EVALUATING URBAN GROWTH TRENDS BY USING SLEUTH MODEL: A CASE STUDY IN ADANA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HÜSEYIN ÇAPAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

JULY 2019

Approval of the thesis:

EVALUATING URBAN GROWTH TRENDS BY USING SLEUTH MODEL: A CASE STUDY IN ADANA

submitted by **HÜSEYIN ÇAPAN** in partial fulfillment of the requirements for the degree of **Master of Science in Geodetic and Geographic Information Technologies Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Zuhal Akyürek
Head of Department, **Geodetic and Geographic Information Technologies** _____

Prof. Dr. Zuhal Akyürek
Supervisor, **Civil Engineering, METU** _____

Examining Committee Members:

Assoc. Prof. Dr. Uğur Murat Leloğlu
Geodetic and Geographic Information Technologies, METU _____

Prof. Dr. Zuhal Akyürek
Department of Civil Engineering, METU _____

Assoc. Prof. Dr. Aslı Özdarıcı Ok
School of Land Registry and Cadastre, Ankara HBV University _____

Assist. Prof. Dr. Meltem Şenol Balaban
Department of City and Regional Planning, METU _____

Assist. Prof. Dr. Semih Kuter
Department of Forest Engineering, Çankırı Karatekin University _____

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Hüseyin Çapan

Signature :

ABSTRACT

EVALUATING URBAN GROWTH TRENDS BY USING SLEUTH MODEL: A CASE STUDY IN ADANA

Çapan, Hüseyin
M.S., Department of Geodetic and Geographic Information Technologies
Supervisor: Prof. Dr. Zuhal Akyürek

July 2019, 125 pages

In this study, various urban growth scenarios are simulated for Adana city by using SLEUTH urban growth model. In addition, the impact of the planned road in 2020 was investigated. The results of this study were compared with other SLEUTH model applications using a SLEUTH-specific comparison method. SLEUTH is a cellular automata simulation model developed for urban growth modelling. It is written in the C language and has an open source library. Since its first occurrence as an urban growth model, SLEUTH model has become the most popular one. It has been applied to more than 50 cities, with various scales, around the world. Superiority of the SLEUTH against other cellular automata approaches has been proven in some studies in the literature. Adana has the most fertile lands in Turkey, but it has been facing with immigration and rapid urbanization problems. Landsat imagery acquired in 1990, 2001, 2006, 2011 and 2016 for Adana are used as input dataset for the SLEUTH model. Urban plan for 2016 has been obtained from official sources in order to create a basis for the scenarios. Urbanization in 2050 is estimated using three scenarios representing current trends, and green areas being fully and partially

protected. Model sensitivity has been discussed and the changes in the impervious surfaces in the area are obtained. Prediction results show that the model responds each scenario in a different manner. Between 2016 and 2050 percent urban area increase for the first scenario is 120%, for the second scenario it is 62% and for the last and most protective third scenario, 50% increase in urban areas have been predicted. Detailed urbanization analyses have been performed to find green and agricultural area losses due to urbanization in the future.

Keywords: Urbanization, Urban growth modelling, SLEUTH, Supervised classification

ÖZ

KENTSEL BÜYÜME TRENDLERİNİN SLEUTH MODELİ İLE DEĞERLENDİRİLMESİ: ADANA'DA BİR DURUM ÇALIŞMASI

Çapan, Hüseyin
Yüksek Lisans, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü
Tez Yöneticisi: Prof. Dr. Zuhal Akyürek

Temmuz 2019 , 125 sayfa

Bu çalışmada, SLEUTH kentsel büyüme modeli kullanılarak Adana için çeşitli kentsel büyüme senaryoları simüle edilmiştir. Ayrıca 2020 yılında yapılması planlanan yolun etkisi araştırılmıştır. Bu çalışmadan elde edilen sonuçlar, SLEUTH'e özgü bir karşılaştırma yöntemi kullanılarak diğer SLEUTH modeli uygulamalarıyla karşılaştırılmıştır. SLEUTH, kentsel büyüme modellemesi için geliştirilen bir hücresel otomata simülasyon modelidir. C dilinde yazılmış ve açık kaynak kodlu bir kütüphaneye sahiptir. Kentsel bir büyüme modeli olarak ilk ortaya çıkışından bu yana SLEUTH, en popüler kentsel büyüme modeli haline geldi. Dünyada çeşitli ölçeklerde 50'den fazla şehre uygulandı ve farklı sürümleri geliştirildi. SLEUTH'nin diğer hücresel otomata yaklaşımlarına karşı üstünlüğü de literatürdeki bazı çalışmalarda kanıtlanmıştır. Adana, Türkiye'nin en verimli topraklarına sahiptir, ancak göç ve hızlı kentleşme sorunları ile karşı karşıyadır. Adana için 1990, 2001, 2006, 2011 ve 2016 yıllarında elde edilen Landsat görüntüleri SLEUTH modeli için girdi veri kümesi olarak kullanılmıştır. Senaryolar için bir temel oluşturmak amacıyla 2016 yılına ait kentsel plan resmi kaynaklardan alınmıştır. 2050 yılındaki kentleşme tahmini, mevcut eğilimleri, kısmi

korumayı ve tam korumayı temsil eden üç senaryo kullanılarak gerçekleştirilmiştir. Model duyarlılığı tartışılmış ve bölgedeki geçirimsiz yüzeylerdeki değişiklikler elde edilmiştir. Tahmin sonuçları, modelin her senaryoya farklı bir şekilde yanıt verdiğini göstermektedir. 2016 ve 2050 yılları arasında kentsel alanlarda birinci senaryoda %120 , ikinci senaryoda %62 en son ve en koruyucu üçüncü senaryoda yüzde %50 artış gözlenmiştir. Gelecekte kentleşme nedeniyle yeşil ve tarımsal alan kayıplarını bulmak için ayrıntılı kentleşme analizleri yapılmıştır.

Anahtar Kelimeler: Kentleşme, Kentsel Büyüme Modeli, SLEUTH, Kontrollü Sınıflandırma

To my parents Vefaret and Mehmet Çapan

And to my sister İrem

Thank you for your love, support and tolerance

ACKNOWLEDGMENTS

I would like to express my sincere thanks especially to my parents for their never ending love. You've gave me everything I needed and encouraged me to stand behind my decisions.

Also, I would like to thank my supervisor Prof. Dr. Zuhal Akyürek for believing in me and never missing her support. Without her continuous encouragement and guidance, this study wouldn't be finished successfully. I feel so lucky to have the opportunity to work with you.

One of the deepest thanks goes to former secretary general of the Adana Metropolitan Municipality, Şaban Dişli for allowing me to make this research by helping me to collect information about urban plans. He instilled in me his love for Adana and encouraged me to do an academic study for Adana.

I would like the express my thanks to Melih Gürçay for sharing his experience on running the SLEUTH model. Even if I didn't use your method, without your experiences, model run could have been a nightmare.

Thank you my dear sister İrem, as a city and regional planner your help with the data manipulation is much appreciated.

I would like to express my thanks to Mr. Keith C. Clarke for writing such a model and being very kind in answering questions from the students.

Dear Feyza Sakıl, your supports on the thesis formatting was very precious, thank you for your patience and friendship.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

GDP	Gross Domain Product
UN	United Nations
LUT	Land Use Transportation
ABM	Agent Based Model
CA	Cellular Automata
USGS	United States Geological Survey
IFOV	Instantaneous Field of View
CATS	Chicago Area Transportation Study
DN	Digital Number
GIS	Geographic Information Systems
SRTM	Shuttle Radar Topography Mission
OSM	Open Street Map
AWS	Amazon Web Services
DEM	Digital Elevation Model
TÜİK	Turkish Statistical Institute

CHAPTER 1

INTRODUCTION

Due to the rapid urbanization in the last few decades, environmental and ecological problems have become more visible and have threatened both humans and wild habitat. Population trends in the past and the projection according to the United Nations, Population Division (UN, 2017) can be seen in Figure 1.1. In 2008 for the first time in history, number of people living in urban areas has exceeded the number of people living in the rural areas. It has been estimated that 70% of world population will be living in cities by 2050 (UN, 2017). Cities embody a lot of things that people need. Various job opportunities, better health care and social facilities are making cities charming than rural areas on the people's perspective. Especially after industrial revolution, migration to cities has become a global phenomenon. But since the urbanization is an irrevocable process, accurate future planning is a must. Otherwise results of bad decisions will be transferred from generation to generation and became tangled.

Usually urbanization process converts natural lands to agriculture first, then agricultural land becomes urban ones (Clarke et al., 1996). When the urbanization process reaches natural borders such as water or steep land, vertical urbanization increases and agricultural areas become more vulnerable to illegal constructions. In this study, urbanization is defined as the expansion of the city in the horizontal direction.

Keith C. Clarke describes urbanization as conversion of natural to artificial land cover characterized by human settlements and workplaces (Clarke et al., 1997). He also states that urbanization speed will be so fast that, people from North America and Europe will not be able to recognize places of their childhood. According to United Nations Department of Economic and Social Affairs, Population Division; world pop-

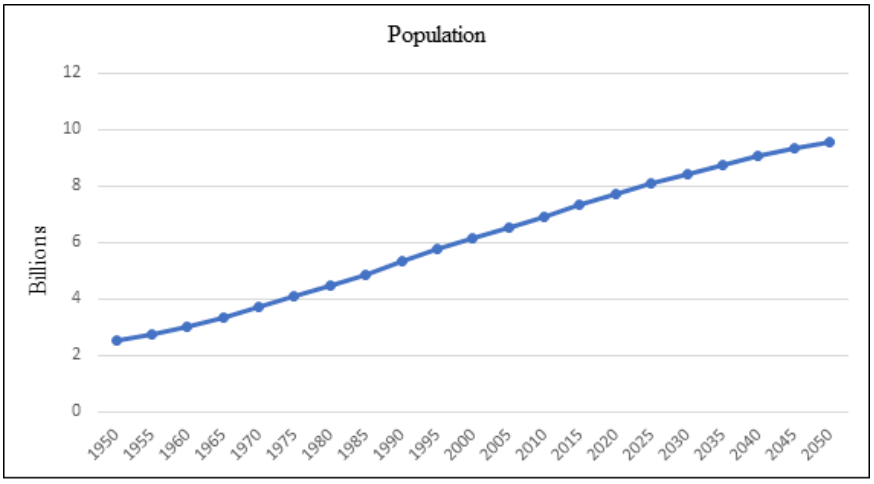

Figure 1.1: Population change between 1950 and 2050 (United Nations, 2016)

ulation increase will be concentrated in urban areas. Additionally to these estimations even if population increase stopped, a deceleration in urbanization would not be guaranteed. Some cities in Europe and North America which have a balanced population growth shows no sign of slowing down in urbanization speed.

On global scale most populated ten cities as of 2016 are as follows (United Nations, 2016);

1. Tokyo, Japan (38 millions)
2. Delhi, India (26 millions)
3. Shanghai, China (24 millions)
4. Mumbai, India (21 millions)
5. São Paulo, Brazil (21 millions)
6. Beijing, China (21 millions)
7. Mexico City, Mexico (21 millions)
8. Osaka, Japan (20 millions)
9. Cairo, Egypt (19 millions)
10. New York-Newark, USA (18 millions)

In 2030 populations of all these cities are expected to rise, except Tokyo. Istanbul with its 14 million inhabitants, as the most crowded city of Turkey, is the 15th most crowded city in the world by 2016. According to TUIK data Istanbul's population has risen with average 19% annually, between 2007 and 2015. Istanbul's population is expected to be around 17 million by 2030 (UN, 2016). Relation between population and urbanization is very tight in developing countries like Turkey. In high income countries, rates of urban land expansion are slower and increasingly related to Gross Domestic Product (GDP) growth (Seto et al., 2000).

In Turkey, due to rapid urbanization after 1950's, slum housing has become a major problem for both citizens and the government. Some laws have been enacted to legalize illegal buildings and unplanned squatter areas. In 1983, because of unplanned urbanization for many years, the law of 2981 was enacted (Prime Ministry Legislation Development and Publication, 1986). With this law, illegal slum buildings turned into apartments until the end of 2000's (Komurcuoglu, 2013). Because of the periodic recurring zoning laws and policies that encourage migration to the cities, urbanization trend is expected to continue countrywide in the future.

Adana, being the 5th most populated city of Turkey has shown a 10% population increase between 2007 and 2015. With this rate Adana has the 15th fastest increase of population in Turkey.

Majority of the urbanization studies focused on specific models like ecological modelling, land use transportation (LUT) models, cellular automata (CA)-based models, agent-based models (ABMs), integrated models, urban economic models (Li and Gong, 2016). Within all these approaches CA-based models have been extensively utilized due to their simplicity, flexibility and intuitiveness (Santé et al., 2010). Among all cellular automata based approaches the most remarkable one is SLEUTH according to some comparative studies (Berberoglu et al., 2016). SLEUTH is an acronym consisting of initials of slope, land-use, excluded, urban, transportation and hill-shade representing the inputs of the model. These models are important because in many developing countries, urban expansion is observed on primary agricultural land (Seto et al., 2000). Land-use changes must be carefully planned to protect this natural economic source.

Aim of this study is building a SLEUTH land-use model to simulate urban growth of Adana and calibrating the patterns to illuminate the growth pattern of Adana and presenting the applicability of SLEUTH to predict the urban development. This study is important in several aspects; first, it shows that urbanization scenario comparison using SLEUTH model can create very detailed spatial information. Secondly, the urbanization pattern of Adana was defined by the comparison of the applications of the model at different scales in different geographies. Thirdly, this study shows that the size of the SLEUTH model is important for the consistency of the model.

In Chapter 2 historical evolution of the urban modelling has been discussed and first examples of the urban modelling are investigated. The shortcomings in the first models and inaccuracies in city modeling approaches are discussed. New generation models from the literature are also investigated. Detailed information about the SLEUTH urbanization model can also be found in this section. In Chapter 3, materials and methodology used to train SLEUTH urban model has been explained. Preparation of three different urban scenarios from the urban plans and satellite imagery classifications has been discussed in this chapter. In Chapter 4, the approach in selecting the coefficients between the phases of the SLEUTH urbanization model and the preparation of the computer environment are explained. In Chapter 5, SLEUTH urban model results are investigated for each scenario, the effect of planned road in year 2020 for the year 2050 is revealed and SLEUTH model results of this study are compared with other studies such as Hawaii and Tampa. In chapter 6, conclusions are presented.

CHAPTER 2

HISTORY OF URBAN GROWTH DETECTION AND PREDICTION

2.1 Monitoring of Urbanization by Remote Sensing

The number of urbanization tracking studies by means of remote sensing has peaked in 70's and then decreased over the time. However, remote sensing studies are still a popular topic due to technological progresses. Google's Ngram Viewer could provide insight about the interest raising over the time about a keyword. How the total number of research topics have changed through the years on 'Urbanization' and 'Remote Sensing' fields can be seen in Figure 2.1. Blue line represents the number of remote sensing books and publications and red line represents the number of the urbanization books and publications over the given period. From this chart one can understand, from the beginning of this century till mid-2000's urbanization mentioned and many books about urbanization have been published. From mid-1970's until now it seems to be it is losing popularity. Remote sensing on the other hand has become very popular when first remote sensing satellites were sent to the space. Through the 1960's urbanization has earned exponential popularity until 1975. Remote sensing researches increased slightly, after the mid-1960's. Technological breakthroughs forced by the cold war has a positive effect on the space science and evolving sensors and low-cost rockets helped scientists to obtain massive data about the Earth.

Despite today's technological progresses, some problems are keeping researchers busy for a long time. The most prominent problem is the mixed pixel problem. Mixed pixels occur when satellite's IFOV (Instantaneous Field of View) covers more than one land cover type (Choodarathnakara et al., 2012). Ideally, it is expected that one pixel represents one land cover type, but when the satellite's spatial resolution is

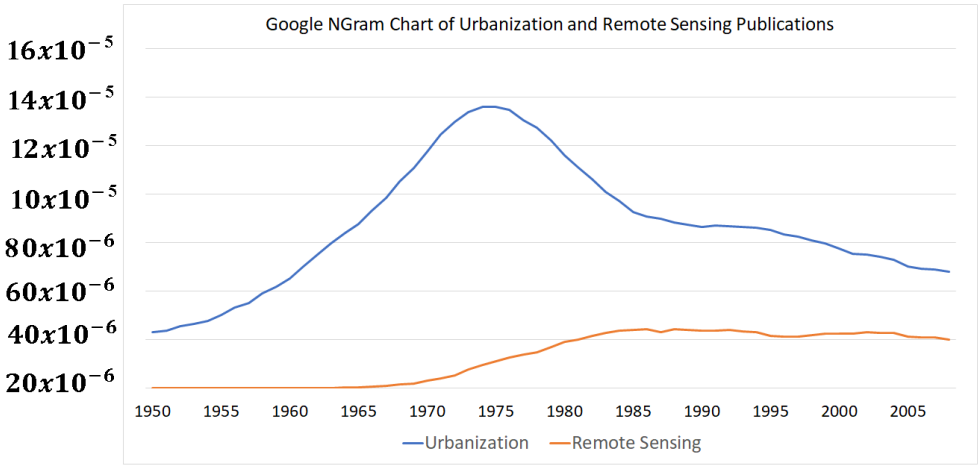

Figure 2.1: Urbanization and remote sensing publications from 1950 to 2008 according to Google Ngram search (Google, 2019) (Y axis is percentage of publications contains the key words)

not enough to understand the difference, pixels are mixed, and numerical counterpart of that pixel is a mixture of two or more land cover type. Remote sensing provides a unique data source for the urbanization studies. The data with high spatial and temporal resolution guide us to understand the development of urbanization. Classification techniques are used to derive information from the remote sensing data. Classification assigns the land-use classes to pixels according to their digital number values. Accurate classification is a key factor for a better spatial prediction.

Different classification approaches can be used for satellite image processing but fundamentally there can be two way of assigning the classes to the pixels such as;

1. **Hard Classification**

   Class memberships are strict, each pixel is assigned to its closest statistical cluster.

2. **Soft Classification**

   Soft classification assigns membership of the pixels to each class. For example, a pure pixel contains only forest land type has a 0% water membership, 100% forest membership and 0% urban membership. If this pixel was mixed with urban and forest, memberships could have changed such as; 2% Water, 60% Forest and 38% Urban.

Shackelford and Davis (2003) have compared traditional maximum-likelihood classification with hierarchical fuzzy classification with a high-resolution imagery for detecting and distinguishing the urbanized areas. Due to complexity of the urban areas it is hard to distinguish features like buildings, asphalt etc. In their study Shackelford and Davis have used Ikonos imagery, a commercial high-resolution satellite imagery, of Columbia, Missouri. Before the classification they have orthorectified the image and applied a colour normalization method to fuse PAN data to MS bands. Maximum likelihood classification was done with seven classes; Road, Building, Grass, Tree, Bare Soil, Water and Shadow. Shadow class was used to minimize problem of shaded pixels in the urban environment. In Shackelford and Davis's study, it has been clearly stated that conventional classification techniques produce high amount of error even with high spatial resolution satellite imagery. Fuzzy membership calculation of hardly distinguishable areas as grass-tree, road-building and water-shadow increases the accuracy of the classification (Shackelford and Davis, 2003). But the SLEUTH model requires only urban areas for future urbanization predictions. Therefore, generally distinction between road and settlements is not very important. Other SLEUTH studies in the literature such as Yakup et al. (2018), Agyemang and Silva (2019), Saxena and Jat (2019), Sahana et al. (2018) also use hard classification of Landsat imagery to train SLEUTH model due to it's simplicity.

2.2 Urban Growth Modelling Approaches

Urban growth modelling is a popular research field and applications all around the world can be found at various scales. Because, an urban growth model that can make dependable projections and predictions is valuable for both scientific and administrative purposes (Berling-Wolff and Wu, 2004). Politicians, policy makers, city planners and finally public can benefit from such a model. Model results can be used by decision makers as a supportive source or different scenarios can be simulated to analyse different policies. Urbanization modelling is only a specialized version of modelling frameworks. Once the modelling framework has been established, not just urbanization but many natural or anthropogenic processes such as; hydrological, infectious diseases and wildfires can also be modelled.

First urban models aimed to model some specific features such as transportation and employment. Theory behind these models were gravity theory, linear or optimizing mathematics and analysis of molecular behaviour. Due to costly procedures and grand expectations that couldn't be satisfied, first urbanization models were disappointment, even state support was withdrawn (Lee, 1973). Independent of the study field, these first unsuccessful attempts were the basis for all future modelling work (Berling-Wolff and Wu, 2004). Eighties and nineties are referred to as the renaissance of the city modelling due to increased spatial data resources and advanced computers (Dietzel and Clarke, 2004). In this era modelling results began to be more consistent and accurate.

2.2.1 Pioneers of Urban Growth Modelling

Roots of urban growth modelling goes back to 60's. Traffic problem, caused by increasing number of vehicles, and concerns of federal government about urban problems gave a birth to demand for more scientific highway impact statements. The complexity and size of the problem forced decision makers to approach the problem scientifically.

Early pioneers of the urban modelling have sought to model either to predict the economic and size relationships between cities or to discover the internal social and economic patterns within the limits of the city (Clarke et al., 1997). For some specific reasons these early systems didn't work as expected and most of them were not able to pass the experimental phase.

Very valuable information learned from the failure of the first systems. Lessons learned from the failures have helped researchers to build better models in the future. Procedural mistakes according to Lee (1973) are:

1. **Excessive Comprehensive Modelling**
   Excessive expectations from the models have resulted with abnormal scopes. Models tried to simulate huge and complicated systems at once and results were expected to answer problems from distinct fields. Evolving computers and algorithms have a big share on these big expectations. Every additional

layer increase uncertainty and it was realized that our understanding about urbanization doesn't evolve as we stack different data layers on a model.

2. **Rough Modelling Assumptions**

   While the models need huge datasets to provide highly detailed information, decision makers need much coarser data to support decision process. For example, when planning a trade centre, planners don't need to use a land-use model. These kind of changes can be done with less effort and much less data. Models should have been very specific about what they are modelling.

3. **Glutton Data Models**

   Regardless of what was modelled, any realistic model needed large data sets. Collecting these datasets for each application could be very time consuming. This can be avoided if the scope of the model is narrowed down. To narrow the scope, the results sought should be determined from the beginning.

4. **Complexity of Modelling**

   Since modelling has many variables and internal connections between the variables, it has a complex structure. Number of possible interactions between terms increases by square of the term quantity. For example from this formula; $n * (n - 1)$, if we have twenty variables, there are 380 possible interactions if bilateral interaction is allowed. Predefining which variables can interact with each other can reduce the number of interactions between terms, even so relations must be aggregated to understand the effects of the relations.

5. **Insisting on False Assumptions About Model Behaviour**

   The behavioural estimation of the model could be misleading, even for the ones who designed the model. Unforeseen relations between variables may alter the results. These relations aren't explicit and hard to define.

6. **Extreme Dependence on Computers**

   Creating models using huge datasets can be accomplished only by using computers. Back in 60's programming wasn't a widespread practice as it is today. Writing a program or debugging was a real deal. Besides difficulties of programming, calculation reliability of the computers was not accurately known. Numerical errors produced by algebraic and repeating operations such matrix

inversion must be handled. But desperately, computers that day couldn't even approach single-digit accuracy.

The problems listed above are not specific to urban growth modelling but modelling problems in general. First decade of urban growth modelling started with optimism and grand expectations. As the models tested and results begin to emerge, expectations are lowered, and fundamentals of modelling began to be questioned. Researchers have realized that size and complexity of urban model have profound influence on knowledge gained from the model. Although there aren't many clear limits, the knowledge gained from the model is transformed from information about urbanization structure to information about large model information as the size and complexity of the system increase (Lee, 1973). CATS (Chicago Area Transportation Study) in 1960 and Detroit Metropolitan Area Transportation Study in 1955 are examples of early urban models.

2.2.2 New Generation Modelling Approaches

The first models aimed to model specific economic activities by displaying narrowly comprehensive approaches. One of the popular approaches was central place theorem which fundamentally assumes each household travels to closest centre to satisfy their needs. This approach has been abandoned in the new generations, as it is possible to apply very few sectors and has simple assumptions about human behaviour. As urban growth modelling evolves as a research area, different approach ideas were put forward.

The maximum entropy approach began to be used in transportation and urbanization models since the 1970's. Entropy maximizing systems proved some entities about urbanization, which are thought to be separated, were related (Berling-Wolff and Wu, 2004). As a result; entropy maximizing systems have undergone internal fractures. But characteristics of the new models were shaped by two things; lessons learned from the failure of the first systems and computer revolution. Modularity were kept in the forefront and some algebraic problems that thought to be unsolvable or unpractical to solve was solvable due to increased computing power. Examples of new generation

models are FUTURES (Meentemeyer et al., 2013) and GeoSOS-FLUS (Liu et al., 2017). These models use the stochastic patch-growing algorithm and an improved cellular automata allocation, respectively.

2.3 Urban Growth Model SLEUTH

Urban growth model SLEUTH was first published in 1996 by K.C Clarke (Clarke et al., 1996). Principles of this model were based on Clarke's previous work about wildfires. The model introduced in this very first paper was relied on cellular automata and Monte Carlo simulations. Purpose of this paper was to see if it is possible to predict urban growth with a self-modifying cellular automata model. SLEUTH works in growth cycles and in every cycle, it assigns statistical values to each pixel. Cycles are years in real life. Calculation of these statistical values are triggered by initial values. These values are commonly called coefficients. SLEUTH coefficients are; Slope, Land-cover, Exclusion, Urbanization, Transformation and Hillshade. Then five factors control the behaviour of the system;

1. Diffusion (Factor of overall outward dispersiveness)
2. Breed (This factor specifies if a new growth pixel will start its own growth cycle.)
3. Spread (Controls how much organic expansion occurs from existing settlements.)
4. Slope (A resistance factor, influences settlements growth through steeper environment.)
5. Road Gravity (A factor that attracts new settlements along the roads.)

In the demo scenario file, which can be found inside the SLEUTH model files, self-modification parameters are written. The values of these parameters are found after long studies as Clarke has stated in his SLEUTH video (Clarke, 2016).

SLEUTH model have been calibrated for vast range of cities and accuracy assessments have been done. It accepts images in GIF format and outputs are also in the

same format. Clarke states they've selected GIF as their main input format because it is easier to read and write. In future, different formats are planned to be added such as GEOTIFF and PNG.

Because of its popularity (Li and Gong, 2016) and reliability (Berberoglu et al., 2016) when compared to the other models, SLEUTH model was applied to almost all major cities of the United States as well as some of the European, African and Asian cities. Applications had different purposes such as determining the likelihood of urban growth around the waste disposal sites and investigating effects of different urban growth scenarios. Different methodologies were developed to understand outcomes of SLEUTH model. For example Sahana et al. (2018) have used an urban sprawl matrix method to understand urban sprawl in the sub-regions. Saxena and Jat (2019) have modified the self-modification parameters of the SLEUTH in order to successfully capture heterogeneous urban growth in the Pushkar Town in India. They've successfully modified the model and stated that self-modification parameters play more important role in the areas which urbanization is fragmented. Critical low and critical high of self-modifying parameters are adjusted accordingly and calibration has been run. It was understood that cities with different forms of urbanization should have different self-modification parameters. Agyemang and Silva (2019) have tested SLEUTH model's sensitivity against a very rapid and spontaneously urbanizing Accra city-region in Ghana. Their study suggested SLEUTH model can be used as decision tool even in overwhelmingly irregular urbanizing areas. In another attempt to predict future urbanization and modelling land-cover change in İstanbul, Yakup et al. (2018) have used SLEUTH model and found if precautions are not taken, all green areas near Sancaktepe district will become urban by 2070. SLEUTH model was also used by Nucci et al. (2016) to study urbanization in Mashhad city after Islamic Revolution of Iran. Findings in this study suggest that Mashhad is facing rapid urbanization and proposed solutions for the possible urbanization challenges may be experienced in the future.

2.3.1 Cellular Automata

Original idea of the cellular automaton dates back to 1940's. Two scientist of Los Alamos National Laboratory, Stanislaw Ulam and John von Neumann have laid the foundations (Berling-Wolff and Wu, 2004). Self-replication problems have driven scientist search about cellular automata. In 1970 Martin Gardner, an American mathematician and science writer has published John Conway's studies about cellular automata in a popular science magazine. Two-dimensional cellular automaton become widely known in early computer society after this publication. Generally cellular automata have similar rules. For example; a simple two-dimensional cellular automaton consists of four primary components;

1. **State**
   The state of the cell at a given time. For example, if the system is replicating urbanization, states of the cells could be; urban and non-urban.
2. **Cells**
   Keeps the track of spatial state of states. They can be thought as scenes of satellite imagery.
3. **Neighbourhoods**
   Neighbourhoods define spatial relation between cells.
4. **Transition Rules**
   Transition rules define the characteristic of replicated system. For instance; well-known two-state cellular automata, Conway's Game of Life has the following transition rules;
   - Any live cell with fewer than two live neighbours dies. (Underpopulation)
   - Any live cell with two or three live neighbours lives on to the next generation.
   - Any live cell with more than three living neighbours are kept.
   - Any dead cell with exactly three live neighbours becomes a live cell.

An example of how neighbours and states are defined in cellular automata can be seen in Figure 2.2.

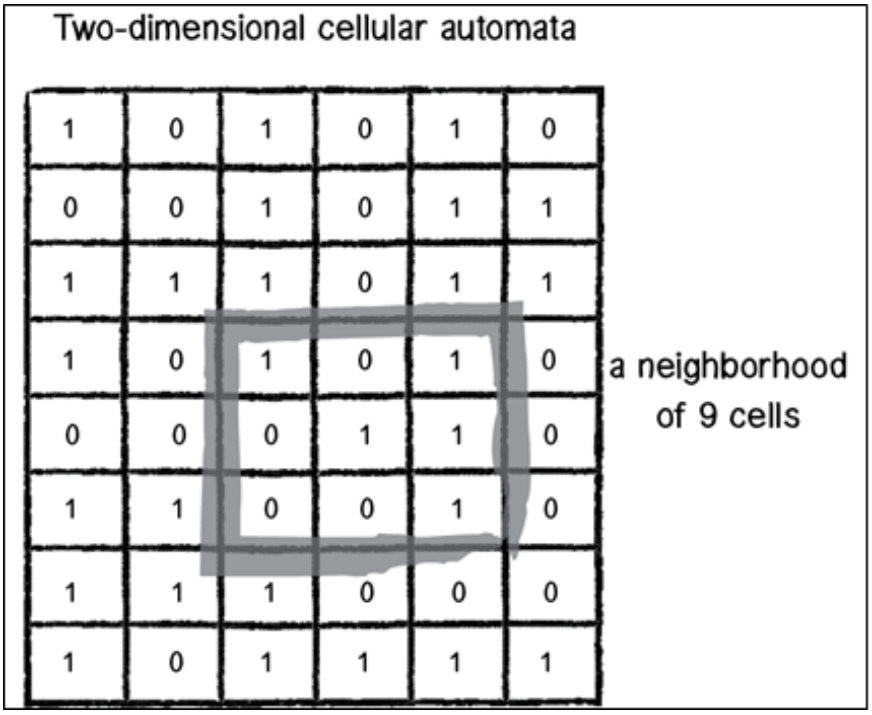

Figure 2.2: Simple 2-D cellular automata (Schiffman, 2019)

Cellular Automaton based on iterative Monte Carlo simulations forms the main framework of SLEUTH. Stephen Wolfram stated on cellular automata; "This universality implies that many details of the construction of a cellular automaton are irrelevant in determining its quantitative behaviour. Thus, complex physical and biological systems may lie in the same universality classes as the idealized models provided by cellular automata. Knowledge of cellular automaton behaviour may then yield rather general results on the behaviour of complex natural systems" (Wolfram, 1984). Complex phenomena could have similar evolving patterns as simple phenomena and cellular automata could help us to understand their behaviour.

2.3.2 SLEUTH model coefficients

Applications cover many continents and areas with different geographical structure, also different satellite imagery with various spatial resolution have been used by researchers such as Sevik (2006), Silva and Clarke (2002) and Berberoglu et al. (2016). These studies have helped researchers to answer some fundamental questions about SLEUTH model. One mind-stirring question is how the input satellite imagery's reso-

lution affects the simulation accuracy and Silva and Clarke (2002) have answered this question in their research. Their conclusion was that the results with high resolution imagery did not affect the simulation accuracy. Since urbanization is a phenomenon affecting large areas, trying to catch small changes did not contribute much to the model (Dietzel and Clarke, 2004). Another important question is about the performance of the cellular automata approach. This question was answered by Berberoglu et al. (2016) by comparing different modelling approaches such as artificial neural networks, regression tree, markov chain and the SLEUTH model. His findings show SLEUTH model performs best across all urban modelling approaches. Another complicated question is assessment of SLEUTH's metrics that are used to measure goodness of fit. These metrics are;

1. **Product**
   Multiplication of all scores.
2. **Compare**
   Modelled population for final year/actual population for final year.
3. **Population**
   Least squares regression score for modelled urbanization compared to actual urbanization for the control years
4. **Edges**
   Least square regression score for modelled urban edge count compared to actual urban edge count for the control years.
5. **Clusters**
   Clusters Least squares regression score for modelled urban clustering compared to known urban clustering for the control years
6. **Cluster Size**
   Least squares regression score for modelled average urban cluster size compared to known average urban cluster size for the control years
7. **Lee-Salee**
   A shape index, a measurement of spatial fit between the model's growth and

the known urban extent for the control years

8. **Slope**
   Least squares regression of average slope for modelled urbanized cells compared to average slope of known urban cells for the control years

9. **% Urban**
   Least squares regression of percent of available pixels urbanized compared to the urbanized pixels for the control years

10. **X-mean**
    Least squares regression of average x_values for modelled urbanized cells compared to average x_values of known urban cells for the control years

11. **Y-mean**
    Least squares regression of average y_values for modeled urbanized cells compared to average y_values of known urban cells for the control years

12. **Rad**
    Least squares regression of standard radius of the urban distribution, i.e. normalized standard deviation in x and y

13. **F-Match**
    A proportion of goodness of fit across landuse classes.(Only if land-use and land cover is being modelled)

For each calibration run, SLEUTH calibrates these metrics according to the input data. On each run calibration parameter range gets narrower. Also, the parameters of the prior calibration run are reviewed and parameters of the next run are determined except the coarse calibration. Researchers have used different parameters on the above to determine best fit values for the calibration (Dietzel and Clarke, 2007). These different pre-acceptances prevented the formation of a consensus about assessment of the SLEUTH metrics. Dietzel and Clarke (2007) have eliminated the confusion by creating a metric specified for the SLEUTH by testing the effect of each metric. Growth rules of SLEUTH are explained in the equations below:

1. **Spontaneous Growth Rule**

   Determines possibility of random urbanization of the cells in cellular automata framework. Spontaneous growth is a function of dispersion and slope coefficients. Formulation of the rule can be seen in Equation 2.1. A random number is used in spontaneous growth function so that stochasticity of the process is ensured.

$$U(i,j,t+1) = F(\text{dispersion coef.}, \text{slope coef.}, U(i,j,t), \text{random}) \quad (2.1)$$

2. **New Spreading Centre Growth Rule**

   Determines newly urbanized pixel's possibility of becoming new spreading centre. Breed and slope coefficients define the probability of each new urbanized pixel to become new spreading centre. At least two neighbouring pixels should be available for urbanization for this process. Slope coefficient is used to determine if adjacent pixels are available for urbanization. A random number is used in this function too, so that stochasticity of the process is ensured. This rule doesn't affect excluded areas and already urbanized cells.

3. **Edge Growth Rule**

   Edge growth determines urbanization near the growing centres. Spread and slope coefficients are variables of edge growth function. Formulation of this rule can be seen in Equation 2.2. In the Equation 2.2, $U(k,l)$ defines neighbour pixels.

$$U(i,j,t+1) = F(\text{spread coef.}, \text{slope coef.}, U(i,j,t), U(k,l), \text{random}) \quad (2.2)$$

4. **Road Influenced Growth Rule**

   Road influenced growth is determined by the existing and future road layers and breed, road gravity, dispersion and slope coefficients. If a road pixel is found in vicinity of the selected pixel, a temporary urban cell is placed. Vicinity of the pixels are determined by road gravity coefficient. Later this temporary pixel is randomly selecting new pixels along the road, this process is called road trip. Through the walk number of steps are conducted and at the last step

temporary pixel becomes permanently urban. In Equation 2.3 R(m,n) is a road pixel, U(k,l) and U(i,j) are urban pixels.

$$U\prime(k,l,t+1) = F(U(i,j,t+1), \text{road gravity coef.}, R(m,n), random) \quad (2.3)$$

At the end of the random walk, new urban spreading centre can be defined by the formula in Equation 2.4;

$$U\prime\prime\prime(i,j,t+1) = F(U\prime(p,q,t+1), R(m,n), \text{slope coef.}, \text{random}) \quad (2.4)$$

Two neighbouring cells can be urbanized with the function in Equation 2.5.

$$U\prime\prime\prime\prime(i,j,t+1) = F(U\prime\prime\prime(p,q,t+1), U(k,l,t+1), \text{slope coef.}, \text{random} \quad (2.5)$$

2.3.3 Self-modifying Cellular Automata

Self-modifying is a distinctive feature of SLEUTH. It refers to preventing uncontrolled exponential growth as system increases in overall size. System modifies itself to the circumstances it generates (Clarke et al., 1997). SLEUTH decreases multiplier factors every subsequent growth cycle (Clarke et al., 1996). In SLEUTH, subsequent growth cycles are years. SLEUTH takes all the main factors that affect urbanization as input and provides scientifically comparable predictions about the future urbanization. Coefficient values produced by SLEUTH simulations are considered as DNA of the regions (Clarke et al., 2007). Just like in a laboratory environment, urbanization of different cities can be examined and the comparisons between them can be made. How urbanization can be different within the framework of different rules or effect of different environmental factors can be predicted by SLEUTH simulations.

CHAPTER 3

MATERIALS AND METHODOLOGY

3.1 Study Area Definition

Resting on the western end of the fertile crescent, Adana has been a very charming place for agricultural activities through the history. Geography, soil type and abundance of wetlands are main reasons for this unique feature. Today Adana still has the most fertile lands in Turkey but also facing with immigration and rapid urbanization in the last few decades. Figure 3.1 shows the study area and Figure 3.2 gives insights about Adana's population change through years. Turkish Statistical Institute (TÜİK) publishes historical data about population of provinces, but urban and rural populations can be found separately only before the year 2000. The data after year 2000 include only the total population. Figure 3.2 shows that Adana is not exception about what is happening on global scale; urban population is on the rise. Study region defined with a rectangular box can be seen in Figure 3.3.

Although the study area does not completely cover the provincial boundaries, it includes about 80% of the population of Adana. Urbanization is clustered in the southern part where the highway and agricultural areas exist.

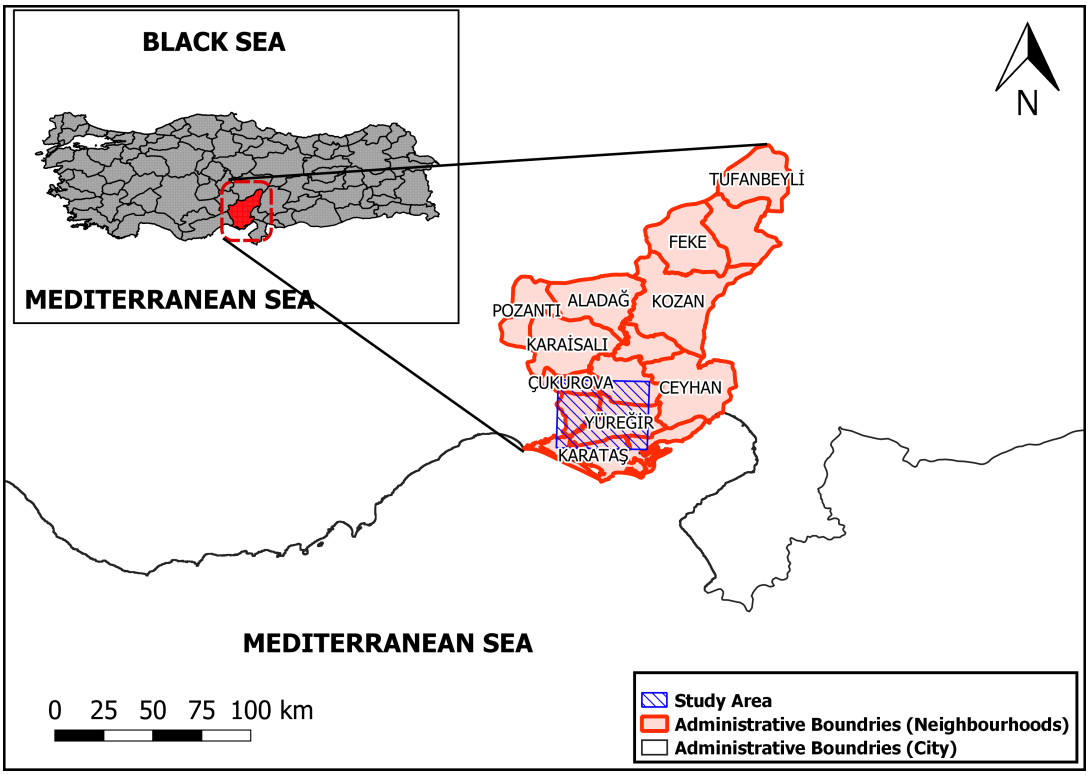

Figure 3.1: Adana administrative boundaries with the study area

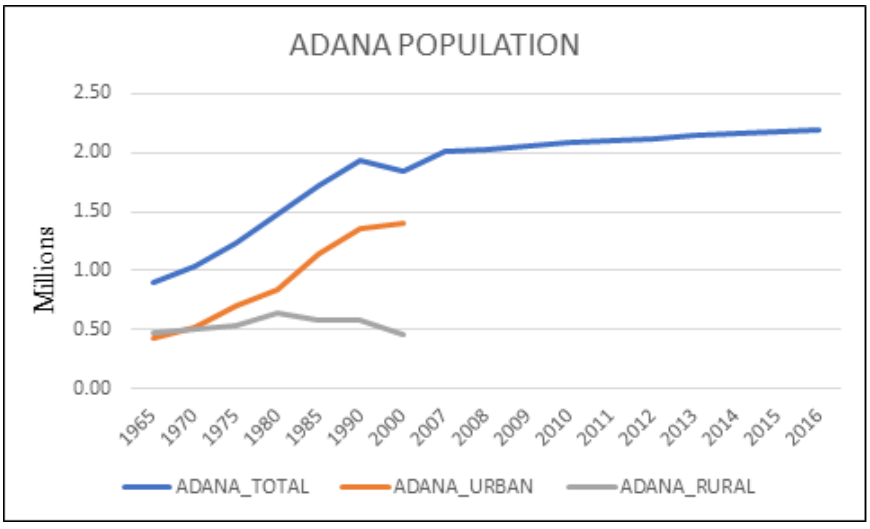

Figure 3.2: Adana population change (TUIK, 2018)

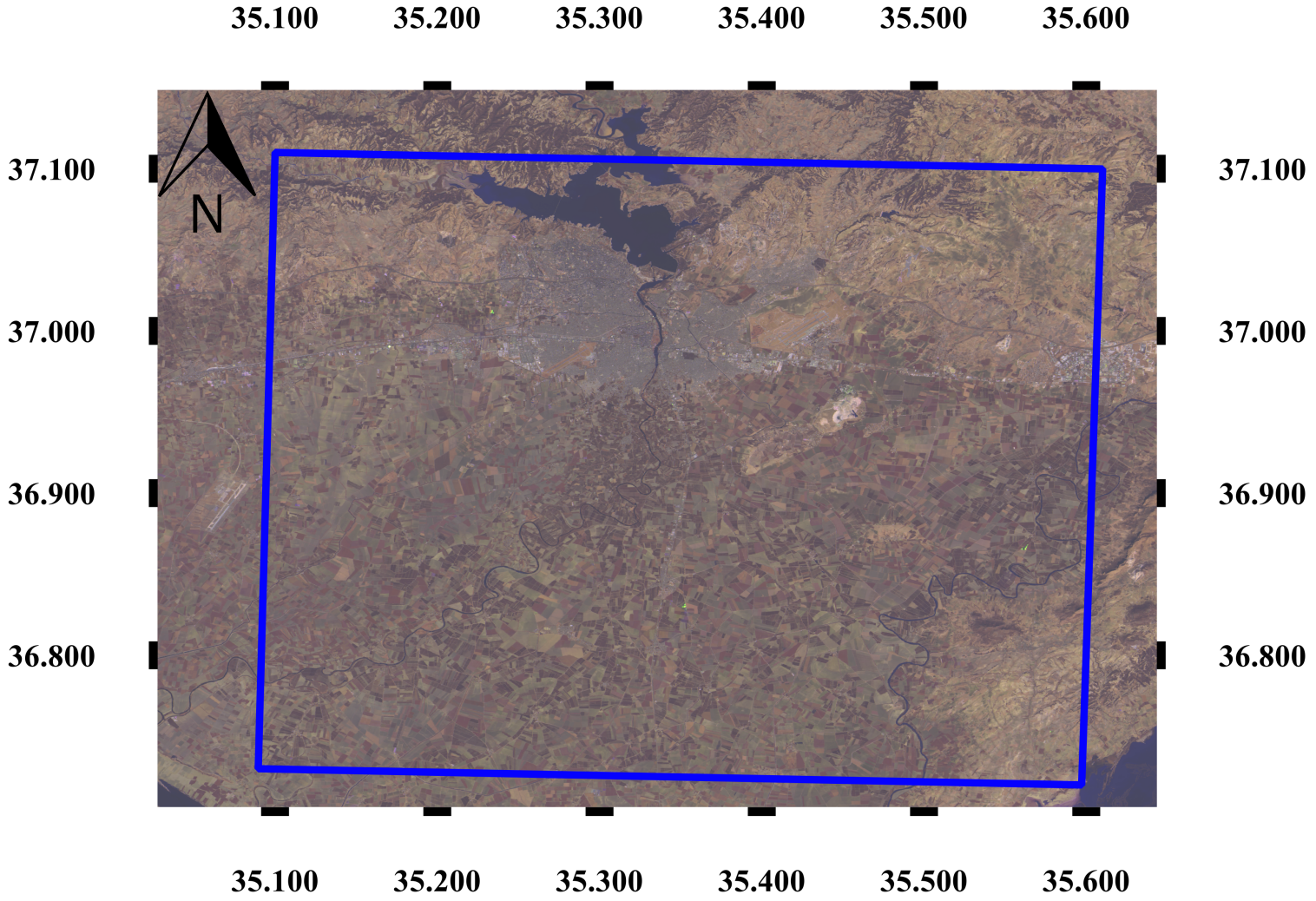

Figure 3.3: Study area shown on Landsat imagery (Satellite image acquisition date is 27-Sep-2018)

While selecting study area, city centre is positioned in the middle of the study area. Most of the agricultural areas which are spread through the south part of the city have been included too. Northern side of study area consists of mostly dam lake and mountains meaning urbanization is naturally limited in this part.

3.2 Methodology

In this section, data sources used in this study and methodologies used to extract information from the input data are explained. Flowchart of SLEUTH application can be seen in Figure 3.4.

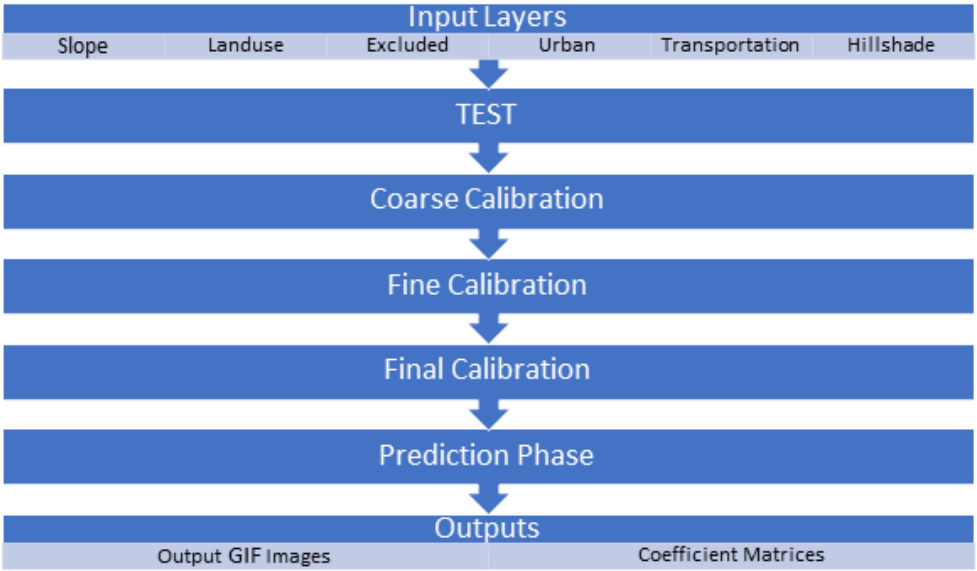

Figure 3.4: Flowchart of SLEUTH

3.2.1 Data Sources Used in the Study

I. **Satellite Imagery**

In this study Landsat images were used to obtain land cover information of the study area. Landsat has wide usage by urban researchers and data library includes 30 years of satellite imagery. Practical data access making it superior to other non-commercial earth observation satellites. Landsat imagery has 30 meters spatial resolution in colour bands and 15 meter spatial resolution in panchro-

matic band. Since urbanization/imperviousness phenomena affects wide areas as discussed in Section 2.3, its spatial resolution is found sufficient for this study.

Landsat images on Worldwide Reference System path number 175 and row number 34 are downloaded from USGS Earth Explorer web site. Landsat 5 (29-Aug 1990, 24-July 2006, 20-June 2011, 20-Aug 2016) and Landsat 7 (18-July 2001) data acquired in summer time are used to create urban growth model input data in this study. Also, a Landsat 8 (27-Sep 2018) imagery has been used as validation data for the model predictions. Time-line of the spatio-temporal data used in this study can be seen in Figure 3.5.

II. **Approved Plans of Adana**

In Turkey planning process of cities and regions follows a procedure. Firstly 1/25.000 Landscape plans are prepared, then based on this plan 1/5.000 or 1/2.000 scale plans are prepared and finally 1/1.000 scale development plans are prepared. As the scale grows, plans become more detailed. Plans are prepared in separated sections. Each section can be prepared by a different company and thus causes notation differences. Using plan data directly in a GIS software is nearly impossible. Layer aggregation is a solution for the multiple layers and different notations in plans. Aggregated version of 1/5.000 scale plan can be seen in Figure 3.6. SLEUTH's Excluded areas are obtained from these urban plans.

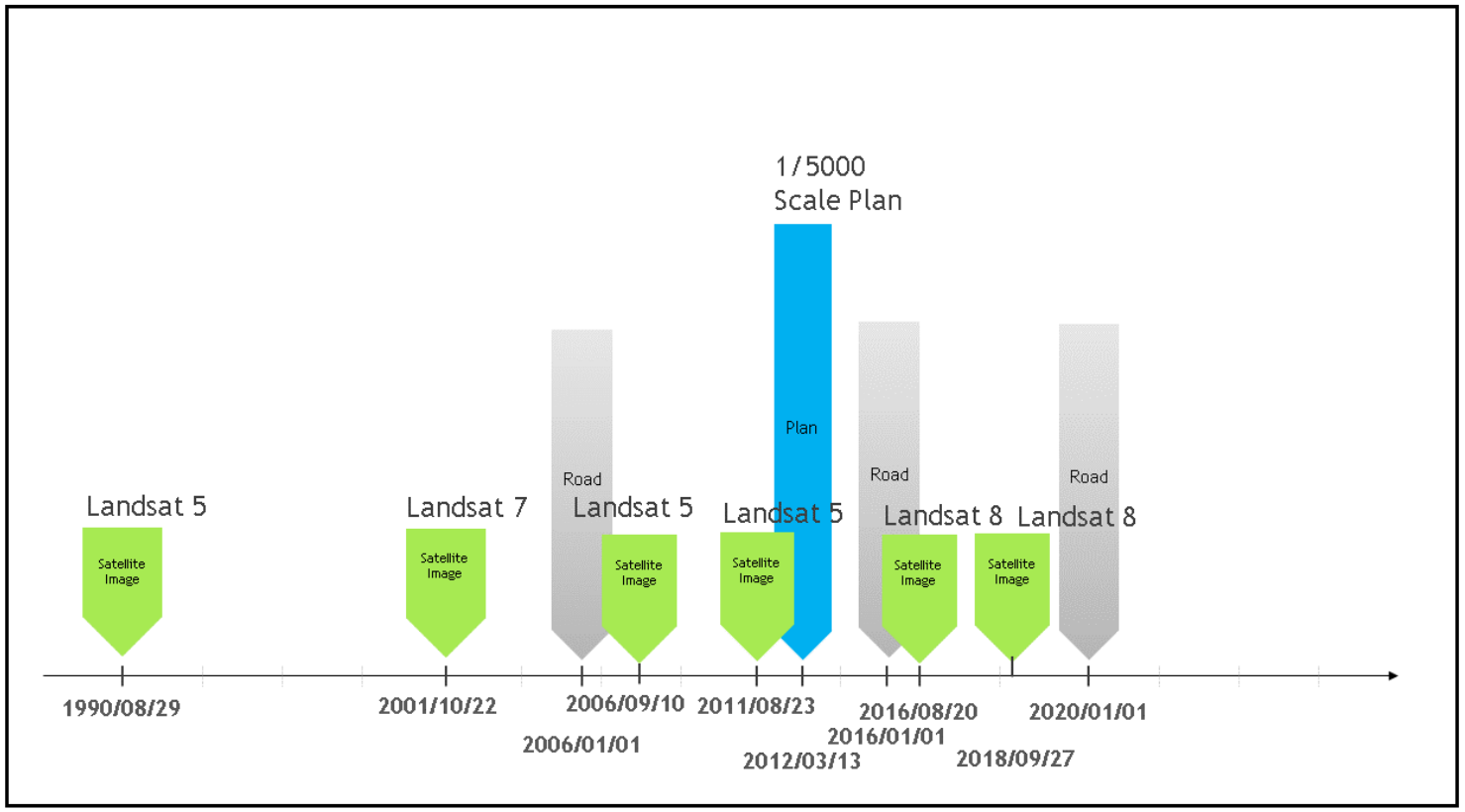

Figure 3.5: Timeline of spatio-temporal data sources

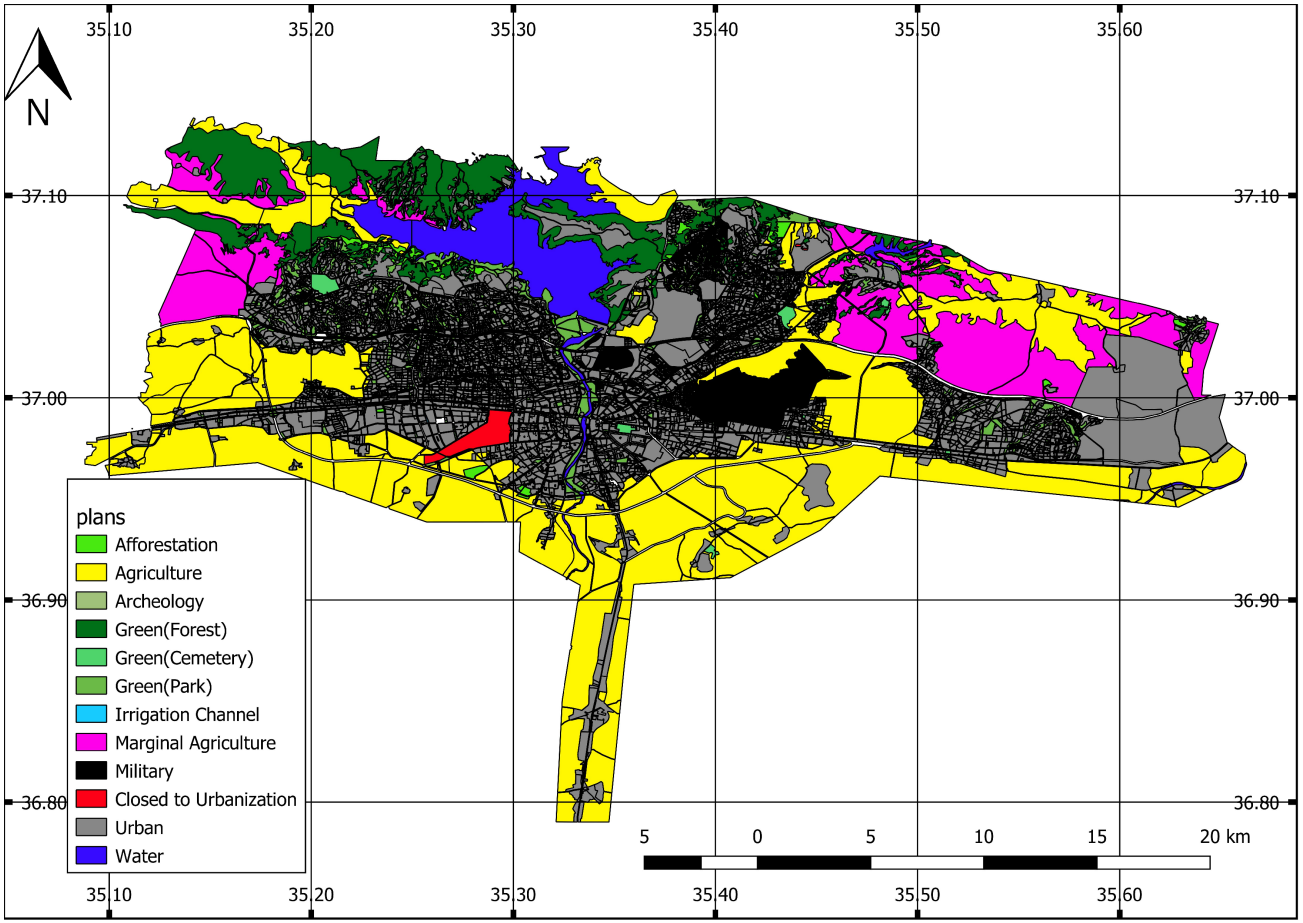

Figure 3.6: Aggregated 1/5000 scale plan of Adana in 2016

III. **Digital Elevation Model**

Digital Elevation Model (DEM) data was downloaded from USGS's web-site. Resolution of this data is 1 arc second i.e., roughly 30 meters. The DEM is produced from downloaded SRTM data using ArcGIS. It can be seen in Figure 3.7. Hillshade and slope layers are calculated from this DEM data.

IV. **Road Data**

Road data are digitized from master plans. If road layer is prepared correctly, SLEUTH can handle distinct types of roads. 1/1000 scale plans contain road width information and road widths vary from 5 meter to 100 meter. Only the roads which are 10 meter or broader are digitized because smaller roads are do not have a significant effect on urbanization. Small roads usually can be seen in already urbanized areas. Road data can be seen in Figure 3.7.

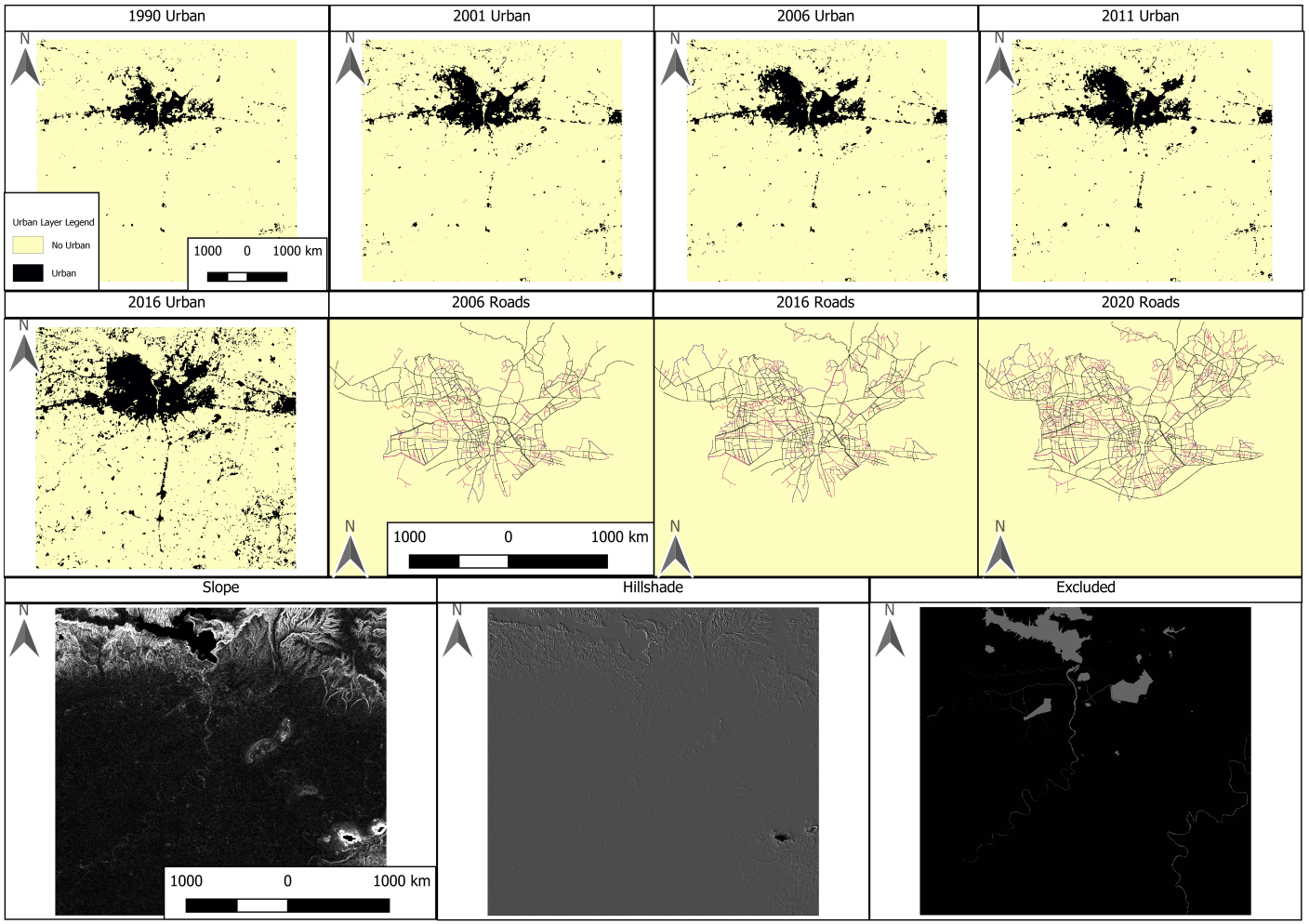

Figure 3.7: Input Data Layers of SLEUTH (First row: Urban layers from 1990 to 2011, second row: 2016 urban layer and road layers from 2006 to 2020, third row: slope, hillshade and excluded area layer)

3.2.2 Processing Steps of Satellite Imagery

Satellite imagery used in this study needs to be processed before being used in the SLEUTH model. Basically, all imagery have been clipped according to the study area extent and digital numbers were transformed to top of atmosphere reflectance values in order to have comparable values among years. After the transformation, the imagery were classified using both supervised and unsupervised classification. Finally, accuracy assessment has been applied to the classified images.

I. **Transformation of Digital Numbers to Top of Atmosphere Reflectance**
Image files downloaded from USGS web site are already processed in various levels. This process includes corrections such as; geographic correction, radiometric correction and inter-calibration of other Landsat instruments. Corrected data subjected to a quality test and classified as Tier1(T1), Tier2(T2) or Real-Time(RT). Tier1 is the available data that having the highest data quality. Scenes can't meet quality of Tier1 are classified as Tier2. Scenes newly acquired and processed with predicted values are classified as RealTime data. The highest quality data level is also divided within itself into six groups. Figure 3.8 shows the classification of Landsat data according to their quality.

Landsat Level-1 Processing Levels		
Pre-Collection	**Collection 1**	**Description**
L1T	**L1TP**	Radiometrically calibrated and orthorectified using ground control points and digital elevation model (DEM) data to correct for relief displacement. These are the highest quality Level-1 products suitable for pixel-level time series analysis.
L1GT	**L1GT**	Radiometrically calibrated and with systematic geometric corrections applied using the spacecraft ephemeris data and DEM data to correct for relief displacement.
L1G	**L1GS**	Radiometrically calibrated and with only systematic geometric corrections applied using the spacecraft ephemeris data.

Figure 3.8: Landsat Level-1 product table (USGS, 2018a)

L1T and L1TP are the top-class processing levels in Tier1. All image files which are used in this study are L1T class. Each Landsat data distributed through the

USGS web site is a compressed file which contains each individual band as an image file and a metadata file in the text format. Images are expressed as digital numbers. These numbers may differ from satellite to satellite due to radiometric characteristic of the sensors. Angle of the Sun is another factor that affects these numbers. The effect of the angle of the Sun can be investigated by comparing the pixel value of the same pixel in the imagery acquired in different times with the same satellite sensors. When looking at these images by stacking RGB bands on top of each other by an image processing software, such as ENVI, scenes look normal but when the pixel values investigated, they have high values (in thousands). Even the pixels representing the same area in different scenes have very different DN values provided that the areas do not change. Different DN values in different imagery may cause classification problems.

Table 3.1 shows the cursor location values of three Landsat images. First two data are from Landsat OLI 8 sensor, the last one is from Landsat 5 TM sensor. Digital number transformation to top of atmosphere reflectance transformation for Landsat 5 and Landsat 7 imagery can be done automatically through an ENVI module called ENVI Landsat Calibration. But Landsat 8 imagery can't be automatically transformed. However, manual transformation can be done by using the following transformation methodology given in web site of USGS (USGS, 2018b).

Table 3.1: Comparison of DN values of the same pixel

Acquistion Year	Sensor	Red Band	Green Band	Blue Band
2018	Landsat 8 OLI	8105	8591	9037
2016	Landsat 8 OLI	8391	9651	10479
2006	Landsat 5 TM	22	31	81

It is possible to find each variable from the metadata file of the downloaded Landsat scene. Transformation methodology for TM and ETM sensors used in Landsat 5 and Landsat 7 is available in Yale University's Centre of Earth Observation web site (Yale University, Center for Earth Observation, 2018) and

given in Equation 3.1.

$$\rho'_\lambda = \mathbf{M}_\rho \mathbf{Q}_{cal} + \mathbf{A}_\rho \tag{3.1}$$

Definition of the variables in the equation 3.1 are as follows;

- ρ'_λ is TOA planetary reflectance, without correction for solar angle.
- $\mathbf{M}_\rho$ is band-specific multiplicative rescaling factor from the metadata
- $\mathbf{Q}_{cal}$ is quantized and calibrated standard product pixel values (DN).
- $\mathbf{A}_\rho$ is band-specific additive rescaling factor from the metadata.

Landsat 7 and Landsat 5 data are automatically converted by using metadata file. Landsat 8 image is converted manually using formula in USGS's web site (USGS, 2018a).

Before the transformation, value of the same pixel in five images can be seen in Table 3.2.

Table 3.2: Value of a pixel in different images before calibration

Acquistion Year	Sensor	Red Band	Green Band	Blue Band
1990	Landsat 5 TM	17	30	81
2001	Landsat 5 TM	42	60	82
2006	Landsat 5 TM	23	30	79
2011	Landsat 7 TM	20	30	79
2016	Landsat 8 OLI	7453	8901	10141

After the transformation, values for the same pixel can be seen in Table 3.3;

Table 3.3: Value of a pixel in different images after calibration

Acquistion Year	Sensor	Red Band	Green Band	Blue Band
1990	Landsat 5 TM	0.057624	0.091639	0.120769
2001	Landsat 5 TM	0.043372	0.077817	0.104778
2006	Landsat 5 TM	0.057345	0.088103	0.118629
2011	Landsat 7 TM	0.0675508	0.113544	0.143870
2016	Landsat 8 OLI	0.042032	0.082162	0.109381

Since a successful calibration has been done, all pixel values are normalized and pixel values in the same location have similar values.

II. **Classification of Satellite Imagery**

Before the classification, all satellite images are clipped according to the study area domain. Study area is defined in ENVI and it is exported as ESRI's shapefile format. There is no ideal classification of land use and land cover, and it is unlikely that one could ever be developed (Anderson, 1976).

All satellite imagery are classified using supervised classification with the maximum likelihood method. Parametric methods such as maximum likelihood classification and unsupervised clustering assume that remote sensing data are normally distributed and require the knowledge about the forms of the underlying class density functions (Duda et al., 2001). Since a priori knowledge about the study area already acquired, training pixels are chosen based on this knowledge. Useful thematic information may be obtained using supervised classification algorithms if the general steps are understood and applied (Beitzel et al., 2005). General steps used to extract thematic land-cover information from digital remote sensing data can be seen in Figure 3.9.

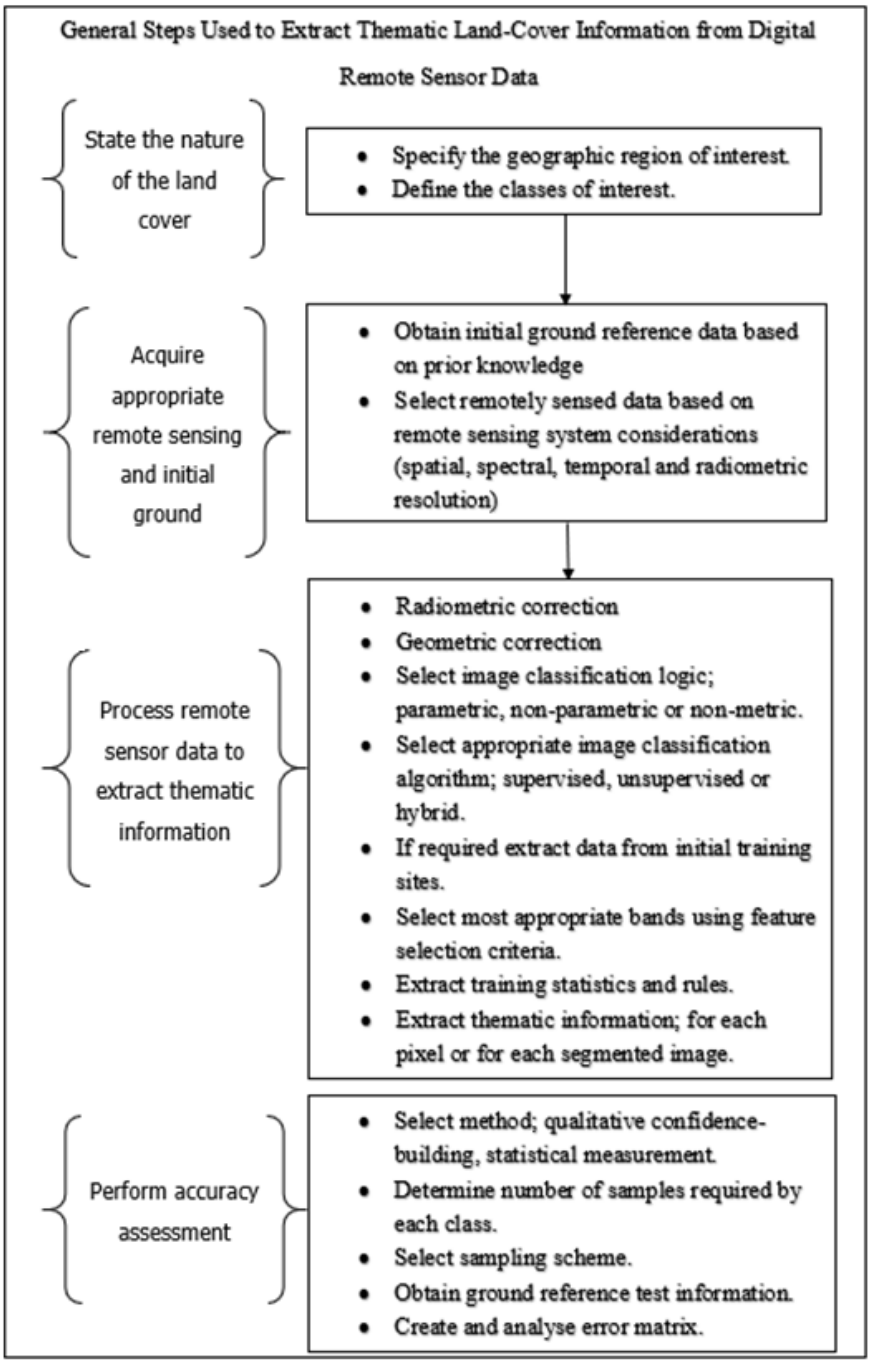

Figure 3.9: General Steps to Extract Thematic Land-Cover Information from Digital Remote Sensor Data (Anderson, 1976)

SLEUTH model can predict only urbanized areas, not the urban classes. Urban classes are needed only when land cover transition in the past are wanted to be modelled. In this study land cover transition are not studied and therefore traditional classification methods are used instead of intensive and complex

classification methods.

The study area is defined according to the planned area of the city. Possible land classes in satellite imagery scene are visually investigated and 11 classes are determined in the scene;

1. Green1: Forest areas especially on high altitudes in the scene.
2. Green2: Coniferous trees especially planted in graveyards.
3. Green3: City parks.
4. Urban1: Dense urbanization areas.
5. Urban2: Roads.
6. Urban3: Industrial areas. These areas are brighter than the residential areas.
7. Bareland1: White calcareous soil.
8. Bareland2: Brown soil.
9. Agriculture: Agricultural areas.
10. Dried Agriculture: Agricultural areas but dried.
11. Ploughed soil: Not green but agricultural soil according to plans and visual interpretation.

Only water class is missing in the above classification list because this class can be more accurately extracted from satellite image by using unsupervised classification. ISODATA classification with 20 classes have been applied to all satellite imagery. Water mask created from this classification is applied to the images on supervised classification.

Results are visually investigated and problematic areas are inspected. If visual inspection seemed to be successful, qualitative confidence has been built, which means statistical measurement can be made. Statistical measurement can be done using GIS programs. Classification accuracy assessment must be done rigorously in order to avoid bias in the results. Accuracy assessment can be done easily with ENVI, but this is a black box operation where user can only give an input classification image and retrieve resulting confusion matrix that ENVI creates. In this study, manual accuracy assessment has been applied to each classification result in Quantum GIS.

For each classification result random samples are created in ENVI within the classification areas. When creating random samples, stratified random sampling option is selected because stratified random sampling is considered as the most efficient sampling way in remote sensing applications (Jensen, 2005). Advantage of the stratified sampling is; even the smallest proportion of classes could have samples because this method divides classified image into strata (classified areas) and collects samples from these areas. If complete random sampling were chosen, random samples may not represent small classification areas.

In manual accuracy assessment multinomial distribution formula has been used. Before collecting the samples, a 7x7 median filter has been applied to the classified image and classification is aggregated according to Anderson's Classification Scheme (Anderson, 1976). This schema can be seen in Figure 3.10. When this schema has been applied to classification, class numbers are reduced to five classes. Each class in classification and city plans has a corresponding class in Anderson Schema except wild life; these areas include both forest and barren land. When the plans are being made both actual land usage and desired land usage are considered. Sample pixels inside the plan areas such as military, no-urban, archaeology and wild-life have been carefully examined. Classification of 2016 imagery and aggregated 2016 plans can be seen in Figure 3.11.

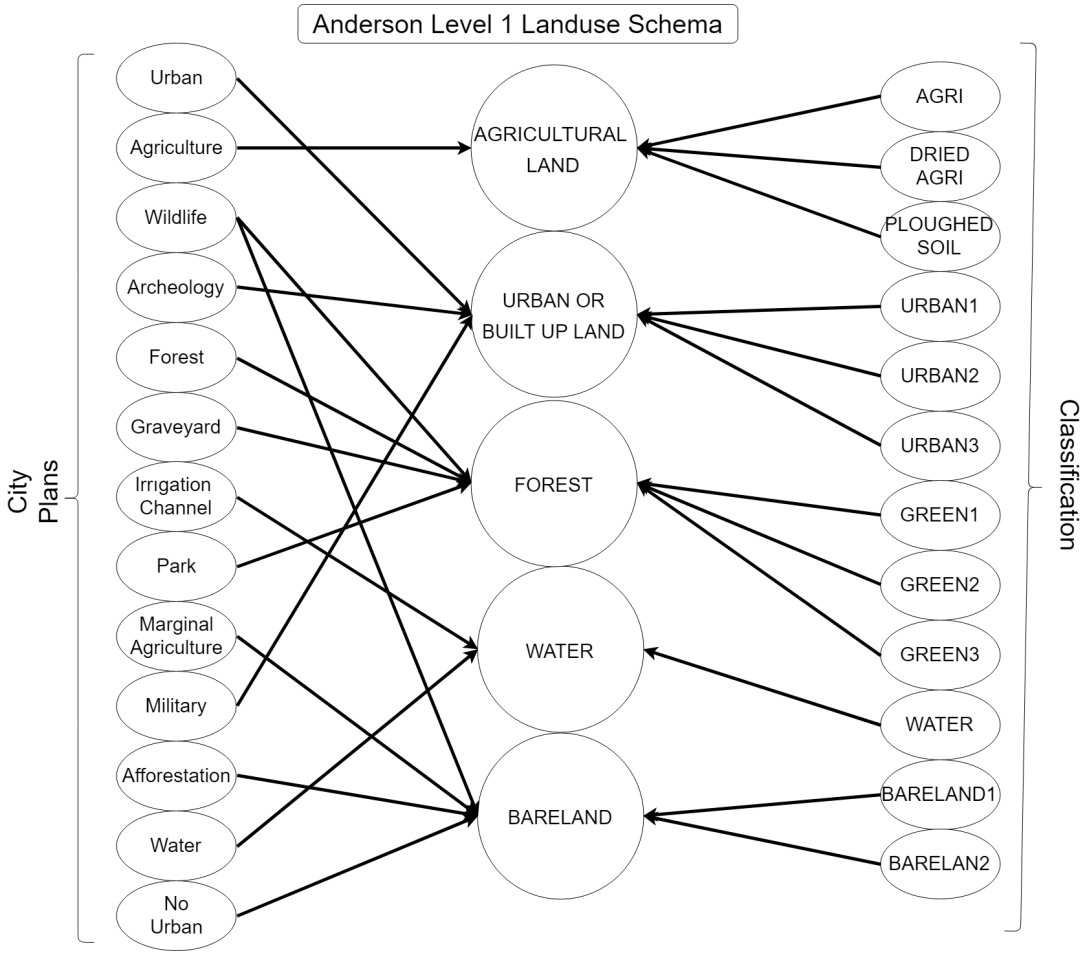

Figure 3.10: City Plans Landuse Classes, Anderson's Classification Schema and Classes Derived From Classification of Landsat Imagery

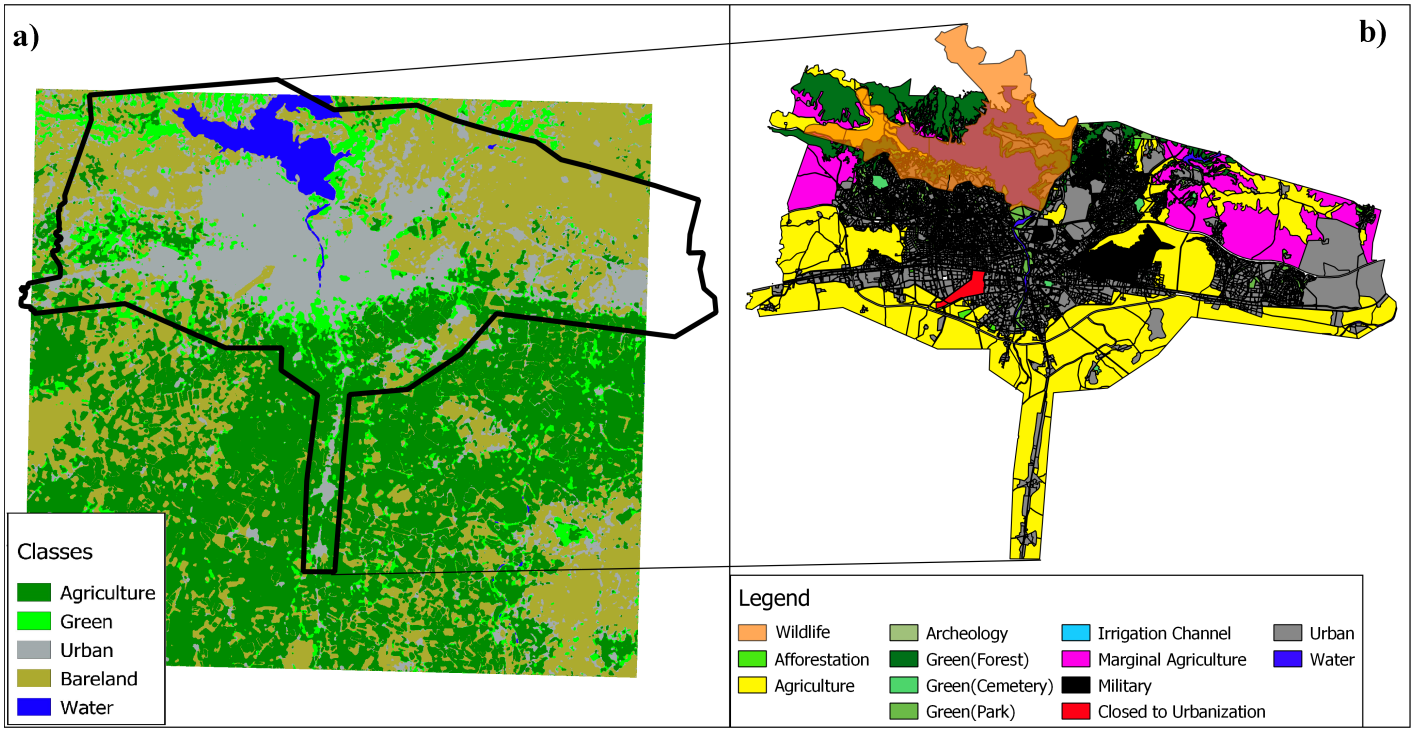

Figure 3.11: 20-Aug 2016 Landsat imagery classification (a) and aggregated urban plans of the year 2016 (b)

Sample size (N) derived from a multinomial distribution is based on equation 3.2;

$$N = \frac{(B * \pi_i * (1 - \pi_i))}{(\mathbf{b}_i)^2} \tag{3.2}$$

- Where π_i is the proportion of a population in the i^{th} class out of K classes that has the proportion closest to 50%. In this study agriculture has the proportion closest to fifty percent of the scene.
- $\mathbf{b}_i$ is the desired precision. (5% in this case)
- B is the chi square distribution. $B = x^2 * (1 - (\frac{\alpha}{k}))$ where α is confidence interval, k is the number of classes.

According to Jensen (2005) 95% confidence interval might be unrealistic. Relaxing the confidence interval to 85% is a standard procedure for many land-use and land-cover mapping products.

According to these assumptions sampling quantity should be as below;

- $1 - (\frac{\alpha}{k}) = 1 - (0.15/6) = 0.975$
- $\pi_i^2(0.975) = 5.02$
- $N = \frac{5.02 * 0.44 * (1 - 0.44)}{(0,05)^2} \approx 494$
- $\frac{494}{5} \approx 100$ samples for each class.

Distribution of sample points and classification result can be seen in Figure 3.12.

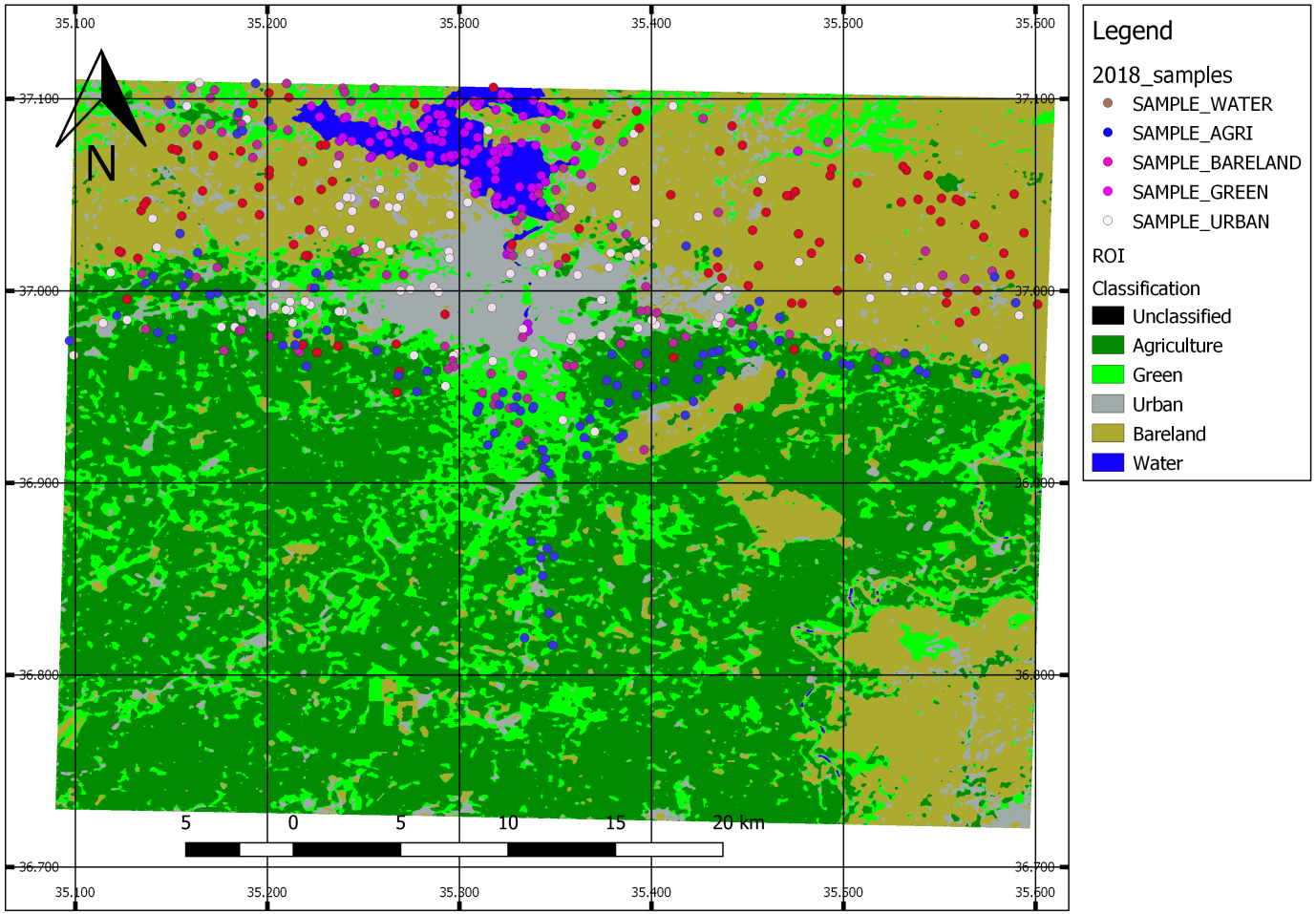

Figure 3.12: Random samples selected on classified image of 20-Aug 2016

In the next step all the sample points and sampling polygons are loaded into PostgreSQL to quantitatively check if the sample points are inside their own sample polygons. This is an important check because if a sampling point is inside the sampling polygon of its own class, the bias can't be avoided. Classification information written to sample polygons using Point Sampling Tool plug-in of Quantum GIS.

Having both random points and their classification results of each sample point, it is possible to prepare the confusion matrix. Confusion matrices compare two sources of information; pixels or polygons in a remote sensing-derived classification map and ground reference test information.

Relationship between these two sets of information is commonly summarized in an error matrix. All values are copied to Microsoft Excel, and the confusion matrix is prepared according to description of Jensen (2005). Plans are accepted as ground truth data and accuracy assessment made without any interference to plans. Detailed information about classification accuracy assessments of the satellite imagery can be found in Appendix A.

Kappa coefficients and overall accuracy results of the confusion matrices are given in Appendix B.

3.2.3 Assessment of the Classification Results

Final products of classifications can be seen in Figure 3.13. Confusion matrices for the classification of the years 1990, 2001, 2006, 2011, 2016 and 2018 can be seen in the Tables 3.4 - 3.9, respectively. In these tables class quantities are shown in percent. When the confusion matrices were examined, there was no systematic error.

Comparison table of classifications can be seen in Table 3.10. According to Anderson (1976), accuracy statistics over 85% are adequate. Accuracy assessment result of the 2016 image is relatively low when compared to the other classifications. Main reason for this is the correctness of the ground truth data gathered from the actual city plans. As it has been discussed in classification section of the 2016 image, plans represent "desired" and "predicted" land uses. Illegal urbanization and bare agricultural areas are the two important factors causing lower classification accuracy.

Table 3.4: Classification Accuracy of the 1990 Imagery (A: Agriculture, B: Bareland, F: Forest, U: Urban, W: Water)

		PREDICTED							
		A	B	F	U	W	Total	%Correct	%Error
REFERENCE	A	18	1.4	0.4	0.2	0	20	90	10
	B	0	18.2	0.4	1.4	0	20	91	9
	F	0	0	19.4	0.6	0	20	97	3
	U	2	0	1.2	16.8	0	20	84	16
	W	0	0	0	0	20	20	100	0
	Total	20	19.6	21.4	19	20	100	92.4	
	%Correct	90	92.9	90.7	88.4	100		Kappa:0.91	
	%Error	10	7.1	9.3	11.6	0			

Table 3.5: Classification accuracy of the 2001 imagery (A: Agriculture, B: Bareland, F: Forest, U: Urban, W: Water)

		PREDICTED							
		A	B	F	U	W	Total	% Correct	% Error
REFERENCE	A	16	3.4	0.2	0.4	0	20	80	20
	B	0	18.6	1.2	0.2	0	20	93	7
	F	0	0.2	19.6	0.2	0	20	98	2
	U	0	0	1	19	0	20	95	5
	W	0	0	0	0	20	20	100	0
	Total	16	22.2	22	19.8	20	100	93.2	
	% Correct	100	83.8	89.1	96.0	100			
	% Error	0	16.2	10.9	4.0	0		Kappa:0.92	

Table 3.6: Classification accuracy of the 2006 imagery (A: Agriculture, B: Bareland, F: Forest, U: Urban, W: Water)

		PREDICTED							
		A	B	F	U	W	Total	% Correct	% Error
REFERENCE	A	15.4	3.8	0.4	0.4	0	20	77	23
	B	0.4	18.4	0.6	0.6	0	20	92	8
	F	0.6	0.2	19.2	0	0	20	96	4
	U	0.8	1	2	16.2	0	20	81	19
	W	0	0	0	0	20	20	100	0
	Total	17.2	23.4	22.2	17.2	20	100	89.2	
	% Correct	89.5	78.6	86.5	94.2	100			
	% Error	10.5	21.4	13.5	5.8	0		Kappa:0.87	

Table 3.7: Classification accuracy of the 2011 imagery (A: Agriculture, B: Bareland, F: Forest, U: Urban, W: Water)

		PREDICTED							
		A	B	F	U	W	Total	% Correct	% Error
REFERENCE	A	15.2	2	2.6	0.2	0	20	76	24
	B	1.8	17	1.2	0	0	20	85	15
	F	0.4	0.4	19.2	0	0	20	96	4
	U	0.2	0.4	0.6	18.8	0	20	94	6
	W	0.2	0	0	0	19.8	20	99	1
	Total	17.8	19.8	23.6	19	19.8	100	90	
	% Correct	85.4	85.9	81.4	98.9	100			
	% Error	14.6	14.1	18.6	1.1	0		Kappa:0.88	

Table 3.8: Classification accuracy of the 2016 imagery (A: Agriculture, B: Bareland, F: Forest, U: Urban, W: Water)

		PREDICTED							
		A	B	F	U	W	Total	% Correct	% Error
REFERENCE	A	19.2	0.2	0	0.6	0	20	96	4
	B	3.2	16	0.8	0	0	20	80	20
	F	5.4	1	10.8	2.2	0.6	20	54	46
	U	2	0.8	1	16.2	0	20	81	19
	W	0.8	0	0.6	0	18.6	20	93	7
	Total	30.6	18	13.2	19	19.2	100	80.8	
	% Correct	62.7	88.9	81.8	85.3	96.9			
	% Error	37.3	11.1	18.2	14.7	3.1		Kappa:0.76	

Table 3.9: Classification accuracy of the 2018 imagery (A: Agriculture, B: Bareland, F: Forest, U: Urban, W: Water)

		PREDICTED							
		A	B	F	U	W	Total	% Correct	% Error
REFERENCE	A	19.4	0.4	0	0.2	0	20	97	3
	B	1.8	17.8	0.2	0.2	0	20	89	11
	F	4	0.8	14.6	0.4	0.2	20	73	27
	U	0.6	0.4	0.8	18.2	0	20	91	9
	W	0	0.6	0.4	0.4	18.6	20	93	7
	Total	25.8	20	16	19.4	18.8	100	88.6	
	% Correct	75.2	89.0	91.3	93.8	98.9			
	% Error	24.8	11.0	8.8	6.2	1.1		Kappa:0.86	

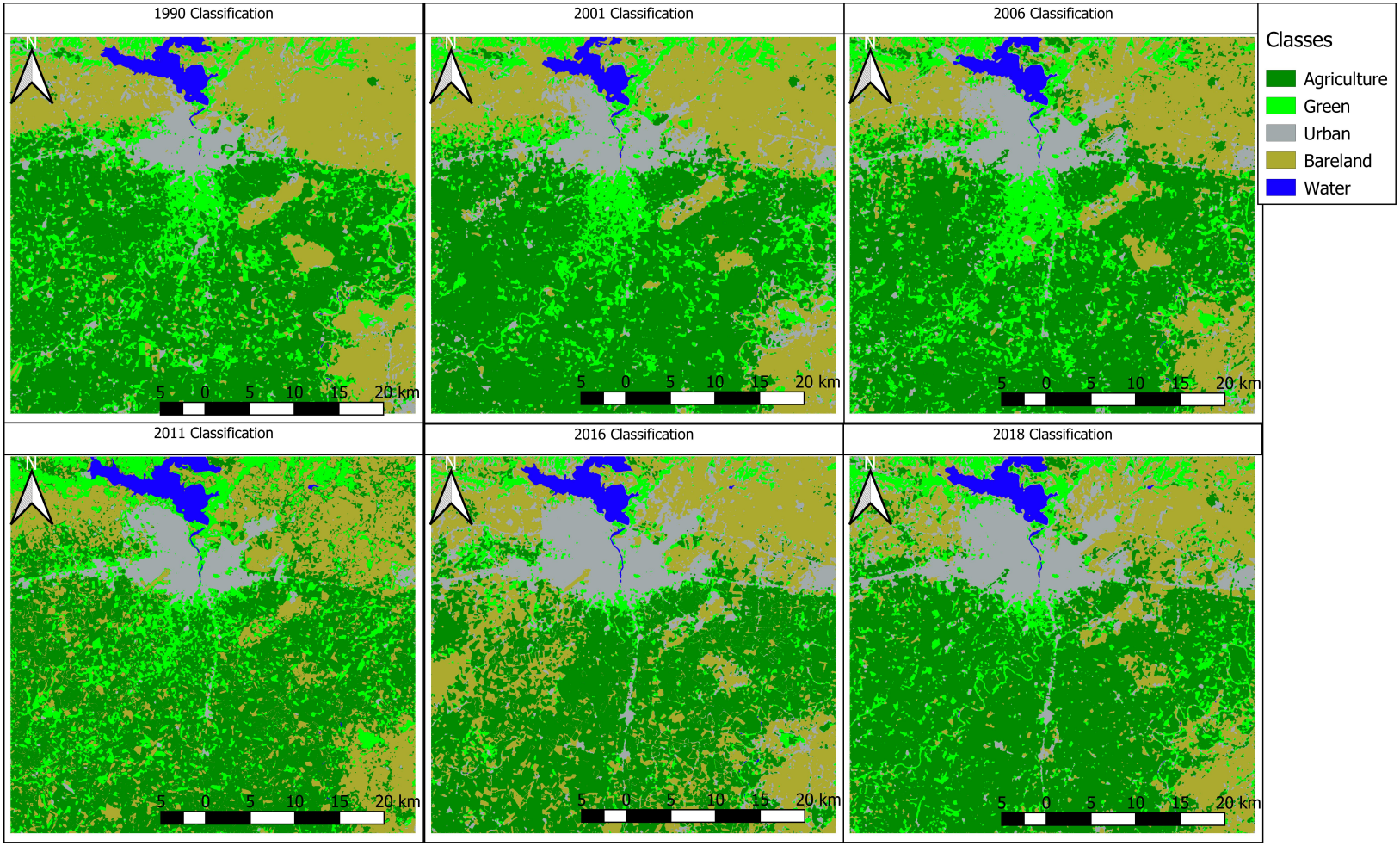

Figure 3.13: Classification results

Table 3.10: Comparison table of classification accuracies

Year	Overall Accuracy	Kappa Coefficient	Satellite
1990	%92	%91	LANDSAT_5
2001	%93	%92	LANDSAT_7
2006	%89	%87	LANDSAT_5
2011	%90	%88	LANDSAT_5
2016	%80	%76	LANDSAT_8
2018	%87	%86	LANDSAT_8

In Table 3.10, 2018 imagery classification result is also presented because additionally Landsat imagery obtained in 2018 is used in the study to validate the model results. Same point locations are used in terms of accuracy for all the satellite images. Prior to the preparation of the SLEUTH input data, three critical issues were highlighted;

- **Gross change of water levels through the years**
  Since there is a large dam reservoir in the scene, water levels change depending on the season. Selected Landsat scenes are close to each other seasonally, so that seasonal colour changes are minimized. Though, water levels are related to drought and water use in the area. Water masks are investigated to find out the one contains the largest water body. Even though visual interpretation shows 2011 image contains largest water body, to verify that, statistics of all water masks are calculated to find which scene contains the largest water body. Quantity of water pixels on each scene can be seen in Table 3.11.

Table 3.11: Quantity of water pixels in each satellite image

Year	Quantity of Water Pixels
1990	41207
2001	31340
2006	36325
2011	49416
2016	43012

According to this comparison, water mask obtained from the classification of the image taken on 20-June 2011 is applied to all the images.

- **Shady pixels over urban areas**

  On the USGS's web site users are allowed to select cloudiness percentage of the imagery. Minimum percentage range can be selected between 0% - 10%. In some cases this option works but it only calculates the percentage cloud coverage of the area not the location of the clouds. If all clouds cumulated over the urban areas even with a cloudiness less then 10%, satellite image scene classification process becomes tangled. In this study, shady pixels over the urban areas are misclassified with water areas when the first satellite imagery data sets were selected. To overcome the problem; the problematic imagery was reclassified to find out cloudy areas but even after successfully finding the shady pixels, assigning these pixels to a class creates another problem. In this perspective a different set of satellite imagery has been chosen for every cloudy scene. While selecting new imagery, images were visually interpreted to avoid cumulated clouds over urban areas.

- **Urban pixels that become non-urban over the next years**

  A cumulative approach was developed to have consistent urban layers. This case considered as a problem because SLEUTH behave abnormally when urban areas become non-urban (Clarke, 2016). Also, possibility of transition from urban to green area is very low especially in developing countries like Turkey, where housing continuously increases. This phenomenon has been discussed in Chapter 3.1. All imagery after 1990 have been compared with previous image (e.g. 2001 compared with 1990, 2006 compared with 2001 etc.). If a pixel is urban at first, then become agriculture, green, bare land or water, this pixel is most likely misclassified and marked as disagreed pixels. Disagreed pixels are considered as non-urban. Agreed pixels of 2001, 2006, 2011, 2016 are found. Flow chart of the method can be seen in Figure 3.14.

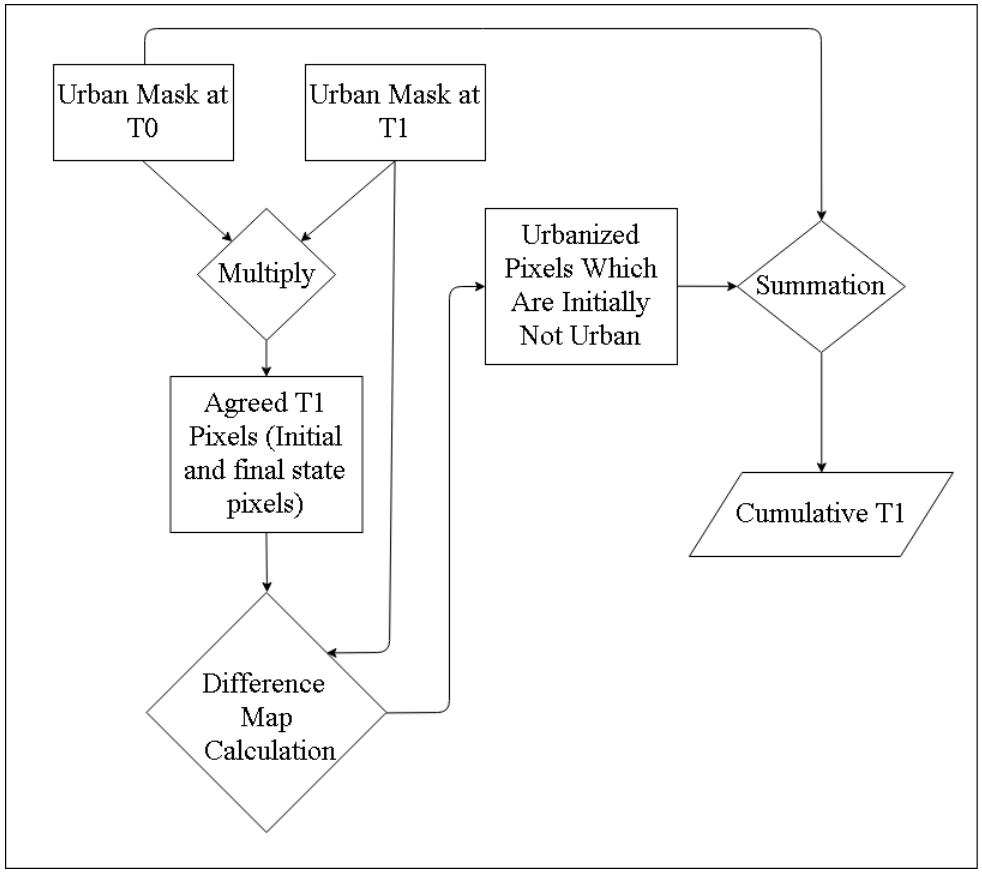

Figure 3.14: Urban layers consistency approach

3.2.4 Model Validation with 2018 Landsat Imagery

SLEUTH creates prediction results for each year between prediction start and prediction stop dates. 2050 prediction results contain a prediction layer for every year between 2016 and 2050. Predictions of 2030 and 2050 have created two prediction results for the year 2018 and predictions are identical both numerically and visually. 2018 Prediction gives 331705 urban pixels in 2018, though 2018 classification contains 296686 urban pixels. Model predicts 11% more urban pixels. This result can be associated to model performance, uncertainty of input data or urbanization prediction threshold used in the study. (Chaudhuri and Clarke, 2014). In this study 90% and above urbanization probabilities are accepted as urban. If a higher threshold value was selected less urbanization could have been observed.

3.2.5 Analysis of Actual City Plans

While city plans keep a record of the land use in the city, they also guide the planned use in the future. In Turkey, spatial plans construction regulation and zoning laws guide city planning process. City plans contain land use function of both the existing and planned land. Information about roads, hazardous areas and protection areas can also be derived from the plans. In this study 1/5.000 scale master building plans are used. Some areas inside the study area don't have a master building plan by the time. Planning information of these areas are derived from 1/25.000 scale environmental plans. Eventually all plans are combined. 1/25.000 and 1/5.000 scale plans those are currently used in Adana have been obtained from Adana Municipality. These plans are dated back to 2012 but they were kept up to date by revisions.

Since the plans are tendered region by region, each tendering region has different annotation of areas. Because each region is prepared by a different company hence naming convention may be different. For example; protected agricultural lands are called $PL\_TARIM\_NITELIGI\_KORUNACAK\_ALAN$ in one region, but in another region, they are called PL_TNKA. At the end when all layers assembled together there are lots of layer names. Combining different named layers is a laborious work but is a must because otherwise excluded layers can't be obtained.

Names of all layers are assessed with a municipality planner and the same areas with different layers are combined. After the layer arrangement, number of the layers are reduced to 12 from 129. Final state of the arranged layers is given in Figure 3.6.

3.2.6 Extraction of the Roads from Plans

Detailed information about roads can be derived from the city plans. Unfortunately, road layers on the plans are in CAD format. Only road widths are written between the cadastre islands. In order to use the road layer in SLEUTH, roads are digitized by creating lines between cadastre islands. Vicinity of roads are naturally suitable for urbanization because of their accessibility (Clarke et al., 1997). When predicting future urbanization, SLEUTH urban model takes road influenced growth into account by using road layers belonging to different years. In order to simulate this effect, factor called road-gravity is used inside the SLEUTH's simulation framework (Clarke et al., 1997). Wider roads affect the road influenced growth more than thinner ones. Only roads wider than 10 meters were digitized from the urban plans. Road notation in plans can be seen in Figure 3.15.

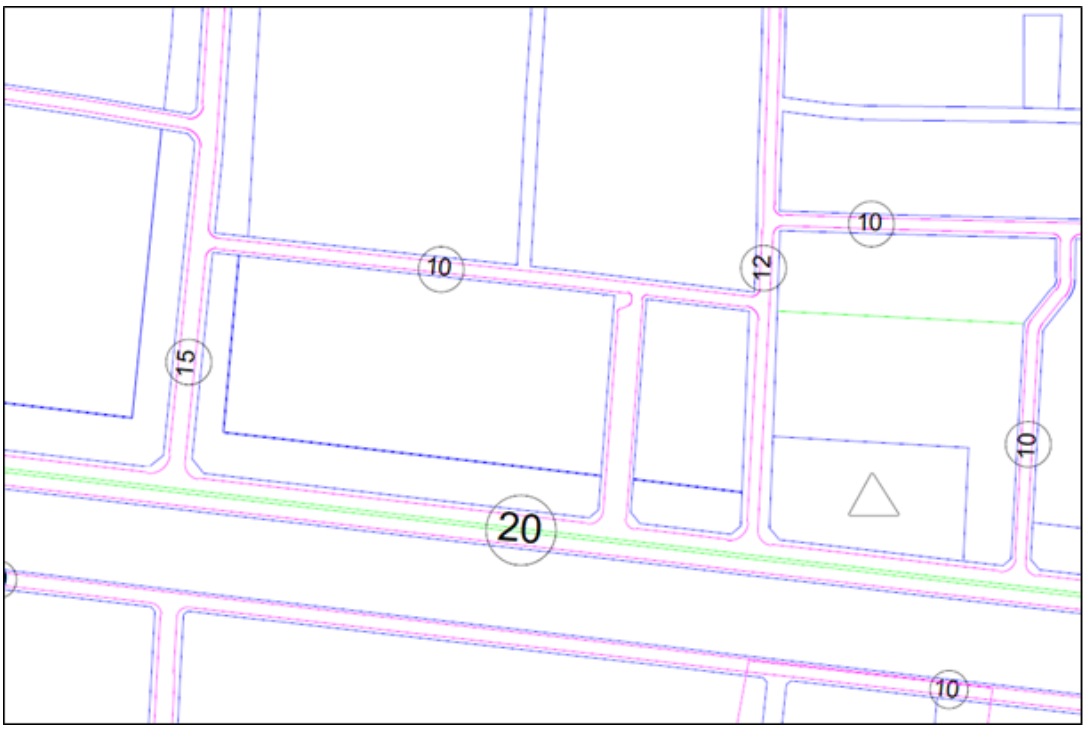

Figure 3.15: Road widths notation in the master plan

These layers are transformed to shapefile in Quantum GIS and rasterized for the study area. After the digitization, 18 different road types appeared. These roads are classified into four classes according to Indian National Programme on Technology Enhanced Learning urban transportation planning online course notes (Arasan, 2014). There are four main types of roads according to their importance;

- Arterial Roads
- Secondary or sub-arterial roads
- Local Roads
- Other Roads

According to this information, all roads are classified into four classes. After shapefile conversion of CAD data, numbers from one to four are given to the roads according to their importance. Number one is arterial and four is other roads. This information is written to new column in shapefile that contains road data. In QGIS, vector to raster conversion is applied to the file. Digitized roads can be seen in Figure 3.16. Road aggregation schema can be seen in Appendix C.

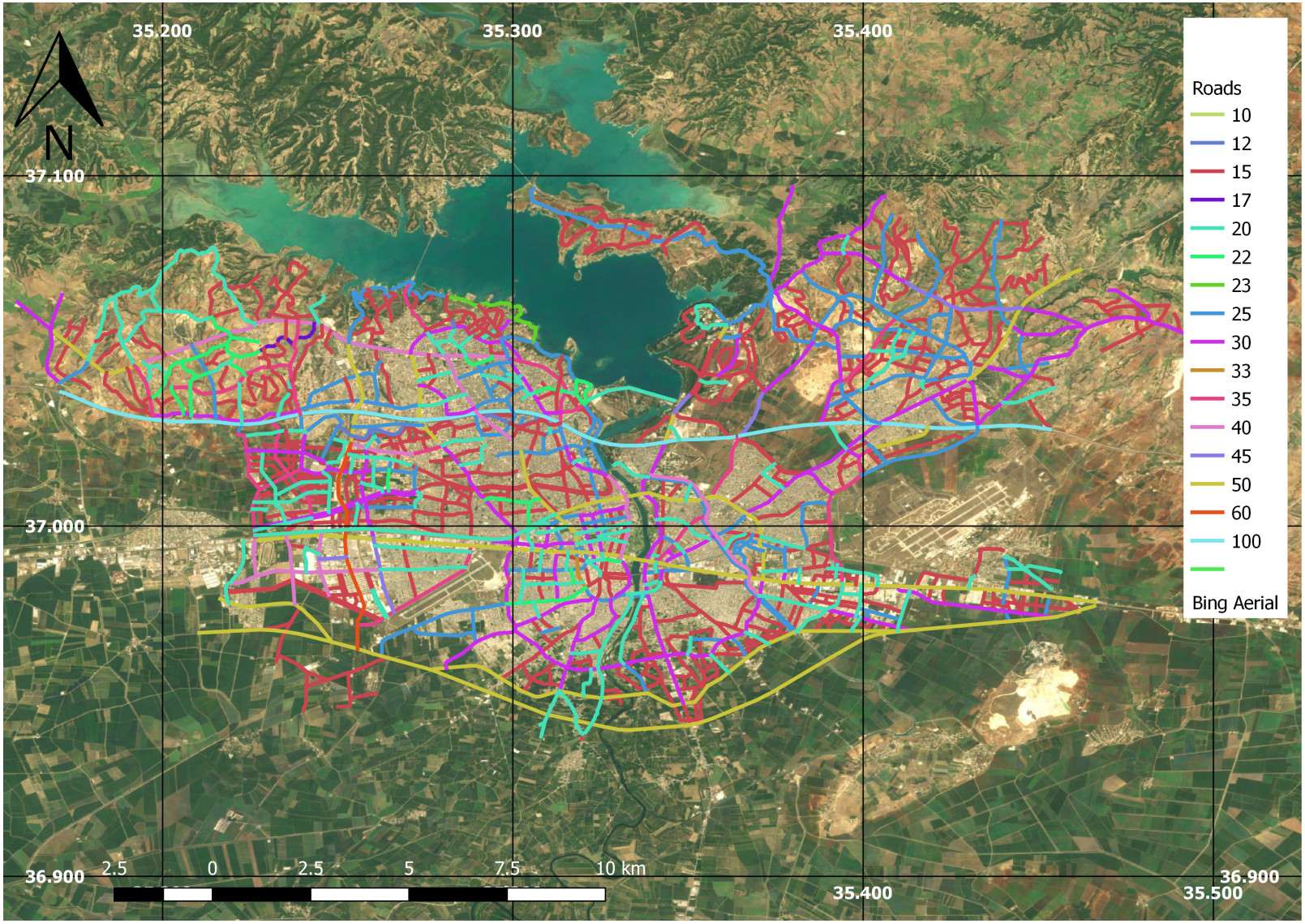

Figure 3.16: Digitized road layer derived from plans.

Road layer contains both actual and planned roads because data were taken from a master plan. To use this road data in SLEUTH, temporal changes must be considered. Classifying digitized road data into three groups according to their existence in the past and present is the way to create road layers.

Google Earth provides historical data very quickly. All vector road data opened in Google Earth's desktop application and by using the time bar, oldest data sets are used to eliminate non-existing roads by that time. For this study area, oldest Google Earth tiles were dated back to 2006. All roads which don't exist by 2006 are deleted and vector layer is re-digitized to obtain 2006 roads layer. Same procedure applied to vector road data for the year 2016. Roads of 2006 and 2016 can be seen in Figure 3.7.

Roads that currently do not exist are planned to be finished by 2020. These roads can be seen in Figure 3.7. Finally, three road layers are prepared for 2006, 2016 and 2020 in GIF format.

3.2.7 Extraction of the Excluded Areas from Actual City Plans

City plans those are obtained from Adana Municipality are used to obtain the excluded areas. Layers are selected and rasterized in Quantum GIS. SLEUTH model accepts values between 0 and 100 on excluded layer files. If pixel value is 100, that pixel is considered as under 100% protection against urbanization. If an area is under partial protection, according to its protection level a value between 0 and 100 can also be given. In this study three different scenario files are prepared.

In the first scenario; water bodies, military and civil airports, cemeteries and military zones are protected from urbanization. This scenario is the most flexible one in terms of protection. On the contrary; forest, green and agricultural areas are not protected. First scenario file can be seen in Appendix D.

Second scenario is used to see how model behave under partial protection. In this scenario additionally to the first scenario agricultural areas are taken under 100% protection. Also, forest areas and afforestation areas are taken under partial protection. Pixel values of forest areas are set to 60% and afforestation areas pixel values

area set to 80% protection for these areas respectively. The file of this scenario can be seen in Appendix D.

In the third scenario, excluded areas include all farming areas (agriculture and marginal agriculture), all green areas. This is the strictest scenario because most urbanization vulnerable areas like marginal agriculture and afforestation are under 100% protection. Scenario file can be seen in Appendix D. Excluded areas of three scenarios are given in Figure 3.18. Using these scenario files future predictions can be done. SLEUTH requires a exclusion layer to be used in calibration phases. Which means it calculates the on going trends using this layer. One can use the same layer in prediction after calibration or it can be changed to predict distinct future scenarios. In this study all agricultural areas, parks, military areas, airports, forests and water bodies are protected from urbanization. This protection schema is presented in the third scenario. So, the third scenario file has been used in calibration phase, then for predictions other two exclusion files (scenarios) are used as input to see effects of different protection schemas on urbanization.

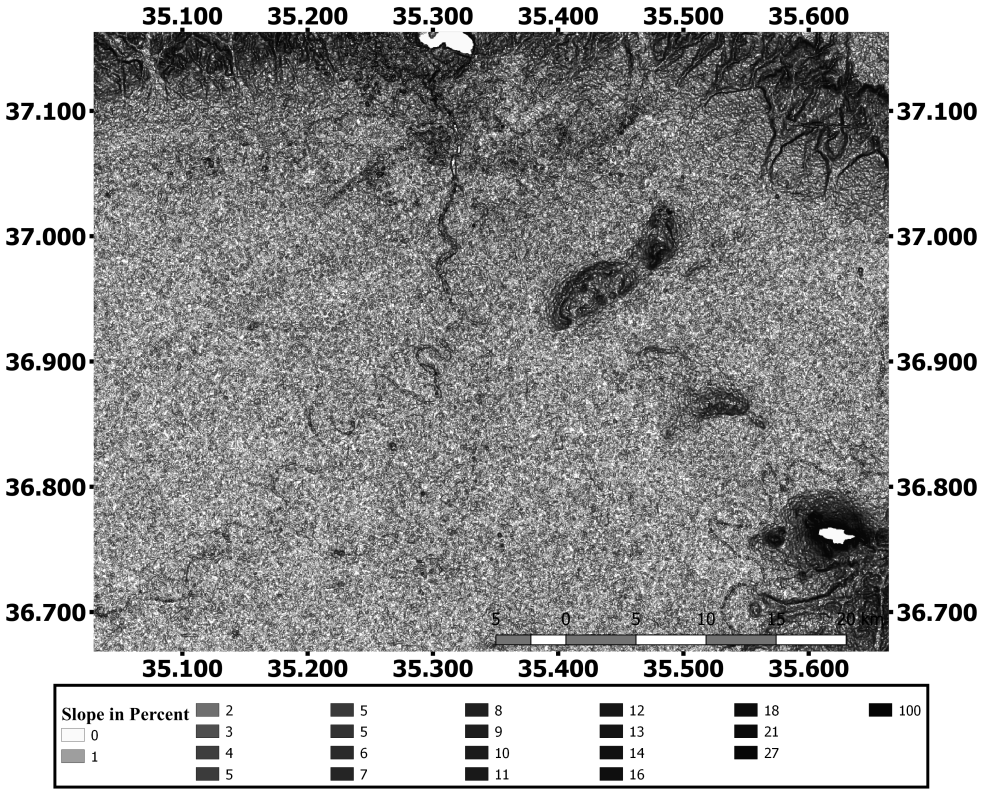

Figure 3.17: Slope map of the study area

3.2.8 Preparation of the Slope and Hillshade Layers

SLEUTH model takes shape of the Earth into account when modelling urbanization. Because of construction costs; areas with high slope will urbanize relatively slower than areas with a lower slope. When preparing the SLEUTH's scenario file, it is also possible to limit urbanization above a specific slope. In this study slope values above 25% considered as the maximum slope for urbanization. When preparing the slope layer to be used in the SLEUTH, like the other layers it must be prepared as 8-bit unsigned integer. Crucial point about the slope layer is; slope values must be in percent.

Study area coincides with two scenes in SRTM scenes. The two scenes are downloaded and combined. Combined scenes are sub-setted in ENVI according to study area. Slope layer has been visualized as hillshade to give a better perspective about the study area. DEM of the study area can be seen in Figure 3.17.

3.2.9 Resampling the Layers

All layers have 30 meters spatial resolution as default. Regardless of their default resolution, input layers must be re-sampled to half and quarter resolution to be used in SLEUTH model. In this case original resolution have been resampled to 60 meters and 120 meters using nearest neighbourhood method. Examples of the resampled layer of the third scenario can be seen in Figure 3.19.

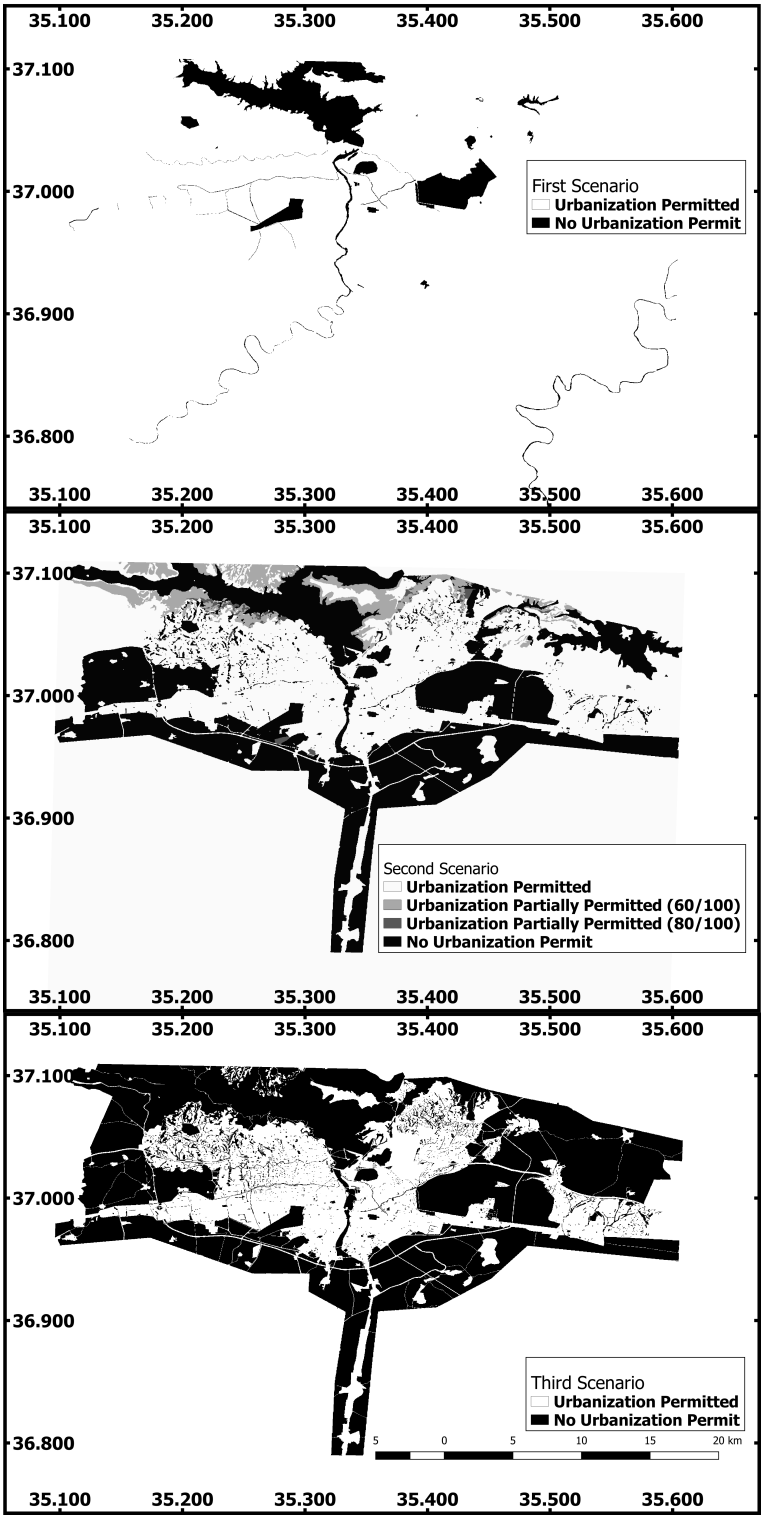

Figure 3.18: Scenario files used in prediction phase(from up to down; first, second and third scenario respectively)

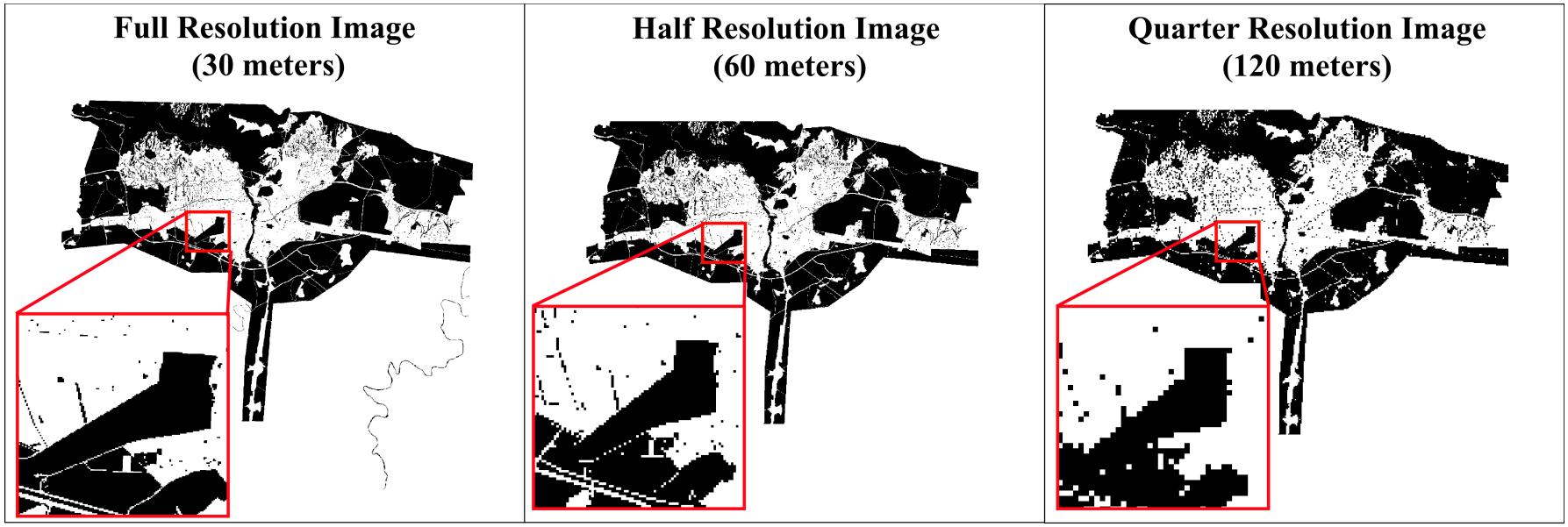

Figure 3.19: Resampling example of the third scenario

3.2.10 Converting Raster Layers to GIF format

Conversion of TIFF to GIF was done with IDL but could also be done in ArcGIS or Quantum GIS. In IDL it is possible to write GIF's with a short script like in Figure 3.20.

```
pro envi2gif
  e = ENVI()
  File =
FILEPATH(image.tif, RootDir=C/full/path/to/input/file)
  image = READIMAGE(File)
  image = REVERSE(image, 2)
  file2 =
FILEPATH(image.gif, RootDir=C/full/path/to/output/file)
WRITEGIF, file2, bytscl(image)
DELVAR, File, file2
```

Figure 3.20: Simple IDL Script to Write Images in GIF Format

3.2.11 Renaming the Input Layers

Input data naming is important for the SLEUTH since dates are read from file names. Naming format has been described in detail in SLEUTH's web site (Clarke, 2018).

$$< location > .urban. < date > .[< userinfo >].gif \tag{3.3}$$

In this convention location is relative folder path, date is four-digit year, user info is optional user data. All data are also named according to this convention. Table showing the reorganized layer names can be seen in Appendix E.

CHAPTER 4

URBAN GROWTH ANALYSES

In this chapter, preparation of the environment where SLEUTH was run, calibration phases and selection of required coefficients on each phase are explained in detail.

4.1 Preparation of Computer Environment for Model Run

As described in official web site of SLEUTH; SLEUTH model is written in the C language under UNIX and uses standard C compiler (gcc). Running SLEUTH can be a challenge if the running environment isn't set properly.

One option for running SLEUTH is using Cygwin on Windows environment. Cygwin is an enormous collection of GNU and Open Source tools which provide similar functionalities of Linux (Anonymous, 2018). Cygwin contains lots of packages for various purposes. Selecting appropriate packages is a vital step.

Another option is setting a Linux environment individually and running the commands in this environment. This environment can be attained on a virtual machine such as Oracle VM or by rebooting a computer. After the reboot or running the virtual machine, libraries must be updated using Linux commands.

In this study a cloud computer from Amazon Web Services is used. Cloud machines provide high performance and stability rather than local machines. Especially long lasting final calibration took more than three days. An electric interruption during the calibration could result restarting the calibration. Also, performances of cloud machine can be increased easily. Model of the Kali Linux distribution is selected as T2.medium model. All updates are installed. GCC version of the machine was gcc

(Debian 7.3.0-5). SLEUTH files copied to the machine and the libraries compiled as explained in SLEUTH web site (Clarke, 2018).

4.2 Data Validation and Selecting Coefficient Ranges

Running SLEUTH model to make predictions about future can be achieved through a rigorous calibration phase.

Over the years SLEUTH model has been using various parameters to determine these values from $control\_stats.log$ file. When there was no clear consensus on which metrics to use on calibration phase, SLEUTH users have used different metrics to assess statistics created by SLEUTH. Most popular ones are Lee-Salee, compare, population, weighted sum of all metrics (Dietzel and Clarke, 2004). In 2007 Dietzel and Clarke have published an article about how to calculate best metrics for SLEUTH calibration. They have noted only eight of the metrics that SLEUTH created through the calibration has influence over the results (Dietzel and Clarke, 2004).

These metrics are;

1. **Compare**
   Ratio of modelled and actual population (urbanization in SLEUTH).
2. **Population**
   Least squares regression score for modelled urbanization compared to actual urbanization for the control years
3. **Edges**
   Least square regression score for the modelled urban edge count compared to actual urban edge count for the control years.
4. **Clusters**
   Clusters Least squares regression score for the modelled urban clustering compared to known urban clustering for the control years

5. **Slope**

   Least squares regression of average slope for modelled urbanized cells compared to average slope of known urban cells for the control years

6. **X-mean**

   Least squares regression of average $xvalues$ for modeled urbanized cells compared to average $xvalues$ of known urban cells for the control years

7. **Y-mean**

   Least squares regression of average $yvalues$ for modeled urbanized cells compared to average $yvalues$ of known urban cells for the control years

8. **F-Match**

   A proportion of goodness of fit across landuse classes. (Only if land-use and land cover is being modelled)

Cluster Size, Lee-Salee, Urban and Rad metrics are not effective in SLEUTH calibration. Without the non-effective metrics, a metric called Optimum Sleuth Metric (OSM) is calculated using the formula in Equation 4.1;

$$OSM = compare * pop * edges * clusters * slope * xmean * ymean \quad (4.1)$$

Using OSM is a good approach when evaluating the coefficient ranges because it is a special statistics that is created only for assessing the SLEUTH results and metrics that don't affect SLEUTH calibration have been excluded when calculating the SLEUTH metric. Through the calibration phase five coefficient values are re-evaluated in Monte-Carlo Simulations, to compare the evaluated statistics Pearson r^2 statistics is used. One can calculate OSM by downloading and compiling source code from SLEUTH web page or by a formula by an assistance program such as Microsoft Excel. Formula for OSM can be found in the readdata3.c file.

To ensure the validation of the input data a test mode run has been performed and input data has been validated. Importance of the test phase is; it allows users to see if there is any problem with the input data. On the test run user can get a feedback about whether SLEUTH data have been prepared correctly or not. Data which are not

copied with the output folder means that program can't read that specific file. In this case, this data must be specifically examined and corrected. Before the calibration phase a test-run has been performed to the input data by using data with the lowest resolution. Since test run takes less than 5 minutes, a low Monte Carlo iteration number, such as five, can be selected.

4.3 Calibration Phase

SLEUTH has three calibration phases;

1. **Coarse Calibration**
   Calibration coefficients are evaluated on a wide range and step value. Start, stop and step values for all coefficients on the coarse calibration are selected as; $START\_VALUE = 0$, $STOP\_VALUE = 100$, $STEP\_VALUE = 25$
   After each calibration phase, ranges get narrower. During the coarse calibration lowest resolution inputs are used. This phase has the shortest duration. In this study coarse calibration lasted 2 hours on average for a data of 378x352 size.

2. **Fine Calibration**
   Coarse calibration narrows down the coefficient ranges and these narrow downed values are used in the fine calibration. During the fine calibration mid resolution inputs are used. In this fine coarse calibration lasted 6 hours on average for a data of 756x753 size.

3. **Final Calibration**
   Fine calibration results are used, and a narrower interval and a smaller step value selected for each coefficient. This is the phase that takes the longest time. In this study final calibration lasted five days on average for a data of 1512x1406 size.

On the calibration phase, five coefficients that control four type of growth rule are calculated. These rules are mentioned in Chapter 2.3, detailed information can be found in Section 2.3.

4.3.1 Coarse Calibration of Input Data

Images which have been re-sampled to 120 meters are used in coarse calibration. Before running the prepared scenario; input directory is specified, input image names are written where output files' flags are set to true. As default $WRITE\_COEFF\_FILE$, $WRITE\_AVG\_FILE$, $WRITE\_STD\_DEV\_FILE$ and $LOGGING$ flags were already true as default. These flags are kept as default.

Only $WRITE\_MEMORY\_MAP$ flag is changed and set to true so that in case of memory heap problem logs can be assessed. Output files are coefficient logs of the cellular automata framework life cycles. Prediction start and stop dates are selected as 2016 and 2030 respectively. For the first run prediction date has been set to 2030 then 2021 urban prediction has been used with scenarios to predict 2050. Monte Carlo iterations flag has been set to five. Coarse calibration coefficient values are left default. Coefficients can be seen in Table 4.1.

On each calibration step after the coarse calibration, Optimum SLEUTH Metric values are calculated using $control\_stats.log$ file using readdata3.c file. This file calculates OSM values and creates a file which contains coefficients with highest OSM values. In SLEUTH's web site; it's recommended to select ranges according to minimum and maximum values of the first ten rows with the highest OSM value (Clarke, 2001).

Table 4.1: Coarse calibration coefficients

Coefficient Type	START	STEP	STOP
Diffusion Parameter	0	25	100
Breed Parameter	0	25	100
Spread Parameter	0	25	100
Slope Parameter	0	25	100
Road Parameter	0	25	100

The OSM values obtained from these parameters are given in Table 4.2.

Table 4.2: Top ten coefficients with highest OSM values of coarse calibration

OSM	Diffusion	Breed	Spread	Slope	Road
0.08690265	1	75	50	50	50
0.07092816	1	75	25	50	25
0.06378988	1	75	25	50	75
0.05659298	1	75	25	50	100
0.05626166	1	100	25	50	100
0.05587526	1	75	50	50	25
0.05527364	1	100	50	50	50
0.05514022	1	100	50	75	75
0.05446653	1	100	50	50	100
0.05292995	1	100	50	50	1

4.3.2 Fine Calibration of Input Data

OSM values are calculated using $control\_stats.log$ file inside the output directory. A C program can be downloaded from SLEUTH's web site named readdata3 which can calculate OSM values automatically. Like SLEUTH this program must be compiled in Linux using "make" command. This program writes coefficient values of first 50 set of runs with the highest OSM. Result of the calibration can be seen in Table 4.3.

Table 4.3: Fine calibration coefficients

Coefficient Type	START	STEP	STOP
Diffusion Parameter	1	6	25
Breed Parameter	75	5	100
Spread Parameter	25	5	50
Slope Parameter	25	5	50
Road Parameter	25	15	100

Top ten coefficients with highest OSM values of the fine calibration are given in Table

4.4.

Table 4.4: Top ten coefficients with highest OSM values of fine calibration

OSM	Diffusion	Breed	Spread	Slope	Road
0.052319	7	90	45	25	55
0.051949	7	75	50	25	55
0.051278	7	80	45	25	100
0.050757	7	100	45	25	25
0.050196	7	80	50	25	85
0.049978	7	90	40	25	85
0.049929	7	75	50	30	70
0.049748	7	100	45	25	55
0.049713	7	80	40	25	70
0.049668	7	100	45	25	100

4.3.3 Final Calibration of Input Data

Similar to previous calibration, first ten rows with highest OSM values are selected from the previous calibration result and calibration is run.

Table 4.5: Final calibration coefficients

Coefficient Type	START	STEP	STOP
Diffusion Parameter	2	2	12
Breed Parameter	80	4	100
Spread Parameter	40	2	50
Slope Parameter	25	1	30
Road Parameter	25	15	100

Top ten coefficients with highest OSM values of the final calibration are given in Table 4.6.

Table 4.6: Top ten coefficients with highest OSM values of final calibration

OSM	Diffusion	Breed	Spread	Slope	Road
0.04856685	4	80	50	26	70
0.04840338	4	88	50	25	70
0.04819809	4	80	48	25	100
0.04789131	4	84	48	25	70
0.04784823	4	80	50	25	25
0.04780863	4	80	50	25	40
0.04774734	4	80	48	25	70
0.04771273	4	80	50	25	85
0.04759206	4	80	50	26	25
0.04754788	4	80	50	25	55

When the final calibration ends; calibration phase is over. Forecasting and prediction phases follow the calibration phase. Calibration phase finds best coefficients that can simulate historical growth in the region. Self-qualification feature may alter coefficient values at the start date of the simulation. Though, at the forecasting phase, stop date coefficients from best calibrated coefficients are used. Forecasting phase produces a single set of coefficients for the last calibration year and this value can be used for prediction. When random variability of the model is considered, then averaged coefficient results of multiple Monte-Carlo iterations can be used to get more robust forecasting coefficients.

4.4 Forecasting Phase

Self-modification alters the coefficients while running Monte-Carlo simulations between START-DATE to STOP-DATE. Since the aim of the calibrations is finding the best initial coefficient values for the study area with given historical data; final calibration coefficients are used to find single STOP-DATE values for each coefficient. Unlike the previous steps, start and stop values are settled to the same value and step values are selected as one. So, the coefficient probability interval becomes the nar-

rowest. Forecasting phase doesn't produce any visual products. Initial values for forecasting phase coefficients can be seen in Table 4.7.

Table 4.7: Forecasting coefficients

Coefficient Type	START	STEP	STOP
Diffusion Parameter	4	1	4
Breed Parameter	80	1	80
Spread Parameter	50	1	50
Slope Parameter	25	1	25
Road Parameter	70	1	70

4.5 Prediction Phase

Prediction phase produces final products of the future urbanization estimations with calibrated parameters. This phase can be run with any random parameters selected by the user but result would be irrelevant if parameters are not calibrated. Calibration creates best fitting values for the study areas, thus best prediction results can be derived only by calibration. For this study, best fitting parameters of the prediction can be seen in Table 4.8.

Table 4.8: Prediction best-fitting coefficients

Coefficient Type	Best Fitting Value
Diffusion	5
Breed	100
Spread	64
Slope	1
Road	72

Prediction phase has been run for each scenario individually. Also, to be able to have more consistent results, predictions for 2030 have been made for each scenario without 2020 road layer. Then with 2020 road data, each scenario file has been used

to predict 2050 urbanization.

From the information in paragraph above, prediction phase considered as composed of two steps; 2030 prediction and 2050 prediction. In order to use 2020 road data in the estimations, such a method was followed. SLEUTH model does not read data that is not within the date range of urban inputs. The purpose is to create a future urban layer which dated to 2020 or a later date. Then with new produced urban layer and 2020 road layer prediction range expanded to 2050. Flow chart of of prediction phase can be seen in Figure 4.1.

Three scenarios and prediction results according to these scenarios for the year 2050 are depicted in Figures 4.2 - 4.4. In the figures below only 2050 predictions are included to avoid presenting too many figures. Intermediate process result of 2030 urbanization without road layer can be seen in Appendix G. Colours in prediction maps present probability of urbanization of the pixels. Pixels that have an urbanization probability of less than 50% are shown transparent on the map. Legend for probability maps can be seen in beneath the figures.

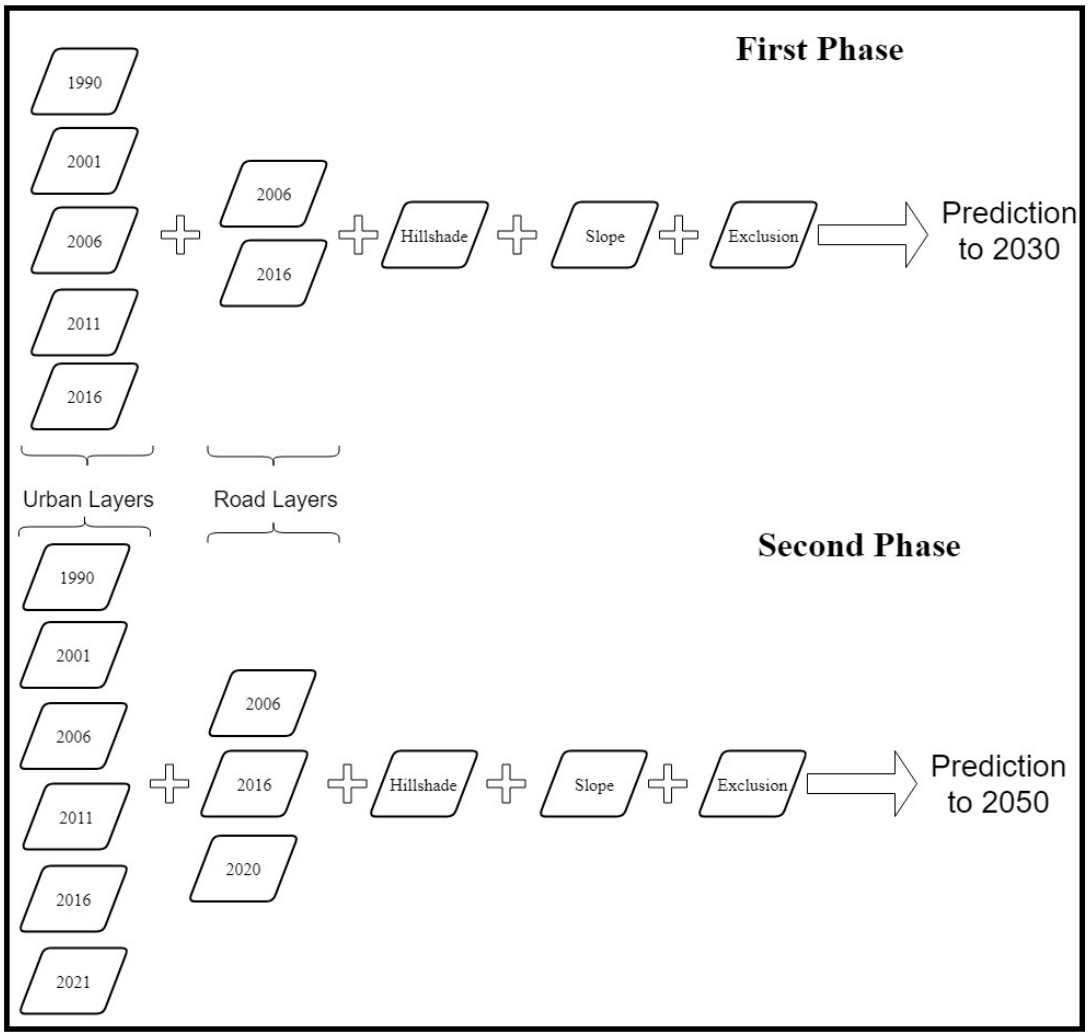

Figure 4.1: Flow chart of prediction phase

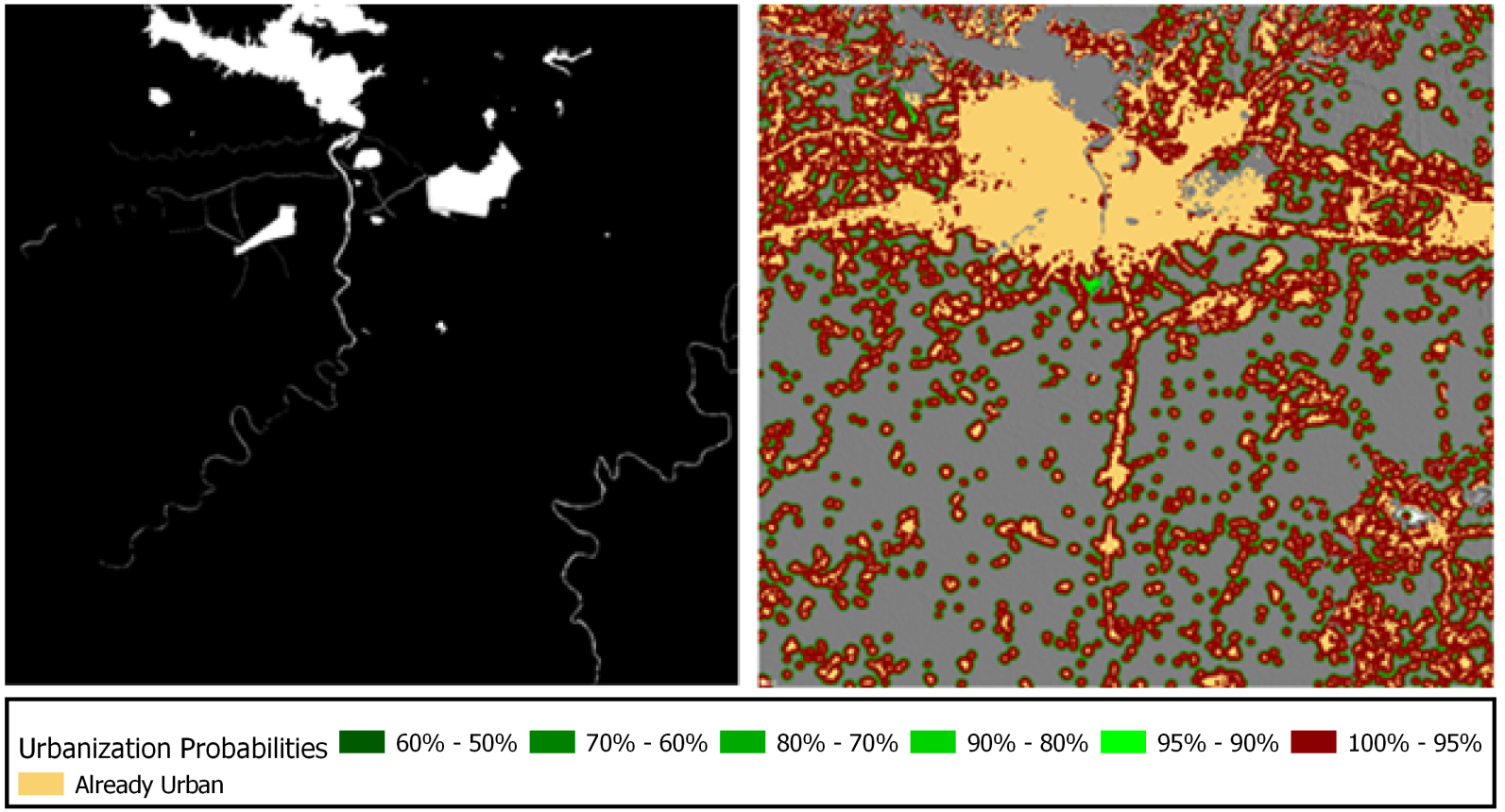

Figure 4.2: Scenario 1, excluded areas (left image white areas) and 2050 urbanization prediction

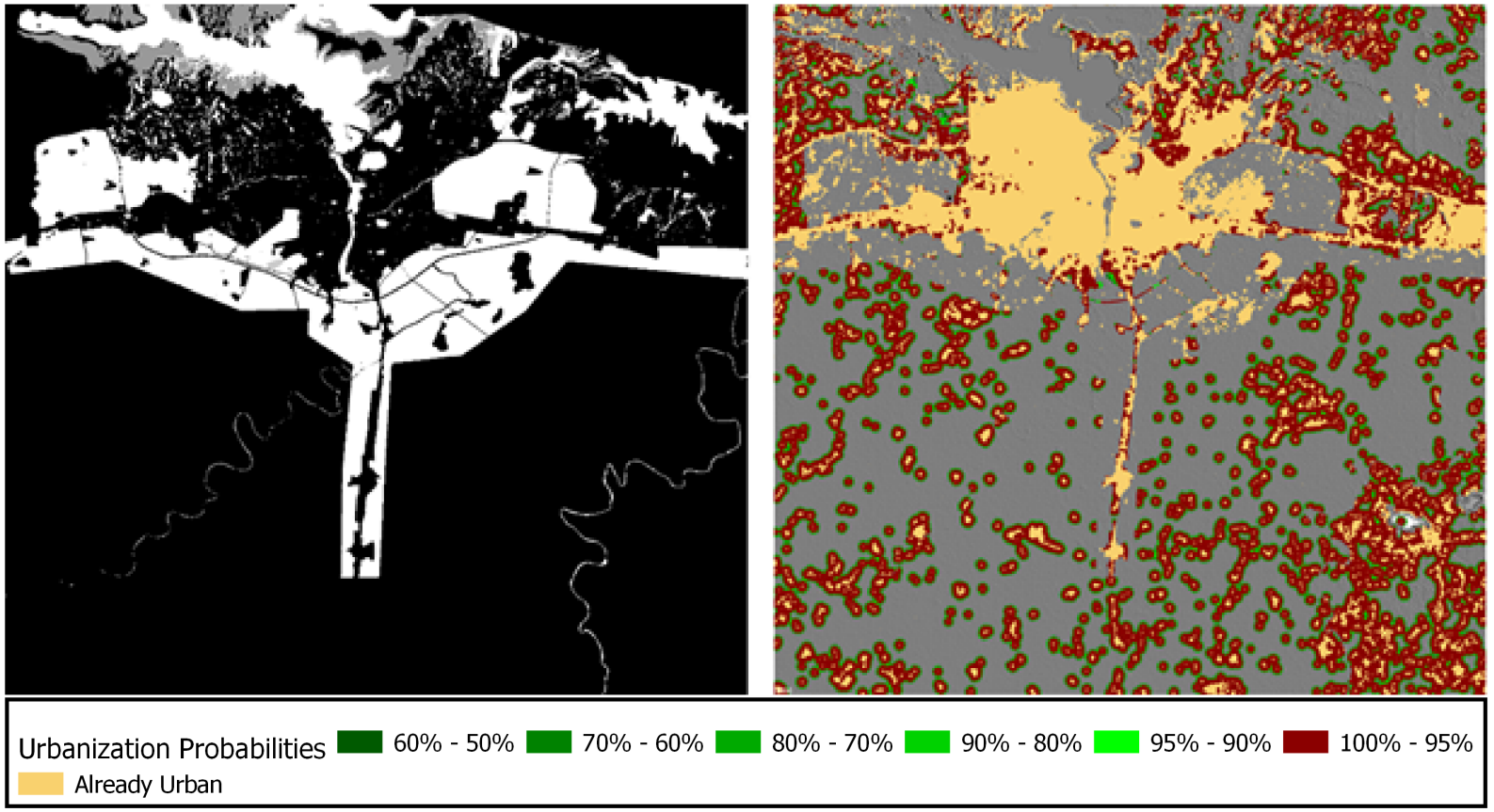

Figure 4.3: Scenario 2, excluded areas (left image, white and gray areas) and 2050 urbanization prediction

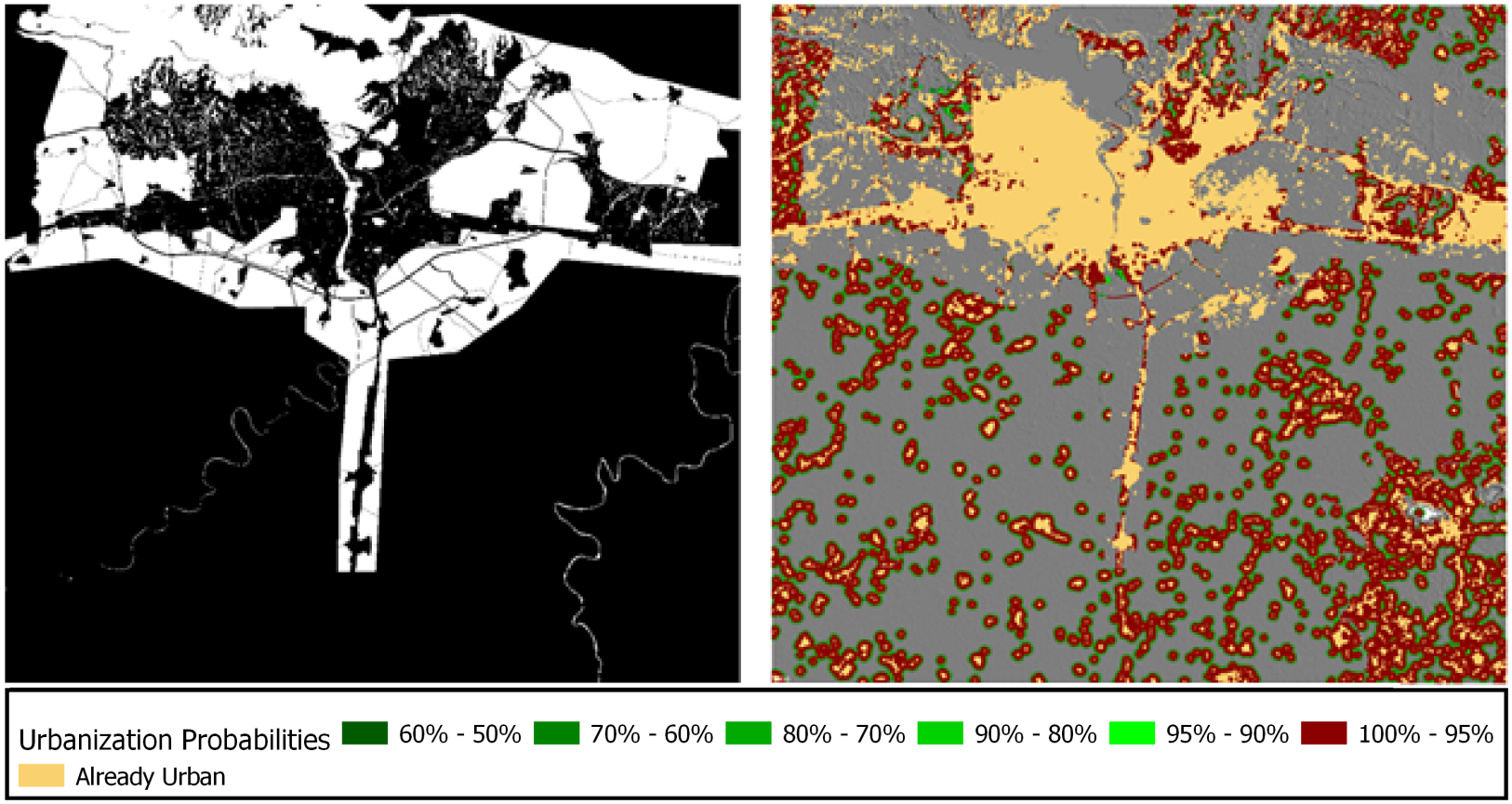

Figure 4.4: Scenario 3, excluded areas (left image white areas) and 2050 urbanization prediction

CHAPTER 5

DISCUSSION OF THE RESULTS

In this chapter, outcomes of the predictions are discussed but before that, all probability maps are processed, pixels which have an urbanization probability over 90% are considered as urbanized in 2050. SLEUTH produces predictions in GIF format, to be able to process these images, all 2050 probability maps are transformed to TIF format. The effect of new planned road of 2020 to urbanization is also discussed in this chapter.

5.1 Assessment of First Scenario Results for 2050 Prediction

As the least strict scenario; visually, most urbanization is observed in the first scenario. Statistically, 1084186 pixels (51% of the study area) are expected to be urban in 2050. 732819 of 1084186 urban pixels are new urbanized areas. Urban area mask can be seen in Figure 5.1.

New urban pixels are also found by calculating difference map between 2050 prediction and 2016 classification urban pixels. New urbanization can be seen spreading homogenic through the scene. Small urban spreading centres on the North side of the city and non-urban pixels near the urban pixels are spread unimpeded. Recently urbanized areas prediction for the year 2050 for the first scenario can be seen in Figure 5.2.

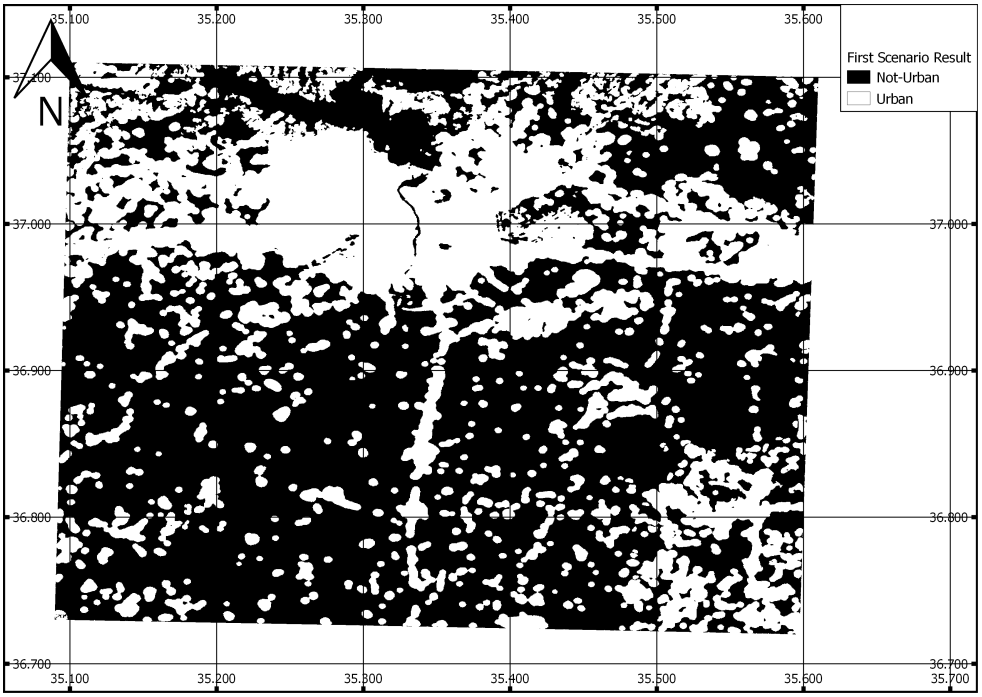

Figure 5.1: 2050 urbanization prediction for the first scenario

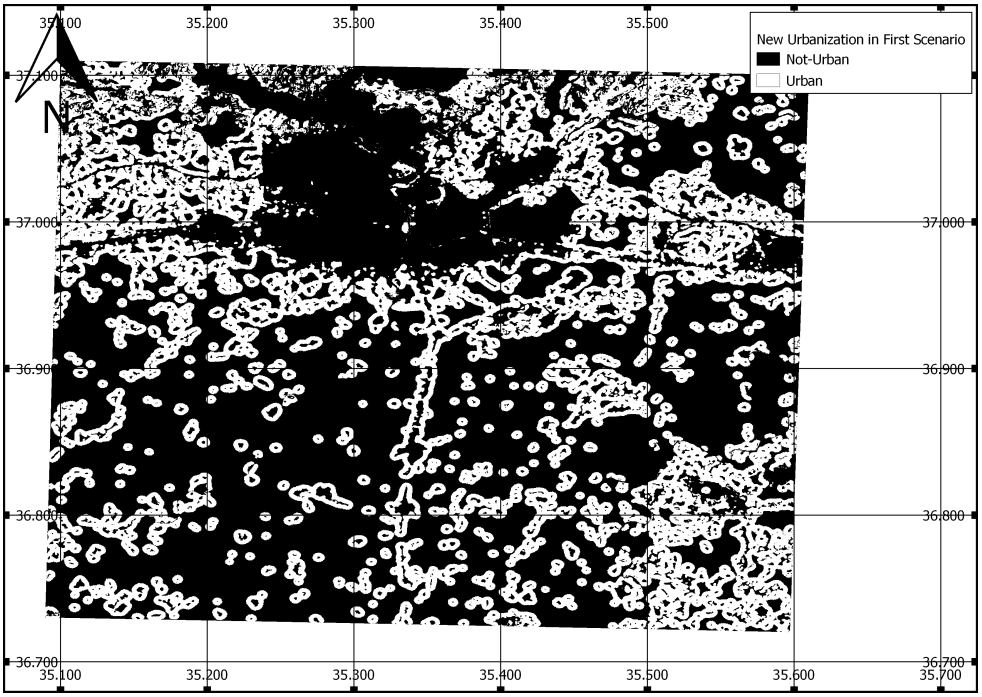

Figure 5.2: Recently urbanized pixels for the first scenario

Another important information is the transition of the pixels between agriculture to

urban or green areas to urban. Using ENVI's band math option, how many agriculture and green pixels are transformed to urban are found. In Figure 5.3 urbanization on agricultural and green areas can be seen. Quantities on different areas can be also seen in Table 5.1. Percent urbanization of classes are calculated by dividing number of transformed pixels to the total quantity of pixels in the scene.

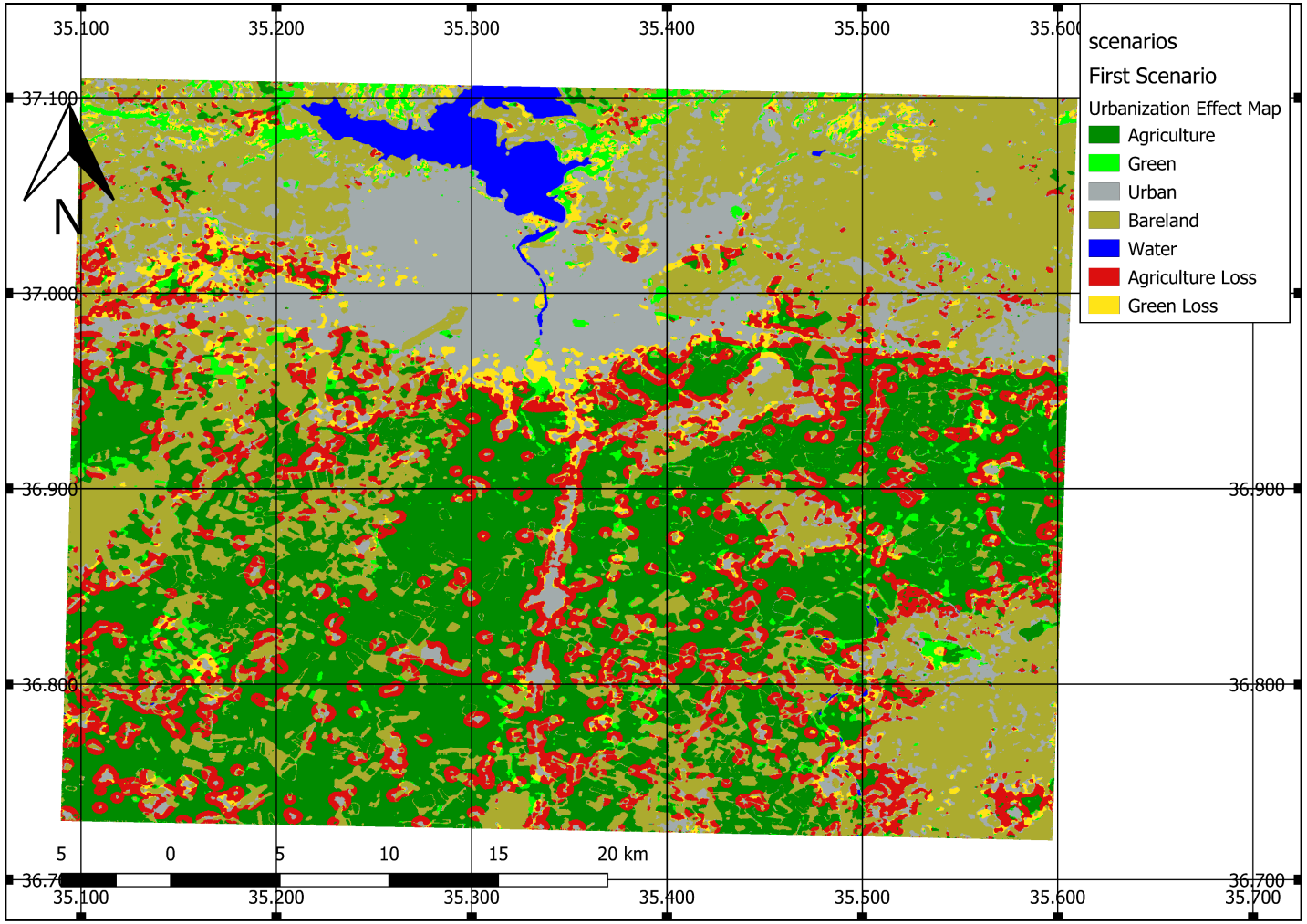

Figure 5.3: Agriculture and green area loss against urbanization highlighted for the first scenario

Table 5.1: Urbanization quantities of non-urban areas in the first scenario

Transition	Pixel Quantity	Kilometer Sq.	Percent Urbanization of Class
Agriculture to Urban	292862	∼263.58	31.57%
Green to Urban	91252	∼82.13	58.72%
Bareland to Urban	348705	∼313.84	52.42%
Total	732819	∼659.54	34.47%

5.2 Assessment of Second Scenario Results for 2050 Prediction

Second scenario has partial conservation on forest and afforestation areas. Marginal agriculture is not protected. Detailed description about second scenario can be found in, Section 3.2.7. Visually, less urbanization is observed when compared to the first scenario. 887665 pixels or 41.76% of the study area is expected to be urban in 2050. 541662 of 887665 pixels are new urbanization. Predicted urban area mask can be seen in Figure 5.4.

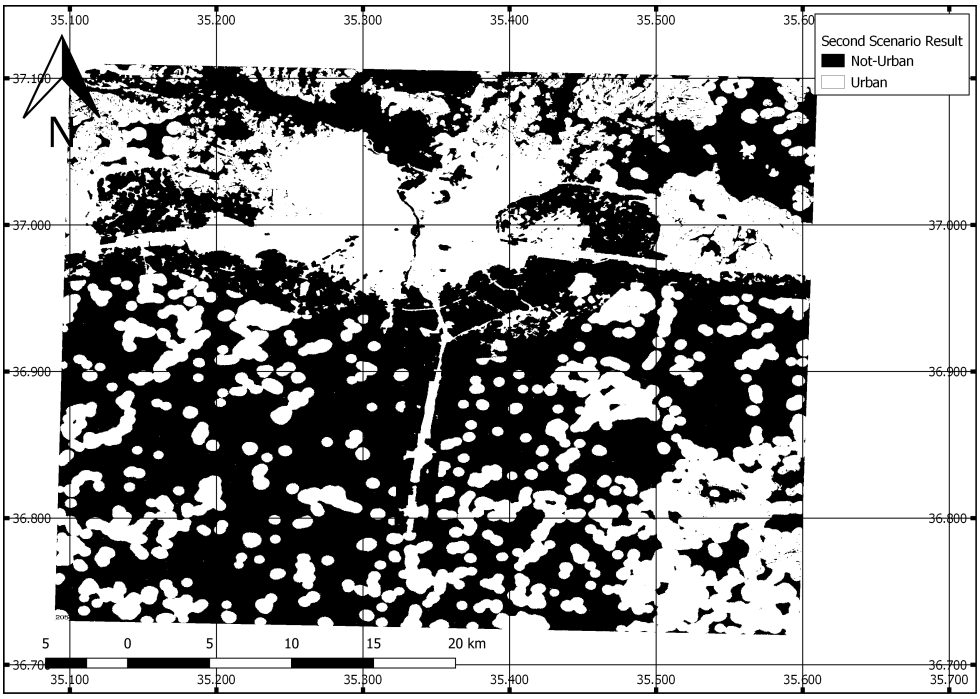

Figure 5.4: 2050 Urbanization prediction for the second scenario

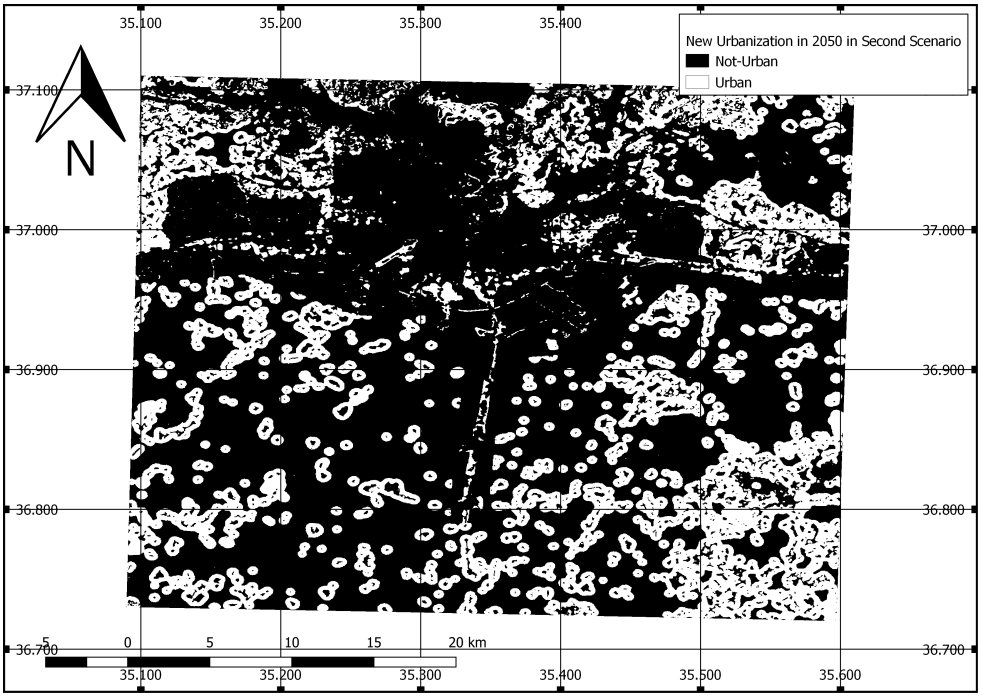

Figure 5.5: Recently urbanized pixels for the second scenario

Protected areas can be clearly seen in the Figure 5.5. Agricultural areas and green areas mostly protected but the green sides in the North part of the study area converted into urban due to partial 60% protection. Another noteworthy point is urbanization on the south, near planned road, even under protection. Unprotected areas are similarly urbanized in all three scenarios.

Urbanization effect on agricultural and green areas can be seen in Figure 5.6. Quantitative analysis for second scenario can be seen in Table 5.2.

Table 5.2: Urbanization quantities of non-urban areas for the second scenario

Transition	Pixel Quantity	Kilometer Sq.	Urbanization as Percent
Agriculture to Urban	222597	∼200.33	24%
Green to Urban	52383	∼47.14	33.71%
Bareland to Urban	266682	∼240.01	40.09%
Total	541662	∼487.50	25.45%

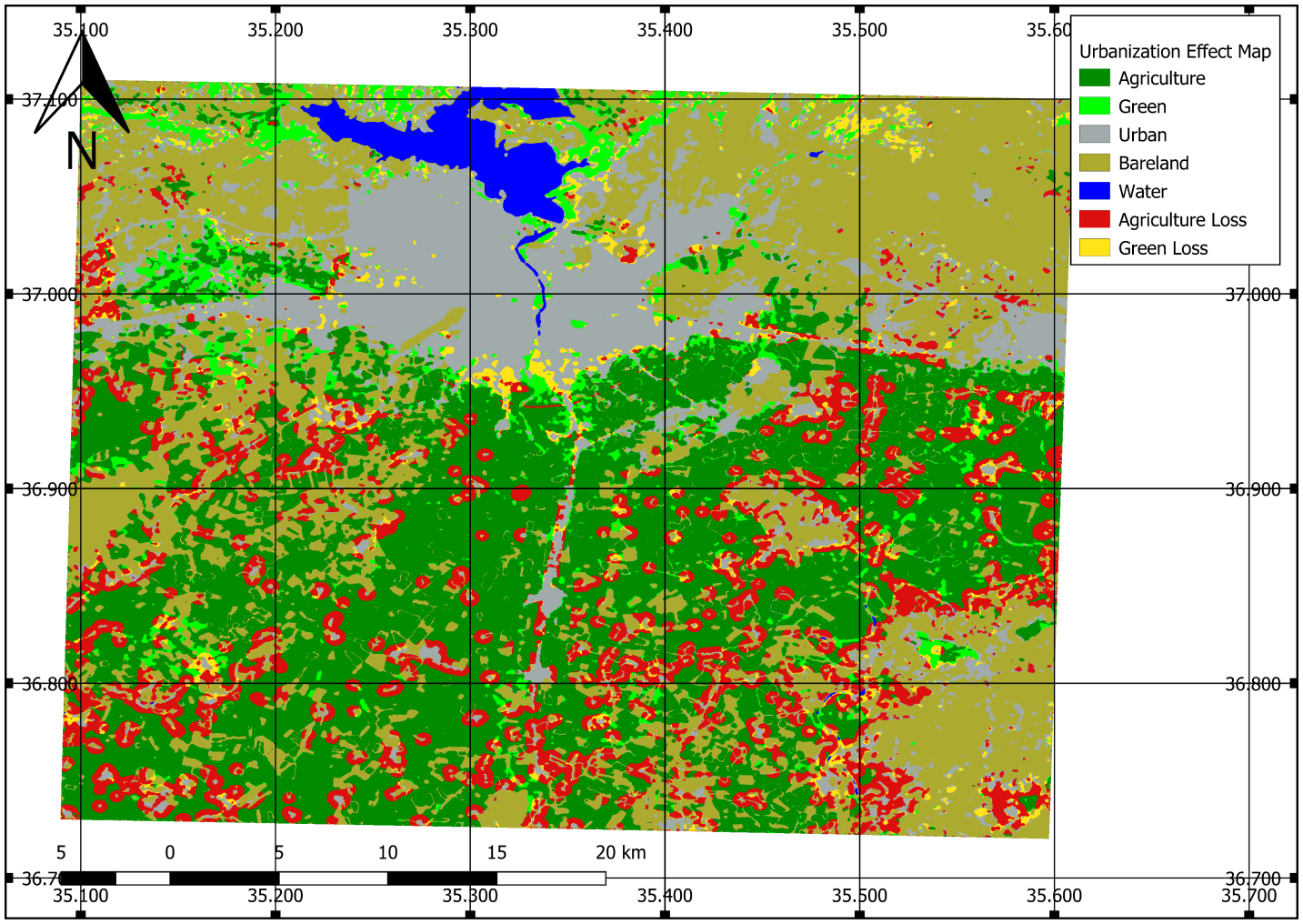

Figure 5.6: Agriculture and green area loss against urbanization highlighted for the second scenario

5.3 Assessment of Third Scenario Results for 2050 Prediction

Third scenario is the strictest scenario, regardless of their type all green, agricultural areas with necessary protection areas such as water and military areas are under protection. Urban mask of prediction result can be seen in Figure 5.7. Recently urbanized pixels can be seen in Figure 5.8.

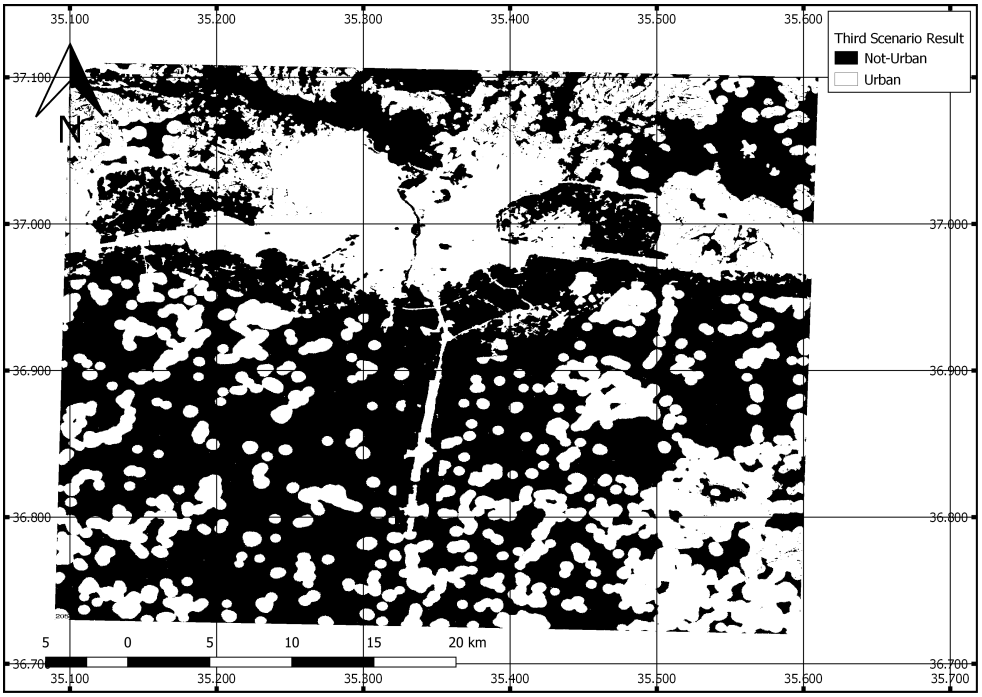

Figure 5.7: Agriculture and green area loss against urbanization highlighted for the third scenario

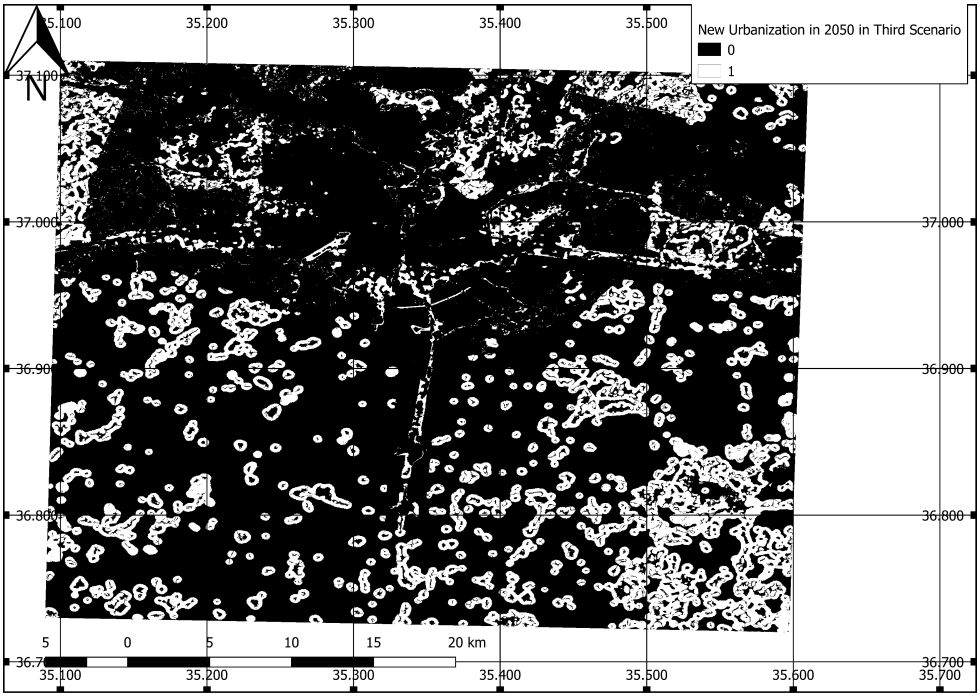

Figure 5.8: Recently urbanized pixels for the third scenario

Planned 2020 highway leads urbanization within its vicinity even under full protection. Unprotected areas are similarly urbanized as in the other scenarios. Urbanization in agriculture and green areas can be seen in Figure 5.9. Quantitative analysis of the third scenario can be seen in Table 5.3.

Table 5.3: Urbanization quantities of non-urban areas in the third scenario

Transition	Pixel Quantity	Kilometer Sq.	Urbanization as Percent
Agriculture to Urban	219293	∼197.36	23.64%
Green to Urban	51067	∼45.96	32.86%
Bareland to Urban	227904	∼205.12	34.26%
Total	498264	∼448.44	23.44%

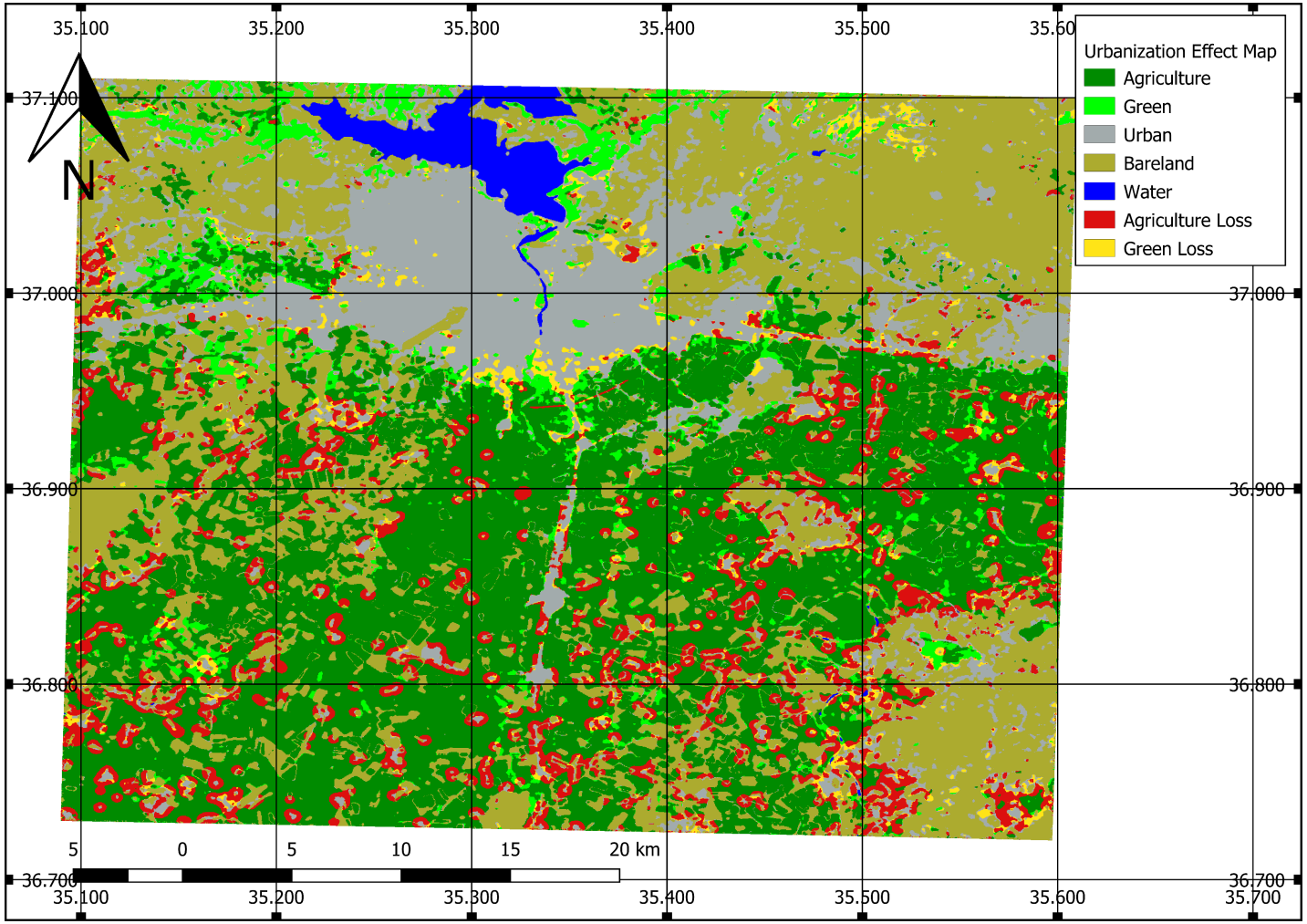

Figure 5.9: Agriculture and green area loss against urbanization highlighted for the third scenario

According to these results comparative urbanization predictions can be found in the Table 5.4.

Table 5.4: Quantities of agriculture, green and bareland losses against urban by 2050 for each scenario

Scenario/Transition	Agriculture to Urban	Green to Urban	Bareland to Urban	Total
Scenario 1	292862	91252	348705	732819
Scenario 2	222597	52383	266682	541662
Scenario 3	219293	51067	227904	498264

When all these quantities compared, it can be understood that strict protection scenario can protect the agricultural and green areas relatively better then partial protection or no-protection scenarios. Green area loss is relatively less than agriculture and bareland areas. Main reason of having less loss of green areas are due to being located in steep and higher locations. Urbanization is rather harder in steep and higher areas. In contrast to forest areas, agricultural and bareland areas are more flat and has lower altitude. Urbanization is focused on these areas.

In all scenarios agriculture to urban transformation has been observed in different volumes. Even in the most strict scenario (third scenario) agricultural areas near to urban edge has been swallowed substantially by urban construction. Therefore future land use planning should emphasize farmland protection to avoid the urbanization taking over high quality farmland.

When these forecast maps are examined together with a current map that showing the settlements, an important point emerges; continuous expansion of urbanization in rural areas. When there is no restriction, current urban areas spread their vicinity over the time. Self-modifying feature of the SLEUTH model slows down the urbanization to avoid an exponential growth but linear increase forces urban areas to spread. Comparison of the second scenario prediction for the year 2050 with current settlements can be seen in the Figure 5.10. These scenes are selected outside of the protection area. They are are located at the southern part of the study area where agricultural

activities are intense.

The number of the habitants of these settlements varies between 100 to 3000 (TUIK, 2018). These small settlements have agricultural areas within them and rural roads connects them. Estimates of urbanization show that the agricultural areas between the distant settlements are becoming urbanized and gradually connect these small settlements. This prediction seems to be unrealistic. Due to immigration, population in these areas decreases or slightly increases over the time. Urbanization dynamics of these areas are different from the planned areas. Since there is only one type of urban class in SLEUTH model, all urban pixels in the scene are treated in the same way. However, the spread of rural settlements and the spread of urbanized areas cannot be expected to be the same.

Modelling the sprawl of urbanized and rural areas separately could provide better modelling of settlements with two different behaviours in one scene. The fact that the urban layer has two classes as rural and non-rural instead of being binary can provide a solution to this situation. Instead of the city layer prepared for SLEUTH model being binary, if rural and urban settlements can be expressed individually this problem can be overcome.

Partial and full protection scenarios produced almost similar results. From 2020 to 2050 urbanization changes on every decade for each scenario can be seen in Figure 5.11. Urbanization quantities comparison of the scenarios can be seen in Figure 5.12.

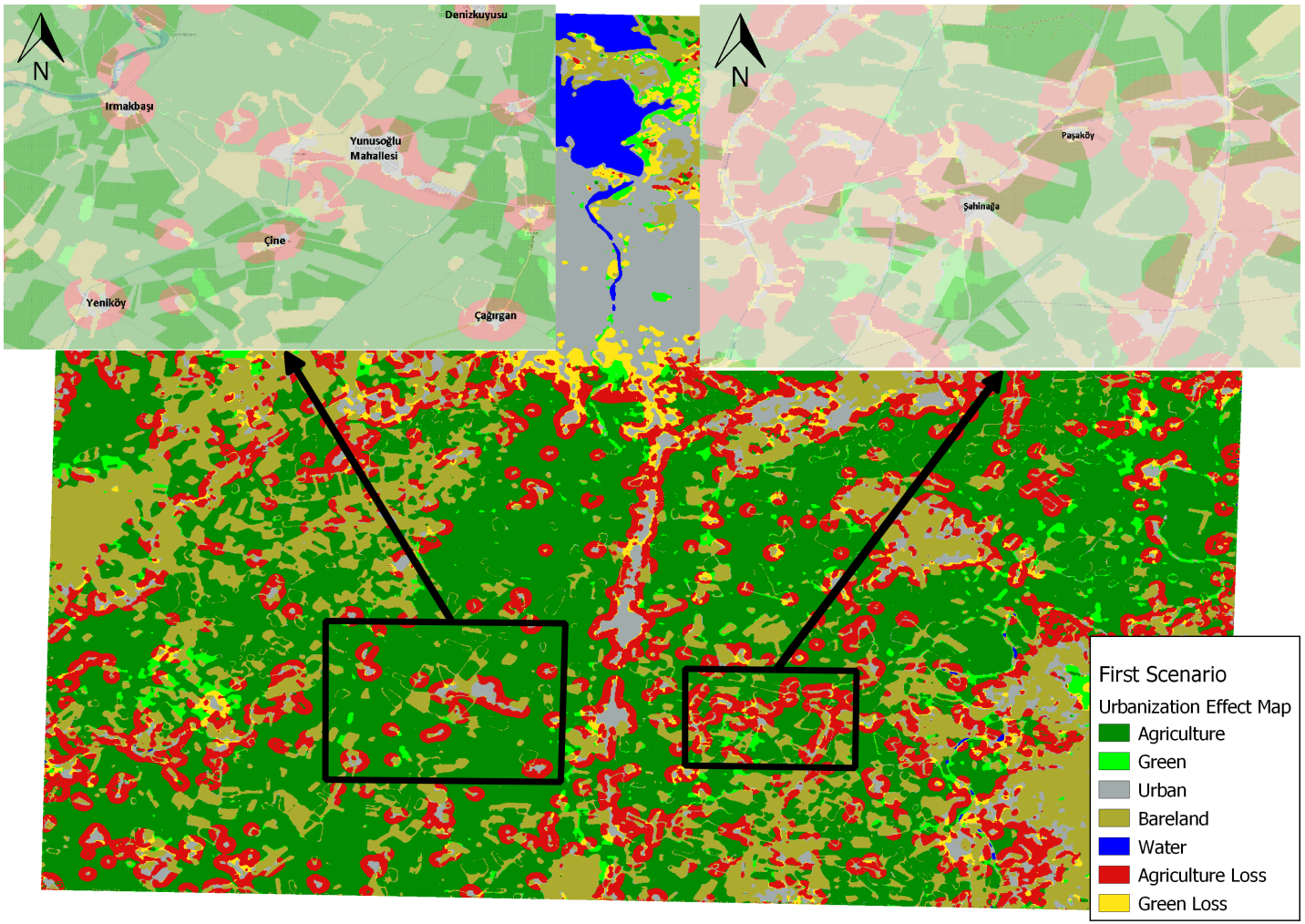

Figure 5.10: Urbanization prediction for the first scenario on rural areas with settlements

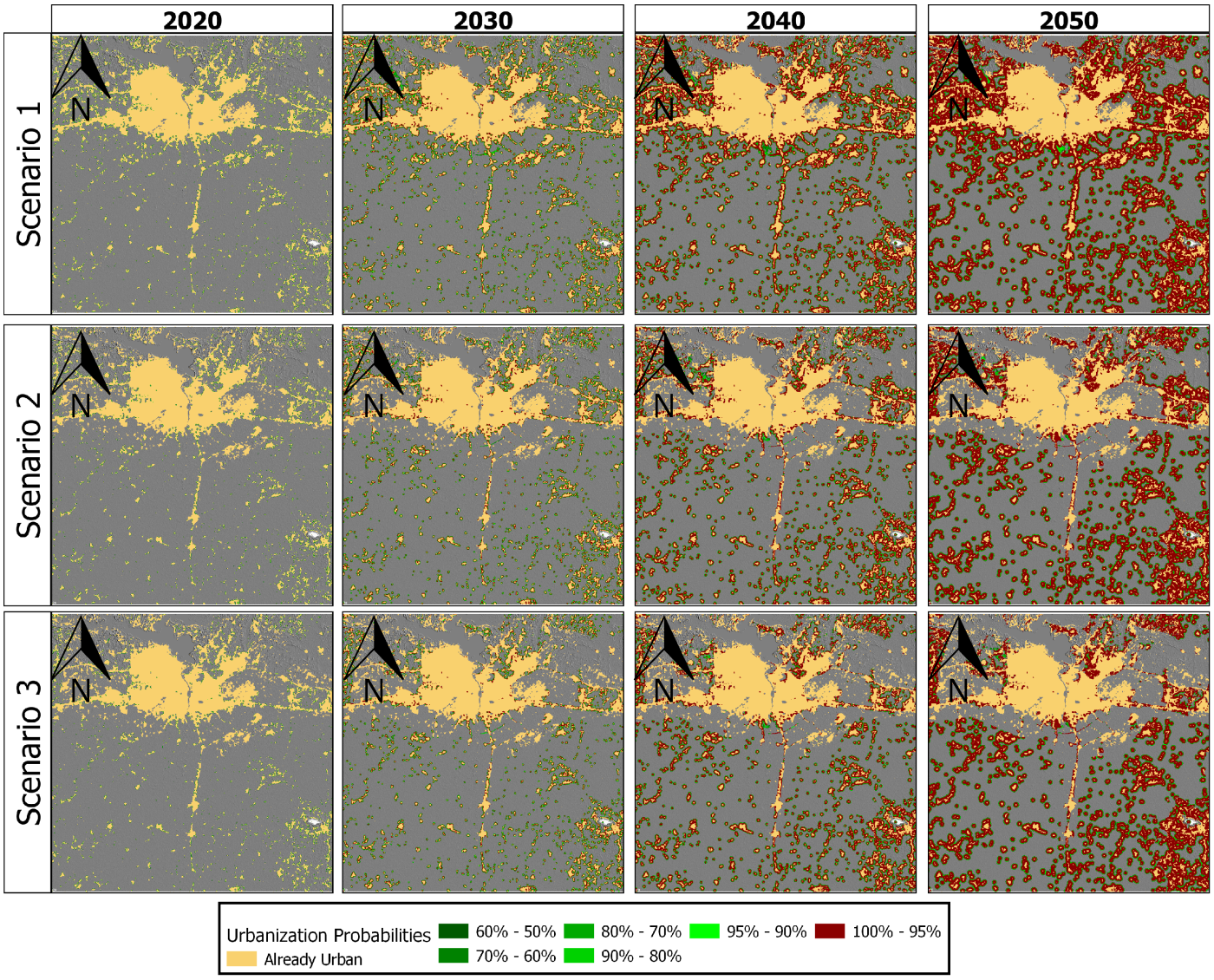

Figure 5.11: Urbanization predictions between 2020 to 2050 for each scenario

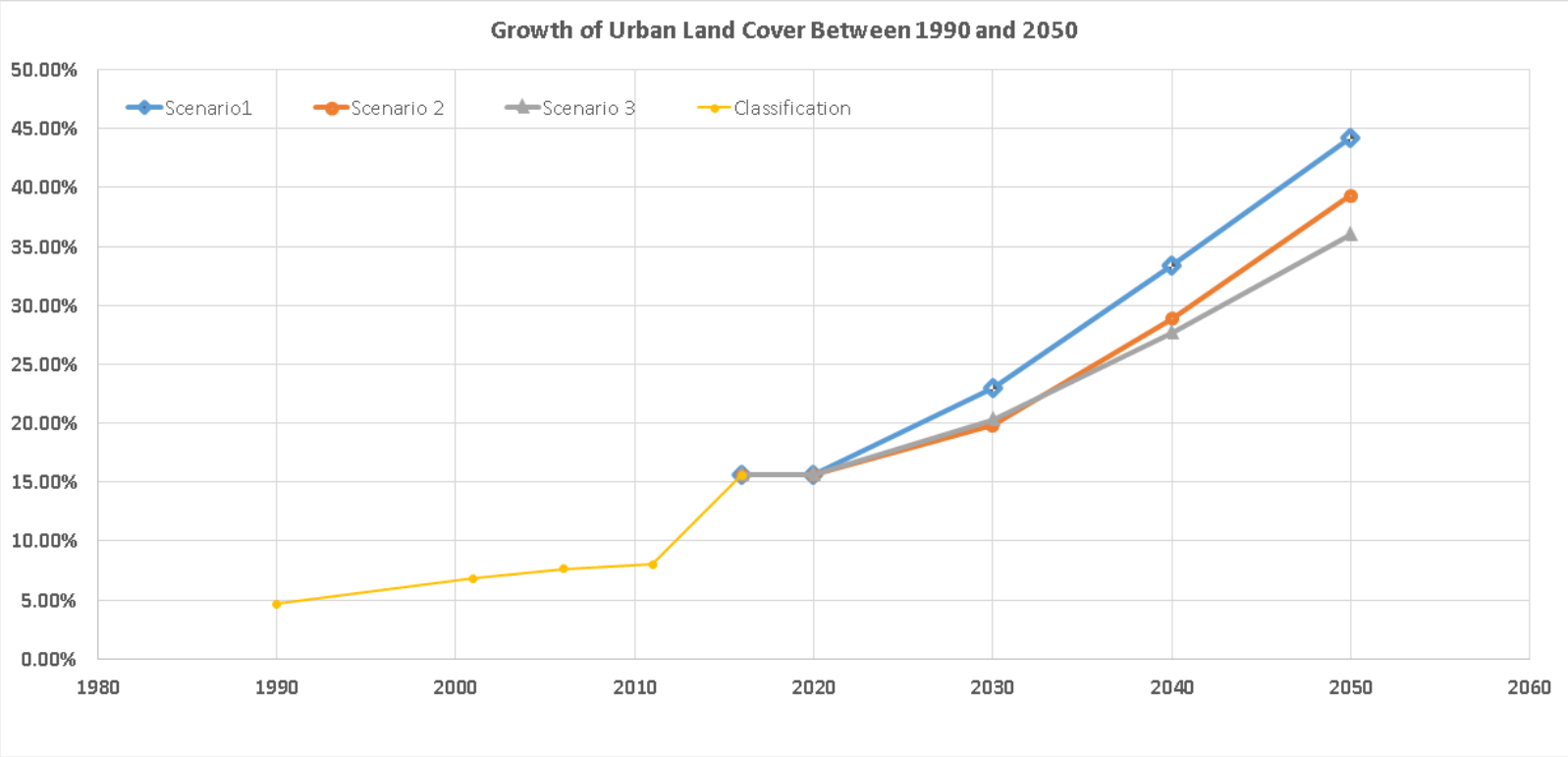

Figure 5.12: Urban growth prediction quantities from 1990-2050

From Figure 5.13, it can be understood that partial and full protection scenarios protect the green areas most.

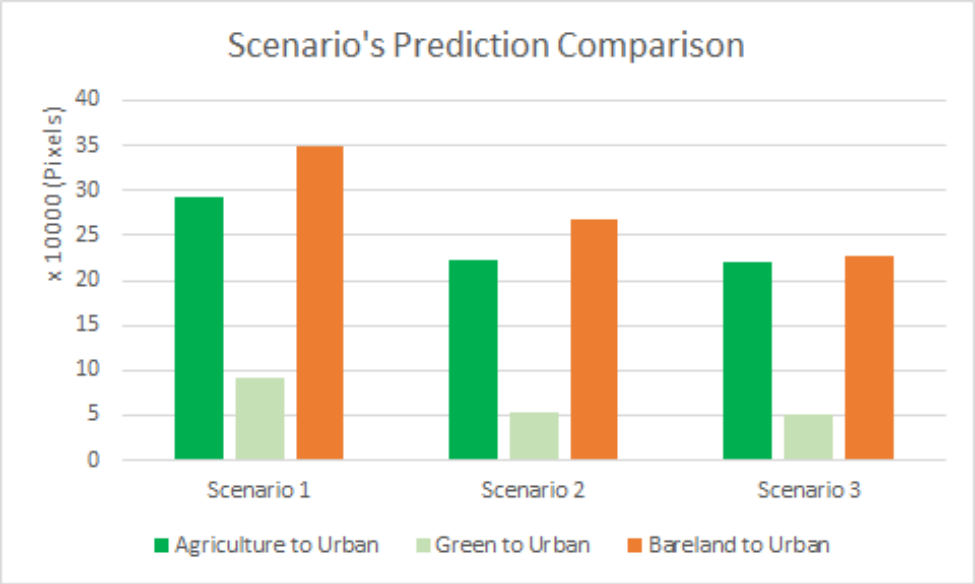

Figure 5.13: Quantitively comparing of scenario predictions

5.4 Investigation of Road Layer Effect for Year 2030

Urban layer coverage is between 1990 and 2016. Input data of five years in this period are used to predict urbanization patterns. In other words, SLEUTH fills the gaps between input data years. Road layer 2020 is out of coverage and can't be used directly for calculations. SLEUTH works by taking into account only the layers that dated between the dates of the urban layers. Using 2020 road data could only be possible if an urban layer of 2020 or a later date is given as an input. Future urban layer can be created using SLEUTH with existing data. Running the prediction to any year later than 2020 and using predicted urban layer as input for another prediction could be a method for predicting effects of a future (planned) road layer.

In order to obtain a future urban layer to use 2020 road layer; existing urban data (from 1990 to 2016) used to predict year 2030. SLEUTH creates outputs for each year from the recent urban data (2016) to a specified prediction date (e.g: 2030). Products of this prediction are used in 2050 predictions. Eventually, two layers for

2030 are obtained, one without road data and one with the road data. Comparison of two layers can be seen in Figure 5.14.

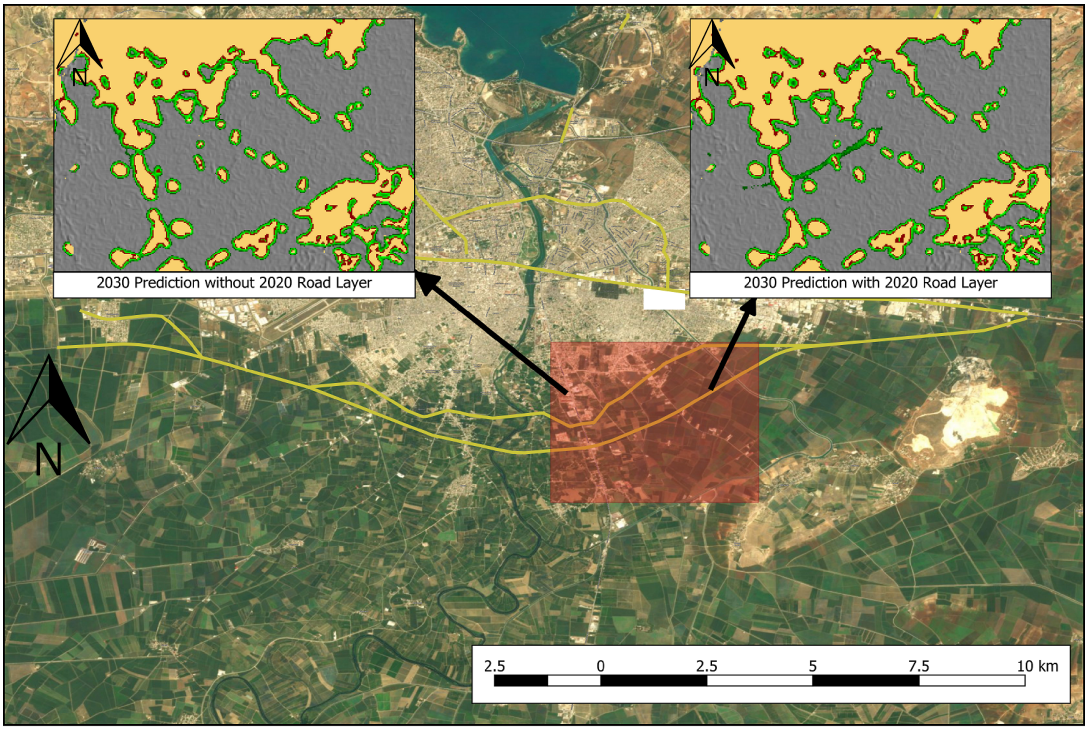

Figure 5.14: 2030 Prediction with and without planned road

In this study all predictions are aggregated and binary maps are created for predictions according to a threshold which is 90% or above probability of urbanization. When this threshold is applied to the prediction maps, two binary images are obtained. Comparison of these imagery showed slight differences in terms of urbanization. 2020 road layer's effect to 2030 predictions is insignificant. Neither a cumulation on the vicinity of planned road nor a general increase in study area were observed. Decreasing threshold value could have been created a difference but for this study 90% urbanization probability is considered as an acceptable threshold. Road effect become visible in 2050 prediction even with 90% probability threshold.

According to predictions road layer effect become significant after the year 2030. Each scenario reacts differently as expected; in the first scenario significant urbanization begin to form in year 2031, in the second and third scenarios significant urbanization begin to form in year 2034. These urbanization changes for eight year periods

can be seen in Figures 5.15 - 5.17.

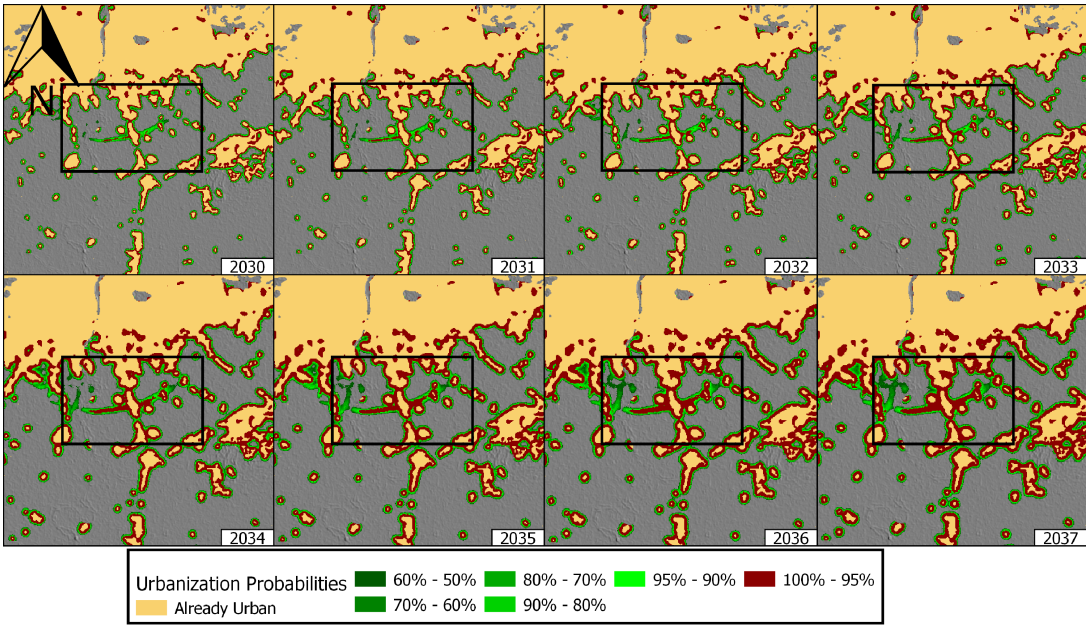

Figure 5.15: Road effect on urbanization in the first scenario from 2030 to 2037

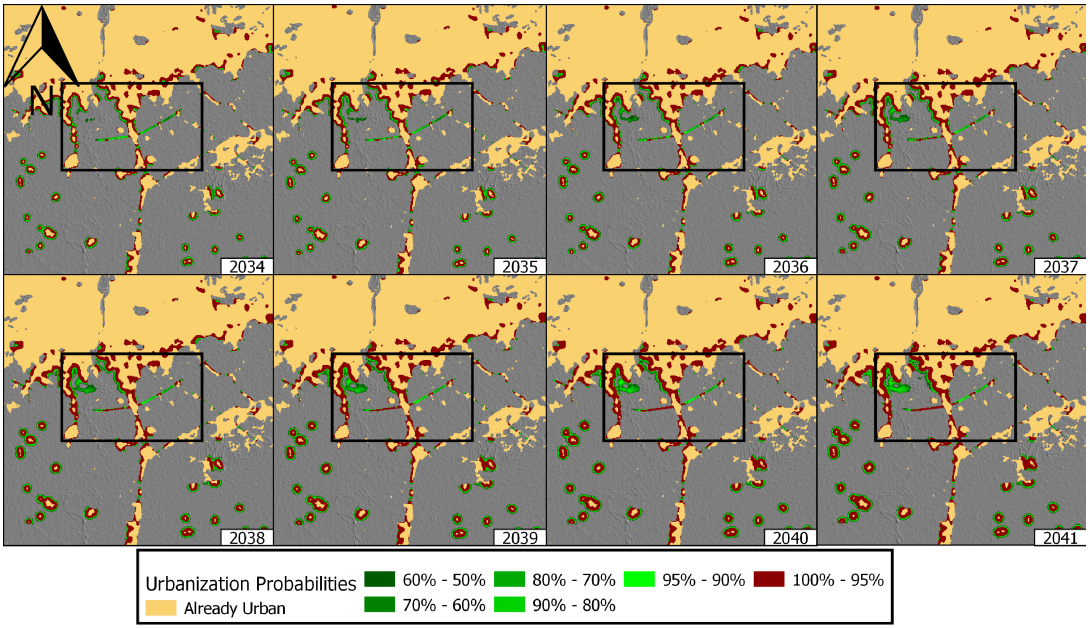

Figure 5.16: Road effect on urbanization in the second scenario from 2034 to 2041

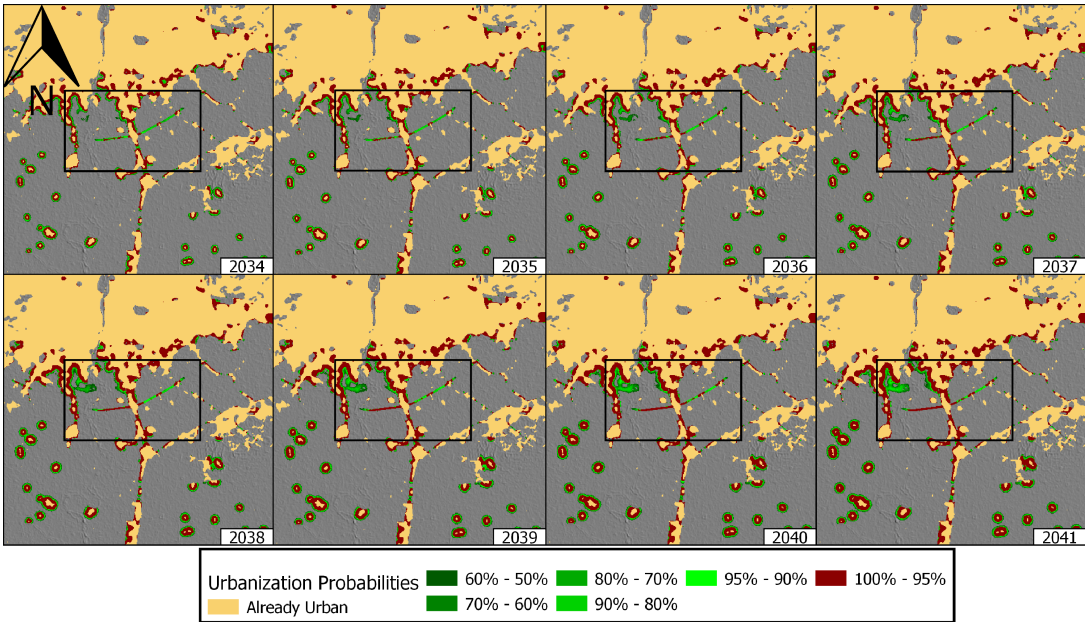

Figure 5.17: Road effect on urbanization in the third scenario from 2034 to 2041

5.5 Comparison of the Study Results with Other SLEUTH Applications

All predictions until this point are the results of a compact prediction framework. Comparing these results with other SLEUTH applications or establishing a logical relation between the applications is not an easy task due to large number of parameters. Each region has it's own characteristic urbanization pattern due to environmental factors. Even so, five parameters that SLEUTH model produces can capture the urbanization pattern of the area when it is calibrated successfully (Dietzel and Clarke, 2007).

Gazulis and Clarke (2006) offered a method which considers SLEUTH coefficients as DNA of modelling region. With this approach urbanization patterns of the cities can be compared and classifications according to these patterns can be made. First inter-city comparison of the SLEUTH application were made by Dietzel and Clarke (2007). In the study, they've used parameter values of 21 SLEUTH applications that have been known by the time. In their study they've pointed cities located near the sea coast in North America show similar urbanization patterns, cities in countries which apply a very careful planning process such as Netherlands, show very low diffusion

parameters.

Instead of parametric comparison like Dietzel and Clarke (2007), a similar approach in the study of Gazulis and Clarke (2006) has been adopted. Four sample input layers prepared as input for petri-dish experiment. Explanation of input data can be seen below;

1. **Urban layer**
   Consist of only one urban pixel in the middle of the image.
2. **Slope Layer**
   Test data generated using Gaussian point spread function. Value interval between 0-100. Slope value increases through the edges.
3. **Road Layer**
   Each middle pixel on the rows marked as road.
4. **Excluded Layer**
   Input extent divided into two areas as North and South from the middle. Pixels inside the Northern part marked 50% resistive to urbanization. Southern parts shows no resistance to urbanization.

Since Gazulis and Clarke (2006) don't make any clear statement about the size of input layers, each input layer is prepared in 1001x1001 size. These layers can be seen in Figure 5.18. Results of the studies in Turkey have been compared with the method. Only, prediction best fitting values are taken from the studies (please look at the Table 5.5). Using the prediction values and synthetic data, SLEUTH was run in prediction mode.

Petri-dish experiment comparison of this study with Tampa and Oahu can be seen in Figure 5.20. When the Berberoglu's and Sevik's SLEUTH applications in Turkey investigated, these applications showed a different characteristics when compared with this study. In both studies even after 100 years of prediction run, pixels couldn't reach urbanization probability higher than 50%. Petri-dish experiment results of these two studies are not shown with the other studies. They can be seen separately in Figure 5.21, because their petri-dish experiment results have different (lower) probability

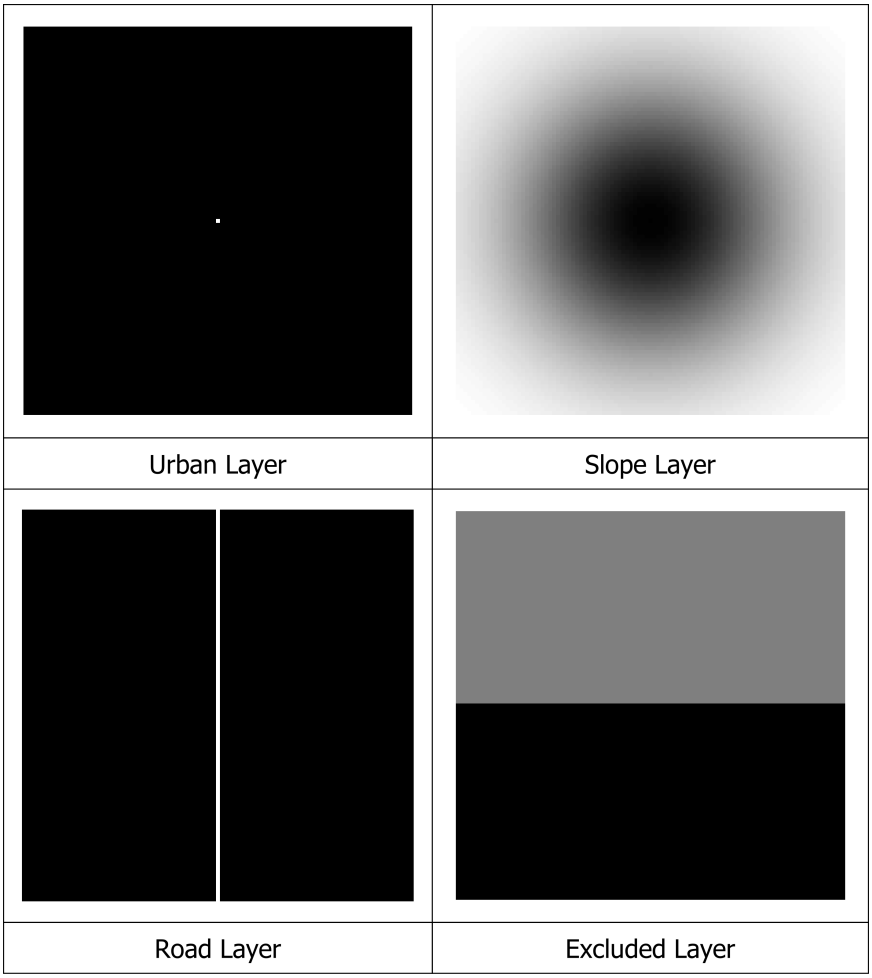

Figure 5.18: Synthetic input data for prediction comparison

range. SLEUTH creates output images for each year at the prediction range. These images only show pixels with urbanization probability higher than 50%.

Table 5.5: Prediction coefficients of the studies compared in petri-dish experiment

Reference Study	**City**	**Diffusion**	**Breed**	**Spread**	**Slope**	**Road Gravity**
This Study	Adana, Turkey	5	100	64	1	72
Gazulis and Clarke (2006)	Tampa, Florida	90	95	45	50	50
Gazulis and Clarke (2006)	Oahu, Hawaii	5	96	12	1	50
Berberoglu et al. (2016)	Adana, Turkey (Prior study)	1	30	62	48	25
Sevik (2006)	Antalya	1	37	23	52	27

These prediction coefficients are important because they show which coefficient type is more effective on urbanization of a city. From these coefficients urbanization patterns of cities can be compared. Breed coefficient seems to be the most efficient coefficient for the urbanization of Adana, in contrast to the prior study, in which spread coefficient seems to be the most effective coefficient. This change could be explained by the change of the urbanization characteristics in the area or the difference of the study area extend. Prior study area for Adana includes smaller area extent which includes mostly planned areas. Since the urbanization is constrained, to by the planned area extent less urbanization is observed and coefficients for this area have been changed accordingly. Study area extents are compared in Figure 5.19, shaded area refers to prior study area. Besides the breed coefficient, spread and road gravity seemed to be important for Adana. Diffusion and slope have the least effect. Urban edges are not necessarily become urban, instead randomly selected new urban spread centres will urbanize the unprotected areas.

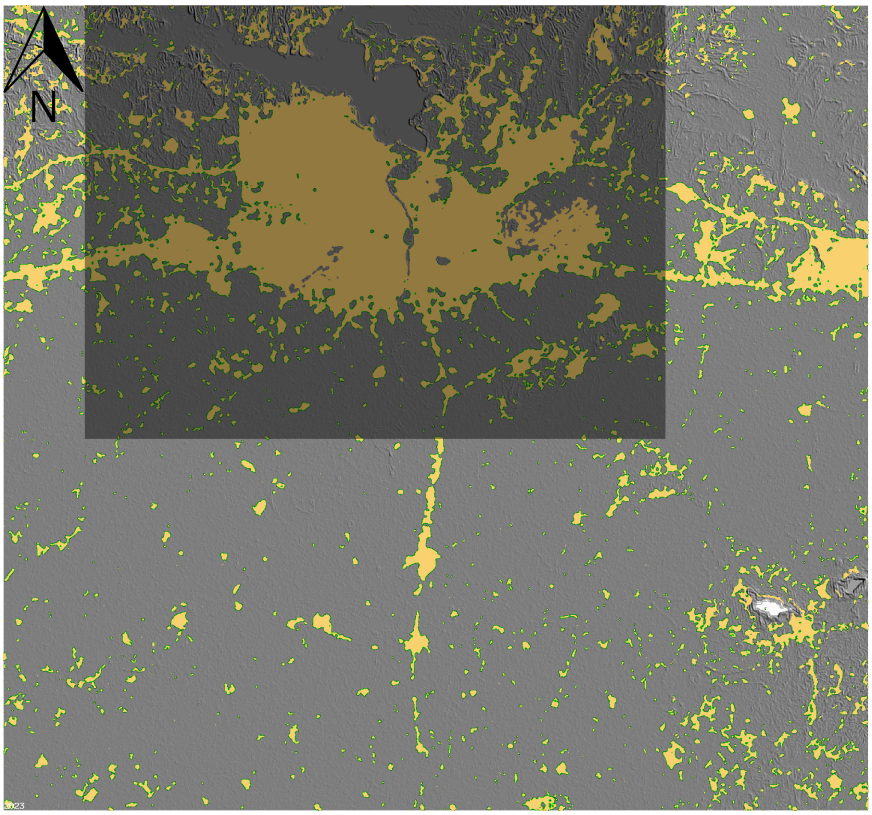

Figure 5.19: Comparison of study area extents of this study and prior study (Shaded area is prior study area)

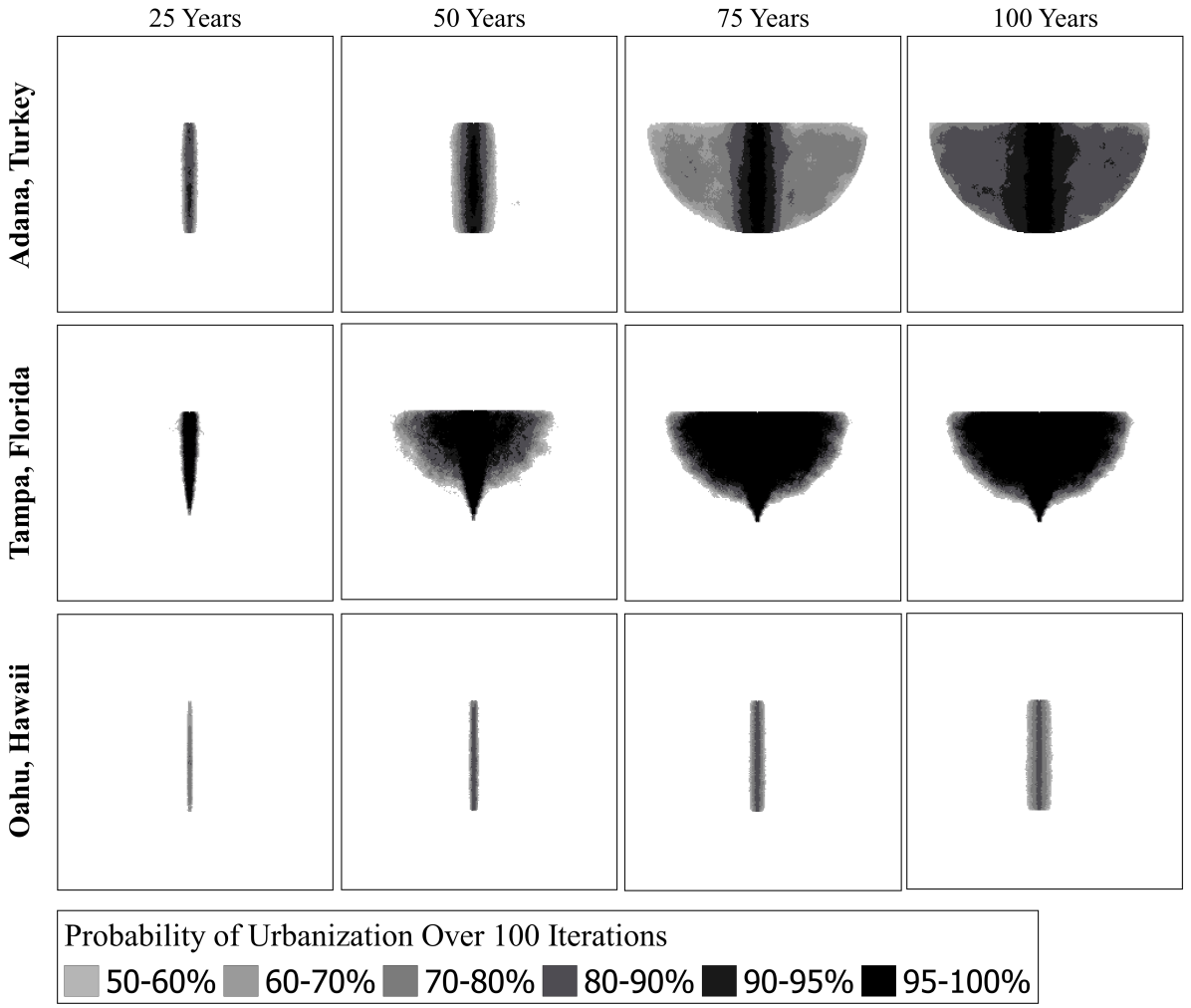

Figure 5.20: Comparison of Adana urbanization characteristics with Oahu and Tampa

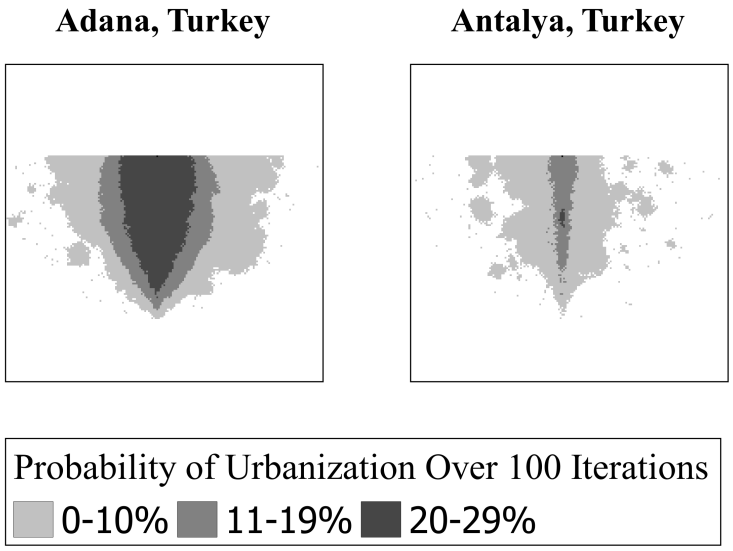

Figure 5.21: Previous Adana study (Berberoglu et al., 2016) and Antalya (Sevik, 2006) studies urbanization characteristics

5.6 Discussion of 2018 Data Verification

As stated in the Section 3.2.4, a validation method was applied and 11% more urbanization was detected for 2018. To be able to run a validation process that uses the probability maps generated by the SLEUTH model, a threshold range must be determined. This threshold value makes it possible to generate a binary map from a probability map. In this study, as previously mentioned, the urbanization probability threshold value was determined as 90%. In this case, if a higher value is selected, an urbanization of less than 11% can be predicted.

CHAPTER 6

CONCLUSIONS

In this chapter, the common and non-common aspects of this study with other studies in the literature are discussed and how the results obtained in this study can be used in future is explained.

Based on the growth of Adana province in 26 years, starting from 1990 to 2016, urbanization predictions for the year 2050 are obtained for each of three scenarios. These predictions are the results of a two dimensional space and predictions about vertical urbanization is not a matter of SLEUTH application. The effects of vertical urbanization have not been discussed in this study, since horizontal urbanization is the main reason for the disappearance of agricultural and green areas around the cities. SLEUTH urban growth model predicted the 2018 urbanization using the data between 1990-2016 11% more compared to the classification image of 2018.

Model validation with a 2018 satellite image of the SLEUTH model increased confidence to the model indicating that the model gives realistic results. Slightly more (11%) urbanization observed in 2018 prediction when compared to the classification results. There could be two different reasons responsible for this over-prediction; classification accuracy or self-modification parameters. In a future research; self-modification parameters can be edited for the study area.

According to predictions; urbanization is observed on three types of land classes; agriculture, green and bareland. Bareland is the most affected class from the urbanization. After barelands, agricultural areas are the second most affected land class. Loss of agricultural areas could also have an economical impact because of crops grown in that part of soil. If urbanization continue with the ongoing trends (with-

out any restriction like in Scenario 1) 264 km^2 agricultural area and 82 km^2 green area will become urban by 2050. If agricultural areas and green areas are partially protected (e.g: urbanization is forbidden, but no strict control) 200 km^2 agricultural area and 47 km^2 green area will become urban by 2050. Finally, if urbanization is fully restricted on green, agricultural areas 197 km^2 agricultural area and 46 km^2 area will become urban by 2050. Although the urbanization of agricultural and green areas under full protection is expected to be impossible, urbanization restriction is applied only inside the planned areas. As a result, unplanned areas continued to be urbanized.

As Clarke (2016) points out, the meticulous preparation and adaptation of the data has been the most labour-intensive part of using SLEUTH model. If satellite imagery classification and preparation of input raster layers could have been done easily or retrieved from a pre-prepared data source, researchers can spend more time on model behaviour. High Resolution Layers (HRL) at this point could be an option if urbanization after 2006 is going to be modelled in Europe. HRL is prepared using Sentinel-1 and Sentinel-2 data. Impervious area layer of HRL data can be used as urban layer in SLEUTH. Though if agricultural and green area loss is aimed to be found; satellite imagery classification must be done for the study area.

One of the issues that confuses the mind of the SLEUTH researchers most is determination of calibration parameters. In this study, similar to other academic studies, OSM has been used as a calibration metric. Comparison of this study's outcomes with other SLEUTH applications has been done using Gazulis and Clarke (2006)'s methodology. Similar experiments are carried out using the prediction parameters of the other SLEUTH studies. As a result of these experiments, it was revealed that Adana is one of the cities with road dependent growth pattern. Same experiments have been repeated for the Oahu and Tampa data from Gazulis and Clarke (2006)'s study to verify experiments correctness. Results showed that the algorithm works correctly. For the previous studies conducted by Sevik (2006) in Antalya and Berberoglu et al. (2016) in Adana, the same petri-dish test was carried out, but the urbanization was very low. Therefore, the cumulative urbanization results of the applications instead of the urbanization prediction maps obtained in 25-year periods are added to the study.

This study shows the possible impact of the road planned to be constructed in 2020 in terms of urbanization to agricultural and green areas and the extent to which the same areas can be affected under different growth scenarios. The results of the study were generally in the expected direction. After the road is constructed in 2020, a small amount of urbanization is observed around the new road, even in a fully preserved scenario. As expected; in the third most conservative scenario, minimum urbanization is observed; urbanization is suppressed. In the case of no protection, more urbanization is observed compared to the maximum protection scenario. Unexpected results are also obtained; for example, in the second scenario, where there is partially protected areas, more urbanization is observed when compared to the no-protection scenario.

One of the important results of this study is that urban sprawl trend in urban areas affects the urban sprawl trend in rural areas. As discussed in the Chapter 5.3, when a scene selected which has both urban and rural settlements is selected urbanization trends are mixed together. Urban and rural areas do not urbanize in the same way. The fact that the results of the previous study Berberoglu et al. (2016) in Adana are different from the results of this study can be shown as evidence for this situation. As shown in the Figure 5.19 previous study has been done in a smaller part of the region. This shows two important points; first, if the SLEUTH model studies in the same region are to be compared, the same study extents should be selected. Secondly, the city layer of the SLEUTH model has two different classes, urban and rural, considering different urbanization behaviour of these classes may lead to more accurate results.

Similar studies to this one can be applied to other major cities such as Ankara, İstanbul and İzmir to find urbanization characteristics of these cities. Possible agricultural and green area losses at the edges of these cities can be forecasted and used in planning and decision making. To protect the ecological environment and promoting sustainable development of the region, relevant decision makers should create effective strategies to control urban sprawl.

BIBLIOGRAPHY

Agyemang, F. S. and Silva, E. (2019). Simulating the urban growth of a predominantly informal Ghanaian city-region with a cellular automata model: Implications for urban planning and policy. *Applied Geography*, 105(August 2017):15–24.

Anderson, J. R. (1976). *A land use and land cover classification system for use with remote sensor data*, volume 964. US Government Printing Office.

Anonymous (2018). Cygwin. `https://www.cygwin.com/`. Online; accessed 24 July 2018.

Arasan, V. T. (2014). Urban transportation planning. `https://nptel.ac.in/courses/105107067/`. Online; accessed 25 June 2018.

Beitzel, S. M., Jensen, E. C., Frieder, O., Lewis, D. D., Chowdhury, A., and Kołcz, A. (2005). Improving automatic query classification via semi-supervised learning. *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 42–49.

Berberoglu, S., Akin, A., and Clarke, K. C. (2016). Cellular automata modeling approaches to forecast urban growth for adana, Turkey: A comparative approach. *Landscape and Urban Planning*, 153:11–27.

Berling-Wolff, S. and Wu, J. (2004). Modeling urban landscape dynamics: A review. *Ecological Research*, 19(1):119–129.

Chaudhuri, G. and Clarke, K. C. (2014). Temporal Accuracy in Urban Growth Forecasting: A Study Using the SLEUTH Model. *Transactions in GIS*, 18(2):302–320.

Choodarathnakara, A. L., Kumar, D. A., Koliwad, D., and Patil, D. (2012). Mixed Pixels : A Challenge in Remote Sensing Data Classification for Improving Per-

formance. *International Journal of Advanced Research in Computer Engineering & Technology*, 1(9):261–271.

Clarke (2016). SLEUTH Master Class Video. `http://www.geog.ucsb.edu/{~}kclarke/Public/SLEUTHClass.mp4`. Online; accessed 03 June 2018.

Clarke, K., Hoppen, S., and Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment & Planning B-Planning & Design*, 24:247–261.

Clarke, K. C. (2001). Project Gigalopolis, NCGIA. `http://www.ncgia.ucsb.edu/projects/gig/Imp/calOSM.htmurl:http://www.ncgia.ucsb.edu/projects/gig/`. Online; accessed 25 June 2018.

Clarke, K. C. (2018). SLEUTH Naming Convention. `http://www.ncgia.ucsb.edu/projects/gig/Imp/imSetUp.htm{#}namingConvention`. Online; accessed 25 June 2018.

Clarke, K. C., Gazulis, N., Dietzel, C., and Goldstein, N. C. (2007). A decade of SLEUTHing: lessons learned from applications of a cellular automaton land use change model. *Classics in IJGIS Twenty years of the International Journal of Geographical Information Science and Systems*, pages 413–425.

Clarke, K. C., Hoppen, S., and Gaydos, L. J. (1996). Methods And Techniques for Rigorous Calibration of a Cellular Automaton Model of Urban Growth. *Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, pages 21–25.

Dietzel, C. and Clarke, K. C. (2004). Spatial differences in multi-resolution urban automata modeling. *Transactions in GIS*, 8(4):479–492.

Dietzel, C. and Clarke, K. C. (2007). Research Article Toward Optimal Calibration of the SLEUTH Land Use Change Model. *Transactions in GIS*, 11(1):29– 45.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). Pattern Classification.

Gazulis, N. and Clarke, K. (2006). Exploring the DNA of our regions: classification of outputs from the SLEUTH model. *Cellular Automata*, pages 462–471.

Google (2019). Google Ngram results of urbanization and remote sensing.

Jensen, J. R. (2005). Thematic map accuracy assessment. *Introductory Digital Image Processing: A Remote Sensing Perspective*, pages 476–482.

Komurcuoglu, M. (2013). The reasons of immorality of land development in turkey. *Turkish Journal of Business Ethics May*, 6(1):26–34.

Lee, D. B. (1973). Requiem for Large-Scale Models. *Journal of the American Planning Association*, 39(3):163–178.

Li, X. and Gong, P. (2016). Urban growth models: progress and perspective. *Science Bulletin*, 61(21):1637–1650.

Liu, X., Liang, X., Li, X., Xu, X., Ou, J., Chen, Y., Li, S., Wang, S., and Pei, F. (2017). A future land use simulation model (flus) for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning*, 168:94–116.

Meentemeyer, R. K., Tang, W., Dorning, M. A., Vogler, J. B., Cunniffe, N. J., and Shoemaker, D. A. (2013). FUTURES: Multilevel Simulations of Emerging Urban-Rural Landscape Structure Using a Stochastic Patch-Growing Algorithm. *Annals of the Association of American Geographers*, 103(4):785–807.

Nucci, L., Rahnama, M. R., and Heydari, A. (2016). Analysis the Process and Priorities of Urbanization in the Mashhad city After Islamic Revolution of Iran. *International Journal of Humanities and Cultural Studies ISSN 2356-5926*, 2(4):653–668.

Prime Ministry Legislation Development and Publication (1986). 24.2.1984 tarih ve 2981 sayili kanunun bazi maddelerinin degistirilmesi ve bu kanuna bazi maddeler eklenmesi hakkinda kanun. `http://www.resmigazete.gov.tr/arsiv/19130.pdf`. Online; accessed 24 July 2018.

Sahana, M., Hong, H., and Sajjad, H. (2018). Analyzing urban spatial patterns and

trend of urban growth using urban sprawl matrix: A study on Kolkata urban agglomeration, India. *Science of the Total Environment*, 628-629:1557–1566.

Santé, I., García, A. M., Miranda, D., and Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, 96(2):108–122.

Saxena, A. and Jat, M. K. (2019). Capturing heterogeneous urban growth using SLEUTH model. *Remote Sensing Applications: Society and Environment*, 13(December 2018):426–434.

Schiffman, D. (2019). Cellular automata.

Seto, K. C., Kaufmann, R. K., and Woodcock, C. E. (2000). Landsat reveals China's farmland reserves, but they're vanishing fast. *Nature*, 406(6792):121–121.

Sevik, O. (2006). Application of Sleuth Model in Antalya. Master's thesis, Middle East Technical University, Ankara.

Shackelford, A. K. and Davis, C. H. (2003). A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *IEEE Transactions on GeoScience and Remote sensing*, 41(10):2354–2363.

Silva, E. A. and Clarke, K. C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems*, 26(6):525–552.

TUIK (2018). Population Statistics. `http://www.tuik.gov.tr/VeriTabanlari.do?ust_id=109&vt_id=27`. Online; accessed 24 July 2018.

UN (2016). United Nations, Department of Economic and Social Affairs, Population Division (2016). The World's Cities in 2016 – Data Booklet (ST/ESA/ SER.A/392). `http://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf`. Online; accessed 27 December 2018.

UN (2017). United Nations, Department of Economic and Social Affairs, Population Division (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. ESA/P/WP/248. https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf. Online; accessed 27 December 2018.

USGS (2018a). Landsat Collection 1 Level-1 Processing Levels. https://www.usgs.gov/media/images/landsat-collection-1-level-1-processing-levels. Online; accessed 24 July 2018.

USGS (2018b). Using Landsat-8 Product. https://landsat.usgs.gov/using-usgs-landsat-8-product. Online; accessed 03 June 2018.

Wolfram, S. (1984). Universality and complexity in cellular automata. *Physica D: Nonlinear Phenomena*, 10(1-2):1–35.

Yakup, A., Gul, F., Kotay, D., Ayazli, I., and Baslik, S. (2018). Extracting an Urban Growth Model's Land Cover Layer from Spatio-Temporal Cadastral Database and Simulation Application. *Polish Journal of Environmental Studies*, 28(3):1063–1069.

Yale University, Center for Earth Observation (2018). How to convert Landsat DNs to Top of Atmosphere (ToA) Reflectance. https://yceo.yale.edu/how-convert-landsat-dns-top-atmosphere-toa-reflectance. Online; accessed 03 June 2018.

APPENDIX A

CONFUSION MATRICES OF CLASSIFIED SATELLITE IMAGERY

Table A.1: Confusion Matrix of Classification Result of 2018 Imagery

class	Agriculture	Bareland	Forest	Urban	Water	Row Total
Agriculture	97	2	0	1	0	100
Bareland	9	89	1	1	0	100
Forest	20	4	73	2	1	100
Urban	3	2	4	91	0	100
Water	0	3	2	2	93	100
Column Total	129	100	80	97	94	500

Table A.2: Confusion Matrix of Classification Result of 2016 Imagery

class	Agriculture	Bareland	Forest	Urban	Water	Row Total
Agriculture	96	1	0	3	0	100
Bareland	16	80	4	0	0	100
Forest	27	5	54	11	3	100
Urban	10	4	5	81	0	100
Water	4	0	3	0	93	100
Column Total	153	90	66	95	96	500

Table A.3: Confusion Matrix of Classification Result of 2011 Imagery

class	Agriculture	Bareland	Forest	Urban	Water	Row Total
Agriculture	76	10	13	1	0	100
Bareland	9	85	6	0	0	100
Forest	2	2	96	0	0	100
Urban	1	2	3	94	0	100
Water	1	0	0	0	99	100
Column Total	89	99	118	95	99	500

Table A.4: Confusion Matrix of Classification Result of 2006 Imagery

class	Agriculture	Bareland	Forest	Urban	Water	Row Total
Agriculture	77	19	2	2	0	100
Bareland	2	92	3	3	0	100
Forest	3	1	96	0	0	100
Urban	4	5	10	81	0	100
Water	0	0	0	0	100	100
Column Total	86	117	111	86	100	500

Table A.5: Confusion Matrix of Classification Result of 2001 Imagery

class	Agriculture	Bareland	Forest	Urban	Water	Row Total
Agriculture	80	17	1	2	0	100
Bareland	0	93	6	1	0	100
Forest	0	1	98	1	0	100
Urban	0	0	5	95	0	100
Water	0	0	0	0	100	100
Column Total	80	111	110	99	100	500

Table A.6: Confusion Matrix of Classification Result of 1990 Imagery

Class	Agriculture	Bareland	Forest	Urban	Water	Row Total
Agriculture	90	7	2	1	0	100
Bare land	0	91	2	7	0	100
Forest	0	0	97	3	0	100
Urban	10	0	6	84	0	100
Water	0	0	0	0	100	100
Column Total	100	98	107	95	100	500

APPENDIX B

OVERALL ACCURACY ASSESSMENT TABLES OF SATELLITE IMAGERY

Table B.1: Overall Accuracy Report of 2018 Classification

overall acc.	class	producers acc	User's accuracy
0.87	agriculture	0.75	0.97
	bareland	0.89	0.89
Kappa	forest	0.91	0.73
0.86	urban	0.94	0.91
	water	0.99	0.93

Table B.2: Overall Accuracy Report of 2016 Classification

Overall Accuracy	Class Type	Producer's Accuracy	User's Accuracy
0.81	agriculture	0.63	0.96
	bare land	0.89	0.8
Kappa	forest	0.81	0.54
0.76	urban	0.85	0.81
	water	0.97	0.93

Table B.3: Overall Accuracy Report of 2011 Classification

overall acc.	class type	producers accuracy	user's accuracy
0.9	agriculture	0.85	0.76
	bare land	0.86	0.85
	forest	0.81	0.96
kappa	urban	0.99	0.94
0.88	water	1	0.99

Table B.4: Overall Accuracy Report of 2006 Classification

overall acc.	Class Data	producer's accuracy	user's accuracy
0.89	agriculture	0.90	0.77
	bareland	0.79	0.92
kappa	forest	0.86	0.96
0.87	urban	0.94	0.81
	water	1	1

Table B.5: Overall Accuracy Report of 2001 Classification

overall acc.	Class	Producers Accuracy	User's Accuracy
0.93	agriculture	1	0.8
	bareland	0.84	0.93
kappa	forest	0.89	0.98
0.92	urban	0.96	0.95
	water	1	1

Table B.6: Overall Accuracy Report of 1990 Classification

Overall Accuracy		Producers Accuracy	Accuracy
0.92	agriculture	0.9	0.9
	bareland	0.93	0.91
kappa	forest	0.91	0.97
0.91	urban	0.88	0.84
	water	1	1

APPENDIX C

ROAD AGGREGATION SCHEMA

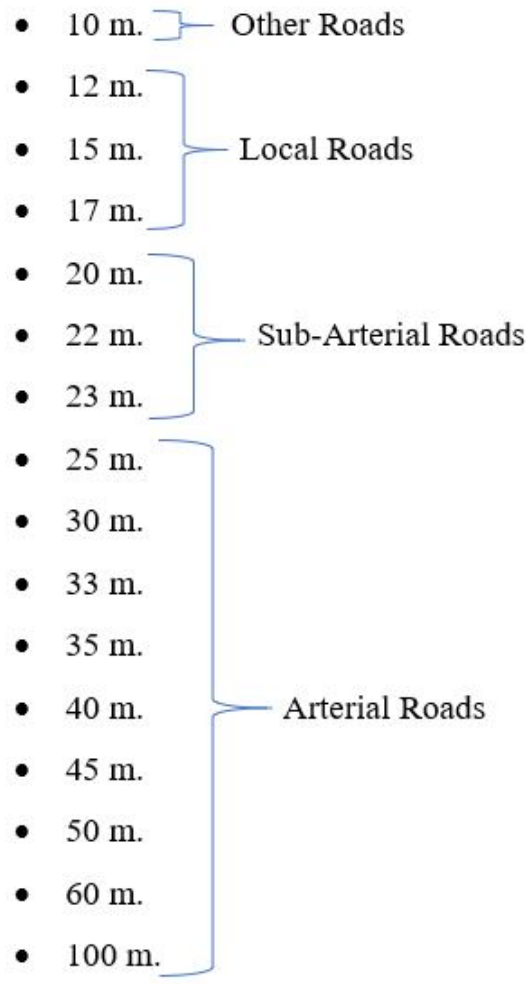

Figure C.1: Aggregation Schema of Roads

APPENDIX D

SCENARIO FILES

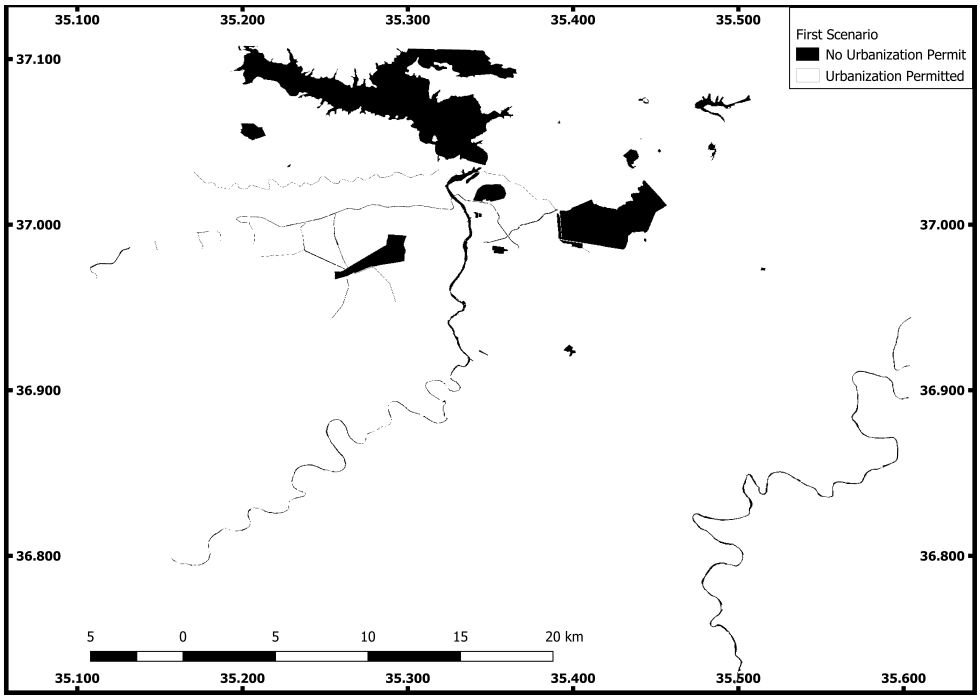

Figure D.1: First Scenario File

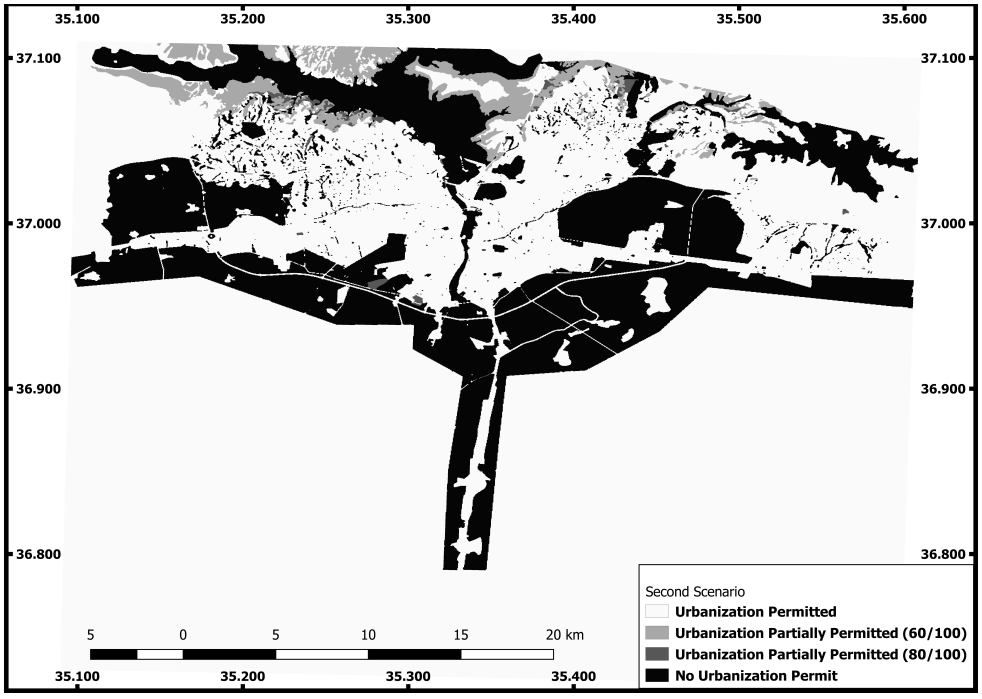

Figure D.2: Second Scenario File

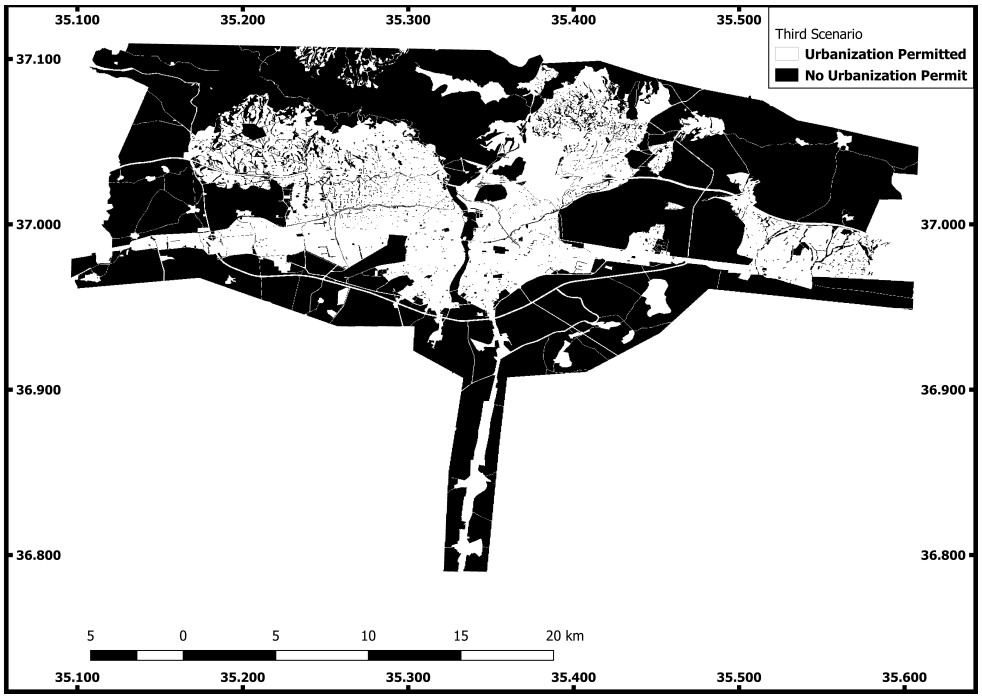

Figure D.3: Third Scenario File

APPENDIX E

FILE NAMES ACCORDING TO SLEUTH NAMING CONVENTION

adana30.urban.2016.gif
adana30.urban.2011.gif
adana30.urban.2006.gif
adana30.urban.2001.gif
adana30.urban.1990.gif
adana30.slope.gif
adana30.hillshade.gif
adana30.excluded.gif
adana30.roads.2006.gif
adana30.roads.2016.gif

Figure E.1: Reorganized Layer Names According to SLEUTH Naming Style

APPENDIX F

CALIBRATION COEFFICIENTS (OSM) OF COARSE, FINE AND FINAL CALIBRATION

1	OSM	Diff	Brd	Sprd	Slp	Road
2	0.163352	75	1	50	1	1
3	0.163005	75	1	50	50	25
4	0.153587	75	1	50	50	1
5	0.150084	75	1	50	50	75
6	0.150084	75	1	50	50	100
7	0.149731	75	1	50	50	50
8	0.149165	50	1	75	25	1
9	0.147058	75	1	50	1	50
10	0.143872	75	1	50	1	25
11	0.142844	1	100	25	25	75
12	0.140352	100	1	50	75	75
13	0.140352	100	1	50	75	100
14	0.139389	75	1	50	75	1
15	0.137109	100	1	50	100	1
16	0.136953	75	1	50	25	50
17	0.136953	75	1	50	25	75
18	0.1363	50	1	100	75	50
19	0.13394	75	1	50	75	50
20	0.13394	75	1	50	75	75
21	0.13394	75	1	50	75	100
22	0.133473	50	1	100	75	75
23	0.133473	50	1	100	75	100
24	0.13023	50	1	75	75	25
25	0.130149	75	1	50	25	1
26	0.12964	25	25	25	50	1
27	0.129627	100	1	50	50	1
28	0.128907	25	25	25	25	50
29	0.128735	100	1	50	100	50
30	0.128205	50	1	100	100	75
31	0.128205	50	1	100	100	100
32	0.12803	100	1	50	25	1
33	0.128016	100	1	50	100	75
34	0.128016	100	1	50	100	100
35	0.12618	25	25	25	75	100
36	0.126001	25	25	25	1	50
37	0.125404	75	1	50	1	75
38	0.125404	75	1	50	1	100
39	0.124085	50	1	75	50	1
40	0.122881	75	1	50	25	25
41	0.122177	50	1	75	100	75
42	0.122177	50	1	75	100	100
43	0.121891	75	1	50	100	50
44	0.12091	100	1	50	75	50
45	0.119075	75	1	75	50	100
46	0.11718	100	1	50	75	1
47	0.116937	100	1	50	1	25
48	0.116598	25	25	25	100	75
49	0.11457	50	1	100	1	50
50	0.11389	75	1	50	75	25
51	0.113679	75	1	50	100	1

Figure F.1: Coarse Calibration Coefficients with Highest OSM Values

1	OSM	Diff	Brd	Sprd	Slp	Road
2	0.106977	60	16	65	29	19
3	0.104427	50	16	65	1	25
4	0.103707	50	16	65	1	1
5	0.102765	50	16	65	36	13
6	0.102616	50	16	60	1	13
7	0.102461	50	16	70	1	1
8	0.102378	55	16	75	15	19
9	0.102222	50	16	75	36	25
10	0.101868	50	16	75	50	7
11	0.10092	65	16	55	8	7
12	0.100778	75	16	50	8	1
13	0.1007	55	16	75	1	1
14	0.100433	55	16	70	43	13
15	0.099554	50	16	70	1	7
16	0.0995	55	16	65	8	7
17	0.099413	55	16	70	29	13
18	0.09938	50	16	75	22	13
19	0.099225	65	16	55	1	7
20	0.099153	55	16	75	8	7
21	0.099109	50	16	75	1	1
22	0.099028	55	16	60	29	13
23	0.098952	65	16	60	1	1
24	0.098372	65	16	60	1	19
25	0.098343	50	16	65	15	13
26	0.098162	55	16	65	29	19
27	0.097866	60	16	60	15	25
28	0.097843	60	16	70	1	13
29	0.097828	70	16	60	36	25
30	0.097781	50	31	50	8	19
31	0.097646	50	16	70	36	13
32	0.097507	50	16	75	8	19
33	0.097494	65	16	70	8	7
34	0.097394	50	16	70	43	13
35	0.097302	50	16	70	8	13
36	0.097093	75	16	55	43	7
37	0.097003	60	16	65	36	19
38	0.096976	55	16	65	8	1
39	0.096914	60	16	70	1	1
40	0.096606	55	16	65	1	13
41	0.096527	60	16	55	22	25
42	0.096444	55	16	65	1	25
43	0.096431	60	16	60	22	7
44	0.096355	65	16	65	43	13
45	0.096297	50	16	75	22	19
46	0.096266	50	16	75	50	13
47	0.096057	55	16	75	29	25
48	0.096049	55	16	70	43	19
49	0.09593	65	16	55	22	7
50	0.095869	50	16	70	1	19
51	0.095736	50	16	70	15	1

Figure F.2: Fine Calibration Coefficients with Highest OSM Values

1	OSM	Diff	Brd	Sprd	Slp	Road
2	0.0485669	4	80	50	26	70
3	0.0484034	4	88	50	25	70
4	0.0481981	4	80	48	25	100
5	0.0478913	4	84	48	25	70
6	0.0478482	4	80	50	25	25
7	0.0478086	4	80	50	25	40
8	0.0477473	4	80	48	25	70
9	0.0477127	4	80	50	25	85
10	0.0475921	4	80	50	26	25
11	0.0475479	4	80	50	25	55
12	0.0472364	4	84	48	25	85
13	0.0471768	4	80	50	25	100
14	0.0470849	4	84	50	26	100
15	0.0470725	4	80	50	27	100
16	0.0469732	4	80	42	25	100
17	0.0469547	4	84	48	25	40
18	0.0468946	6	80	50	25	40
19	0.0468912	4	80	50	26	85
20	0.0466931	4	84	50	25	70
21	0.0466912	4	80	46	25	100
22	0.0466705	4	80	46	25	25
23	0.0465973	4	80	46	26	100
24	0.0465865	4	88	50	26	40
25	0.0465535	4	92	50	25	40
26	0.0465486	4	84	50	25	25
27	0.0465283	4	96	50	26	40
28	0.0464868	4	88	48	26	85
29	0.0464851	4	92	50	26	55
30	0.0464091	4	88	50	25	55
31	0.0464001	4	80	48	26	85
32	0.0463874	4	80	48	25	40
33	0.0463439	4	80	50	27	55
34	0.0463361	4	96	50	25	25
35	0.046333	4	88	50	25	100
36	0.0463226	4	88	40	25	70
37	0.0462848	4	84	50	28	55
38	0.0462565	4	96	50	25	85
39	0.0462322	4	80	50	26	55
40	0.0462212	4	84	44	25	70
41	0.0461814	4	84	50	27	40
42	0.0461794	4	80	50	27	25
43	0.04615	41	0	50	25	100
44	0.0461495	4	84	50	26	25
45	0.0460551	4	92	48	25	70
46	0.0460543	4	92	48	25	40
47	0.0460229	4	88	48	25	100
48	0.0459926	4	80	42	25	85
49	0.0459813	4	80	46	25	40
50	0.0459714	41	0	50	25	70
51	0.0459603	4	80	48	25	85

Figure F.3: Final Calibration Coefficients with Highest OSM Values

APPENDIX G

2030 PREDICTIONS

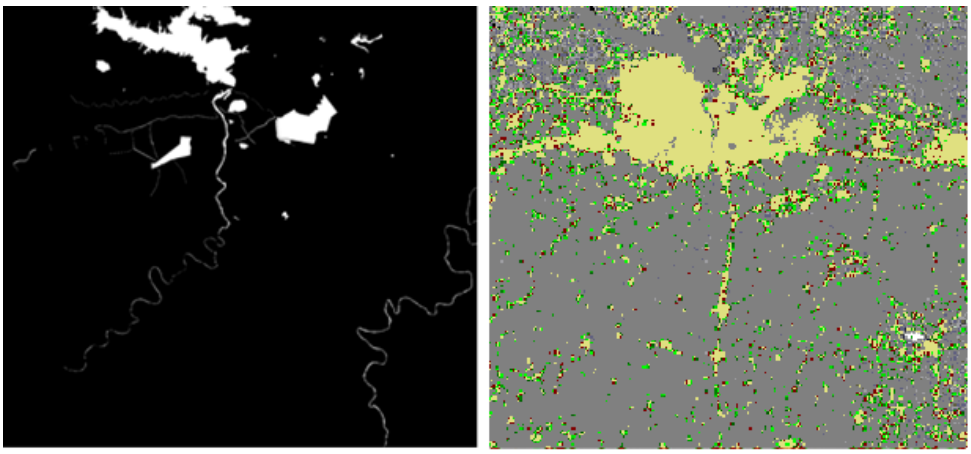

Figure G.1: 2030 Prediction Without 2020 Road Layer