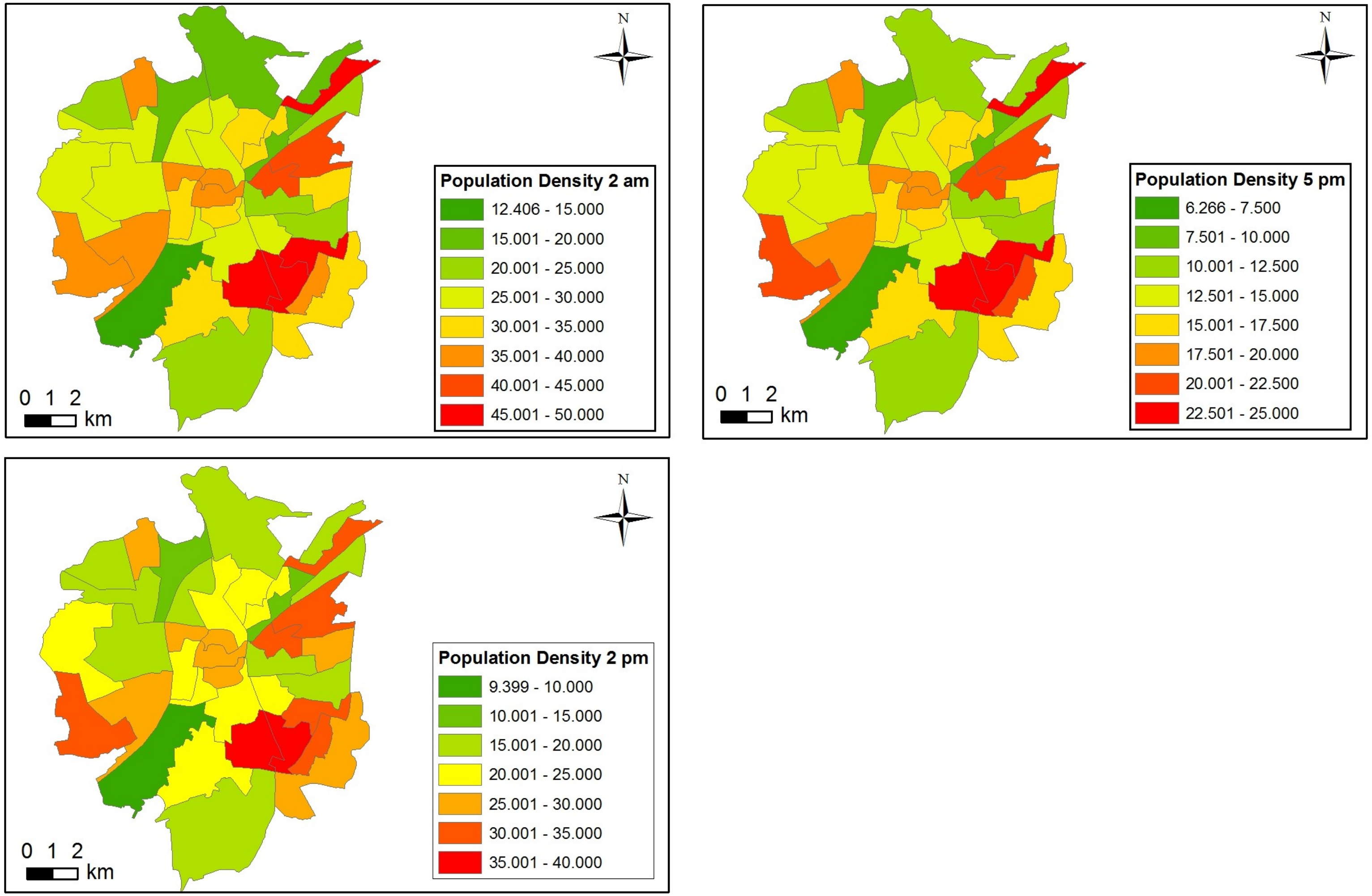


Figure 22: Population density maps (people/km^2) for residential occupancy for three time scenarios (2 am, 5 pm and 2 pm) based on the HAZUS relationships.

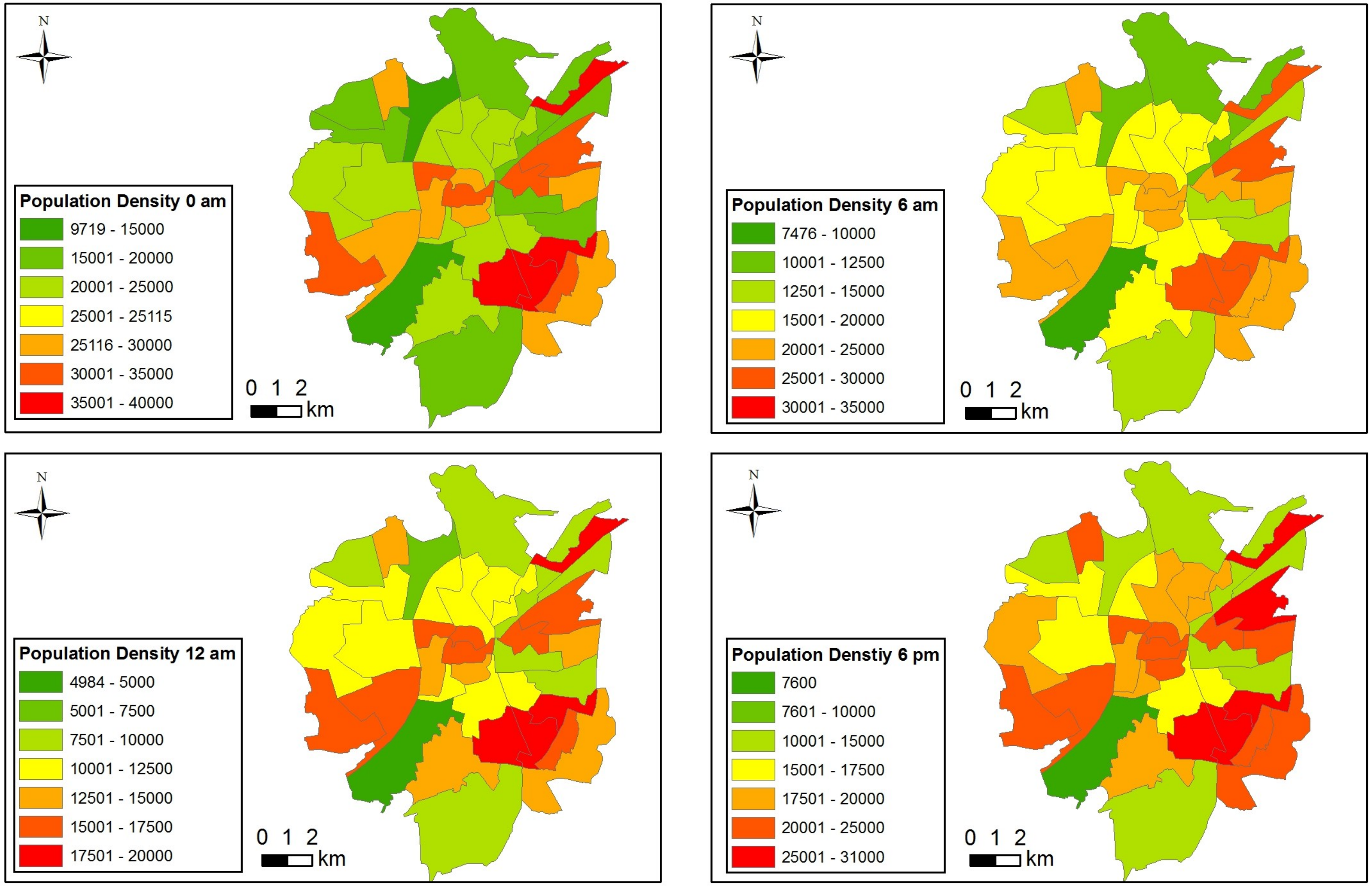


Figure 23: Population density maps (people/km²) for residential occupancy for four time scenarios (0am, 6 am, 12 am, and 6 pm) based on the occupancy curve provided by Coburn and Spence (2002).

5.4 Findings

In this section, the findings of the population estimation and population modelling on district level are discussed. Model IV generates population and population density maps based projected district population values for 2008 and the administrative district area (see section 2.3.4). The generated population maps revealed that most districts with the low population density are located at the periphery of the AMC area (see Figure 19). This is because of the larger district area of the peripheral districts. For example, the Isanpur district located in the southern periphery of the AMC encompasses an area of 12,93km^2 and the Sabarmati district located in the northern periphery of the AMC encompasses 12,10km^2. In contrast, the Jamalpur district located in the centre of the AMC area has an area of 1,01km^2. Although the population of these large districts is considerably higher than the population of the smaller districts, the large difference in area leads to low population densities. In 2008, the population density in Isanpur is 4.127 people/km^2 and in Sabarmati the population density is 6.615 people/km^2, compared to this the population density in Jamalpur is very high 6.8451 people/km^2.

Model V: urban district population is based on the assumptions that people usually reside in built-up areas. The resulting population and population density maps depict a less distinct trend of the population distribution. Using the built-up area extracted from Quickbird image, most of the high population density districts are located in the northern part of AMC with population densities ranging from 80.000 to 120.000 people/km^2. For the Landsat 5 TM, most of the high population density districts are located in the centre part of AMC with population densities ranging between 60.000 and 100.000 people / km^2. In general, the population densities calculated using Quickbird results are higher than using Landsat results because the extracted built-up area is larger for Landsat than for Quickbird (see section 4.2.1 for a detailed discussion). For example, for the Isanpur district a population density of 27.815 people / km^2 is calculated using Quickbird based built-up areas and 15.110 people/km^2 using Landsat based built-up area. The similar magnitude of deviation can be observed for smaller districts. Figure 24 displays exemplary the night time population estimated for district number 1 to 14. It is obvious that no distinct trend for over- or underestimation the population using the urban extent rather than the administrative area can be observed. This is also in accordance with the results obtained in section 4.2.3.

Model VI employs information on the different occupancies for each district and generates population information for different times of day. Two methods are used: HAZUS occupancy relationships for three time scenarios and occupancy curves after Coburn and Spence (2002). With model VI, district level population maps can be developed displaying population densities for each occupancy or different times of day. The deviations between the district population densities calculated the two different methods originate from the different percentage of population assumed to be present in a particular occupancy for a specific time of day.

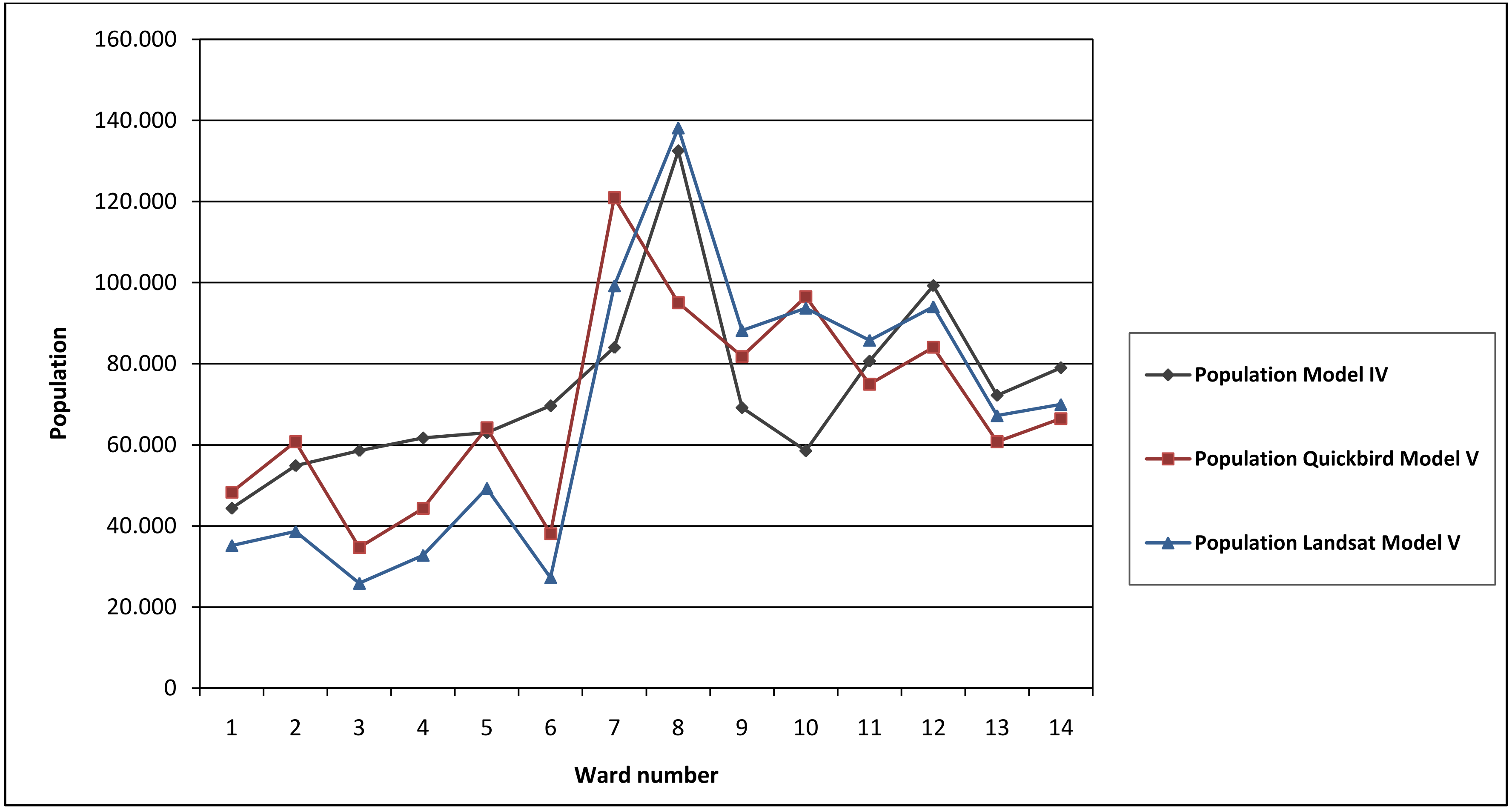


Figure 24: Population for district number 1 to 14 estimated using Model IV and Model V (Landsat and Quickbird).

Table 33 shows the comparison the population densities calculated for 5:00pm and 2:00am (HAZUS) and 6:00pm and 0:00am (Coburn & Spence, 2002) for two selected districts. The calculated population densities reveal the same trend; the population is decreasing from night to day time in residential occupancy. From this it can be concluded, that both districts have a higher percentage of residential occupancy than non-residential occupancy where people work during day time. For Jamalpur, 48% of the district is categorized as residential, for Isanpur 37% of the district is residential. From Table 33, it becomes obvious that the resulting population density is quite similar. The difference ranges between 2730 people/km^2 for Isanpur at commuting time and 6236 people/km^2 for Jamalpur at night time.

Some difficulties arise from the similarity of the population estimates for different times of day. Especially with a decreasing total population of the considered unit of analysis, the difference between the categories become smaller and thus, displaying these results in a population distribution maps is rather difficult. Displaying the results in graduated colours to show the difference is difficult for small changes. Therefore, a sub-district level map is recommended to be used to display minor changes in population over the course of day. A methodology for generating population maps on sub-district level is presented in chapter 6.

Table 33: Population densities estimated for the Jamalpur and Isanpur district using the HAZUS occupancy relationships and the occupancy curves by Coburn and Spence (2002).

District	HAZUS		Coburn & Spence	
	2:00 am	5:00 pm	0:00 am	6:00 pm
Jamalpur	29.463	14.880	23.227	18.165
Isanpur	22.650	11.439	18.118	14.169

6 Tier 3 – Building level

At tier 3, the developed models operate on building level (objective 6). The models allow for developing maps which display the population distribution within individual districts. Two models are developed at tier 3: Model VII: Building block population and Model VIII: Building population. With model VII, buildings with the same occupancy are digitized as building blocks and the population is calculated using the occupancy based population densities calculated in section 5.3.

In this study, model VII is tested exemplary for the Naranpura district (no. 12) located in the north-western part of Ahmedabad. The occupancy based population densities derived on district level at tier 2 are transferred to building block level at tier 3. For each building-block polygon in the district, the population can be calculated based on its occupancy using a top-down approach. The top-down approach is used on the occupancy based population data calculated using the HAZUS relationships and on the population data calculated using the occupancy curves by Coburn & Spence (2002). It has to be noted that with no information on the number of floors or building height available, the polygon area had to be used instead. So only the absolute number of people in an individual building block is calculated and no information on the distribution of the population within a building block is obtained. Figure **25** shows the population map for the Naranpura district which displays the population distribution at 2:00 am. The population is also modelled for the Naranpura district using the occupancy curves by Coburn and Spence (2002) (see Figure **26**). For each occupancy polygon, the population is estimated for the four selected times of day using the population densities derived in section 4.3.3. Both maps reveal that the least people are in residential occupancy at night time in the north-eastern part of the district, whereas a large number of people are estimated to reside in the south-western part of the district. According to the HAZUS relationships 99% of the population reside in residential occupancy at night time (2 am), therefore the population estimated per building block is larger than from the occupancy curves, with 78% residing in residential occupancy at night time.

Model VII provides a methodology to develop per building population estimates. In this study, information on building level for the city of Ahmedabad from the Indian census 2001 is not available from the census bureau. Instead, commercial companies offer building data sets for the AMC area. A sample data set is available for download from a company's website. The sample data set covers 744 surveyed building in the districts Gandhigram (district no. 9) and S P Stadium (district no. 11) in the north-western part of the AMC area. The following attributes are provided for each building: Occupancy category, number of floors, building height, number of apartments, and number of shops. Information on the number of people per building is not included in the attribute tables.

To initially test the quality of the included attributes, a visual comparison with the Quickbird image is conducted. This inspection revealed that the information provided on building height is questionable. According to the attribute table, some buildings have a height of 1 to 2 m, but from the image it becomes obvious that the building has at least 4 storeys and a height of at least 10m. In addition, the attribute table is incomplete. Height information is missing for 41 out of 744 buildings.

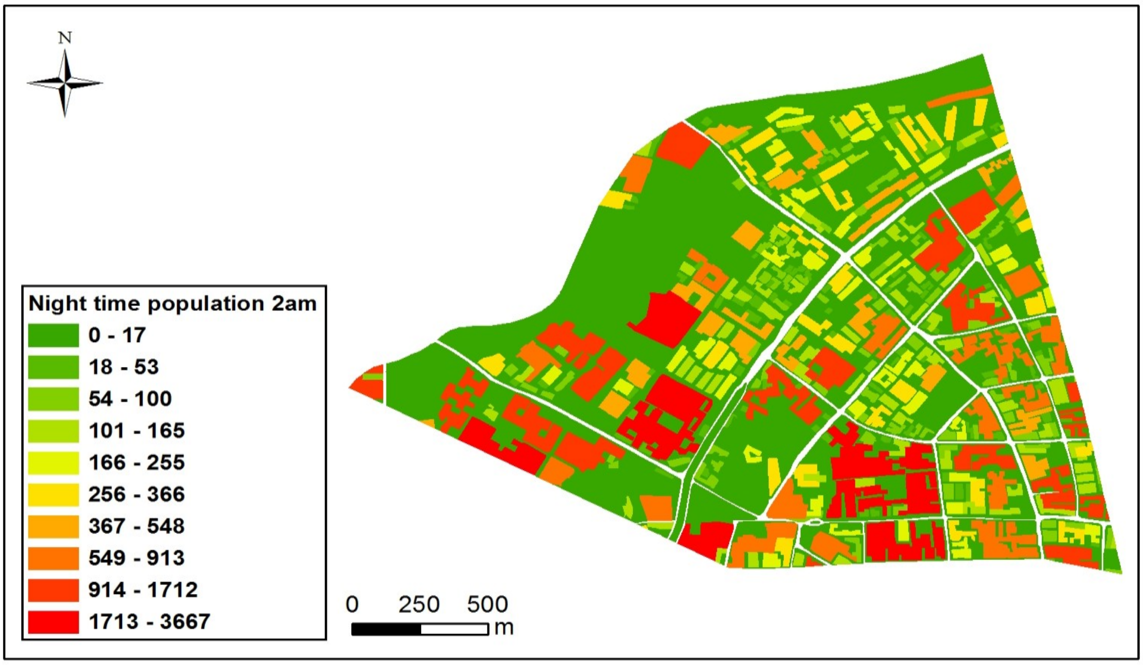


Figure 25: Population distribution for the Naranpura district for the HAZUS night time scenario at 02:00 am.

Figure 26: Population distribution for the Naranpura district for 00:00 am following Coburn and Spence (2002).

The up-to-dateness of the building outlines can only be guessed by comparing the buildings present in the satellite image acquired in 2008 with the building polygons in the sample data set. This is hindered by the fact that the sample data set provided in an India specific coordinates system and the polygon are skewed during the transformation. The visual comparison shows that some building included in the data set are demolished while new constructed buildings are not included. Therefore, the data set must be older than the image acquisition date. Taking the mentioned uncertainties into account, it is concluded that the sample data is not suitable to serve as basis for the per-building analysis. The problem of unavailable digital building level maps for the AMC was already encountered in previous research studies. Katuri et al. (2007) conducted a risk assessment study for the AMC area and used urban mapping units which conist of a number of buildings with similar land use pattern instead of single buildings. Katuri et al. (2007) points out that detailed building level information for the whole city is lacking for each individual building.

Considering the lack of building information, the methodology to calculate the number of people per building cannot be tested for the city of Ahmedabad and can only be presented as a concept. In general, there are two major approaches for estimating building population: (1) Top-down, and (2) bottom-up. With the top-down approach, the average population per m^2 for each occupancy category is calculated based on the results generated with model VI for individual districts. For each building, the usable space is calculated by multiplying the footprint area with the number of floors derived from the building height. The usable space in m^2 is then multiplied with the average population density to calculate the number of people in each building in every occupancy category.

To overcome the limit availability of build level data, a number of techniques to semi-automatically extract the building information from very high resolution satellite images such as Quickbird exist (see section 3.4.3). However, for this case study an automated extraction of building attributes is only feasible for specific building types under certain conditions. The most vital condition is that the spectral characteristics of building roof are sufficiently different from the spectral characteristic of the surrounding objects. In order to generate representative building footprints, all parts of the building roof have to be spectrally similar, so they can be assigned to the same class. Another solution is to classify the building roof in different classes and join these classes later. However, this procedure may lead to confusion with spectrally similar objects and should be avoided. Another problem arises from high density of buildings in some areas of Ahmedabad, for example the historical city centre with its pol structure or the low income areas. This leads to problems in separating adjacent buildings. With the bottom-up approach, information collected for single building by survey is interpolated for the whole study areas. From the information available, generalizing building categories are developed. For example, if for the majority of 3-storey residential buildings with a building footprint area between $100m^2$ and $80m^2$ and a resulting useable space of approximately $270m^2$, 20 inhabitants are enumerated, this is assumed to be the case for all buildings in the study area that fit into this category.

7 Summary and outlook

The overarching goal of this study was to **develop a tiered estimation method for urban population that combines optical satellite imagery and census data to provide population information for large cities**. Models for population estimation and population distribution modelling were developed at city level (Tier 1), district level (Tier 2), and building level (Tier 3). As a case study for the methodology development the city of Ahmedabad located in northwest India was selected.

7.1 Summary

Exhaustive population surveys such as the official census are conducted by official authorities forms the main source of population information. In many countries, the collected data are attributed to the administrative units for which they were collected and then provided as geocoded data sets for the respective census year. Due to the time- and cost-consuming nature of this procedure, this kind of data is only collected on a decennial basis, in some countries every five years. Facing the rapid development of today's cities, especially in developing countries such as India, the census data fail to monitor the dynamic of the urban population. From this time gap for which no population is available, the need for population projection and estimation and for the employment of modern technologies such as satellite images and GIS arises. To date, a number of studies on population estimation and remote sensing have been conducted. However, the majority of these studies consider only small test site. The suitability of existing techniques for population estimation in large cities has not been explored yet. In contrast to these sample site studies, a considerable amount of research has been focused on global population estimation. The review of existing population data and population estimation techniques in section 3.1 and 3.3 of this dissertation revealed the need to develop a methodology that fills this research gap. A methodology for population estimation and for modelling the spatial population distribution of large cities without relying on detailed survey data, but still fine grained enough to represent the complexity of the city's population.

A common feature of all population models developed in this study is the employment of Indian census and other statistical data. In order to be able to assess the inherent uncertainties associated with these data, a number of sources of uncertainties were identified (see section 3.3.1). Strictly speaking, the census information is only valid for the census moment, which was 00.00 hours of 1st March 2001 for the last Indian census. However, for the annual population projection the information is assumed to be valid for the year 2001 in general. So the population changes which occur in the remaining month of 2001 cannot be considered. A second source of uncertainty arises from the aggregated form in which Census information is available. In this study, the district level is the most detailed unit of analysis available. Therefore, the population estimation on sub-district level is more dependent on the validity of the underlying assumptions. The quality of the underlying assumptions, however, is related to the availability of secondary information and on expert knowledge. This subjective influence is very difficult to assess, but also facilitates the integration of local knowledge. For example, the analysis of the occupational pattern of Ahmedabad in section 4.3.2 revealed that the informal sector was not considered at all in the employment statistics. The information on the number of people working in specific sectors was therefore assumed to be minimum estimates. This information could be significantly improved using local expert knowledge.

In chapter 4, the population estimation models were developed for the city level at tier 1. For 2008, a population projection was available from the AMC statistical department. To verify the quality of this projection, the vital rate procedure was applied to generated a validation population projected for 2006 and compare this projected the projection by the AMC statistical department (see section 2.3.3). The comparison revealed a very small deviation of 1,6% (3.958.423 this study, 3.894.430 AMC). From this, it was concluded that the official projection for 2008 was sufficiently accurate to be used as input information. This indirect way of validating the input data has the advantage of moderate data requirement. The vital statistics used in this procedure are available for a large number of cities and thus, this procedure could be applied to verify population data for other cities.

At tier 1 with model I on city level, this verified population information was used to calculate the population density of the administrative AMC area in 2008 using an area overlay methodology based on the simplified assumption that population is uniformly distributed across the city. For an extent of the administrative boundary of 136,51 km^2, a population density of 23.091 people/km^2 was calculated. This very simple model allows for calculating the population density with very limited data requirements. The area of the administrative unit and a population count is most often provided by the authorities (see section 3.3). A limitation of model I is that no information on the spatial distribution of the population within the selected, geographic region can be obtained.

The underlying assumption of a uniform distributed population is very generalizing and to obtain a more realistic picture of the population distribution only the build-up areas were considered with model II (see section 4.2). The extent of the built-up areas was extracted from satellite images using an NDVI – based procedure (see section 4.2.1). This procedure was selected because it can be easily applied on the large data volumes which come along with the extent of large cities. Two types of satellite images were tested: (1) Commercial, very-high resolution Quickbird images, and (2) freely available, moderate resolution Landsat 5 TM images. The image analysis showed that the extent of the built-up area from Quickbird is smaller than the extent extracted from Landsat. From the Quickbird image, a total built-up extent of 55,50 km^2 was extracted, from the Landsat image 91,50km^2 were extracted as built-up area. The larger extent for the Landsat image is caused by the mixed pixel problem which arises from the larger pixel size of the Landsat image compared to the Quickbird pixel size. The accuracy assessment based on visually classified reference pixel (1472 for Quickbird and 656 for Landsat) showed an overall accuracy of 81,00% for Quickbird and 78,45% for Landsat. Real ground truth data were not available since a ground survey in Ahmedabad was beyond the scope of this study. The accuracy assessment is therefore influenced by the subjective interpretation, but offers, however, an alternative for validation.

The population density for the built-up areas was calculated using the projected population for 2008 3.152.108. Using the Quickbird derived built-up extent, a population density of 56.791 people/km^2 was calculated; using the Landsat derived built-up extent, a population density of 34.446 people/km^2 was calculated. From this, it can be concluded that using Landsat images for population estimation results in an underestimation of the population density. The estimate can still be considered to depict a more realistic population density than the population density calculated with model I. With Landsat image being available for free download on the internet and the image analysis being also feasible with open source image analysis software, model II enables an improvement of existing population counts without any additional expenses. It is important to note that the application of Landsat image for population estimation is limited to coarse, large-scale population studies because of the pixel size of 900m^2. The extraction of individual buildings is not feasible and therefore for studies using a smaller unit of analysis higher resolution imagery (Quickbird, Ikonos) have to be employed. Therefore, Quickbird imagery was used for the more detailed analysis on district and building level in chapter 5 and 6 of this dissertation.

The use of satellite images lead to the need of integrating a spatial reference into the model. As a digital map for the AMC areas was not available at a reasonable price, an open source method to generate a geocoded, administrative data set for the city of Ahmedabad was developed in section 2.4.4. In order to assure, that the resulting data sets was of sufficient accuracy, the resulting district areas were compared to the official district areas provided by the AMC. The deviations ranged between 0,89km^2 and 1,03km^2. This results show that in areas with no geocoded data available an administrative boundary map can be generated and facilitates therefore a GIS-based analysis of the population. This procedure can also be used to verify and refine existing GIS data sets which are available in a coarse resolution of many regions of the world.

To overcome the simplification of an equally distributed population within the considered unit of analysis and to enable the distinction of day and night time population, model III considers occupancy categories which allow for disaggregating the population based to occupancy relationships (see section 0). The following occupancies were identified for the city of Ahmedabad: residential (low income, middle and high income, pols), industrial, commercial, and service (see section 4.3.1). The number of people employed in the three sectors was determined from statistical data provided by non-governmental organizations and statistical offices in Ahmedabad. In addition, the share of non-working population including unemployed people, children less than 6 years, school children and students was calculated. The analysis of the occupational pattern showed that in 2008 796.747 people were registered as employed in various sectors in the AMC area. Considering that 1.447.364 people were assumed to be not working and a total population of 4.023.106 people, it was concluded that the remaining 2.051.780 people were working in the informal sector. The employment in the informal sector constitutes a major source of uncertainty in occupancy based population estimation. Information on population distribution in different occupancies is not available for almost half of Ahmedabad total population. Therefore, it is very likely that the number of people employed in the industrial, commercial and service sector was underestimated since only regularly employed people were considered in this study. The estimation of children younger than 6 years can be assumed to be comparably accurate as the registration of new borne is obligatory in India. In this study, the number of people which are retired and therefore not working is not considered because the share of the population is expected to be very small due to an average retirement age between 58 and 60 years and a life expectancy of about 64 years.

In section 0, two different methodologies are tested to estimate the population in different occupancy categories for different times of day on city level. The occupancy curves provided by Coburn & Spence (2002) were initially developed for calculating the occupancy on building level for different times of day. For the modelling of the population distribution, the city area is binary divided into residential and non-residential areas based on the manually digitized occupancy data set. The estimation using the occupancy curves revealed that more than 3 Million people are in residential occupancy at 0:00am. The least number of people in residential occupancy are calculated for 12:00am. This means that more than 2.8 million people commute to non-residential occupancy from residential between 6:00am and 12:00am. In the non-residential areas, this leads to a very high population density of 133.814 people/km^2 between 6:00am and 12:00am. The share of people assumed in a specific occupancy is derived from various past earthquakes for which survey were conducted including the counting of the number of people in residential and non-residential buildings. To what degree these occupancy percentages are also valid for the city of Ahmedabad could not be verified as information on building level was not available. To verify the occupancy percentages, information on people's sojourn over the course of the day has to be collected in surveys covering a statistically significant number of people.

In contrast to the continuous occupancy relationships by Coburn & Spence (2002), HAZUS provides occupancy percentage for three times of day: night time scenario (02:00am), day time scenario (02:00pm), and commute time scenario (05:00pm). Out of the six occupancy categories within HAZUS, four could be identified to be applicable to Ahmedabad (see section 4.3.1). For each of the four categories (residential, commercial, industrial, and service), the number of people was calculated for the predefined three time scenarios. The results showed that the largest population fluctuation occurs in residential occupancy between 2:00am and 5:00 pm with a total residential population of 3.965.033 people for the night time scenario and 2.002.543 people for the commute scenario. As a consequence of the informal sector not being considered, the population in residential occupancy is assumed to be overestimated, especially for the 5:00pm, commuting time scenario. Population estimation for the 2:00am, night time scenario can be assumed to be more realistic the people's employment only plays a marginal role for the population distribution at this time of day.

Comparing the two occupancy based population estimation approaches, it becomes obvious that the binary method by Coburn and Spence (2002) is very generalizing and does not allow for including secondary information for example on the employment pattern. This can be an advantage as the method can be easily applied to any administrative boundary the total population is known for and for which a binary occupancy map is available. With the HAZUS approach integrating secondary information is feasible and this approach is the better choice for study areas for which employment statistics are available. An important difference between the two approaches is that the HAZUS methodology is limited to three times of day whereas the occupancy curves by Coburn and Spence can be applied continuously over the course of day.

Following the system applied on city level at tier 1, at tier 2 (chapter 5): District level, three different models were developed. Model IV generates population and population density maps based projected district population values for 2008 and the administrative district area. The generated population maps revealed that the larger districts with lower population density are located at the periphery of the AMC area whereas the districts with higher population density are located in the central parts of the AMC. Model V is based on the assumptions that people usually reside in built-up areas (see section 5.2). The resulting population and population density maps depict a less distinct trend of the population distribution than model IV. In general, the population densities calculated using Quickbird results are higher than using Landsat results because the extracted built-up area is larger for Landsat than for Quickbird. This can be explained by the fact that urban areas in Ahmedabad are too complex and heterogeneous to be captured in detail by a moderate resolution image like Landsat with 900m^2 pixel size.

With model VI, district level population maps can be developed displaying population densities for each occupancy or different times of day. The deviations of the district population densities calculated using the two tested methods originates from the different percentage of population assumed to be present in a particular occupancy for a specific time of day. The same deviations observed on city level also occur on district level. Some difficulties with displaying the results arise from the similarity of the population estimates for different times of day. Especially with a decreasing total population of the considered unit of analysis, the difference between the categories become smaller and thus, displaying these results in a population distribution maps is rather difficult. Displaying the results in graduated colours to show the difference is difficult for small changes. Therefore, a sub-district level map is recommended to be used to display minor changes in population over the course of day. A methodology for generating population maps on sub-district level is presented in chapter 6.

At tier 3, the developed models operate on building level. The models allow for developing maps which display the population distribution within individual districts. Two models are developed at tier 3: Model VII: Building block population and model VIII: Building population. Model VII allows for calculating the distribution of the population on sub-district level without relying on per-building information. This procedure is very useful in areas with limited data availability because it nevertheless provides an insight into the population distribution within the districts. The population of each building block is calculated based on the occupancy and the related population density calculated with model VI. The lack of building level data results in absolute population estimate for each building block. The limited data availability also hindered the testing of model VIII. Commercially available data were test, but the quality proved to be inconsistent and out-date and therefore not suitable for the analysis.

In this study, different population estimation models are developed which allow for generating information on the population and its spatial distribution on spatial levels. In order to be able apply this method to other study sites than the city of Ahmedabad, certain data requirement have to be met. Table 34 provides an overview of the data requirements of each developed population estimation model. This table facilitates the choice of the most suitable model to be used for a selected application.

7.2 Conclusion and outlook

Considering the overall goal on this dissertation "**developing a tiered estimation method for urban population that combines optical satellite imagery and census data to provide population information for large cities**", a number of conclusions can be drawn.

Population estimation and spatial population distribution modelling for large cities for inter-census years without relying on detailed survey data is only feasible on city and district level. With the developed methodology, for both units of analysis, administrative city boundary and district, population information can be generated with different degree of detail. For sub-district level, the presented models VII and VIII only generate results for which the accuracy matches the large scale of the analysis with detailed building data available. In this study, with model VII only coarse population estimation for individual building block could be made and the testing of model VIII was not feasible at all.

The occupancy relationships used with model III and VI are not adjusted to the local conditions in Ahmedabad. A regional adjustment of the occupancy relationships would certainly improve the modelling of the population distribution. For this, detailed, reliable information on occupational pattern of the study site is required. This and the above conclusion lead to the more general problem of data availability for population studies. One approach might be the development of employment questionnaires and to conduct surveys in different regions to collect the necessary information. This information could be used to develop regional occupancy relationships. In order to assure that the collected information is assigned to the correct building, one possible solution is to use field data collection tools which allow recording the coordinates of each building using GPS. This kind of field survey technology is already used to collect damage information in the post-earthquake phase. The potential for pre-event data collection has yet to be explored. The use of this kind of technology for field data collection could also improve the availability of validation data. Especially, georeferenced validation data are of great value in the validation phase.

Nevertheless, in many countries very detailed information is collected in decennial censuses and other statistical surveys. In many cases, this information not available for research purposes due to privacy, confidentiality and sometimes political reasons. An international agreement of data exchange would be a major improvement to this problem. This data could be stored in an international data base for research purposes. A first step towards this direction form data base projects like the World Housing Encyclopedia hosted by the EERI and IAEE. This kind of projects enables researchers and other professional to make their local knowledge about building types available to the research community. In order to be able to use this kind of information for modeling and validation purposes, it would be very helpful to for example introduce geocoded building example as obligatory elements of a housing report. However, the development towards more open population data availability is hindered by media reports on the misuse of this kind of data. Ironically, the detailed data can in many cases be purchased from official data resellers.

Facing the limited availability of suitable reference data and no survey included in the research project, the validation had to be carried out indirectly by comparing trends and results obtained with different methods. The lack of reliable validation data and the inherent uncertainties associated with population estimation results in population data that should be conceived as the most likely status, but not the truth. A necessary step would include a standard methodology for population validation data generation, including a protocol for each data collection step. Until now, validation efforts for population distribution are limited to crosschecking the results with aggregated population counts on higher administrative levels provided by official sources or other institutions. With this, generalizations are unavoidable due to for example, limited data availability or time constraints and the influence of these generalization have to be considered in the interpretation of the results. If the results are used as input data for example in casualty estimation, the implications of the generalization have to be considered as well. An example for the impact of unreliable population data are the far out estimates of the affected population in the direct aftermath of earthquake.

In the research field of applied remote sensing, population and other social parameters still receive less attention than other physical parameters such as building types. One reason for this is that the relation between population structure and settlement pattern of urban areas is too complex and diverse to be comprised by simple generalizations. Another reason is the limited spectral resolution of existing optical images. In the last 20 years, the technical advances have lead to a decreasing spatial resolution of the optical satellite image but did not improve the limited spectral resolution of these images (see Table 35). The launch of the EnMap satellite in 2013 is the first step into the direction of providing hyperspectral, space born images for large scale analysis. Although the spatial resolution is limited to 30 m and does not allow for detailed urban analysis, it is sufficient for a city wide analysis. Considering the short time periods in which the spatial resolution of optical image improved, the launch of EnMap is seen as the first step towards very high resolution, hyperspectral satellite images in the near future. The potential of the EnMap image for urban application is immense for example different material can be distinguished which could be used for a more detailed classification of the urban areas.

Existing GIS – data and satellite images are complementary source of information which can be used for the generation of population data for specific applications. In this study, the data generation is limited to the three administrative levels. However, the developed methodology has to be tested for generating population for other spatial units used by other application. For example, some flood risk models operate on catchment areas which naturally intersect administrative boundaries. This is a challenging task as catchment areas can vary significantly in size.

Table 34: Population estimation models developed for city, district and building level, population data products and data requirements for each model.

Tier	Model	Spatial unit	Population data	Data requirement
1	I	City level: admin. area	Night time, residential population	Population count for source zone, area of source and target zone
1	II	City level: built-up area	Night time, residential population	Population count for image acquisition year, built-up extent of study site
1	III	City level: occupancy categories	Population distribution for different time scenarios and occupancy categories	HAZUS (1999): Urban population, people employed in the commercial and industrial sector, number of students in school, number of students in college, geocoded map with considered occupancy categories
				Coburn & Spence (2002): Urban population estimate, geocoded map with residential and non-residential categories
2	IV	District level: admin. area	Night time, residential population	Population count for individual districts, administrative area of individual districts
2	V	District level: built-up area	Night time, residential population	Population count for individual districts, built-up area for each district
2	VI	District level: occupancy categories	Population distribution for different time scenarios and occupancy categories	HAZUS (1999): Urban population for each district, people employed in the commercial and industrial sector, number of students in school, number of students in college, geocoded map with considered occupancy categories for each district
				Coburn & Spence (2002): Urban population estimate for each district, geocoded map with residential and non-residential categories for each district
3	VII	Building level: building block	Population distribution for different time scenarios	Population density for different occupancies, geocoded map with building blocks for different occupancy categories
3	VIII	Building level: individual buildings	Night time, residential population / population distribution for different time scenarios	Population count for higher spatial unit of analysis, geocoded map with single building and the following attributes: footprint area, building height, and number of floors
				Population information for sample buildings on number of people per building, geocoded map with single building and the following attributes: footprint area, building height, and number of floors

8 References

Ahmad, M. & Chattopadhyaya, B., 2001. *Structure and Dynamics of Urban Economy: Study of Linkeages between formal and informal Sectors in Ahmedabad and Vishakhapatnam*. Research Study Series Nr. 80. New Delhi: National Institute of Urban Affairs.

Ahmedabad Municipal Cooperation, 2008. *Slum Networking Project*. Pro Actice Disclosure as per RTI Act 2008.

Alavi, H., 1980. India: Transition from Feudalism to Colonial Capitalism. *Journal of Contemporary Asia*, 10(4), pp.359 - 399.

Alessandrini, E., 2008. Ahmedabad - Gandhi Ashram Area: Project for houses and craftsmen's workshop by the Sabarmati River. In *Proceedings of the 4th International Semiar on Vernacular Settlement*. CEPT University, Ahmedabad, 2008.

Al-Garni, A., 1995. Mathematical predictice models for population estimation in urban areas using space products and GIS Technology. *Math. Comput. Modelling*, 22(1), pp.95 - 107.

Al-Garni, A., 1996. Urban photogrammetric data base for multi-purpose cadastral-based information systems: the Riyadh city case. *International Journal of Photogrammetry and Remote Sensing*, 51, pp.28 - 38.

Ali, A., 1997. The potential of Almaz Sar Imagery for Population Estimation. *Sudan Engineering Society Publications* , pp.35 - 44.

Almeida, C. et al., 2007. Multilevel object-oriented classification of Quickbird images for urban population estimates. In Samet, H., Schneider, M. & Shahabi, C., eds. *Proceeding of the 15th International Symposium on Advances in Geographic Information Systems*. Seattle, 2007.

AMC, AUDA, CEPT, 2002. *Ahmedabad City Development Plan 2006 - 2010*. Ahmedabad.

AMC, 2003. *Ahmedabad Municipal Cooperation E-governance*. [Online] Available at: http://www.egovamc.com/ [Accessed March 2009].

AMC, 2004. *Statistical Outline AMC*. Ahmedabad: Ahmedabad Municipal Cooperation.

Ameri, B., 2000. *Automatic Recognition and 3D Reconstruction of Buildings from Digital Imagery*. Deutsche Geodätische Kommission C 526. München.

Anas, A., Arnott, R. & Small, K., 1998. Urban Spatial Structure. *Journal of Economic Literature, American Economic Association*, 36(3), pp.1426 - 1464.

ATC, 1985. *Earthquake Damage Evaluation Data for California*. Report ATC-13. Redwood City: Applied Technology Council.

Balk, D., Pozzi, F., Yetman, G., Deichmann, U., & Nelson, A. et al., 2004. *The Distribution of People and the Dimension of Place: Methodologies to Improve Global Estimation of Urban Extents*. New York: Center for International Earth Science Information Network, Columbia University.

Balk, D. & Yetmann, G., 2004. *The global distribution of population: Evaluating the gains in resolution refinement*. Documentation for GPW Version 3. New York: Center for Earth Science Information Network, Columbia University.

Baltsavias, E., Pateraki, M. & Zhang, L., 2001. Radiometric and geometric evaluation of Ikonos images for 3D building evaluation. In *Proceedings of the ISPRS Workshop on High Resolution Mapping from Space*. Hannover, Germany, 2001. International Society for Photogrammetry and Remote Sensing.

Balz, T. & Haala, N., 2005. Interpretation of high resolution SAR data using existing GIS data in urban areas. In *Proceedings of the Joint Workshop of ISPRS and DAGM CMRT*. Vienna, Austria, 2005.

Berry, B. & Spodek, H., 1971. Comparative ecologies of large Indian cities. *Economic Geography*, 47, pp.266 - 285.

Bharti, M., 2008. Willingness to Pay: In Urban Slums of Ahmedabad. In *Proceedings of the 14th ASEAN Regional Forum*. Manila, 2008.

Bhatt, M., 2003. Urban Slum Reports: The case of Ahmedabad, India. In *The challenge of slums*. Nairobi: United Nations Center for Human Settlements.

Binsell, R., 1967. *Dwelling Unit Estimation from Aerial Photography*. Illinois : Northwestern University, Department of Geography.

Biond Software Technologies , 2008. [Online] Available at: http://www.biondsoftware.com/index.html [Accessed February 2009].

Breman, J., 2001. An Informalised Labour System: End of Labour Market Dualism. *Economic and Political Weekly*, 36(52), pp.4804 - 4821.

Briggs, A., Wonderling, D. & Mooney, C., 1997. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Economics*, 6(4), pp.327 - 340.

Brockway, J. & Wurdock, C., 1981. *Kentucky Demographics: Preliminary Subcounty Population Estimation and Projection for Kentucky 1980 – 1985*. Louisville: Urban Studies Center, Univerity Louisville.

Brunn, A. & Weidner, U., 1997. *Discriminating buildings and vegetation areas with digital surface models*. Technical Report. Bonn: Institute of Photogrammetry, University Bonn.

Bryan, T., 2004. Population Eestimates. In J. Siegel & D. Swanson, eds. *Methods and Material of Demography*. Elsevier Academic Press. pp.523 - 560.

Canny, J., 1986. Computational approach to edge detection. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 8(6), pp.679 - 698.

CE Info Systems (P) Ltd, 2008. *Map my india*. [Online] Available at: http://www.mapmyindia.com/index.html [Accessed February 2009].

Census of India, 2007. *Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, India*. [Online] Available at: http://www.censusindia.gov.in/ [Accessed February 2008].

Centeno, J. & Miqueles, M., 2004. Extraction of buildings in Brasilian urban environment using high resolution remote sensing imagery and laser scanner data. In *Proceedings of the 2004 ISPRS Congress*. Istanbul, 2004. International Society for Photogrammetry and Remote Sensing.

CEPT / GIDB, 2005. *Ahmedabad Bus Rapid Transportation System*. Report no. 1. Ahmedabad: Gujarat Infrastructure Development Board; Center for Environmental Planning and Technology.

Chandhoke, N., Priyadarshi, P., Tyagi, S. & Khanna, N., 2007. The displaced of Ahmedabad. *Economic and Political Weekly*, 42, pp.10 - 13.

Chang, T., 2009. Improving Slum Conditions with Public- Private Partnerships. In *Proceedings of the 14th International Conference on Urban Planning and Regional Development in the Information Society*. Catalonia, 2009.

Chauhan, U. & Lal, N., 1999. Public-Private Partnerships for Urban Poor in Ahmedabad - A Slum Project. *Economic and Political Weekly*, 34(10 / 11), pp.636 - 642.

Chawdhury, S., 1996. Industrial restructuring, unions, and the state; textile mill workers in Ahmedabad. *Economic and Political Weekly* , 31(8).

Chung, C. et al., 2004. Remote sensing for building inventory update and improved loss estimation in HAZUS 99. In *Proceedings of the 2nd International Workshop on Remote Sensing for Postdisaster Response*. Newport Beach, California, 2004.

CIESIN; SEDAC; Columbia University, 2010. *Gridded Population of the World*. [Online] Columbia University. Available at: http://sedac.ciesin.columbia.edu/gpw/global.jsp [Accessed April 2009 2009].

Coburn, A. & Spence, R., 2002. *Earthquake Protection*. 2nd ed. Wiley.

Collins, W. & El-Beik, A., 1971. The acquisiton of urban land use information from aerial photographs of the city of Leeds (Great Britain). *Photogrammetria*, 27(2), pp.71 - 92.

Congalton, R. & Green, K., 1998. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 1st ed. CRC Press.

Core Consultants, 1983. *Survey of Slums in Seven Cities of Gujarat*. Ahmedabad.

Daniell, J., 2009. *Open source Procedure for Assessment of Loss using Global Earthquake Modelling (OPAL - GEM Project)*. CEDIM Earthquake Loss Estimation Series Research Report No. 09/01. Karlsruhe.

Danko, D., 1992. The Digital Chart of the World Project. *Photogrammetric Engineering and Remote Sensing*, 58, pp.1125 - 1128.

Deichmann, U., 1996. *A review of spatial population database design and modeling*. Technical Report TR-96-3. Santa Barbara: National Center for Geographic Information and Analysis.

Deichmann, U., Balk, D. & Yetman, G., 2001. *Transforming Population Data for Interdisciplinary Usages: From Census to Grid*. [Online] Socioeconomic data and application center. Available at: http://sedac.ciesin.columbia.edu/gpw/analapps.jsp [Accessed November 2009].

Desai, A., 1985. *Environmental perception - the human factor in urban planning*. New Delhi: S.B. Nangia, Ashih Publishing House Department of Geography, Gujarat University.

Digital Globe, 2007. *Product Guide*. [Online] Hatfield Group Available at: http://www.hatfieldgroup.com/ [Accessed March 2009].

Dobson, J. et al., 2000. LandScan: a global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, 66, pp.849 - 857.

Doniger, W., 1994. *The rig veda*. Penguin Classics.

Dutta, 2000. Partnerships in urban development: a review of Ahmedabad's experience. *Environment and Urbanization*, 12.

Dutta, D. & Serkerr, N., 2004. Urban building inventory for Bangkok City with very-high resolution remote sensing data. *Monthly Journal of Institute of Industrial Science (University of Tokyo)*, pp.203 - 206.

Eidinger, J., 2001. *Gujarat (Kutch) India earthquake of January 26, 2001 - Lifeline performance*. Monograph No. 19. Technical Council of Lifeline Engineering.

Erdik, M. & Aydinolgu, N., 2002. *Earthquake Performance and Vulnerability of Buildings in Turkey*. Report prepared for the World Bank Disaster Management Facility. Washington DC.

Erdik, M. et al., 2003. Istanbul Earthquake Rapid Response and Early Warning System. *Bulletin of Earthquake Engineering*, pp.157 - 164.

Erdik, M. & Fahjan, Y., 2006. Damage Scenarios and Damage Evaluation. In *Assessing and Managing Earthquake Risk - Geo-scientifc and engineering knowledge for Earthquake Risk Mitigation: development, tools, techniques*. Oliveira, C.; Roca, A.; Goula, X. ed. 2006: Springer.

European Centers of Excellence in Earthquake and Geotechnical Engineering , 2007. *Less Loss*. [Online] Available at: www.lessloss.org [Accessed March 2009].

European Space Agency, 2006. *Kathrina (Florida) Hurricane, August 2005*. [Online] Available at: http://earth.esa.int/ew/cyclones/Kathrina_Hurricane-aug05/ [Accessed May 2009].

FEMA, 2008. *HAZUS - MH Estimated Annualized Earthquaje Loss for the United States*. Washington DC: Federal Emergency Management Agency.

Fitzpatrick-Lins, K., 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map. *Photogrammetric Engineering and Remote Sensing*, 47(3), pp.343 - 351.

Foody, G., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, pp.185 - 201.

Frankena, M., 1978. A bias in estimating urban population density functions. *Journal of Urban Economics*, 5, pp.35 - 35.

Fraser, C., Hanley, H. & Yamakawa, T., 2001. Sub-metre Geopositiong with Ikonis Geo Imagery. In *Proceedings of the ISPRS Workshop on High Resolution Mapping from Space*. Hannover, Germany, 2001. International Society of Photogrammetry and Remote Sensing.

GAF AG, 2008. *GAF AG*. [Online] Available at: http://www.gaf.de/content/quickbird [Accessed February 2009].

Gamba, P., Dell'Acqua, F. & Houshmand, B., 2002. SRTM data characterization of urban areas. In *Proceedings of ISPRS Symposium Commission III Photogrammetric Computer Vision*. Graz, Austria, 2002. International Society of Photogrammetry and Remote Sensing.

Geotechnical Earthquake Engineering Server, 2001. [Online] Available at: http://gees.usc.edu/GEES/ [Accessed April 2010].

Gillion, K., 1968. *Ahmedaba: A study in Indian urban history*. Berkeley: University of California Press.

Global Land Cover Facility , 2009. [Online] Available at: www.landcover.org [Accessed February 2009].

Goodchild, M., Anselin, L. & Deichmann, U., 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning*, 25, pp.383 - 397.

Green, N., 1956. Aerial photographs analysis for residential neighborhoods: An evaluation of data accuracy. *Social Forces*, 35, pp.142 - 147.

Green, N. & Monier, R., 1957. Aerial Photographic Interpretation and the Human Geography of the City. *The Professional Geographer*, 9, pp.2 - 5.

GSHAP, 1999. [Online] Global Seismic Hazard Assessment Program Available at: http://www.seismo.ethz.ch/gshap/ [Accessed August 2009].

Gugler, J., 1996. *The Urban Transformation of the Developing World*. Oxford University Press.

Guo, T. & Yasuoka, Y., 2002. Snake-based approach for building extraction from high-resolution satellite images and height data in urban areas. In *Proceedings of the 23rd Asian Conference on Remote Sensing*. Kathmandu, 2002.

Haala, N., 1994. Detection of buildings by fusion of range and image data. In *Proceedings of the ISPRS Congress Commission III*. Munich, 1994. International Society of Remote Sensing and Photogrammetry.

Haala, N. & Brenner, V., 1999. Laser data for virtual landscape generation. In *Proceedings of Workshop for Mapping Surface Structure and Topography by Airborne and Spaceborne Lasers*. La Jolla ed. 1999. International Society of Photogrammetry and Remote Sensing.

Halla, N. & Brenner, C., 1999. Extraction of buildings and trees in urban environments. *International Archives of Photogrammetry and Remote Sensing*, 43, pp.139 - 137.

Hadfield, S., 1963. *Evaluation of Land Use and Dwelling Unit Data derived from Aerial Photography*. Chicago Areas Transportation Study. Chicago: Urban Research Section.

Hamesse, J., 1983. *Sectoral and spatial interrelations in urban development. A case study of Ahmedabad - India*. Herodot ed.

Hardin, P., Jackson, M. & Shumway, J., 2007. Intraurban Population Estimation Using Remotely Sensed Imagery. In R. Jensen, J. Gatrell & D. McLean, eds. *Geo-spatial Technologies in Urban Environments - Policy, Practice and Pixels*. Springer.

Harvey, J., 2002. Estimating census district populations from satellite imagery: some approaches and limitations. *International Journal of Remote Sensing*, 23(10), pp.2071 - 2095.

Harvey, J., 2002a. Population estimation models based on individual TM pixels. *Photogrammetric Engineering & Remote Sensing*, 72(2), pp.187 - 196.

Haverkamp, D., 2004. Automatic building extraction from IKONOS imagery. In *Proceedings of ASPRS 2004 Conference*. Denver, 2004. American Society of Photogrammetry and Remote Sensing.

Henderson, F. & Xia, Z., 1997. SAR applications in human settlement detection, population. *IEEE Transactions On Geoscience and Remote Sensing*, 35(1), pp.79 - 85.

Henricsson, O. & Grün, A., 1996. Overview of research activities at ETH-Zürich in automated 3-D reconstruction of buildings from aerial images. In *Tagesband zur Jahrestagung des Deutschen Gesellschaft für Photogrammetrie and Fernerkundung*. Oldenburg, 1996.

Herold, M., Gardner, M. & Roberts, D., 2003. Spectral resolution requirement for mapping urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9), pp.1907 - 1919.

Hough, S. & Bilham, R., 2006. *After the Earth Quakes*. New York: Oxford University Press, Inc.

Hyman, G. et al., 2000. *Latin America and Caribbean Population Database Documentation*. [Online] Available at: http://www.na.unep.net/globalpop/lac/intro.html [Accessed March 2009].

Ilhamdaniah, S., Munshi, T. & Amer, S., 2005. Evaluating the planning of social infrastructure in Ahmedabad, India. In *Proceeding of the CUPUM 05, Computer in Urban Planning and Urban Management*. London, 2005.

Imhoff, M., Lawrence, W., Stutzer, D. & Elvidge, C., 1997. A techniques for using composite DMSP / OLS "city lights" satellites data to map urban areas. *Remote Sensing of the Environment*, 61(3), pp.361 - 370.

Jain, S., 2002. *Indian Seismic Code IS: 1893 (Part 1)*. Kanpur: IIT Kanpur; Gujarat State Disaster Management Authority.

Jaiswal, K. & Wald, D., 2008. *Creating a global building inventory for earthquake loss assessment and risk management*. USGS Open File Report, OF 2008 - 1160. US Geological Survey.

Jat, M., Garg, P. & Khare, D., 2008. Modelling of urban growth using spatial analysis techniques: a case study of Ajmer city (India). *International Journal of Remote Sensing*, 29(2), pp.543 - 567.

Jensen, J., 1996. *Introductory digital image processing: A remote sensing perspective*. 2nd ed. Englewood, NJ: Prentice Hall.

Jensen, J. & Cowen, D., 1999. Remote sensing of urban / suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, 65, pp.611 - 622.

Johnson, N. & Kotz, S., 1976. On a multivariate generalized occupancy model. *Journal of Applied Probability*, 13(2), pp.392 - 399.

Joshi, R., 2002. Integreated Slum Development: Case of Pravinnagar - Gutanagar. In K. Amitabh & D. Mahadevia, eds. *Poverty and Vulnerability in a Globalising Metropolis Ahmedabad*. New Delhi: Manak Publications Pvt Ltd.

Katuri, A., Sharifi, M. & van Westen, C., 2007. *Urban mapping and land use characterisation for risk assessment*. [Online] Faculty for Geo-Information Science and Earth Observation, University of Twente Available at: http://www.itc.nl/personal/ajaykaturi/Research.asp [Accessed November 2009].

Kraus, S., Senger, L. & Ryersom, J., 1974. Estimation Population from Photographically Determined Residential Land Use Types. *Remote Sensing of the Environment*, 3, pp.35 - 42.

Kundu, A. & Mahadevia, D., 2002. *Poverty and vulnerability in a globalising metropolis, Ahmedabad*. Kundu, A.; Mahadevia, D. ed. New Delhi: Manak Publications.

Lakha, S., 1988. *Capitalism and Class in Colonial India - The Case of Ahmedabad*. New Delhi: Sterling Publishers Asian Studies Association of Australia.

Langford, M., Maguire, D. & Unwin, D., 1991. The areal interpolation problem: Estimating population using Remote Sensing within a GIS framework. In I. Masser & M. Blakemore, eds. *Handeling Geographical Information: Methodologies and Potential Applications*. London: Longman. pp.55 - 77.

Langford, M., 2006. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems*, 30, pp.161 - 180.

Langford, M., Higgs, G., Radcliffe, J. & White, S., 2008. Urban population distribution models and service accessibility estimation. *Computers, Environment and Urban Systems*, 32, pp.66 - 80.

Lee, D., Shan, J. & Bethel, J., 2003. Class-guided building extraction from IKONOS. *Photogrammetric Engineering and Remote Sensing*, 69(2), pp.143 - 150.

Lee, T. et al., 2004. Day / Night Visible Sensor aboard NPOESS VIIRS. In *Proceedings of the 13th Conference on Satellite Meterology and Oceanography*. Norfolk, 2004. American Meterological Society.

Lillesand, T., Kiefer, R. & Chipman, J., 2008. *Remote Sensing and Image Interpretation*. 6th ed. Wiley&Sons.

Liu, X., Clarke, K. & Herold, M., 2006. Population Density and Image Texture: A comparison study. *Photogrammetric Engineering and Remote Sensing*, 72(2), pp.187 - 196.

Liu, W. & Prinet, V., 2005. Building detection from high-resolution satellite images using probability model. In *Proceedings 2005 IEEE International Geoscience and Remote Sensing Symposium*. Seoul, 2005.

Lo, C., 1986. Accuracy of population estimation from medium-scale aerial photography. *Photogrammetric Engineering and Remote Sensing*, 52(12), pp.1895 - 1869.

Lo, C., 1995. Automated Population and Dwelling Unit Estimation for High-Resolution Satellite Images: a GIS approach. *International Journal of Remote Sensing*, 16, pp.17 - 34.

Lo, C. & Welch, R., 1997. Chinese Urban Population Estimates. *Annals of the Association of American Geographers*, 67(2), pp.246 - 253.

Lo, C. & Mesev, V., 2003. *Zone-based estimation of population and housing units from satellite generated land use / landcover maps*. Lo, C.; Mesev, V. ed. CRC Press.

Lu, D., Weng, Q. & Li, G., 2006. Residential population estimation using a remote sensing impervious surface approach. *Int. J. Remote Sensing*, 27(16), pp.3553 - 3570.

Maas, H. & Vosselman, G., 1999. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sening*, 5, pp.153 - 163.

Mahadevia, D., 2002. Communal Space of Life Space: Saga of Increasing Vulnerability in Ahmedabad. *Economic and Political Weekly*, 37(48), pp.4850 - 4858.

Marangoz, A., Oruc, M., Karakis, S. & Sahin, H., 2006. Comparison of pixel-based and object-oriented classification using Ikonos imagery for automatic building extraction - Safranbolu Testfield. In *Proceedings of 5th International Symposium "Turkish-German Joint Geodetic Days"*. Berlin, 2006. Technical University Berlin.

Mather, P., 2004. *Computer Processing of Remotely - Sensed Images*. 3rd ed. Chicester, West Sussex: Wiley.

Mathur, O., 1999. Fiscal innovation and urban governance. In O. Mathur, ed. *India: the challenge of urban governance*. New Delhi : National Institute of Public Finance and Policy. pp.223 - 263.

McDonald, J., 1989. Econometric studies of urban population density: a survey. *Journal of Urban Economics*, 26, pp.361 - 385.

Metha, D. & Metha, M., 1992. *Metropolitan Housing Market: A Study of Ahmedabad*. Sage Publications.

Morrison, P., 1971. *Demographic information for cities: A manual for estimating and projecting local population characteristics*. Report No. R-618-HUD. Santa Monica: Rand Cooperation.

National Institute of Urban Affairs , 2001. *Structure and dynamics of urban economy: Study of linkages between formal and informal sectors in Ahmedabad and Visakhapatnam*. Research Study Series No. 80. New Delhi: National Institute of Urban Affairs (NIUA).

Nelson, 2004. *African population database documentation*. [Online] Available at: http://na.unep.net/globalpop/africa/ [Accessed March 2009].

Newell, C., 1988. *Methods and Models in Demography*. New York: The Guilford Press.

NIBS / FEMA, 2002. *HAZUS 99 Earthquake Loss Estimation Methodology*. Service Release 2 (SR2) Technical Manual. Washington, DC: Federal Emergency Management Agency, National Institute of Building Science.

NOAA, 1972. *A study of earthquake loss in the San Francisco Bay Area; Data and Analysis prepared for the Office of Emergency Preparedness*. Washington DC: National Oceanic and Atmospheric Administration.

NOAA, 1973. *A study of earthquake loss in the Los Angeles, California Area; Data and Analysis, prepared for the Federal Disaster Assistance Administration*. Washington, DC: Department of Housing and Urban Development by the National Oceanic and Atmospheric Administration.

Nordbeck, S., 1956. *The law of allometric growth*. Michigan Inter - University, Community of Mathematical Geographers.

Operations Research Group , 1973. *Slums in Gujarat (A Study of Seven Urban Centres)*. Ahmedabad: ORG.

Pandya, Y., 2001. *The Ahmedabad Chronicle - Imprints of a millennium*. Ahmedabad: Vastu-Shilpa Foundation for Studies and Research in Environmental Design.

Patel, S., 1987. *The Making of Industrial Relations*. Oxford University Press.

Pathak, P., 2001. *Structure and dynamics of urban economy: Study of linkages between formal and informal sectors in Ahmedabad and Visakhapatnam*. Research Study Series 80. New Delhi, India: National Institute of Urban Affairs.

Pathan, S. et al., 1993. Urban growth trend analysis using GIS-techniques - a case study of the Bombay metropolitan region. *Int.J. Remote Sensing*, 14(17), pp.3169 - 3179.

Peek-Asa, C. et al., 2000. Fatal and Hospitalized Injuries Resulting from the 1999 Northridge Earthquake. *International Journal of Epidemiology*, 27(3), pp.459 - 465.

Pillai, V., 1987. Correlates of "Decision to Move or Stay" in Ahmedabad, India. *Population and Environment*.

Plane, D. & Rogerson, P., 1994. *The Geographical Analysis of Population with Application to Business and Planning*. New York: Wiley.

Pozzi, F. & Small, C., 2002. Vegetation and population density in urban and suburban areas in the USA. In *Proceedings of th 3rd International Symposium of Remote Sensing of Urban Areas*. Istanbul, 2002.

Pradhan, A., Pandya, J. & Gosai, S., 2009. *Employment and Human Resources Development*. [Online] Ahmedabad: Sardar Patel Institute of Economic and Social Research Available at: http://www.jgpandya.com/articles.htm.

Prasad, J., Singh, Y., Kaynia, A. & Lindholm, C., 2009. Socioeconomic clustering in seismic risk assessment of urban housing stock. *Earthquake Spectra*, 25(3), pp.619 - 641.

Preston, S., Heuveline, P. & Guillot, M., 2008. *Demography - Measuring and Modeling Population Processes*. EPZ ed. Wiley - Blackwell.

Pursell, D., 1970. Improving population estimates with the use of dummy variables. *Demography*, 7, pp.87 - 91.

Raman, S., 2003. Communities and spatial culture in a communally diverse city: Ahmedabad, India. In *Proceedings of the 3rd International Space Syntax Symposium*. Atlanta, 2003.

Ramani, K. et al., 2005. *Urban health status in Ahmedabad city: GIS based sutdy of Baherampura, Kubernagar, and Vasna wards*. IIMA Working Papers. Ahmedabad: Indian Institute of Management Ahmedabad, Research and Publication Department.

Rao, R., 1990. *Social organisation in an Indian slum: study of a caste slum*. New Delhi: K.M. Mittal for Mittal Publications.

Rashed, T., Weeks, J., Gadalla, M. & Hill, A., 2001. Revealing the Anatomy of Cities through Spectral Mixture Analysis of Multispectral Satellite Imagery: A Case Study of the Greater Cairo Region Egypt. *Geocarto International*, 16(4), pp.5 - 16.

Raymondo, J., 1992. *Population estimation & projection - methods for marketing, demographic, and planning personnel*. New York: Quorum Books.

Regional Employment Exchange Office, 2007. *Nummber of employment seekers by level of education & sex in Ahmedabad city (except physically handicapped persons)*. Statistics reports. Ahmedabad: Regional Employment Exchange Office.

Revi, A., 2008. Climate change risk: an adaptation and mitigation agenda for Indian cities. *Environment and Urbanization*, 20(1), pp.207 - 229.

Rice, A., 1958. *Productivity and social organization: The Ahmedabad experiment: technical innovation, work organization and management*. London: Tavistock.

Rindfuss, R., Walsh, S., Mishra, V. & Dolcemascolo, G., 2003. Linking Household and Remotely Sensed Data: Methodologies and Practical Problems. In J. Fox, R. Rindfuss, S. Walsh & V. Mishra, eds. *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*. Dordrecht: Kluwer Academic Publisher. pp.1 - 30.

Risk Management Solutions, 2009. [Online] Available at: http://www.rms.com [Accessed February].

Rives, N. & Serow, W., 1984. *Introduction to Applied Demography: Data Sources and Estimation Techniques (Quantitative Applications in the Social Sciences)*. Newbury Park, California: Sage Publications.

Robinson, J., Ahmed, B., Gupta, P. & Woodrow, K., 1993. Estimation of Population Coverage in the 1990 US Census based on Demographic Analysis. *Journal of the American Statistical Association*, 88(423).

Rosenberg, H., 1968. Improving current population estimates through stratification. *Land Economics*, (44), pp.331 - 338.

Saito, K. & Spence, R., 2004. Image classification methods and post-earthquake damage assessment. In *2nd Workshop on remote sensing for disaster response*. Newport Beach, 2004.

Saito, K., 2008. *High-resolution optical satellite images for post-earthquake damage assessment*. Dissertation thesis. Cambridge: Department of Architecture, University Cambridge.

Sapovadia, V., 2007. A critical study of urban landowership by an individual vis-à-vis institutional (or community) based ownership - The impace of type ownership on spatial growth, efficiency and equity: A case study of Ahmedabad, India. In *Proceedings of the 4th Urban Research Symposium*. Washington DC, 2007. Urban Research Network.

Scharfe, H., 1993. *Investigation on Kautalya's manual of political science*. Harrassowitz Verlag.

Seligson, H. & Shoaf, K., 2003. Human impacts of earthquakes. In W. Chen & C. Scawthorn, eds. *Earthquake Engineering Handbook*. Boca Raton: CRC Press.

Shahidullah, M. & Flotow, M., 2005. Criteria for selecting a suitable method for producing post-2000 county population estimates: A case study of population estimates in Illinois. *Population Research and Policy Review*, 24, pp.215 - 229.

Shakhramanjyan, M. et al., 2000. Seismic and Complex Risk Assessment and Management for the Kamchatka Region. In *Proceedings of the XII World Conference on Earthquake Engineering*. Auckland, New Zealand, 2000.

Shan, J. & Lee, D., 2002. Generalization of building polygons extracted from Ikonos imagery. In *Proceedings of the ISPRS Commission IV Symposium Joint Symposium of Geospatial Theory, Processing and Applications*. Ottawa, 2002. International Society of Photogrammetry and Remote Sensing.

Shoaf, K., Sareen, H., Nguyen, L. & Bourque, L., 1998a. Injuries as a result of California earthquake in the past decades. *Disasters*, 22(3), pp.218 - 235.

Shops & Establishment Department , 2007. *Employment in Shops & Establishments in Ahmedabad City as on 31 st March 2007*. Statistics Report. Ahmedabad: Shops & Establishment Department.

Shukla, P., 2009. *Low Carbon Society Vision 2035 - Ahmedabad*. Ahmedabad: Indian Institute of Management Ahmedabad, Kyoto University, Mizuho Information & Research Institute, National Institute for Environmental Studies.

Smith, S. & Mandell, M., 1984. A Comparison of Population Estimates Methods: Housing Unit versus Component II, Ratio Correlation and Administrative Records. *Journal of the American Statistical Association*, 79(386), pp.282 - 289.

Smith, S. & Cody, S., 1999. Evaluating the housing unit method: a case study of 1990 population estimates in Florida. In *Proceeding of the Population Estimates Methods Conference*. Washington, 1999. US Census Bureau.

Smith, K., Nogle, J. & Cody, S., 2000. A Regression Approach to Estimating the Average Number of People per Household. *Demography*, 39(4), pp.697 - 712.

Sohn, G. & Downman, I., 2001. Extraction of buildings from high resolution satellite data. In *Proceedings of the 3rd International Workshop on automatic extraction of man-made objects from aerial and space images*. Ascone, Switzerland, 2001.

Souza, I., Pereira, M. & Kurkdjian, M., 2003. *Evaluation of High Resolution Satellite Images for Urban Population Estimation*. Instituto Nacional de Pesquisas Espaciais, Brazil.

Spence, R., Baxter, P. & Zuccaro, G., 2004. Building vulnerability and human casualty estimation for a pyroclastic flow: a model and its application to Vesuvius. *Journal of volcanology and geothermal research*, 133, pp.321 - 343.

Spence, R., 2007. Saving lives in earthquakes: successes and failures in seismic protection since 1960. *Bulletin of Earthquake Engineering*, 5(2), pp.39-251.

Srivastava, R., 2005. *The informal sector and urban poverty*. [Online] infochange urban India Available at: http://infochangeindia.org/200502056108/Urban-India/Backgrounder/The-informal-sector-and-urban-poverty.html [Accessed March 2009].

Statistics Canada, 2009. *History of the Census*. [Online] Available at: http://www.statcan.gc.ca/edu/power-pouvoir/ch2/history-histoire/5214912-eng.htm [Accessed March 2009].

Sutton, P., Roberts, C., Elvidge, C. & Meij, H., 1997. A comparison of nighttime satellite imagery and population density for the continental united states. *Photogrammetric Engineering and Remote Sensing*, 63(11), pp.1303-13.

Sweeney, L., 2000. *All the data on all the people*. Data Privacy Lab White Paper Series, LIDAP-WP3. Pittsburgh: Carnegie Mellon University, School of Computer Science.

Sweeney, L., 2002. K-anonymity: a model for protecting privacy. *Internationl Journal on Uncertainty, Fuzziness and Knowlegde-based Systems*, 10(5), pp.557 - 570.

Taragi, R., Bisht, K. & Sokhi, B., 1994. Generation of population census data trough aerial remote sensing. *Journal of the Indian Society of Remote Sensing*, 22, pp.131 - 138.

Taubenböck, H. & Roth, A., 2006. Assessment of urban location factors from remote sensing. In *Proceedings of the 25th Urban Data Management Symposium*. Aalborg, Denmark, 2006.

Taubenböck, H. et al., 2009. Urbanization in India - Spatiotemporal analysis using remote sensing data. *Computers, Environment and Urban Systems*, (33), pp.179 - 188.

Taylor, P., 1977. *Quantitative Methods in Geography - An Introduction to Spatial Analysis*. Boston: Houghten Mifflin Company.

Taymann, J., Smith, S. & Lin, J., 2007. Precision, bias, and uncertainty for state population forecasts: an exploratory analysis of time series models. *Population Research Policy Review*, 26, pp.347 - 369.

Tian, J., Wang, J. & Shi, P., 2003. Urban building boundary extraction from Ikonos imagery. In *Proceedings of the 25th Symposium canadien sur la teledetection*. Montreal, 2003.

Tian, J., Wang, J. & Shi, P., 2007. Urban building boundary extraction from Ikonos imagery., 2007.

Tobler, W., Deichmann, U., Gottsegen, J. & Maloy, K., 1995. *The Global Demography Project*. Technical Report TR-95-6. Santa Barbara: National Center for Geographic Information and Analysis.

Tripathi, D., 1999. *Slum Networking in Ahmedabad: The Sanjay Nagar Pilot Project*. Working Paper No. 101. Department of International Development, University College London.

Tveite, H. & Langaas, S., 1995. Accuracy Assessments of Geopgraphical Line Datasets: The Case of the Digital Chart of the World. In Bjorke, T., ed. *Proceedings of the 5th Scandinavian Research Conference of Geographical Information Systems*. Trondheim, Norway, 1995.

UNCHS, 2008. *The State of the World's Cities*. Narobi, Kenya: United Nations Centre for Human Settlements.

UN, 2009. *World Population Prospects: The 2008 Revision*. Working Paper No. ESA/P/WP 210. United Nations, Department of Economic and Social Affairs, Population Division.

UNSD, 2010. *United Nations Statistics Divison*. [Online] Available at: http://unstats.un.org/unsd/syb/ [Accessed March 2009 2009].

US Geological Survey, 2009. *USGS Global Visualization Viewer*. [Online] Available at: http://glovis.usgs.gov/ [Accessed February 2009].

Van Genderen, J., Lock, B. & Vass, P., 1978. Remote Sensing: statistical testing of thematic map accuracy. *Remote Sensing of Environment*, 7, pp.3 -14.

Venkatachalam, P., 2007. *Municipal Finance System in Conflict Cities: Case Studies on Ahmedabad and Srinagar, India*. Working Paper No. 15. London: Crisis States Research Centre.

Vogelmann, J. et al., 2001. Completion of the 1990's National Land Cover Dataset for the conterminous United States.. *Photogrammetric Engineering and Remote Sensing*, 67, pp.650 - 662.

Wald, D. et al., 2006. Challenges in rapid ground motion estimation for the prompt assessment of global urban earthquakes. *Bulletin of the Earthquake Research Institute*, 81, pp.272 - 281.

Watkins, J., 1984. The Effect of Residential Structure Variation on Dwellung Unit Enumeration from Aerial Photographs. *Photogrammetric Engineering and Remote Sensing*, 50, pp.1599 - 1607.

Weeks, J. et al., 2004. The fertility transition in Egypt: Intra-urban patterns in Cairo. *Annals of the Association of American Geographers*, 94, pp.74 - 93.

Weeks, J., 2005. *Population: Introduction to Concepts and Issues*. 9th ed. Belmont, CA: Wadsworth Thomson Learning.

Wenzel, F., Bendimerad, F. & Sinha, R., 2007. Megacities - megarisk. *Nat. Hazards*, 42, pp.481 - 491.

Whiteman, R., J., R. & Hong, T., 1974. Earthquake damage probability matrices. In *Proceedings of the 5th World Conference on Earthquake Engineering*. Rome, 1974.

Wikantika, K., Darmawan, S. & Hadi, F., 2005. Application of Remote Sensing in Demography, Land Use and Land Cover and Disaster: An Indonesian Experience. In *Proceedings of the ICALRD-JIRCAS Workshop on Enhancement of Remote Sensing and GIS Technologies for Sustainable Utilization of Agricultural Resources in Indonesia*., 2005. Japan Internatioanl Research Center for Agricultural Sciences; Indonesian Center for Agricultural Land Resources Research and Development.

Williamson, J. & Zagha, R., 2002. *From the Hindu Rate of Growth to the Hindu Rate of Reform*. SCID Working Paper 144. Stanford Institut for Economic Policy Research.

Woiwoode, C., 2009. Risk reduction and urban development: the case of Ahmedabad, India. *Geographische Rundschau International*, 5(3), pp.20 - 25.

Wu, S., Qui, X. & Wang, L., 2006. Using semi-variance image texture statistics to model population densities. *Cartography and Geographic Information Science*, 33(2), pp.127 - 140.

Wu, C. & Murray, A., 2007. Population Estimation using Landsat Enhanced Thematic Mapper Imagery. *Geographical Analysis*, 39, pp.26 - 43.

Wyss, M., 2006. The Kashmir M7.6 Shock of 8 October 2006 calibrates estimates of losses in future Himalayan earthquakes. In Van de Walle, B. & Turoff, M., eds. *Proceedings of the third International Conference on Information Systems for Crisis Response and Management*. Newark, 2006.

Yaukey, D., Anderton, D. & Lundquist, J., 2007. *Demography - The study of Human Population*. Waveland Press.

Zha, Y., Gao, J. & Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), pp.583 - 594.

Appendix I

In section 2.3.1, the technological characteristics of the optical satellite images employed in this study were introduced. Two types of image were tested: (1) Moderate resolution Landsat 5 TM and (2) very high resolution Quickbird. Table 35 lists the existing and future satellites including their technical characteristics.

Table 35: List of existing and future optical satellites in chronological order.

Launch	Sensor	Country	Operator	Wavelength	Spatial Resolution
1972	Landsat 1 MSS	USA	USGS	R,G, 2 × IR	80m
1975	Landsat 2 MSS	USA	USGS	R,G, 2 × IR	80m
1978	Landsat 3 MSS	USA	USGS	5 bands 0,50 – 0,75µm thermal	30m
1982	Landsat 4 TM	USA	USGS	5 bands 0,50 – 0,75µm thermal	30m
1984	Landsat 5 TM	USA	USGS	RBG, NIR, 2 × MIR	30m
1986	SPOT 1	France	Spot Image	Pan,G,R,NIR	10m, 20m
1990	SPOT 2	France	Spot Image	Pan,G,R,NIR	10m, 20m
1993	SPOT 3	France	Spot Image	Pan,G,R,NIR	10m, 20m
1993	Landsat 6 TM	USA	USGS	RBG, NIR, 2 × MIR	30m 15m (pan)
1995	Orbview 1	USA	Orbital Image Cooperation	1 band	10km
1995	IRS C 1	India	ISRO	G,B, NIR, MIR Pan	5m (pan) 23m / 70m (MIR)
1997	IRS 1 D	India	ISRO	G,B, NIR, MIR Pan	5m (pan) 23m / 70m (MIR)
1997	Orbview 2	USA	Orbital Image Coop.	8 bands	1,1km
1998	SPOT 4	France	Spot Image	Pan, G,R,NIR, MIR	10m, 20m (MIR)
1999	ASTER	USA Japan	JPL ERSDAC	15 bands RGB, NIR, SWIR TIR	15m 30m 90m
1999	Landsat 7 ETM +	USA	USGS	RBG, NIR, 2 × MIR	30m 15m (pan) 60m (IR)
1999	MODIS	USA	NASA USGS	36 bands RGB, NIR, SWIR, MIR, LWIR	250m 500m 100m

Launch	Sensor	Country	Operator	Wavelength	Spatial Resolution
1999	IKONOS	USA	GeoEye	RBG, NIR Pan	4 m 0,80m (pan)
2000	ALI / EO 1	USA	NASA USGS	6 bands (µm) 0,433 – 0,453 0,845 – 0,890 1,20 – 1,30	30m
2000	EROS A	Israel	ImageSat	Pan	1,8 m
2000	Orbview 3	USA	Orbital Image Coopera-tion	RGB, NIR	4m 1m (pan)
2000	Quickbird 2	USA	Digital Globe	RBG, NIR Pan	2,4m 0,60m (pan)
2001	Orbview 4	USA	Orbital Image Coop.	RGB, NIR	4m 1m (pan)
2001	Proba 1	EU	ESA	CHRIS BW - HRC	18m (CHRIS) 5m (HRC)
2002	AISAT 1	Algeria	CNTS	RBG, NIR, 2 × MIR	30m
2002	ENVISAT – MERIS	EU	ESA	15 bands RGB, IR	300 m
2002	SPOT 5	France	Spot Image	Pan, G,R,NIR, MIR	2,5 m, 10m, 20m (MIR)
2004	Formosat 2	Taiwan	NSPO	RGB, NIR Pan	8 m 2m (pan)
2005	Beijing 1	China	Beijing Landview Mapping Inform.	RBG, NIR Pan	32m 4m (pan)
2005	Cartosat 1	India	ISRP	Pan	2,5m
2006	ALOS	Japan	JAXA	RGB, NIR Pan	10m 2,5m (pan)
2006	EROS B	Israel	ImageSat	Pan	1,8 m
2007	Cartosat 2	India	ISRP	Pan	< 1m
2007	World-view 1	USA	Digital Globe	RGB, NIR Pan	2,6m 0,5m (pan)
2008	Orbview 5 Geoeye 1	USA	GeoEye	RGB, NIR Pan	1,65m 0,41m (pan)
2009	Cartosat 3	India	ISRP	Pan	25cm
2009	World-view 2	USA	Digital Globle	8 bands RGB, NIR Pan	1,84m 46cm (pan)
2013	EnMap	Germany	GFZ / DLR	218 bands 420 – 2450nm	30m

Appendix II

Previous studies on the city of Ahmedabad constituted an important source of information for the methodology development. The studies provide information on the city development and its structure. In section 2.3.2, selected studies of major importance are introduced. Table 36 lists all studies available on the city of Ahmedabad.

Table 36: List of previous studies on Ahmedabad 1950 – 2009.

Author / Source	Title
(AMC, AUDA, CEPT, 2002)	City Development Plan Ahmedabad 2006 - 2012
AMC (2008)	Pro Active Disclosure – Slum Networking Project
AMC (2008)	Ahmedabad Municipal Cooperation – Housing Project
Alessandrini (2008)	Ahmedabad – Ghandi Ashram Area
Bharti et al. (2008)	Willingness to pay: In urban slums of Ahmedabad
Bhatt (2003)	The case of Ahmedabad, India
Berry & Spodek (1971)	Comparative ecologies of large Indian Cities
Chandhoke et al. (2007)	The displaced of Ahmedabad
Chang (2009)	Improving slum conditions with public privat partnerships
Chauhan & Lal (1999)	Public-Private Partnership for Urban Poor in Ahmedabad: A Slum Project
Desai (1985)	Environmental Perception: The human factor in urban planning
Dutta (2000)	Partnerships in urban development: a review of Ahmedabad's experience
Gillion (1968)	Ahmedabad: A Study in Indian Urban History
Gugler (1996)	The urban transformation of the developing world
Hamesse (1983)	Sectoral and Spatial Interrelations in Urban Development, a case study of Ahmedabad India
Ilhamdaniah et al. (2005)	Evaluating the planning of social infrastructure in Ahmedabad, India
Kulkarni (1981)	Geography of Crowding and Human Response
Lakha (1988)	Capitalism and Class in Colonial India: The Case of Ahmedabad
Mahadevia (2002)	Communal Space of Life Space: Saga of Increasing Vulnerability in Ahmedabad
Metha & Metha (1989)	Metropolitan Housing Market: A Study of Ahmedabad
National Institute of Urban Affairs (2001)	Structure and Dynamics of Urban Economy: Study of Linkages between Formal and Informal Sectors in Ahmedabad and Visakhapatnam
Pandya (2001)	The Ahmedabad Chronicle: Imprints of a millennium
Patel (1987)	The making of industrial relations
Rice (1958)	Productivity and Social Organization: The Ahmedabad Experiment
Raman (2003)	Communities and spatial culture in a communally diverse city: Ahmedabad, India
Ramani et al. (2005)	Urban Health Status in Ahmedabad: GIS based Study of Baherampura, Kubernagar and Vasna districts
Sapovadia (2007)	A critical study of urban landowership by an individual vis-à-vis institutional (or community) based ownership - The impace of type ownership on spatial growth, efficiency and equity: A case study of Ahmedabad, India
Tripathi (1999)	Slum Networking in Ahmedabad: The Sanjay Hagar Pilot Project
Venkatachalam (2007)	Municipal Finance System in Conflict Cities: Case studies on Ahmedabad and Srinagar, India

Appendix III

In section 2.4.3, the population for the city of Ahmedabad is estimated for 2006. This population data set serves the purpose of validating the population projection for 2006 based on Census 2001 provided by the Ahmedabad Statistical Department. The following tables list the input data for the population estimation using vital rate procedure. In addition, the underlying equations of this procedure are exemplary presented using birth rate as the vital statistic.

Table 37: Vital statistics for the city of Ahmedabad and the state of Gujarat used for the population estimation for 2006. The rates are calculated per 1000 people of the total population.

Area	Birth rate 2001	Birth rate 2006	Death rate 2001	Death rate 2006	No. of birth	No. of death
Gujarat urban	21,5	21,1	5,6	5,9	n.n.	n.n.
Gujarat total	25,0	23,5	7,8	7,3	n.n.	n.n.
Ahmedabad	23,38	23,2	7,49	6,97	93.784	34.250

Table 38: Population of Ahmedabad and the state of Gujarat according to the Indian Census 2001.

Area	Population 2001
Gujarat total	50.700.00
Ahmedabad	3.520.085

Table 39: Population estimation of Ahmedabad 2006 using the urban area of Gujarat and the total area of Gujarat as the reference area. The rates are calculated per 1000 people of the total population.

Area	Ahmedabad Population 2006 (birth rates)	Ahmedabad Population 2006 (death rates)	Ahmedabad Population 2006 (average)
Gujarat urban	4.042.413	4.913.916	3.958.423
Gujarat total	4.131.453	3.874.434	4.522.684

In the following, the exemplary steps for this procedure using birth rates are presented:

1) Calculate birth rate (BR) of the state of Gujarat for 2001

$$BR_{Gujarat\ \_2001} = \frac{B_{Gujarat\ \_2001}}{P_{Gujarat\ \_2001}}$$

Equation 8

2) Calculate birth rate (BR) of Ahmedabad for 2001

Equation 9

$$BR_{Ahmedabad\ \_2001} = \frac{B_{Ahmedabad\ \_2001}}{P_{Ahmedabad\ \_2001}}$$

3) Calculate ratio of birth rate (BR) of Ahmedabad to Gujarat for 2001

Equation 10

$$R_{BR} = \frac{BR_{Ahmedabad\_2001}}{BR_{Gujarat\_2001}}$$

4) Calculate birth rate of the state of Gujarat for 2006

Equation 11

$$BR_{Gujarat\_2006} = \frac{B_{Gujarat\_2006}}{P_{Gujarat\_2006}}$$

5) Calculate birth rate of Ahmedabad for 2006

Equation 12

$$BR_{Ahmedabad\_2006} = R_{BR} * BR_{Gujarat\_2006}$$

6) Estimate the population of Ahmedabad for 2006

Equation 13

$$P_{Ahmedabad\_2006} = \frac{B_{Ahmedabad\_2006}}{BR_{Ahmedabad\_2006}}$$

Where

$B_{Ahmedabad}$ = Number of births in Ahmedabad

$P_{Ahmedabad}$ = Population of Ahmedabad

$BR_{Ahmedabad}$ = Birth rate of Ahmedabad

$B_{Gujarat}$ = Number of births in Gujarat

$P_{Gujarat}$ = Population of Gujarat

$BR_{Gujarat}$ = Birth rate of Gujarat

R_{BR} = Ratio of birth rate in Ahmedabad to Gujarat

Appendix IV

Image analysis for automated occupancy extraction

The objective of this chapter is to evaluate if the occupancy categories identified in the city of Ahmedabad can be extracted from satellite images on city level. The occupancy categories of interest were previously defined in section 4.3.1. The working hypothesis is that each occupancy category is characterized by a certain density of specific feature with explicit pixel values. For example, middle and high come areas are assumed to have a larger density of white roof than low income areas. In a first step a workflow is developed which is used for all tests (see Table 40). The image enhancement techniques tested include radiometric image enhancement, band algebra and geometric image enhancement.

Table 40: Workflow for feature density calculation to distinguish occupancy categories from panchromatic satellite image using radiometric image enhancement techniques with ENVI and ArcGIS.

No.	Processing steps
1	Open panchromatic image in ENVI software
2	Visually identify the image features that can be used to distinguish occupancy categories
3	Identify radiometric image enhancement techniques suitable for supporting the extraction of relevant features identified in step 2
4	Identify data range which represent relevant features from enhanced image using density slice function
5	Create class image from selected data range
6	Create tiff (8bit) from class image and import tiff to ArcGIS
7	Improve raster dataset if necessary using raster functions like fill to fill raster holes
8	Convert raster dataset to polygon shape file and calculate areas for polygons
9	Delete polygons which do not represent relevant features using area attribute
10	Convert polygon shapefile to point shapefile
11	Calculate density distribution of points using Kernel Density function, adjust kernel size and analysis window to analysis scale (100, 100)
12	Use zonal statistics to calculate mean density ranges for each occupancy category
13	Use determined average to classify density dataset
14	Compare output with digitized occupancy category

Radiometric image enhancement

From the point of computation, radiometric enhancement procedures involve determining a new brightness value for a pixel by some specific algorithm from its existing brightness values. Because neighbouring pixels have no influence as they have in geometric enhancement procedure, radiometric image enhancement is less processing intensive (Richards, 1994). Most radiometric image enhancement techniques relate to scalar image, in which each pixel has only a single brightness value associated with it. In this section, the panchromatic Quickbird image is used, stored in 11 bit information depth with a dynamic range from 0 to 2048.

A very powerful technique radiometric image enhancement technique is contrast stretching. If the raw image was directly used in the display device, only a small portion of the full range of possible display levels would be used. A number of stretch types are available and some are tested in the following sections using image analysis software ENVI. The simplest stretching is the linear stretching. It sets the minimum and maximum input values to 0 and 255 respectively and all other values in between are linearly aligned to intermediate output values. The minimum and maximum values can be interactively adjusted depending on the image feature one wants to emphasize. As a first step, a preliminary testing is conducted on a panchromatic subset. Figure 27 displays spectral profiles for a selected section of the panchromatic subset the linear stretch is applied to. From this it becomes obvious that high DN values of the peaks refer to a high reflectance of the object e.g. the bright roofs correspond well with the peaks in the spectral profile. The initial reflectance of these objects can be seen in the distribution of the DN displayed in the upper histogram in Figure 28. It shows a small peak between 900 to 1100 DN which represents high reflectance objects like bright building roofs. The second and larger peak of the histogram between 250 and 525 represents all features with moderate reflectance such as grayish roofs or unpaved streets. The bright roofs can serve as an indicator for middle and high class areas compared to the large portion of the buildings have moderate reflectance. Therefore the minimum and maximum values are set close to the moderate reflectance peaks in the histogram to increase the brightness values of the moderate reflectance roofs. In this case, the values are set to 335 and 579 respectively. In the lower figure displays the resulting histogram of the panchromatic subset. To identify the data range which represents the pixel of interest, in this case roof pixels from middle to high income areas are identified using the density slice function on the stretched image. For the linear stretched image, a data range of 248 – 255 DN is identified from visual inspection.

In a next step, the correlation between pixel value and occupancy categories is analyzed. The working hypothesis is the each occupancy category is characterized by a certain density of specific pixel values. The grid generated using the density slice function which represents the data range between 248 and 255 is converted to a polygon shape file. Then the polygon shape file is converted to a point shape file.

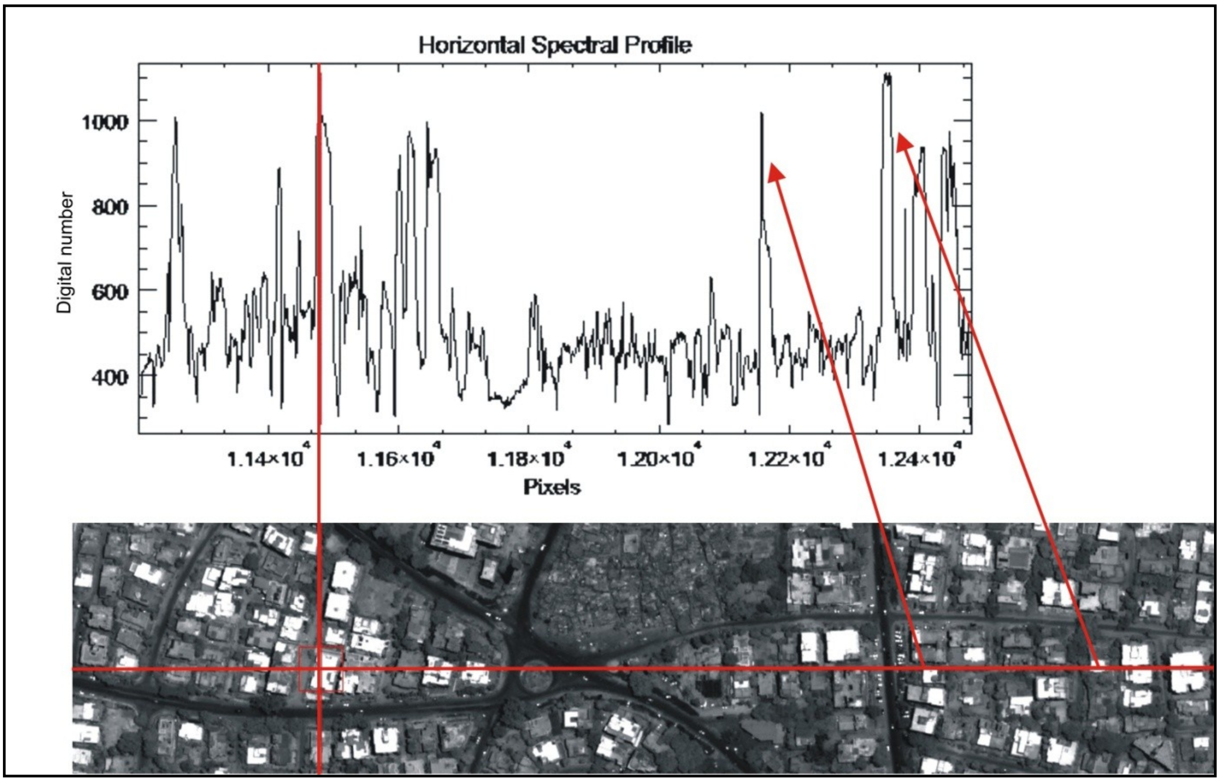


Figure 27: Panchromatic subset from Quickbird image and horizontal spectral profile. The bright roofs correspond well with the peak in the spectral profile.

A kernel density analysis is conducted to calculate the density of the point representing bright features. For each occupancy category, a mean density value is calculated using a zonal statistic function. Because the analysis is conducted is a spatial analysis, an open space occupancy category is introduced to avoid coverage gaps. The mean density value refers to the mean density of bright features in the occupancy categories, mostly bright roofs. These ranges give a first idea of the distribution of the occupancy categories using the density of bright features. The residential occupancy has a significantly higher density of bright features than the other categories. Since the linear stretch technique allows for interactive adjustment of the minimum and maximum values, another test is conducted by selecting a different data range to be stretched. The values are set to 369 and 414 respectively. With this setting only the very low DN values, which represent the low reflectance materials, are preserved and stretched linearly. The same analysis procedure is utilized to calculate the mean density of low reflectance features for each occupancy category.

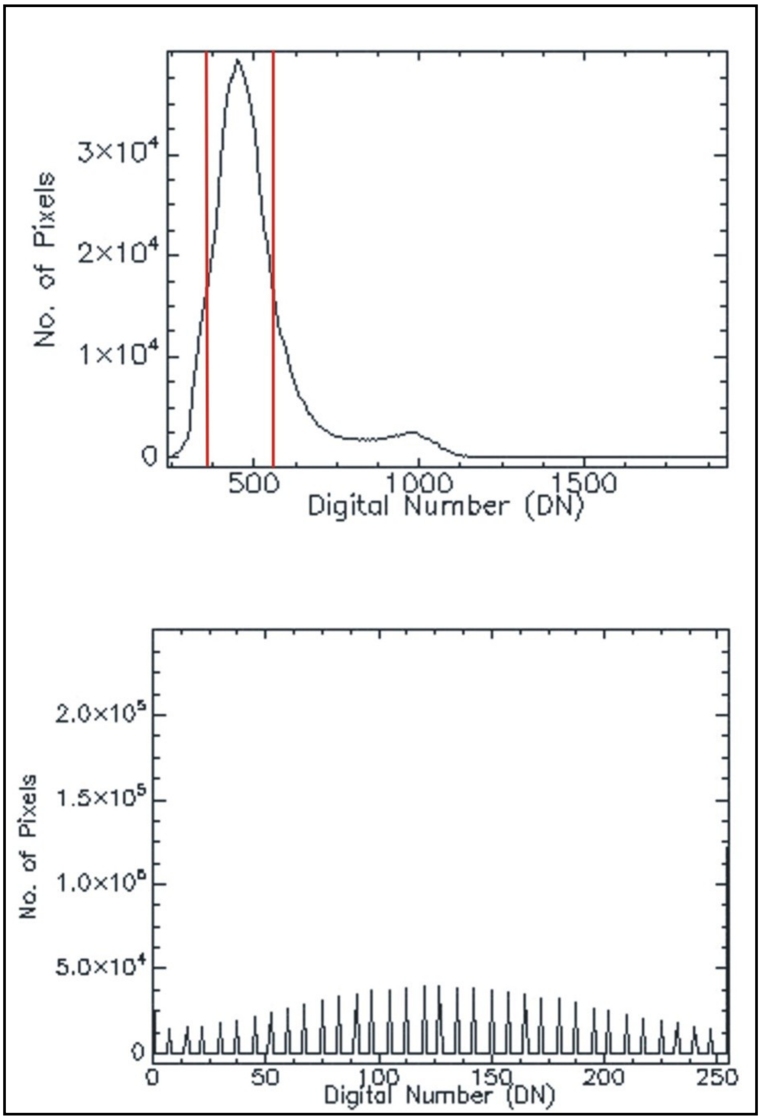


Figure 28: Linear stretch histograms. Upper histogram: The right peak between 900 and 1000 DN represents the high reflectance objects whereas the large peak represents the large portion of moderate reflectance objects. Lower histogram: Linear stretch for DN from 335 to 579 to increase the reflectance of the bright roofs which serve as an indicator for middle and high income areas.

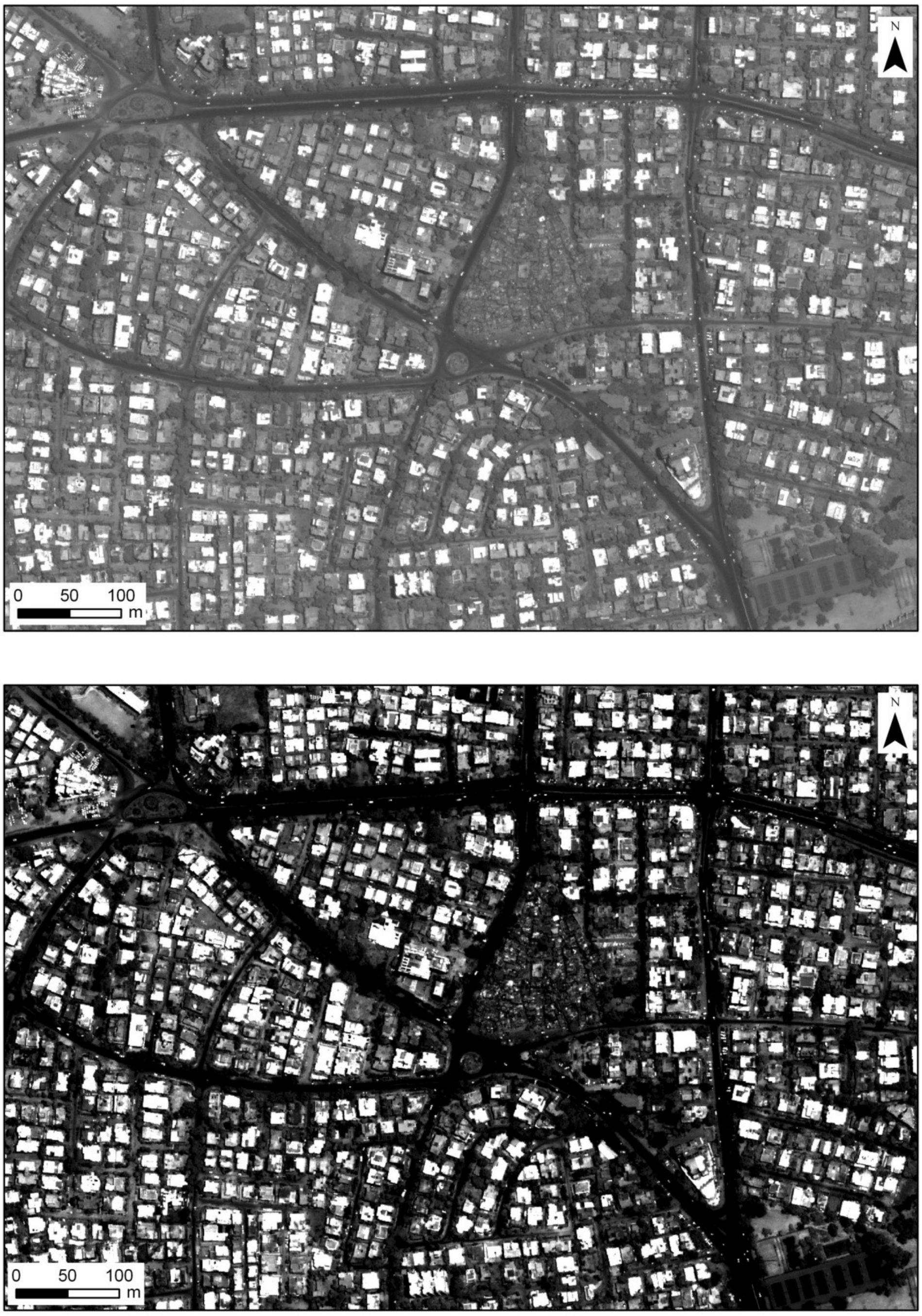


Figure 29: Linear stretch on panchromatic subset. The upper figure shows the original panchromatic image. The lower figure displays the linear stretched image in which the brightness of the moderate reflection objects is increased. From the image, it becomes obvious that less bright roofs can be observed in low income areas then in middle to high income areas.

Table 41: Mean density range for bright features in the data range 248 and 255 extracted using the density slice function on the linear stretched image.

Occupancy Category	Mean Density Range
Residential	1.293,48 – 1805,40
Slum	792,68 - 937,36
Commercial	792,68 – 937,36
Service	709,22 – 792,68
Open Space	625,75 – 709,22
Industrial	0 - 386,49

Table 42: Mean density range for the linear stretched DN values between 369 and 414.

Occupancy Category	Mean Density Range
Residential	0 – 1744,42
Slum	1896,33 – 1957,93
Commercial	2175,53 – 2791,40
Service	1957,93 – 2040, 04
Open Space	1744,42 – 1879,92
Industrial	2040,04 – 2175,53

Figure 30 displays the resulting image. In contrast to the increased contrast between bright roofs and surroundings displayed in Figure 29, Figure 30 displays the inverse of image. However, the mean density ranges for the occupancy categories show that there is no distinct correlation between occupancy categories and the density of pixels representing shadow. In order to generate an occupancy map based on the identified density ranges, the stretched images are classified using the density ranges as class threshold. However, the density ranges revealed to be too close to serve as class thresholds. No explicit occupancy boundaries can be generated.

Gaussian Stretching

The foregoing section addressed the simple expansion or contraction of the histogram of an image. Some radiometric image enhancement techniques are based on a pre-specified shape of an image histogram to give a modified image with a particular distribution of brightness values (Richards, 1994). The reference is a mathematical function that describes the desired shape. The Gaussian stretching technique matches an image histogram to a Gaussian shape. The modified output image has few black and white regions with the most detail contained in the mid-grey range. In ENVI, the default Gaussian stretch is centred at a mean of DN 12 with the data values ± 3 standard deviations set to 0 and 255. For the panchromatic image, the data range to be stretched was set to 512 and 608. This way all brightness values for the high reflectance pixel are increased, whereas the brightness values for the low reflectance pixel are decreased (see Figure 31). Using the density slice function the high reflectance pixel (250 – 255 DN) are extracted and exported as a grid to ArcGIS for statistical analysis (see Table 43).

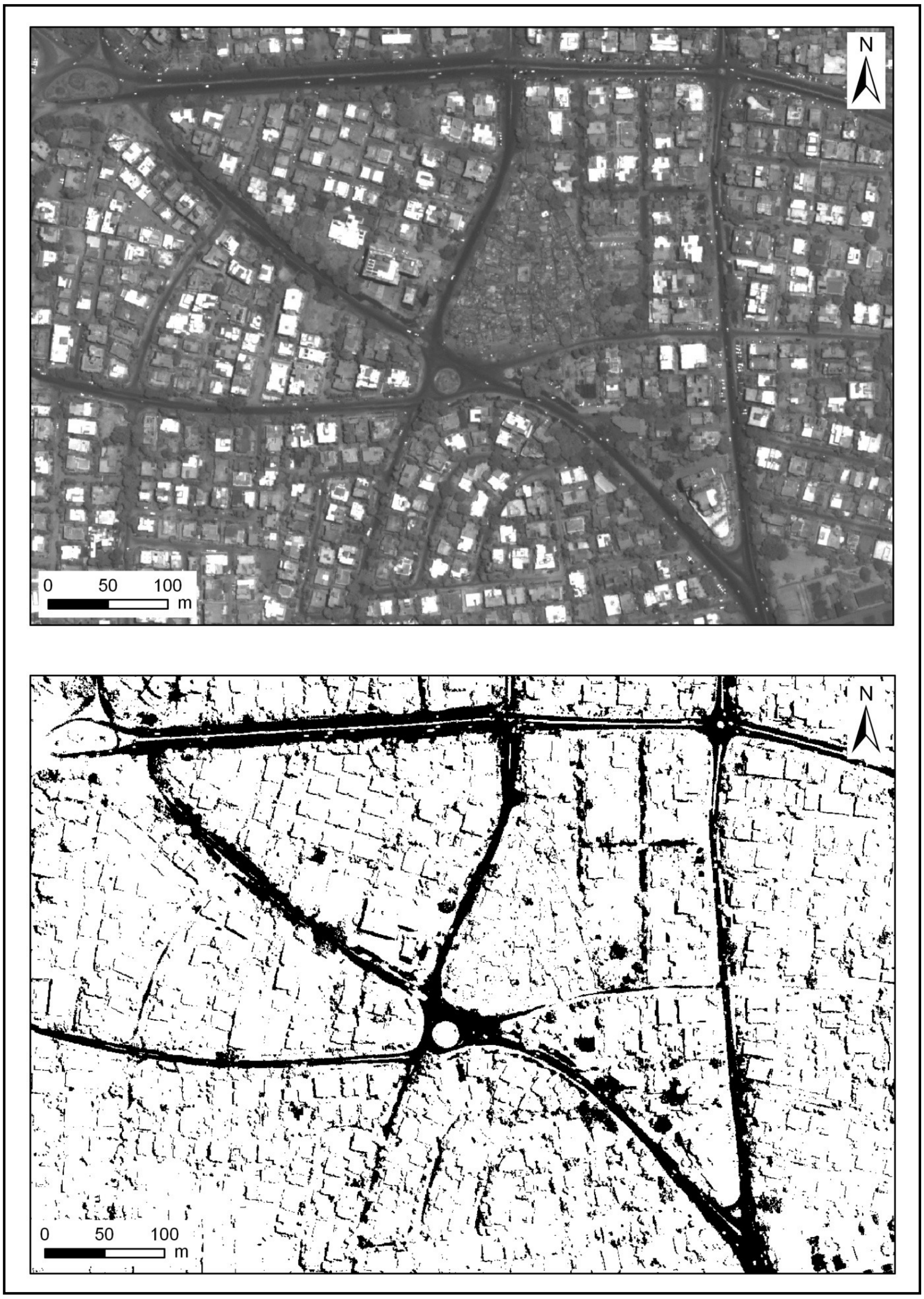


Figure 30: Linear stretch on panchromatic subset. The upper figure shows the original panchromatic image. The lower figure displays the linear stretched image with the data range set to 369 to 414.

Histogram Equalization Stretching

In contrast to linear stretching where as many display level are assigned to the rarely occurring image values as to the frequently occurring values, with histogram equalization image values are assigned to the display level on the basis of their frequency of occurrence. If a brightness value has many pixels associated with it, then it is given a new grey level value that is farther away from its lower neighbor, thus improving their relative contrast. For the panchromatic image, the data range to be stretched is set to 192 and 440. If this small data range is uniformly stretched, the data outside the range will appear very bright if larger than 440 and very dark if smaller than 192. This allows for extracting the low reflectance areas in the image which represent streets, dark roof and shadow (see Figure 32). Using the density function set to 0 – 140, the dark features are extracted and saved in an image file, then exported to ArcGIS for statistical analysis. The density distribution is calculated using a kernel window of 100 ×100m and for each occupancy category the mean density is calculated (see Table 44).

Table 43: Mean density range for bright features in the data range 248 and 255 extracted using the density slice function on the Gaussian stretched image.

Occupancy Category	Mean Density Range
Residential	1396,76 – 2181,96
Slum, Commercial	730,74 – 1046,23
Service	716, 72 – 730,74
Open Space	641,61 – 716,72
Industrial	394,26

Table 44: Mean Density Range for bright features in the data range 192and 440 extracted using the density slice function on the histogram equalization stretching.

Occupancy Category	Mean Density Range
Residential	14411 – 15937
Slum	15937 – 18050
Commercial, Industrial	14176 - 14411
Service	13061
Open Space	13061 – 14137

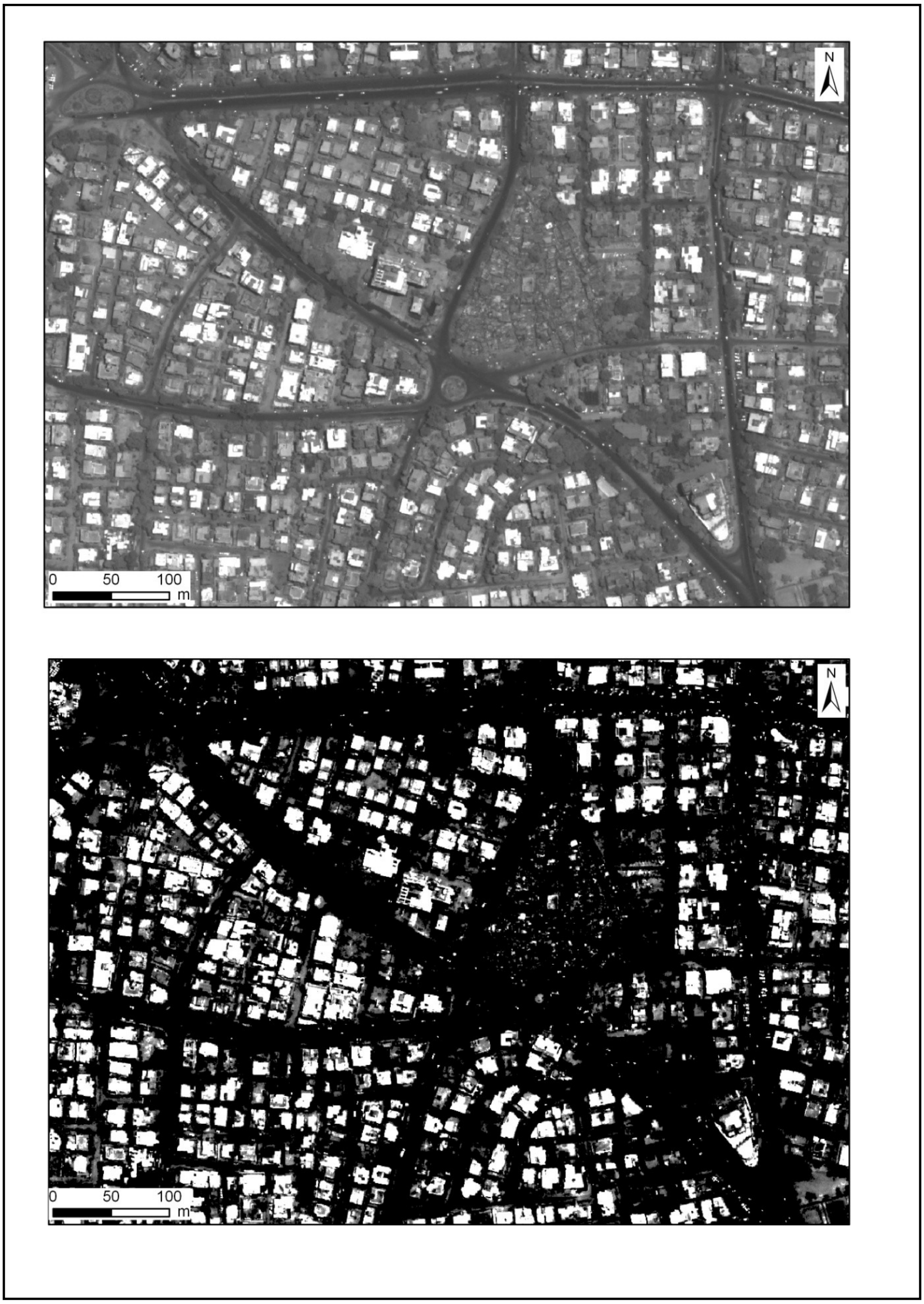


Figure 31: Gaussian stretch on panchromatic subset. The upper figure shows the original panchromatic image. The lower figure displays the linear stretched image with the data range set to 512 to 608.

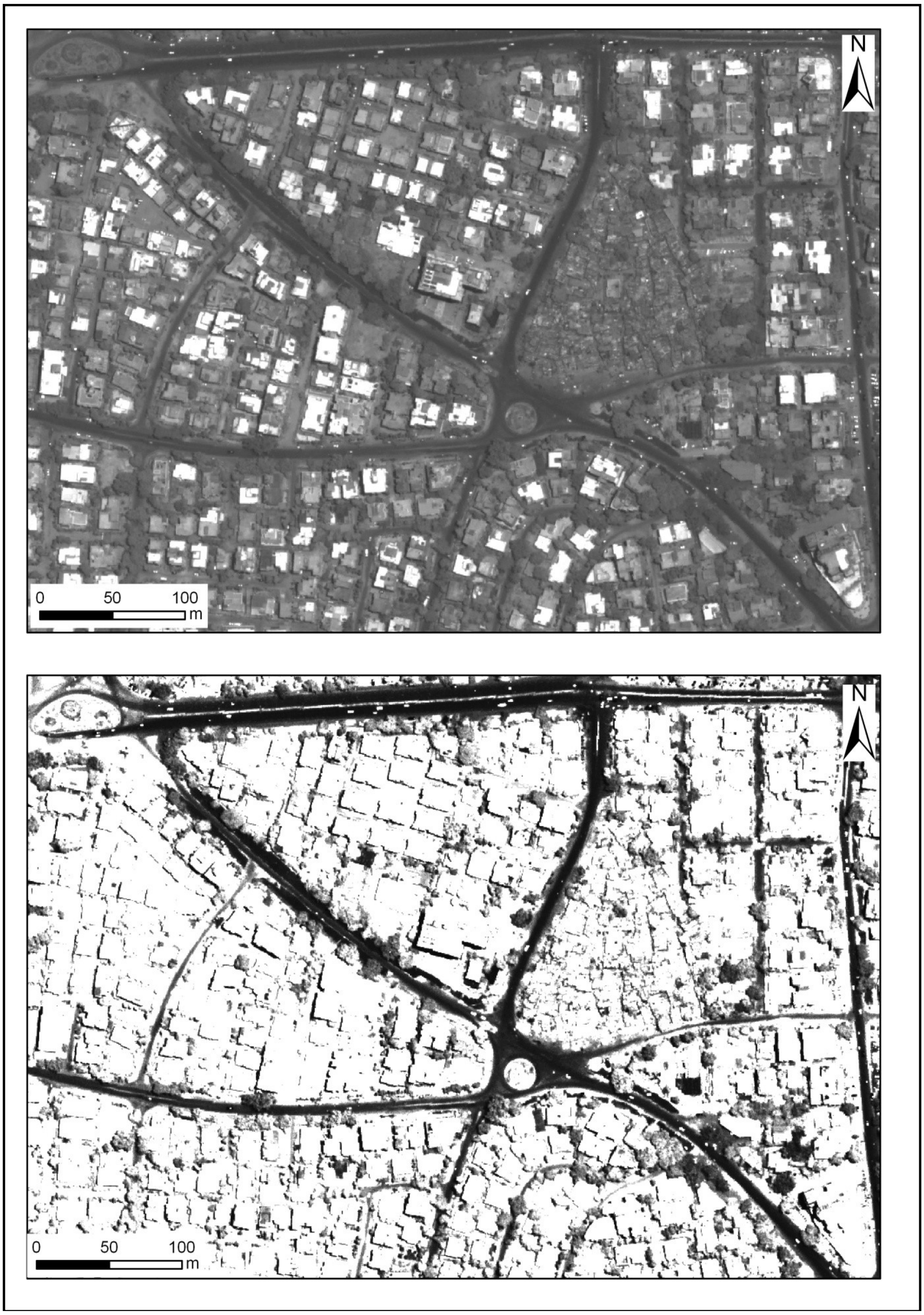


Figure 32: Histogram equalization stretch on panchromatic subset. The upper figure shows the original panchromatic image. The lower figure displays the linear stretched image with the data range set to 192 and 440.

Square Root Stretching

The square root stretching technique takes the square root of the input histogram and applies then a linear stretch. A square root stretch compresses the higher brightness values within an image and disproportionately expands the darker values. It is applied by taking the square root of the original brightness values. Thus the original darker values are given a greater relative contrast than the original values at the higher end of the brightness scale. This would be done in order to allow better visual discrimination among the darker features of the image, while maintaining some limited recognition of the brighter features (see Figure 33). The data range to be stretched was set to 520 and 616. The relevant features were extracted using density slicing with a value range of 250 to 255 (see Table 45).

Table 45: Mean density range for bright features in the data range 520 to 616 extracted using the density slice function on the square root stretching.

Occupancy Category	Mean Density Range
Residential	428,63 – 688,76
Slum	148,50 – 204,07
Commercial	317,4 – 428,63
Service	121,81 – 148,50
Open Space	148,49 – 204,07
Industrial	121,81

In this section, four different contrast stretching techniques (linear, Gaussian, square root and Histogram Equalization Stretching) are tested for semi-automated extraction of occupancy categories on city level. Six different occupancy categories are applied (residential, commercial, industrial, slum, service and open space). For the six categories it turned out that features with distinct occupancy – related density which can be used to distinguish the different occupancies do not exists. Although the different occupancy categories can be clearly distinguished by the human eye, the automatic extraction using the grey values of the panchromatic image only is not feasible. However, the stretched images can be used to assist a manual digitization procedure to extract occupancy categories. In a next step, the suitability of spectral band algebra to extract occupancy categories is tested employing the information available from the spectral bands of the Quickbird image.

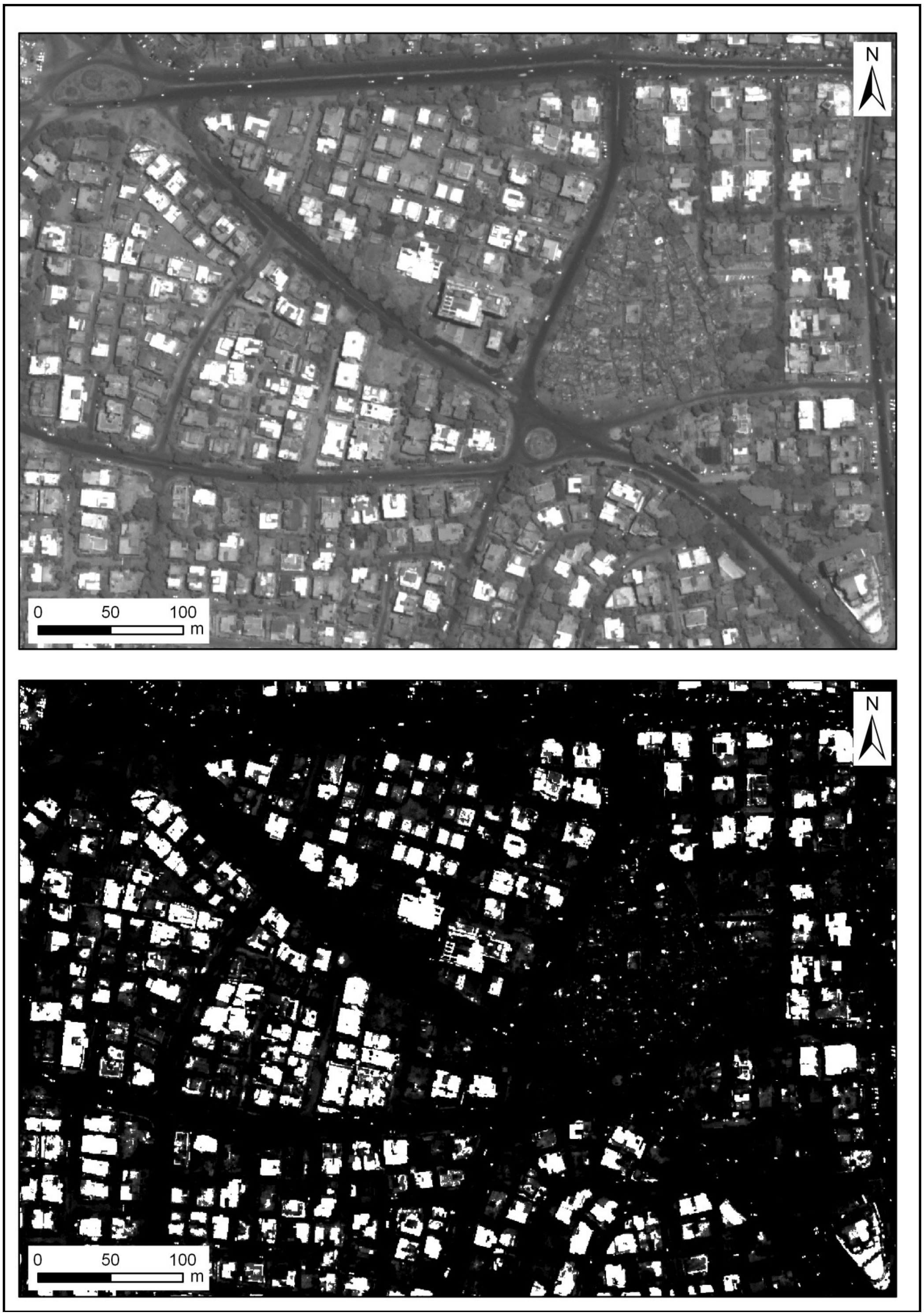


Figure 33: Square root stretch on panchromatic subset. The upper figure shows the original panchromatic image. The lower figure displays the linear stretched image with the data range set to 520 and 616.

Spectral band algebra

Spectral band algebra is the procedure to conduct algebra operations with different spectral bands. With ENVI, the spectral math function is applied on pansharpend, multispectral Quickbird images. First, the ratio between the NIR and B band is calculated and displayed as an image. Using the density slice function, the data representing relevant feature is identified. The pixels within the data range of 0,9 – 1,4 display the built up areas. Table 46 displays the different band ratios calculated and the data range which represent built up urban areas. To ensure that the chosen data range represents the built-up areas with sufficient accuracy, an overlay analysis is conducted using manually digitized occupancy data as a dataset. The percentage of correctly identified built-up urban areas using the band ratio is calculated. Figure 34 Figure 34: Distribution of built-up area in % in different occupancy categories. Table 47 displays the distribution of the detected built-up areas using NIR / B ratio in the individual occupancy categories. The highest percentage of built-up area is detected in slum and residential area. It is surprising that the built-up area in commercial areas with 73,93% is much higher than in industrial (54,43%) areas. From the analysis, it becomes obvious that the lowest percentage can be found in streets and open space. This can be explained by the fact that in Ahmedabad not all streets are paved and therefore might be detected as bare ground rather than impervious surface. The analysis is repeated for the other calculated ratios and the results are compared in the following. From this analysis, it becomes obvious that the results follow the working hypothesis that the occupancy categories are characterised by specific spectral features. For example, the highest percentage of built-up area is observed for the residential categories and built-up areas have a distinct spectral signature. It is important to note that the percentage of built-up areas differs from ratio to ratio, but the trend is similar. Although the built-up area detection showed reasonable results, the similar percentage of built-up area in the different categories makes it unsuitable criteria to distinguish different occupancies. The only categories which can be distinguished is open space because its spectral characteristic are very different from the other categories and open space is compared to the other categories rather homogenous.

Table 46: Band ratios and identified data range for built-up area.

Ratio	Feature	Data Range
NIR / B	Built-up area	0,90 – 1,50
NIR / R	Built-up area	0,99 – 1,10
NIR – R	Built-up area	3,00 – 60,0

Table 47: Occupancy categories and percentage of built-up area obtained from different spectral ratios.

Occupancy category	NIR / B (% of built-up area)	NIR / R (% of built-up area)	NIR – R (% of built-up area)
Slum	76,09	44,02	33,19
Residential	74,67	51,09	32,51
Commercial	73,93	45,13	30,57
Street	63,99	26,44	24,16
Industrial	54,43	25,06	18,73
Service	49,11	25,86	16,82
Open Space	19,87	08,09	05,65

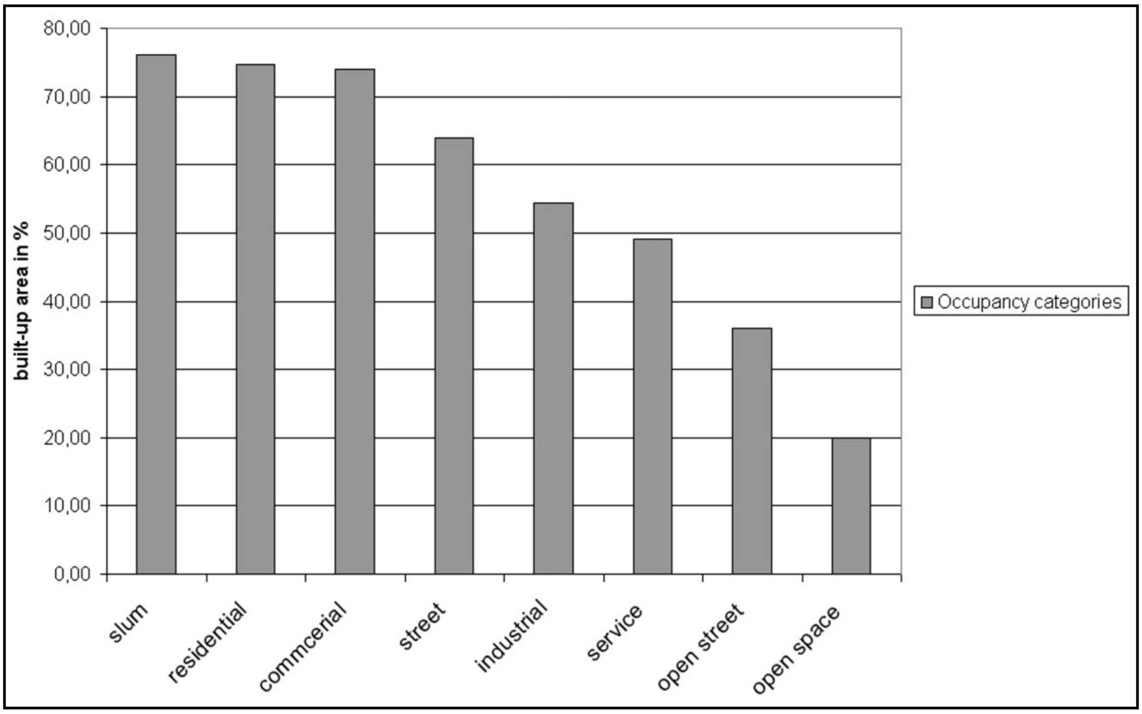


Figure 34: Distribution of built-up area in % in different occupancy categories.

Geometric image enhancement

In the following section, different spatial filtering techniques are tested to delineate occupancy categories from Quickbird image. Spatial filtering is a local operation in that pixel values in an original image are modified on the basis of the grey values of their neighbouring pixel (Lillesand et al., 2008). Two different types of filters are tested: (1) convolution and morphology filters and (2) textural filter.

Convolution and Morphology Filters

A number of convolution and morphology filters are available in the ENVI software. The high pass filter is implemented as a moving window which passes throughout the image. A new image is created whose digital number at each pixel corresponds to the local average of moving window at each of its positions in the original image. The high pass filter emphasizes the high frequent components of an image while removing the low frequency components. In the new image, the grey level changes abruptly over a relatively small number of pixels in these areas (see Figure 35). This way edges between different regions can be enhanced as well as images sharpened. This is accomplished by using a kernel with a central high value, typically surrounded by negative weights. ENVI uses a 3 times 3 kernels with a value of 8 for the central pixel and of -1 for the exterior pixel. In this study a high-pass filter is applied to the panchromatic image using different kernel sizes of 3, 5, 7 and 9 to identify the best size for the moving window. From a visual inspection the image filtered with kernel size 7 is selected because the differences between the occupancy categories are emphasised best. A density slice is applied function to separate relevant data ranges from the filtered grey scale image. From the results of the geometric image enhancement, the same observation as for the other enhancement techniques can be made. The spectral and textural characteristic of the occupancy categories are not distinct enough to be automatically extracted. Especially on city scale, the characteristics of the occupancy categories are too heterogeneous to be detected by a single density threshold.

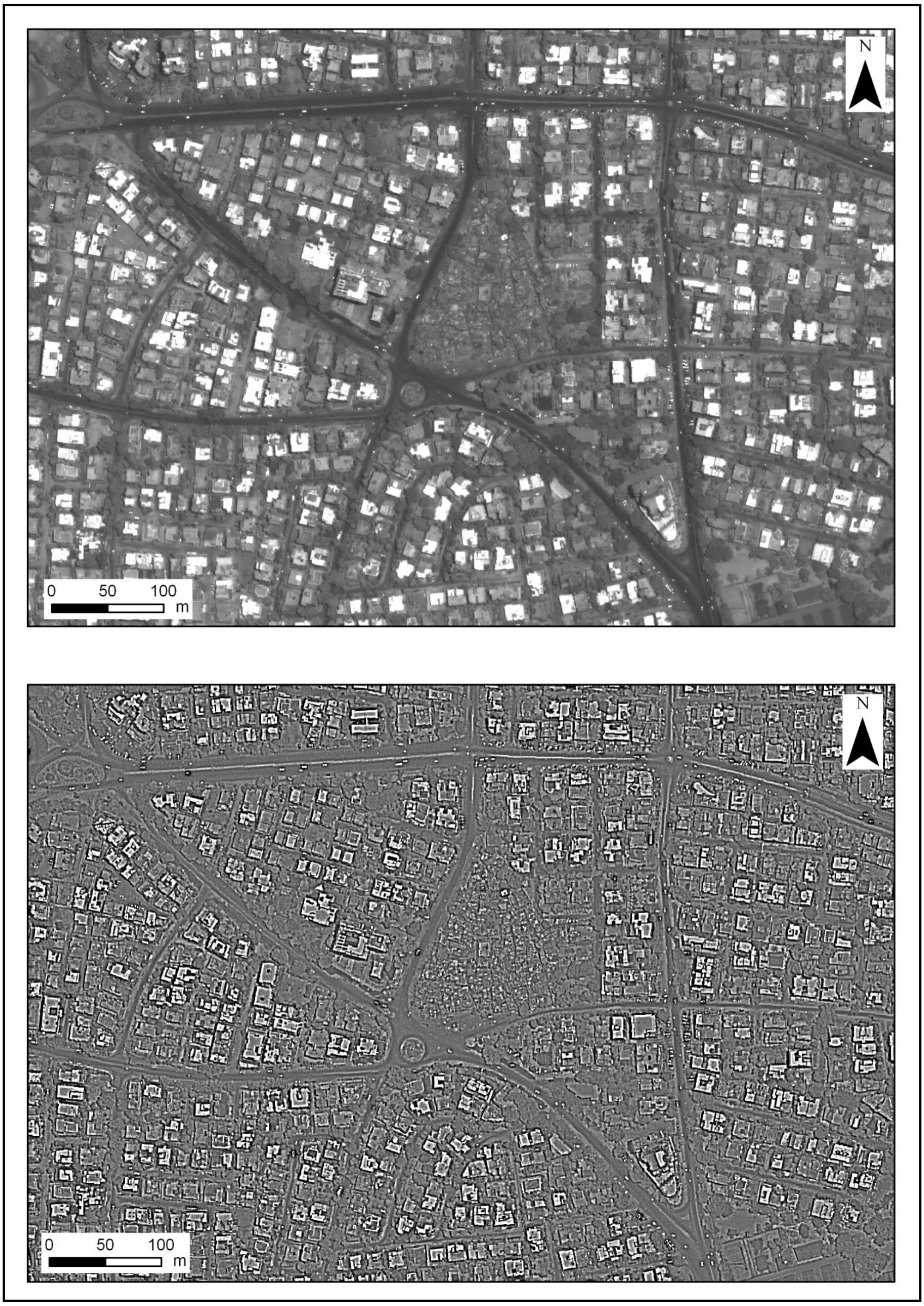


Figure 35: High pass filter with a kernel size of 3 applied to the panchromatic Quickbird image.

Appendix V

Table 48: Projected district population, district area (km²) and population density (people/km²) for 2008 within the 43 individual districts in Ahmedabad (see section 5.1).

District	No.	Population 2008	Area (km²)	Population density 2008 (people/km²)
Khadia	1	44346	1,222	36289,69
Kalupur	2	54847	1,494	36711,51
Dariapur	3	58559	0,873	67077,89
Shahpur	4	61672	1,260	48946,03
Raikhad	5	62972	2,086	30187,92
Jamalpur	6	69615	1,017	68451,33
Paldi	7	83992	5,667	14821,25
Vasna	8	132515	5,684	23313,69
Gandhigram	9	69146	6,914	10000,87
Navrangpura	10	58510	6,765	8648,93
S P Stadium	11	80632	3,594	22435,17
Naranpura	12	99255	3,588	27663,04
New Wadaj	13	72203	2,317	31162,28
Old Wadaj	14	78997	3,840	20572,14
Sabarmati	15	80055	12,101	6615,57
Dudheshwar	16	70561	2,658	26546,65
Madhupura	17	74035	3,524	21008,80
Girdhar Nagar	18	67617	2,321	29132,70
Asarwa	19	44468	1,463	30395,08
Naroda Road	20	93742	1,686	55600,24
Saraspur	21	71736	1,752	40945,21
Potalia	22	88721	3,954	22438,29
Kuber Nagar	23	101267	2,161	46861,18
Sardar Nagar	24	124779	9,888	12619,24
Salipur Bogha	25	88123	2,282	38616,56
Thakkarbapa Nagar	26	177006	3,216	55039,18
Naroda Muthia	27	114833	10,371	11072,51
Bapun Nagar	28	108889	2,352	46296,34
Rakhial	29	79685	2,424	32873,35
Gomtipur	30	69601	1,759	39568,51
Rajpur	31	78350	3,105	25233,49
Amrawadi	32	80490	2,130	37788,73
Bhaipura Hatkeshwar	33	131699	1,728	76214,70
Nikoal Road	34	162293	4,106	39525,82
Odhav	35	140471	7,320	19190,03
Khokhra Maherndabad	36	74053	2,383	31075,54
Mani Nagar	37	102215	3,421	29878,69
Kankaria	38	74901	3,488	21473,91
Baherampura	39	84615	7,788	10864,79
Danilimda	40	141951	6,231	22781,42
Bag e firdosj	41	169399	6,597	25678,19

District	No.	Population 2008	Area (km^2)	Population density 2008 (people/km^2)
Vatva	42	163295	39,567	4127,05
Isanpur	43	148674	12,932	11496,60

Abbreviations

AGR	Annual Growth Rate
AMC	Ahmedabad Municipal Cooperation
ASAG	Ahmedabad Study Action Group
ATC	Applied Technology Council
AVHRR	Advanced Very-High Resolution Radiometer
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
AUDA	Ahmedabad Urban Development Authority
CIA	Central Intelligence Agency
CIESIN	Center for International Earth Science Network
CNSA	China National Space Administration
DCW	Digital Chart of the World
DEM	Digital Elevation Model
DLR	German Aerospace Center
DMA	Defence Mapping Agency
DMS	Digital Surface Model
DMSP OLS	United States Air Force Defence Meteorological Satellite Program -Operational Linescan System
DN	Digital Numbers
DTED	Digital Terrain Elevation Data
ESA	European Space Agency
EMS	European Macro Seismic Scale
EERI	Earthquake Engineering Research Institute
EPEDAT	Early Post-Earthquake Damage Assessment Tool
ESRI	Environmental Systems Research Institute Inc.
FEMA	Federal Emergency Management Agency
GIDB	Gujarat Infrastructure Development Board
GLCC	Global Land Cover Characterization
GPW	Gridded Population of the World
GRUMP	Global Rural Urban Mapping Project
HAZUS	Hazards U.S.
HAZUS MH	Hazards U.S. Multi-Hazard
IAEE	International Association for Earthquake Engineering
IBC	International Building Code
IDB	International Data Base
IFPRI	International Food Policy Research Institute
INLET	Internet-based Loss Estimation Tool
IRS	Indian Remote Sensing Satellite
JRC	Joint Research Centre
KN	Kadiya Nakas
LU	Land Use
LC	Land Cover
M_g	Magnitude
MSK	Medvedev-Sponheuer-Karnik
MODIS	Moderate Resolution Imaging Spectroradiometer

NASA	National Aeronautics and Space Administration
NCGIA	National Center for Geographic Information Analysis
NDVI	Normalized Differenced Vegetation Index
NEHRP	National Earthquake Hazard Reduction Program
NIBS	National Institute of Building Science
NIMA	National Imagery Mapping Agency
NIR	Near-Infrared
NOAA	National Oceanic and Atmospheric Administration
NRC	National Research Council
OECD	Organisation for Economic Co-operation and Development
ONC	Operational Navigation Chart
ORG	Operation Research Group
ORNL	Oak Ridge National Laboratories
PAGER	Prompt Assessment of Global Earthquakes for Response
PGA	Peak Ground Acceleration
RESCUE	RESponding to Crises and Unexpected Events
RGB	Red, Green, Blue
ROI	Region of Interest
SAATH	Gujarati for "Together, Co-operation, a Collective or Support"
SEDA	Socioeconomic Data and Application Center
SEWA	Self Employed Women's Association
SRTM	Shuttle Radar Topography Mission
UNECE	United Nations Economic Commission for Europe
UNECE PAU	United Nations Economic Commission for Europe Population Unit
UNEP	United Nations Environment Program
UNOSAT	UN Institute for Training and Research (UNITAR) Operational Satellite Applications Program
USGS	US Geological Survey
VHR	Very High Resolution
VMap0	Vector Smart Map
WAPMERR	World Agency of Planetary Monitoring and Earthquake Risk Reduction
WHE	World Housing Encyclopaedia

List of Figures

List of Tables