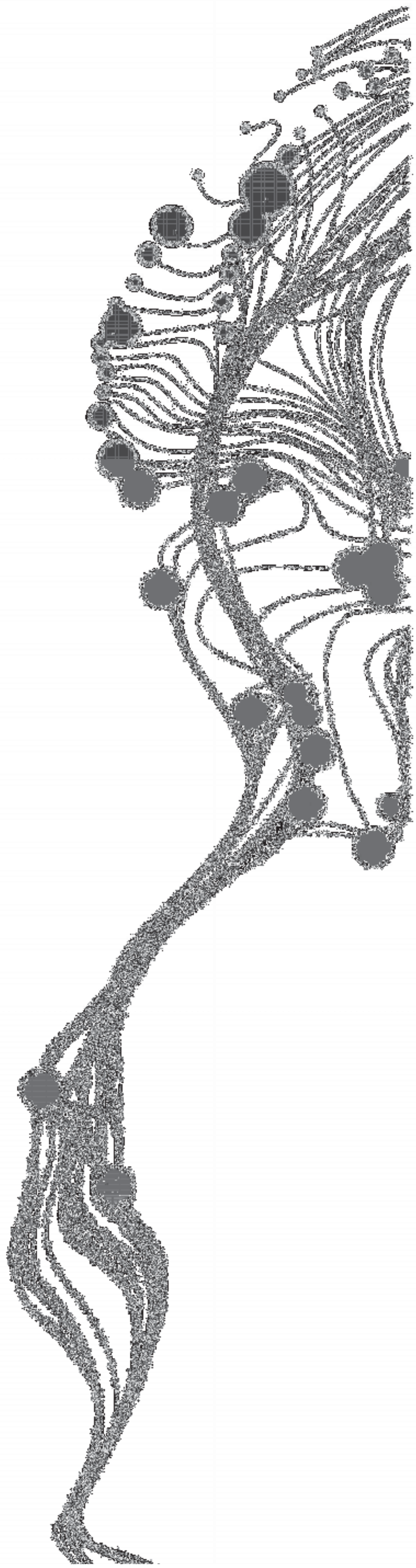
MODELLING URBAN GROWTH IN KATHMANDU VALLEY

SUNITA DUWAL
February, 2013

SUPERVISORS:
Dr. Sherif Amer
Ms. Monika Kuffer, MSc



MODELLING URBAN GROWTH IN KATHMANDU VALLEY

SUNITA DUWAL

Enschede, The Netherlands, February 2013

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Urban Planning and Management

SUPERVISORS:
Dr. Sherif Amer
Ms. Monika Kuffer, MSc

THESIS ASSESSMENT BOARD:
Prof. Dr. Ir. M.F.A.M. van Maarseveen (Chair)
Ms. Olena Dubovyk, MSc (External Examiner, University of Bonn)
Dr. Sherif Amer (1st Supervisor)
Ms. Monika Kuffer, MSc (2nd Supervisor)

ABSTRACT

Kathmandu valley has witnessed rapid urbanization with an annual population growth rate of 4.3%. Due to this, the city is expanding further and further out causing various environmental impacts such as- decrease in fertile and productive agricultural land, loss of forest areas, reduction of water quality, and landscape fragmentation. To understand the spatial and temporal dynamics of these processes and to monitor the growth, factors that drive urban growth should be identified and analyzed especially those factors which can be used to forecast the future urban growth. The main aim of this research is to identify and quantify the relationship between urban growth and its driving factors and to forecast future urban growth based on historical urban growth pattern for Kathmandu valley. This involves the land cover change analysis from Landsat images for 1989, 1999 and 2010 by using RS/GIS techniques and then quantification and identification of spatio-temporal process of urban growth; review of driving factors of urban growth from different literature; construction of binary logistic regression (LR) model using urban against non-urban as a dependent variable for 1999-2010; and forecast future urban growth pattern by using the LR model.

To investigate the major driving forces of different patterns of urban growth, three models (overall urban growth, infill growth and expansion) are constructed. The overall urban growth model correctly predicted 72.5% of the observed urban growth for 1999-2010 while infill growth model and expansion model predicted 97% and 22% of urban growth with huge amount of over-prediction and under-prediction for built-up areas. Population densities, proportion of built-up area in the surrounding and land value were the strongest predictors of overall urban growth. Other driving forces were less significant in explaining overall urban growth process. The most significant driving forces of infill growth were distance to urban centre, proportion of built-up area in surrounding, distance to major road and population density, whereas land value and proportion of built-up area in the surrounding are the major driving forces for urban expansion.

The results showed that the integration of remote sensing, GIS and LR model provide important information regarding the pattern and process of land cover change, relationship between urban growth and its driving factors, and trend of future urban growth. This information can be a taken as a reference for urban planner and decision makers to make planning decision in guiding future urban growth of the Kathmandu valley.

Key words: *land cover change*, *urban growth, driving factors, infill, expansion, logistic regression model, Kathmandu valley*

ACKNOWLEDGEMENTS

I would like to acknowledge all the personnel and organizations that directly or indirectly supported me to conduct this research.

My foremost gratitude goes to the Netherlands organization for international cooperation in higher education (NUFFIC) for granting me scholarship to study in ITC and acquire the knowledge on Geo-information science and Earth observation.

I would like to express my sincere gratitude and acknowledgement to my first supervisor Dr. Sherif Amer and second supervisor Ms. Monika Kuffer for their constant guidance and critical suggestions till the completion of this thesis. The research would not have been possible to go in right direction without their care and valuable suggestions. Their academic and moral support indeed inspired me a lot to work independently throughout the study period.

I would also like to express my especial thanks to Drs. Emile J.M. Dopheide for his valuable suggestions especially to improve my statistical thinking in this study. My appreciation and thanks to Dr. Richard Sliuzas and Dr. Johannes Flacke for their support during the hard time of the research.

I also deeply acknowledge to Prof. Huang Bo who accessed me the Change Analyst software with essential materials without which this research would not have been possible to this stage. Also, my acknowledgment will not be completed without thanking Mr. Daizi, student of Prof.Huang Bo who patiently helped me to resolve the problem related to software.

I must extend my appreciation to Dr. Rajesh Bahadur Thapa and Prof. Barry Haack for providing me their simulated urban growth map, which indeed helped me a lot to get an insight of different modelling approaches in case study area. I am grateful to Mr. Anish Joshi for his kind assistance during my field work providing me essential data and materials. Also, I would like to extend my gratitude to all those personnel who took part in interview and share their ideas and knowledge.

I am very thankful to my fellow student Gezahagn Debebe Fura for his continuous help, encouragement and valuable suggestions. I would also like to thank Sabrina, Reshma, Aidan and Islam for their guidance during tough stages of the research. Many thanks goes to my fellow UPM students for sharing fruitful 18 months together with fun and entertainment despites of continuous study pressure. I highly appreciate Nepalese society for providing homely environment during the stay with the feeling of togetherness.

My heartfelt thanks go to my parents for their love, care and inspiration throughout my life. I would not have reach at this stage without their caring support and encouragement. Finally my love and gratitude goes to my beloved Harish who has been my source of inspiration. His kind attention has always boosted me up for the completion of this thesis.

Sunita Duwal
February 2013

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This research focuses on determining driving factors of urban growth in Kathmandu valley and forecast future urban growth by using a logistic regression model. Kathmandu valley is a fast growing agglomeration area where driving factors can play a major role for urban land cover change. Therefore, it is a need to identify and quantify the process of urban land cover change, to determine its driving factors and to understand the relationship between urban growth and driving factors. This research has applied remote sensing and GIS techniques coupled with a logistic regression model to analyse spatio-temporal process of the urban growth and identify its driving factors to model future urban growth.

1.1. Background and Justification

Urbanization is a process that leads to a spatial concentration of households and economic activities within a limited area. It is the result of social, economic and political developments (Kivell, 1993; Black & Henderson, 1999; Hall & Tewdwr-Jones, 2010) that leads to the land use and land cover change (Belal & Moghanm, 2011) and spatial expansion of the cities. Urbanization is also associated with higher population concentration, which results in rapid changes of urban landscape pattern causing transformation of natural lands (Thapa & Murayama, 2010). Therefore, a growing population and changing household characteristics influence the demand for land that leads to the conversion of non-urban land into urban land.

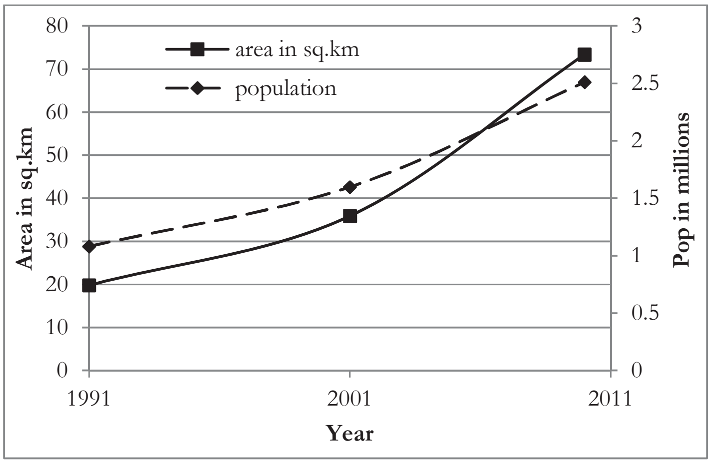


Figure 1.1 Urban population growth and expansion of built-up area in Kathmandu valley (1991-2011)

Source: CBS (2011)

Kathmandu valley, the capital city of Nepal, has an estimated population of 2.51 million in 2011 with an annual growth rate of 4.32% (CBS, 2011). This shows that the valley is undergoing urbanization in a very high rate. Such an urbanization results in gradual change of predominantly agricultural and natural landscape to an urban landscape. From Figure 1.1, it can be observed that both population and built up areas are increasing from 1991 to 2011. During this time period, around 53.54 km^2 of land has been converted to urban built-up areas and it is expected that this trend will accelerate in next decades (Thapa & Murayama, 2011). Similarly, the size of population has almost tripled from approx. 1.08 million in 1991 to around 2.5 million in 2010. Important to state is that, the city is facing considerable unplanned urban sprawl and haphazard growth particularly in the urban fringe area. It is likely that such an uncontrolled growth will eventually threaten agricultural and forest areas around the city. So to develop the city in a more sustainable way, urban growth has to be better managed.

Monitoring of land cover change is crucial to understand and forecast the future urbanization trends and urban land cover pattern. Remote sensing is one of the most commonly used techniques to monitor urban land cover changes. It provides multi-temporal and multispectral data for monitoring land use patterns and processes (Oluseyi, 2006). A number of researchers including Herold *et al.* (2003); De Almeida *et al.* (2005); Maktav & Erbek (2005); Belal & Moghanm (2011) and Xiao *et al.* (2006) have used multispectral remote sensing data of different spatial resolution to evaluate land use/cover changes. All these studies have examined changes in land use pattern over a given period. Satellite remote sensing techniques have, therefore, been widely used in detecting and monitoring urban land cover change.

In recent decades a number of studies have been done to model the urban growth of Kathmandu valley using Earth observation and GIS techniques. For instance, Thapa & Murayama (2011) and Haack & Rafter (2006) used cellular automata (CA) modelling and spatial analysis tools respectively to model the urban growth in Kathmandu valley. By employing a CA approach, Thapa et al. (2011) found that all the municipalities will be aggregated into a greater metropolitan region by 2020. They further concluded that urban growth in Kathmandu valley will continue through both in-filling in vacant pockets of existing urban areas and outward expansion towards the east, south, and west directions in the future. Similarly, Haack & Rafter (2006) concluded that the extensive growth will lead to various negative environmental impacts such as- loss of valuable agricultural lands and increased air and water pollution. Neither of these researches, however, investigated the degree of influence of driving factors on urban growth. Among other modelling techniques logistic regression model has been claimed to be an effective tool for understanding the relationship between urban land cover change and its drivers (see section 2.3). However, logistic regression model have never been applied in Kathmandu valley to model urban growth. Therefore, in order to contribute to the understanding of dynamic processes of urban growth and its drivers this research has used a statistical approach by means of logistic regression model to forecast the future urban growth.

1.2. Research problem

A detail study has been done to investigate driving factors of urban growth in Kathmandu valley by Thapa & Murayama (2010). Yet, the degree of influence of those factors on future urban growth in the valley has not been explored. In addition, a number of studies have modelled the urban growth using spatial analysis tools and cellular automata like-Haack & Rafter (2006) and Thapa & Murayama (2011). But a logistic regression model has never been applied to model urban growth of the valley besides its many advantages (see Table 2.3 in section 2.4). So, this research has used a logistic regression model to understand the relationship between driving factors of urban growth as well as to forecast the future urban growth.

1.3. Objectives and research questions

1.3.1. Aim

The main aim of the research is to identify the primary drivers of urban growth in Kathmandu valley and to forecast future urban growth pattern by using logistic regression (LR) model.

1.3.2. Objectives and research questions

1. To apply GI and Earth Observation techniques for identifying the urban growth pattern of Kathmandu valley during the period 1989, 1999 and 2010.
    - What are the accuracies obtained by using image classification techniques for extracting land cover change?
    - Which types of land cover change are the most prominent in Kathmandu valley?
    - What is the rate of urban land cover changes during the period 1989-1999 and 1999-2010?
    - What is the pattern of urban land cover change during the period 1989-1999 and 1999-2010?
2. To determine the main drivers of urban growth using a logistic regression model.
    - What are the factors that play a key role in urban growth in the scientific literature?
    - What are the main factors that are responsible for urban growth in Kathmandu valley and what is their relative importance according to the local experts?
    - Are the LR method and the expert knowledge arriving at the same conclusion about the main driving factors?
3. To forecast the future urban growth pattern in Kathmandu valley.
    - Which logistic based approaches have been used to model urban growth in literatures?
    - Where are the probable areas of future urban growth?

1.4. Conceptual framework

Urbanization can lead to a gradual transformation of rural landscape into urban-built form in term of infilling, expansion and scattered growth (see section 2.1). The information about trends and patterns of urban growth is commonly determined by using multi-spectral and multi temporal remote sensing techniques. But only having the knowledge of trend and pattern of land cover change is not sufficient to project future urban growth. It is very important to understand the key driving factors that lead to land cover conversion. Among different modeling approaches, logistic regression (LR) is one of the common statistical models that determines the key drivers of urban growth and projects the future urban growth. It also helps to understand relationship between urban growth and its drivers quantifies their degree of influence and identifies the most significant drivers of urban growth.

Figure 1.2 shows the conceptual framework to model urban growth and its driving factors.

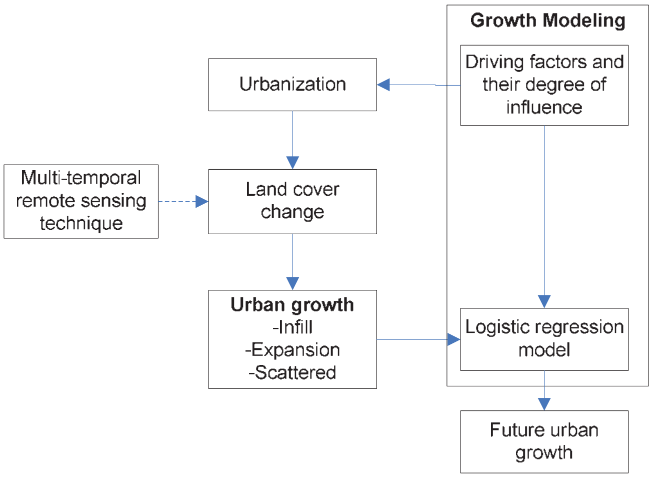


Figure 1.2 Conceptual framework

1.5. Research methodology

The methodology of this research is divided into five phases as shown in Figure 1.3.

The first phase is *problem identification* which was concerned with the establishment of a theoretical concept about urban growth and its consequences, understanding factors that drive urban growth and methods of modeling the urban growth.

The second phase is *data acquisition* through primary and secondary sources. All essential information was gathered from different institutions and field observation. Primary data like- training sample set verifications were gathered via field observation and secondary data like- images, spatial data, population data *etc.* were gathered from different institutions.

The third phase is concerned with *data preparation* which was about preparing land cover maps and factor maps for probable driving factors. The land cover maps were prepared from Landsat imageries for three different time periods 1989, 1999 and 2010 using RS techniques. These maps were further used to prepare binary maps of urban and non urban. In the same phase, factors maps of different driving forces were also prepared from spatial and socio-economic data.

The fourth phase is *Modelling* phase which was about constructing a LR model using GIS tools and data prepared in the third phase. To avoid the multicollinearity among different driving factors, a multicollinearity test was done to eliminate highly correlated factors and only the remaining factors were participated in the LR modeling. Statistical test of the model was done to evaluate the significance of remaining factors. Then, model parameters were set to get a final LR model which was used to forecast urban growth in the valley. At the same time key driving factors of urban growth in the valley was determined.

The last phase is *Communication of result* that includes quantitative analysis of the result. The model was evaluated to assess its performance using statistical measure such as Percentage of Correct Prediction (PCP) and Kappa statistics.

The result of this research can be helpful for urban or regional planners and decision makers since it provides information like- rate, trends and patterns and direction of urban growth along with probable area of future urban growth. In addition, this research provides knowledge of driving factors that have significant influence on the urban growth of the valley so that policy makers can determine which factor needs more concentration to manage future urban growth of the valley.

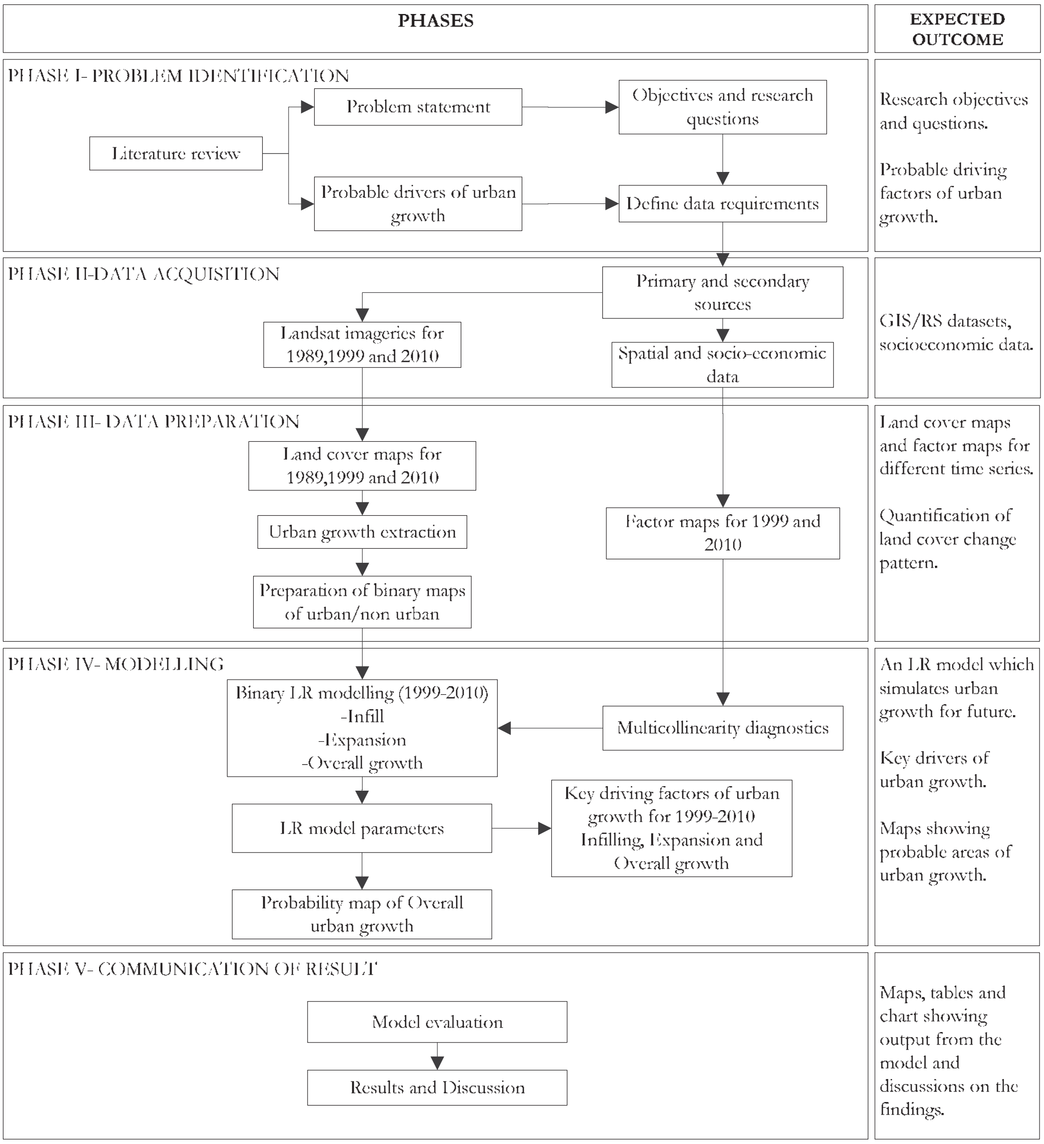


Figure 1.3 Research methodology

1.6. Thesis outline

Chapter 1: Introduction

This chapter provides general information about the urbanization trend in Kathmandu valley followed by research problem, research objectives, research questions and conceptual framework. It also presents a brief methodology of the research.

Chapter 2: Literature review

This chapter reviews the theoretical background on urban growth and its spatial pattern, driving forces of urban growth, methods and techniques to model and forecast urban growth. It also presents a brief introduction to different types of urban growth models including their strength and weaknesses.

Chapter 3: Methodology

This chapter gives a short introduction to the study area followed by overview of data collection and methodological approach to answer the research questions. It mainly focuses on the methods for analysing and quantifying spatio-temporal process of urban growth, identifying major driving factors of urban growth and modeling future urban growth by using logistic regression model.

Chapter 4: Results and discussion

Chapter 4 presents the main findings from the research. It includes the results from spatio-temporal analysis of land cover maps, relationship between urban growth and driving factors and probable area of future urban growth. It continues with discussions of the results and along with the policy implications.

Chapter 5: Conclusions

This chapter concludes the research by summarizing the main findings from the research and pointing out future research directions.

2. LITERATURE REVIEW

This chapter deals with the concept of urban growth, driving factors, methods and techniques of modeling and forecasting the urban growth. It starts with general concept about urban growth, its spatial characteristics and remote sensing technique to detect urban growth followed by driving factors of urban growth in the literature. It presents different types of urban growth models along with their strength and weaknesses giving an insight into the models which assist in understanding the relationship between urban growth and driving factors. Finally, it highlights the logistic regression (LR) model in detail as well as gives few examples of researches that have applied LR model for determining driving factors of urban growth.

2.1. Urban growth and its spatial characteristics

Urban growth is a broad concept that has been defined in several ways by many authors. In general, it is recognized as physical and functional changes of urban landscape due to social, economic and political development (Kivell, 1993; Black & Henderson, 1999; Hall & Tewdwr-Jones, 2010) that leads to transformation of rural landscape to urban forms. It leads to the change of land use and land cover (Belal & Moghanm, 2011) that consist of interaction between several factors like- topography, river, population, economy, plan and policies, infrastructure, *etc.* These factors can have a great influence on the nature of urban growth leading to the different types of urban growth.

Urban growth can occur in a several ways such as- sprawling or compact, scattered or clustered, continuous or leapfrog, planned or organic (Cheng, 2003). Sprawling is generally defined as development of rural land in outskirt of the city into urban land in a dispersed manner (Arbury, 2005) whereas, compact growth occurs within the city in infilling and densification manner. Scattered or leapfrog growth means the formation of new urban patches that are isolated from the existing city while clustered or continuous growth is defined as new growth that are grouped together to form a urban cluster. Organic growth spreads outwards from the existing urban centres that tends city to expand (Clarke *et al.*, 1996) while planned growth always aims to achieve organised or managed growth.

Camagni *et al.* (2002) have distinguished five types of urban growth such as infill, expansion, linear development, sprawl and large scale projects. Whereas, Wilson *et al.* (2002) have identified five types of urban growth like infill, expansion, isolated, linear branch and clustered branch. Berling-Wolff & Wu (2004) have distinguished four different types of urban growth- spontaneous, diffusive, organic and road influenced. The classifications mentioned here primarily focus on spatial pattern of the growth. In some cases, they are interrelated as well. For this research, the urban growth is categorised as follows (see Figure 2.1):

a) **Infilling:** Infilling means conversion of non-urban areas which is surrounded by urban areas into built-up (Xu *et al.*, 2007). It occurs on vacant land parcels within existing developed areas that remained so far as non built-up. This type of growth is very common in core areas and already developed new areas.

b) **Expansion:** This type of growth occurs in the adjacent urban fringe. The newly developed urban areas in the fringe spread outwards in adjacency to existing urban areas. Therefore it is also called urban fringe development (Camagni *et al.*, 2002; Xu *et al.*, 2007). Linear growth can also be included in this growth type. Because in linear growth, urban growth expand along road corridors which stimulates the development of new areas (Berling-Wolff & Wu, 2004).

c) **Scattered growth**: Scattered growth means the formation of new urban patches which have no direct spatial connection with the existing urban patches (Berling-Wolff & Wu, 2004). This type of growth usually occurs away from the existing urban area by forming isolated urban patches.

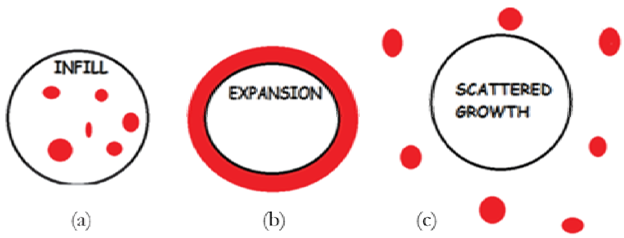


Figure 2.1 Spatial characters of urban growth (a: Infilling, b: Expansion and c: Scatter growth)

2.2. Detecting urban growth using remote sensing techniques

Remote sensing (RS) is the process of data acquisition through airborne or space sensors without having any contact with the object in the earth surface (Richards, 2012). It allows acquisition of multi-temporal, multispectral and multi-resolution data that are required for analysing and modelling of land cover changes. Because of their cost effectiveness and temporal frequency, RS data and techniques are widely used in urban studies to detect, monitor and simulate urban land use changes (Yuan *et al.*, 2005; Araya & Cabral, 2010). The strength of remote sensing techniques is to extract the physical growth while GIS has capacity to analyse other aspects of urban growth.

There are many researches that are based on RS primarily focussed on image classification including Yuan *et al.* (2005); Lu & Weng (2007); Araya & Cabral (2010); Belal & Moghanm (2011). Image classification is the process of assigning pixels of a continuous raster image to predefined land cover classes (Araya & Cabral, 2010). The result of image classification is likely to be affected by various factors for *e.g.* nature of input images, classification methods, algorithm, *etc*. In order to improve the classification accuracy and to enable an analyst to detect changes successfully, the selection of an appropriate classification method is very essential. Basically there are two classifications paradigms-Pixel-based and Object-based (Araya & Cabral, 2010). Pixel based image classification is a powerful technique where each pixel is assigned to certain spectral class which is based upon the surface characteristics and training data. While an object based image classification is an approach where classification is based on homogeneous segments that corresponds to different objects such as-fields, trees, buildings, *etc* (ITC, 2010). An object-based approach is particularly suitable for images of high spatial resolution. For the urban studies at city scale, it is rather expensive to acquire high resolution images for an entire city as well as very high resolution satellite images are only available for recent years. Moreover, most of the urban studies are using moderate resolution images and pixel based classification methods to analyse the land cover/use change (Yuan *et al.*, 2005; Oluseyi, 2006; Belal & Moghanm, 2011; Zaki *et al.*, 2011). Hence, this study has used supervised classification at pixel level.

2.2.1. Supervised classification using maximum likelihood

Supervised classification is developed for satellite image-processing and it has been widely applied to classify the spectral layers. In supervised classification, one of the underlying requirement is that the analyst should have sufficient number of known pixels to allow correct classification (Richards, 2012). It is dependent on the experience and ability of the user in detecting the signature differences accurately between various land covers in the satellite image using his/her bare eyes (Zaki *et al.*, 2011). Moreover, selection of ground truth points and local knowledge on land covers are also important requirements of supervised classification.

The maximum likelihood classifier is commonly used in practice, because of its robustness and its easy availability in almost any image-processing software (Lu & Weng, 2007). Some researches like Booth & Oldfield (1989); Lu & Weng (2007) have compared different image classification methods and used maximum classification as an standard algorithm to compare the result of different other algorithms. Booth & Oldfield (1989) conclude that maximum likelihood classifier is the most accurate algorithm in comparision to dicision tree, minimum distance and deviant distance, although it requires relatively more processing time. Similarly, many urban growth studies like- Maktav & Erbek (2005); Yuan *et al.* (2005); Oluseyi (2006); Xiao *et al.* (2006); Belal & Moghanm (2011); Zaki *et al.* (2011) have applied this method to classify multi-temporal images to get land cover classes. So, maximum likelihood is one of the common methods in the practice.

2.2.2. Accuracy assessment of classified imageries

One of the important steps after performing an image classification is the evaluation of the classification results. Congalton (1991) cited by Zaki *et al.* (2011) defined the accuracy assessment as the agreement between the remote sensing data and the reference information. There are different approaches of accuracy assessment ranging from a qualitative evaluation based on expert knowledge to a quantitative accuracy assessment based on sampling strategies (Lu & Weng, 2007). Among them, an error matrix approach is the commonly used in accuracy assessment (Foody, 2002). The error matrix compares information from a classified image or land cover map to known reference (truth) sites for a number of sample points (Zaki *et al.*, 2011). It tabulates the overall accuracy and determines errors due to omission (producer accuracy) and error due to commission (user accuracy). Error of omission indicates how many pixels on the map for a given class are actually what they are on the ground. It is calculated by dividing the number of correct reference site for a class by total number of reference sites that were classified in that class. Error of commission indicates how many pixels on the map for a given class are correctly classified out of total reference sites. It is calculated by dividing the total number of correctly classified pixels by the total number of reference sites for each land cover class.

A number of researchers have used the error matrix to assess the accuracy level of classified imageries. For instance-Maktav & Erbek (2005); Yuan *et al.* (2005); Oluseyi (2006); Xu *et al.* (2007); Belal & Moghanm (2011); Zaki *et al.* (2011) have used multi-temporal Landsat imageries to examine the changes in land cover pattern over a certain period and all these works have obtained accuracy levels between 90% to 96% after classification using 3 to 7 land cover-classes.

2.2.3. Change detection

Change detection is the process of determining or analysing change in land cover properties based on co-registered multi-temporal remote sensing data (Belal & Moghanm, 2011). There are lots of change detection methods such as- Image differencing, Image regression, Image rationing, Vegetation index differencing, Principal components analysis (PCA) and Post-classification comparison (Singh, 1989). Among them, post-classification comparison is one of the most commonly used methods of change detection (Singh, 1989; Yuan *et al.*, 2005; Belal & Moghanm, 2011; Zaki *et al.*, 2011). The post-classification approach provides "from–to" change information and the kind of land cover transformations that have occurred can be easily calculated and mapped (Yuan *et al.*, 2005).

2.3. Driving factors of urban growth

Urban growth and land use change cause various impacts on social, economic and environmental aspects. So it is very important to understand the spatial and temporal dynamics of these processes. In order to understand these processes, the factors that drive urban development must be identified and analyzed, especially those factors that can be used to forecast future changes (Aguayo *et al.*, 2007). From the literature of Allen & Lu (2002),Fang *et al.* (2005),Hu & Lo (2007),Priyanto (2010) and Thapa & Murayama (2010), similar driving factors can been found which are grouped together into seven major types including biophysical factors, population growth, economic, proximate causes, existing conditions, plans and policies and the political situation as shown in Table 2.1. The degree to which they contribute to urban growth depends upon the local context however it can be seen that all the driving factors affect the urban land cover to a certain extent.

Biophysical factors refer to characteristics and processes of the natural environment, such as climate variation, landform, topography, soil types, drainage patterns and the availability of natural resources. These characteristics usually affect the urban growth pattern based on suitability of land for specific purpose. *Economic factors* include a wide range of economic activities such as job and business opportunities, industries, etc (Priyanto, 2010; Thapa & Murayama, 2010). Generally new economic activities emerge around existing economic centre for the sake of economic agglomeration and market competition whereas companies that need spacious sites such as university campuses and industries often settle in urban fringes (Kivell, 1993). *Proximal causes* are widely mentioned in most of the literature. The area which is located in the proximity of infrastructure, major roads and rivers, central business districts, public services etc. tends to grow in future due to potential benefits like-ease of access, economic opportunities, social services etc. *Population growth* is also seen as one of the major driving factors that guide urban expansion due to new settlement development around the main city centre (Fang *et al.*, 2005; Oluseyi, 2006; Xiao *et al.*, 2006). *Plans and policies* including land use zoning, transportation policy, development control and investment plans are also considered as driving factors that have high capability to direct the future growth pattern (Kivell, 1993). The *proportion of existing urban area* also has a big influence on future land use. The probability of land use change at a certain location is highly determined by proportion of built-up area in the surrounding. For example- existing ecological, cultural, water and transport networks highly influence the choice of location for new developments. *Political situation* identified by Thapa & Murayama (2010) also can influence on safety of the region which can affect the urban growth as well. It is human nature to choose the safe location to build their house there. So if the city is politically unstable and there is the problem of social safety and security, urban growth tends to decrease. Conversely, urban growth tends to increase in safer locations. Therefore, the influence of driving factors on urban growth depends upon the context.

In order to acquire a good understanding of the relationship between driving factors and land use change, Allen & Lu (2002), Fang *et al.* (2005), Hu & Lo (2007) and Priyanto (2010) have analysed the spatiotemporal changes in urban land use by using logistic regression model (see Table 2.2). For instance, Priyanto (2010) used qualitative as well as quantitative approach to identify and analyse the driving factors. The qualitative approach was used to identify the underlying factors of land use change by analysing cause-effect interrelationship between driving forces, pressure, states, impacts and responses (DSPIR). This approach is then followed by logistic regression analysis which is used to examine the relation between land use and possible driving factors. Similarly, Thapa & Murayama (2010) considered the local residents, researchers, urban and regional planners, and academics working in diversified disciplines as local experts and conducted group discussions, interviews, and literature review for identifying the driving factors of urban growth in the last decade in Kathmandu valley. The relative importance of each factor was evaluated using the Analytical Hierarchy process (AHP) framework. AHP is widely used multi-criteria evaluation technique which is used for decision making process including multiple actors, scenarios and criteria (Bagheri *et al.*, 2012). It constructs a ratio scale associated with priorities for the various factors compared.

Table 2.2 shows that different authors have used different spatial resolution depending on the size of study area to analyse the urban growth of the study area. The resolution varies from 30m x 30m to 250m x 250m pixel size and the study area varies from 15 km^2 to 16800 km^2.

Table 2.1 Overview of driving factors

Factors	**(Allen & Lu, 2002)**	**(Fang *et al.*, 2005)**	**(Hu & Lo, 2007)**	**(Priyanto, 2010)**	**(Thapa & Murayama, 2010)**
Bio-physical character	Slope, forest, wetland, waterfront	Degree of steepness, forest, rivers and lakes	Slope percentage	Elevation	Slope, soil, river
Population growth	Population density	Population density	Population density	Annual population growth	High Population influx
Plans and policies	Corporate boundary and protected lands.	City development policy Policy to preserve farmland			Zoning, land reform, investment plans, land development plans
Proximate causes	Distance to major roads, Distance to major nodes, Accessibility to water lines and sewer line Distance to CBD	Proximity to country road, Proximity to major road intersection, Proximity to state highway, Proximity to sewage, water supply and electricity,	Distance to CBD, Distance to active economic centres	Primary road proximity, Secondary road proximity, Public facilities proximity, Educational facilities proximity, Health facilities proximity, Central business district accessibility, Major river proximity,	Public service accessibility like- transportation, electricity, education, drinking water, health services, commercial services, waste disposal, open spaces and recreational facilities.
Existing conditions	Distance to existing urban cluster	No. of immediate houses	Distance to nearest urban cluster, No of urban cells within 7 by 7 cell window,	Existing urban areas, Existing agricultural areas, Existing ponds	
Economic factors			Per capita income, Employment rate, Poverty rate		Jobs and business opportunities, industries
Political situation					Safety

Table 2.2 Method used to determine driving factors

Authors	**(Allen & Lu, 2002)**	**(Fang *et al.*, 2005)**	**(Hu & Lo, 2007)**	**(Priyanto, 2010)**	**(Thapa & Murayama, 2010)**
Methods	Integrated GIS model	Logistic regression analysis	Logistic regression analysis	Logistic regression analysis and DSPIR	Analytical hierarchy process
Spatial resolution	250 x 250 metres	30m×30m pixel	225m×225m pixel	30m×30m pixel	30m×30m pixel
Study Site area		272.25 sq.km	16800 sq.km	15 sq.km	685 sq.km

2.4. Urban growth models

An important step to understand the urban growth pattern of the cities is to quantify the urban growth and its spatio-temporal dynamics (Berling-Wolff & Wu, 2004). Urban growth models can play an important role in this process. Several models have been developed since 1950s by urban planners, geographers, and ecologist (Berling-Wolff & Wu, 2004). For instance, Cellular Automata (CA), spatial statistical models, agent based models (ABM), artificial neutral network (ANN), fractal based modeling and Chaotic and Catastrophe modeling (Cheng *et al.*, 2003) are some of these modeling techniques that are concerned with simulating and forecasting future urban growth (Wu *et al.*, 2009). Basically these models can be roughly grouped together into three types: empirical estimation models, dynamic simulation models and rule based simulation models (Hu & Lo, 2007; Zeng *et al.*, 2008). The brief explanations of the most commonly used modelling techniques are described below along with their strength and weakness.

a) Rule based simulation model: Among rule based models, Cellular automata (CA) is one of the widely used simulation model that consists of a lattice of discrete cells. Based on transition rules and neighbourhood cell state, the cell determines whether or not to change from one state to another. One of the advantages of CA models is that they are inherently spatial and dynamic (White & Engelen, 2000). In addition, they are simple and highly adaptable. However, Wu *et al.* (2009) argue that CA models focus on the simulation of spatial pattern rather than interpretation of spatio- temporal processes of land use change. The output from this type of models gives a map showing simulated land cover pattern.

b) Dynamic simulation models: System dynamics modeling (SD) is one of the common techniques used for understanding complex phenomenon and problems through dynamic simulation. It is an approach that can help the urban planners and managers to meet the challenges of decision-making and policy formulation for the development of a system (Sonar, 2008). The model can arrange as well as describe the complex connections among each element in different levels and it can also deal with dynamic processes with feedback in a system (He *et al.*, 2006a). However, SD model's ability to represent the spatial process of land use is weak because it can not deal with spatial data well and can not effectively describe the detailed distribution and situations of the spatial factors in the land system (Guo *et al.*, 2001). It contains a large amount of uncertainty because of too many assumptions and generalisation of factors while predicting complex urban problems which may lead to wrong interpretation of the result. It helps to predict future development scenarios of urban land use phenomenon and gives output in the form of simulated land cover map.

c) Empirical estimation models: These models use statistical techniques to model the relationship between land use change and the drivers based on historical data (Hu & Lo, 2007). As an empirical estimation model, the logistic regression methods have been claimed as an effective tool for understanding land use change due to its explanatory power and spatial explicitness. It can readily identify the driving factors of land use change and provides information about the degree of confidence regarding their contribution (Hu & Lo, 2007). In addition, computation requirements for this type of models are not as intensive as for CA models and input data requirement are relatively easy to fulfil making them especially useful in those places where data is scare (Dubovyk *et al.*, 2011). However, one of the major drawbacks of these models is that they lack the understanding on simulation process that influences the land use process. The output from this type of models gives a map showing probability of land use/cover change.

Table 2.3 Advantages and disadvantages of urban growth models

Models	Advantages	Disadvantages
Cellular automata (White & Engelen, 2000; Wu *et al.*, 2009)	- Simple and highly adaptable. - Inherently spatial and dynamic.	-Does not interpret spatio- temporal processes of land use change.
System dynamic model (Guo *et al.*, 2001; He *et al.*, 2006b; Sonar, 2008)	-Understands complex phenomenon and problems through dynamic simulation. -Arrange as well as describe the complex connections among each element in different levels.	-Can not effectively describe the detailed distribution and situations of the spatial factors in the land system. -Large amount of uncertainty while predicting complex urban problems.
Logical regression (Hu & Lo, 2007; Dubovyk *et al.*, 2011)	-Requires relatively less data. -Gives an understanding of relationship between urban growth and driving factors. -Computation requirements for this type of model are not as intensive as in CA model.	-Lack the understanding on simulation process that influences the land use process. - Does not incorporate temporal dynamics of urban growth.

According to Table 2.3, statistical model seems to be an effective tool to model urban growth in spite of some limitation because it gives the knowledge of most significant driving factors of urban growth, their degree of influence and their relationship with urban growth. So, in order to understand the relationship between urban growth and driving factors, this research will use statistical based logistic regression model. The following section provides detail explanation of this model.

2.5. Modeling urban growth using Logistic regression model

Logistic regression model is a type of multivariate analysis model which was developed by McFadden (1973). This model is conceptually based on the random utility theory and discrete choice theory in urban economics and behaviour science (Allen & Lu, 2002). In general, there are two models of logistic regression, namely multinomial logistic regression and binomial/binary logistic regression.

a) Multinomial logistic regression

Multinomial logistic regression is typically used when dependent variable comprise of more than two possible cases (Anderson, 1982). The term dependent variable means outcome variables which are going to be analyzed for example types of land covers in land cover change analysis. This model is also referred as *Logit regression model*. Multinomial logistic regression has very similar results to binary logistic regression (Anderson, 1982).

b) Binomial/Binary logistic regression

When the dependent variable is dichotomous (categorical) and the independent variables are either continuous or categorical, then binary logistic regression is used (Anderson, 1982). It is used for binary response variables having only two possible outcomes for *e.g.*- Yes or No, Urban or Not urban, Present or Absent etc. The term independent variables mean explanatory variables or predictor variables.

2.6. Example of researches using logistic regression model

Hu & Lo (2007), in his research have applied binomial logistic regression model to model urban growth and to discover the relationship between urban growth and the driving factors in the Atlanta Metropolitan Area. They found two groups of factors affecting urban growth in the region- (1) population density, distances to nearest urban clusters, activity centres and roads, and high/low density urban uses and (2) distance to the CBD, number of urban cells within a 7 x 7 cell window, bare land, crop/grass land and forest. The model was 85% accurate which was evaluated by Relative operating characteristic (ROC) curve. They have concluded that logistic regression allows much deeper understanding of the forces driving the growth and the formation of urban spatial pattern.

Another example of modeling urban growth is research conducted by Fang *et al.* (2005). The study was conducted on the south-east side of Peoria, Illinois, USA. The main aim of study was to determine influence of individual factors in the prediction of the probability of urban sprawl. It also tested the effect of combining a multinomial logistic model and cellular automata. They used factors such as- weighted travel time to city, country road, proximity to forest, slope, proximity to major road intersection, proximity to lakes and rivers, utilities, historical growth trend, number of immediate neighbours and agricultural protection in the logistic regression model. Although the model showed that the proximity to road and historical growth trend were the most significant factors, it has also verified that all other factors are good to explain urban sprawl of the region. The model was evaluated using ROC curve and it was 71% accurate.

Aguayo *et al.* (2007) also constructed multinomial logistic regression model to quantify the relationship between urban growth and its driving factors and to predict spatial growth pattern for the city of Los Angeles in central Chile. Three different models were constructed for predictive variables that were grouped together into three categories- 1) Distance variables, 2) Neighbourhood variables and 3) Environmental variables. Distance variables included proximity to road network, rivers, amenities and services, railway line etc. Neighbourhood variables included density of urban road network, urban area, industrial area, etc. and environmental variables included elevation, slope and soil type. The models showed that urban growth of Los Angeles is heavily dependent on distance factors. Similarly, it also showed that most of the growths are extending from already existing urban areas in an infilling manner. The accuracies of these models were evaluated from coefficient of determinant (R^2) that scored above 65%.

2.7. Concluding remarks

Urban growth and its spatial pattern can be effectively modelled by RS and GIS techniques. It helps to identify and quantify the spatio-temporal process of the urban growth. Moreover, there are different drivers of urban growth which can have major and minor influence in the urban growth. Urban growth model plays a significant role to structure the relationship between driving factors of urban growth. Among several urban growth model, LR can be used to examine the relationship between urban growth and driving factors including their degree of influence. So this model has been used to support urban growth analysis of the study area.

3. STUDY AREA AND METHODOLOGY

This chapter introduces the study area and provides the information about data sources and their metadata. It presents methods and tools used in this research including the approach of quantifying the urban growth, the way to determine driving factors of urban growth and the identification of probable areas of future urban growth using logistic regression model in detail.

3.1. Study Area

Kathmandu valley is situated between 27°31'55" to 27°48'56" North latitudes and 85°11'11" to 85°31'52" East longitudes. The valley measures about 25 km in length and 15 km in width, with an average altitude of 1,300 m above mean sea level. It consists of three districts-Kathmandu district, Latitpur district and Bhaktapur district covering an area of 518 km^2. The valley consists of five municipal urban centers-Kathmandu, Latitpur, Bhaktapur, Madhyapur Thimi and Kirtipur. The remaining area is made up of a 131 Village Development Committees (VDCs). The only river draining out the valley is the Bagmati River which has been serving as major sources of drinking water and irrigation throughout the year.

Figure 3.2 Location of Kathmandu valley

Figure 3.1 Study area based on watershed boundary

Kathmandu valley has long been centre of power, politics, culture and commerce. In recent years, the valley is inhabited by more than 2.5 million people. The metropolitan region is surrounded by mountainous terrain and steep hills with very limited land for urban expansion. The extent of study area in this study is based on previous two researches-Haack & Rafter (2006) and Thapa & Murayama (2011) which are delineated according to watershed boundary of the valley covering an area 696 sq.km. Based on this boundary, the present study area covers 5 municipal urban centers and 97 VDCs (Figure 3.1). Figure 3.2 is the location map of the Kathmandu valley.

3.2. Urban prospect in Kathmandu valley

The growth of settlements in the valley is generally unregulated, and there is very little planning intervention by the government (Rana, 2008). The government does not have financial resources to acquire a huge amount of land where planned urban development can be promoted. The current constitutional provision does not allow the government to impose any kind of restriction on the use of private property. Hence many areas has been converted into built-up which are declared as restricted areas for the development (Shrestha, 2003).

Besides, private developers face difficulties in assembling land parcels due to land ceiling provisions laid down by the Land Reform Act of 1964 (Shrestha, 2003). Land ceiling is the fixed amount of the ownership of the land holdings. In Kathmandu valley, any individual can hold the ownership of maximum 1.5 ha according to land ceiling proposed in 2001 (Adhikari, 2008). Developers also face difficulties in procuring land parcels from speculative landowners who either demand high prices or simply refuse to sell the land. There is no legal tool that can be used to acquire isolated land parcels from landowners. Notwithstanding, developers cannot buy sufficient land in Kathmandu valley due to the land ceiling (Shrestha, 2003; Rana, 2008).

Kathmandu Valley Town Development Committee (KVTDC) which is responsible for overall planning and regulation of urban development is operating with a land-use plan adopted in 1976. Several efforts were made to revise the land-use plan after, which were not successful. KVTDC is involved in several land pooling projects and guided land development programmes. It also looks after court cases involved in violation of building byelaws and other issues related to planning. Local bodies like municipalities and village development committees are issuing building permits mainly for revenue generation rather than regulating urban development. The technical capability of local authorities to deliver basic urban services is relatively low and people look upon central government agencies for such services (Shrestha, 2003).

3.3. Plans and programs

3.3.1. Kathmandu valley town development plan (1976)

Kathmandu valley town development plan was prepared under the Department of Urban Development and Building Construction (DUDBC) in 1976 that constitute three components-zoning, policy components, and land use regulation. The statutory plan for controlling development within the valley is zoning proposals prepared in 1976, entitled "Instructions for various actions to be taken in different areas in Kathmandu valley town development plan". These instructions lack precise zoning demarcation, for uses, which are not sufficiently defined in terms of permitted or restricted activity, with little or no reference to infrastructure provision, standards or programming. Moreover the zoning does not extend beyond the inner ring road (Rana, 2008).

3.3.2. Kathmandu valley physical development plan (1984)

The Kathmandu valley long term development plan includes proposed zoning regulations that are almost identical to the 1976 instructions. The main improvement by this latest effort is the preparation of zoning map called proposed land use covering the area within the ring road. The map contains some inconsistencies. However it represents the first attempt to think through spatial distribution of desired land use in Kathmandu City (Shrestha, 2003).

The 1984 development concept addresses some of the main issues affecting land use in the valley, including open acknowledgement of the current system of unplanned development by both government and private sectors; the importance of transportation in governing the land use; the need to keep in mind a vision of the economic role of the Kathmandu valley mainly as an administrative, cultural, touristic and agricultural area and not as a major industrial area; and the reality that financial resources are limited restricting what can realistically achieved in the short to medium run (Shrestha, 2003).

3.3.3. National urban policy

The proposed National urbanization policy is expected to be a land mark towards framing positive directions in a planned, integrated and coordinated way in solving the unplanned urbanization process and its challenges (Shrestha, 2003). This policy has maintained three objectives, first, obtain national urban feature by developing infrastructure services and direct the investment, secondly, improve the living standard of the city inhabitants by developing healthy, secured and welfare city environment. Thirdly, by consolidating local agencies in legal and institutional way and develop the cities in a coordinate way and develop the sense of partnership among the concerned agencies and make the city management influential.

3.3.4. Kathmandu valley long term development plan 2020

The long term development plan is prepared by Kathmandu Valley Town Development Committee (KVTDC) and Department of Urban Development and Building Construction (DUDBC) in the year 2002. The development plan of Kathmandu valley tends to provide guidelines to minimise the external influence in the process of urban expansion and development (Rana, 2008). According to the senior urban planner in Bhaktapur municipality, 2020 plan is only a plan that spells out some aspects in urban growth management in Kathmandu valley through orderly rural-urban transition, rural-urban land delineation, agricultural land preservation and land tenure ship.

The Director General of DUDBC states "The 2020 plan mainly focuses in urban growth control through the demarcation of the urban and rural boundary as well as the preservation of agricultural land. This will make land management easier and effective in the urban and rural context".

All these policy eventually affect the future spatial pattern of urbanization. These policies and strategies undertaken by government can help to counter negative impacts of urban sprawl and provide some future perspectives as well if they are implemented in proper way.

3.4. Data sources

3.4.1. Primary data collection

Primary data were collected through interviews and field data collection. Interview was conducted to understand the key driving factors of urban growth in the Kathmandu valley through the local experts' knowledge. The composition of experts was done in such a way that it incorporated academicians, planners in Ministry of urban planning and building constructions, planners in local municipality and private consultants dealing with urban planning issues. An appointment was taken for 5 experts for interview. And a question guideline was prepared for the interview which consisted of three sections. The first section was prepared to understand the most prominent type of urban growth in Kathmandu valley and the driving factors. Second section was prepared to rate the driving factors of urban growth derived from the literature and the third section was prepared to understand the plans and policies related to urban development in Kathmandu valley (See Annexes III).

Field data collection was done to verify the training samples for the accuracy assessment of the land cover classification. Altogether 100 sample points were selected from world view image for 2010. Same sets were used for 1989 and 1999 after verifying from ortho-rectified aerial image 1992 and Ikonos image 2001. The Ikonos image for 2001 is only available for limited areas in the valley; therefore, field verification of training samples were done for 30 urban points and 10 bare points using a handheld (GPS) unit.

3.4.2. Secondary data collection

Secondary data required for the study were obtained from various sources. According to the type, these data were grouped into three categories-Remotely sensed data, Vector dataset and Statistical data as shown in Table 3.1, Table 3.2 and Table 3.3 respectively below.

3.4.2.1. Remotely sensed data

Landsat imageries with moderate resolution were acquired from website of US Geological Survey (USGS, 2012) which were projected to WGS 84, UTM Zone 45. The high resolution imageries for 1992, 2001 and 2010 were acquired from Kathmandu valley town development committee (KVTDC).

3.4.2.2. Vector data set

Vector datasets were obtained from Department of Survey (DOS) and KVTDC for the year 1995 and 2008 respectively. Dataset from Department of survey were prepared from topographic map of scale 1:25000 and dataset from KVTDC were digitized from topographic map of scale 1:10000. The coordinates in vector data set were based on the Modified UTM Projection on Everest Spheroid 1830 (semi major axis, a = 6 377 276.345 m. and semi minor axis, b = 6 356 075.413 m.) having zone width of 3 degrees East-west with Central Meridian 84 degree longitude. The unit of measurement is metre.

3.4.2.3. Statistical data

Statistical data consist of population data for the year 1991 at VDC and municipal level while data for 2001 and 2011 were at ward level. These data were collected from Central Bureau of Statistics (CBS), Nepal. In addition, it also comprised of Land value data for 1999 and 2010 at ward level which was fixed according to each road category (refer section 3.6.2).

Table 3.1 Remotely sensed data

Dataset	Acquisition year	Scale/Resolution	Source	Projection	Purpose
Landsat TM Landsat ETM+ Landsat ETM+	1989/10/31 1999/11/04 2010/03/07	30 meter 30 meter 30 meter	USGS USGS USGS	UTM zone 45(WGS 84)	Landsat Images were used for preparing land cover maps
Ortho-rectified arial image	1992	1 meter	KVTDC	Everest 1830 Modified UTM	High resolution images were used for verification of training sample sets.
Ikonos image	2001	1meter	KVTDC	UTM zone 45 (WGS84)	
World view image	2010	0.5 meter	KVTDC	UTM zone 45 (WGS84)	

Table 3.2 Vector data set

Data	Year	Scale	Source	Description	Purpose
Administrative boundary	1995	1:25,000	Survey department	Valley, District, Municipality, Ward, VDC	Delineating study area extent
Road network	1995	1:25,000	Survey department	Road centrelines categoriesd as national highway, ring road, feeder road, urban road and district road	Road map was used as driving factors of urban growth by using proximity analysis.
	2008	1:10,000	KVTDC		
Rivers	1995	1:25,000	Survey department	River centrelines	River map was used as driving factors of urban growth by using proximity analysis
Public services	1995	1:25,000	Survey department	Facilities like- school, hospital, post office, petrol pump, temple etc.	Facilities map was used as driving factor of urban growth by using proximity analysis.
	2008	1:10,000	KVTDC		
Slope	1995	1:25,000	Survey department	Digital elevation map of cell size 20 m.	Slope map was used as driving factor of urban growth
Designated areas	1995	1:25,000	Survey department	Conservation areas categorised as wild life reserve and national parks	It was used as factor of urban growth.
Land use	1995	1:25,000	Survey department	Land use map with categories-airport, waterbodies, ponds, builtup, etc.	It was used as a reference for preparing land cover categories.

Table 3.3 Statistical data for population and land value

Data	Year	Source	Description	Purpose
Population	1991, 2001& 2011	CBS, Nepal	Population per ward and VDC	To produce population density map at ward level
Land value	1999 and 2010	KVTDC	Land value for each road category in wards.	To produce land value map

3.5. Identifying urban growth pattern

The main concern of the study was to find urban growth pattern of the Kathmandu valley. So an emphasis was given to detect built-up area using image classification. In order to quantify and map the urban growth following methods were adopted.

a) Supervised classification

Using Erdas Imagine Version 11, three multi-temporal Landsat datasets (Landsat TM and ETM+) with moderate resolution were processed to identify the changes in the land cover pattern of the valley for the years 1989, 1999 and 2010. These data were classified separately using a supervised classification technique taking all bands except the thermal band. Four general land cover classes-urban, water, forest and arable lands were used to analyze the changes in land cover pattern as shown in Table 3.4. Maximum likelihood supervised classification was employed in the production of land cover maps for the study area. Figure 3.3 shows some pictures for the land cover classes in the study area.

Figure 3.3 Different land covers of the study area

Table 3.4 Land cover classification scheme

Code	Land cover classes	Description
1	Urban	This class consists of urban fabrics, industrial, commercial, transportation and other built-up areas.
2	Waterbodies	It consists of water related features such as river, stream, lake, ponds *etc*.
3	Forest	It consists of different types of forest such as-deciduous forest, evergreen forests, mixed forest *etc*.
4	Arable land	This land cover class comprises of irrigated and non irrigated cultivable lands along with bare areas.

b) Post classification refinement

To improve the accuracy of the classified results, post classification refinements was conducted. It followed two steps- the first step used a majority filter using 3x3 window size to eliminate the noises of misclassified pixels. In second step, river network and other waterbodies were overlaid with the filtered image to retain the waterbodies in land cover maps (refer section 4.1.1). After applying post classification refinement, several misclassifications were corrected.

c) Accuracy assessment

The ground truth points used for assessing the accuracy of the classifications were selected using high resolution ortho-rectified aerial image of 1992, Ikonos image 2001 and World view image 2010 (see section 3.4 II). The accuracy assessment of the land cover maps include the generation of 100 random reference points for 1989, 1999 and 2010 land cover map from each high resolution images. After the post-classification refinements, accuracy assessment of the land cover maps was performed.

d) Change detection

Land cover change was determined by using three independent results of land cover classification by using post-classification approach. The 'from-to' change information was calculated in ArcGIS environment for each land cover and the rate of change was quantified. Areas of change and no change were then mapped for the built-up class.

In order to examine the urban growth intensity, an indicator called annual urban growth rate (AGR) was adapted for evaluating the rate of growth per unit area. The formula for AGR is expressed as follows:

$$\mathrm{AGR} = \frac{\mathrm{UA}n{+}i - \mathrm{UA}i}{n\mathrm{TA}n{+}i} \times 100\% \qquad (1)$$

Where, $\mathrm{UA}n{+}i$ and $\mathrm{UA}i$ are urban areas in target unit at time $i{+}n$ and i respectively, n is number of year and $\mathrm{TA}n{+}i$ is the total land area of the target unit to be calculated at the time of $i{+}n$ (Xiao *et al.*, 2006). According to Xiao *et al.* (2006), the target calculating unit is generally set to the administrative district so as to link with administration or economic statistics.

3.6. Modeling urban growth using logistic regression

3.6.1. Identifying probable driving factors of urban growth

The preliminary assessment of driving factors of urban growth was based on literature review. From the literature review an overview of driving factors of urban growth was listed which were grouped into seven categories (see Table 2.1). These factors were used as input variables for logistic regression modeling. However, safety factor, distance to electricity and distance to sewerage were not incorporated in the modeling due to data unavailability.

3.6.2. Preparation of input data for logistic regression model

The input data called factor maps were prepared for 1999 and 2010 in ArcGIS environment. Because of the unavailability of data for 1989, the LR modeling was done based on 1999 and 2010 data only. The dataset of 1995 from Department of Survey was used to prepare factor maps for 1999 and dataset of 2008 from KVTDC was used to prepare factor maps of 2010. Because of time difference between study period and data acquired, there might be some degree of uncertainty during modeling. Vector data of road network and public facilities such as-health and educational institutions for 2010 was available for 57 VDCs and 5 municipalities only. So, for remaining 40 VDCs of the valley, it was derived from 1995 dataset. Table 3.5 shows the list of driving factors (factor maps) for 1999 and 2010 that were used in LR model.

Table 3.5 Factors that were included in LR model

<table>
<tr><th>Category</th><th>Factors</th><th>1999</th><th>2010</th></tr>
<tr><td>Biophysical characteristics</td><td>Degree of slope</td><td colspan="2">=</td></tr>
<tr><td>Zoning</td><td>1- forest; 0-not forest</td><td colspan="2">=</td></tr>
<tr><td>Population density</td><td>Population density(person/sq.km)</td><td>●</td><td>●</td></tr>
<tr><td rowspan="9">Proximate causes</td><td>Distance to major roads(m)</td><td>●</td><td>●</td></tr>
<tr><td>Distance to minor roads(m)</td><td>●</td><td>●</td></tr>
<tr><td>Distance to water supply lines(m)</td><td colspan="2">=</td></tr>
<tr><td>Distance to CBD(m)</td><td colspan="2">=</td></tr>
<tr><td>Distance to industrial area(m)</td><td colspan="2">=</td></tr>
<tr><td>Distance to major rivers(m)</td><td colspan="2">=</td></tr>
<tr><td>Distance to urban centres(m)</td><td colspan="2">=</td></tr>
<tr><td>Distance to health facilities(m)</td><td>●</td><td>●</td></tr>
<tr><td>Distance to educational facilities(m)</td><td>●</td><td>●</td></tr>
<tr><td rowspan="2">Existing urban cluster</td><td>Proportion of urban area in the surrounding (a rectangular neighbourhood of 5x5)</td><td>●</td><td>●</td></tr>
<tr><td>Distance to existing urban cluster(m)</td><td>●</td><td>●</td></tr>
<tr><td>Economic factor</td><td>Land value</td><td>●</td><td>●</td></tr>
</table>

●, assumed to be different for each year
=, assume to have same value in each year

The factor maps were used as independent or explanatory variables that can be either dichotomous or continuous. Dependent variable of the model was binary output map that showed urban or non-urban for 1999 and 2010. All data were represented in raster format with cell size of 30 m. The cell size was determined by the spatial resolution of land cover maps that were prepared from Landsat images with 30m resolution. An approach of constructing factor maps is briefly described below:

Biophysical characteristics
Degree of slope was calculated from elevation grid/DEM of 20m which was resampled into 30 m cell size.

Zoning
Land use zoning of Kathmandu valley was not available in digital format. However, designated areas such as forest were derived from land use map of 1995 and assumed that it remains constant throughout the study period. The designated areas of forest were clipped from land use map 1995 and were prepared as dichotomous map of forest /no forest.

Population density
Population density mapping is generally done at an aggregate level for administrative zones, primarily to protect confidentiality and to reduce data volume. But in realism, population density is highly dependent on the geographic nature of an area. So in order to get an accurate information about population distribution the relationship between population density and geographic nature of area should be incorporated (Langford, 2006). In this study the population data were disaggregated into the 30 m cells using land cover maps. This approach was adopted from Langford (2006) which used the following population distribution scheme (see Table 3.6). Thus all urban cells with an administrative unit received 80% of the population.

Table 3.6 Population distribution scheme

Urban	80%
Waterbodies	0%
Forest	5%
Arable land	15%

The population data was available for 2001 and 2011 in tabular forms which was joined with ward boundaries of the valley. Ward boundary with population data was then intersected with land cover maps of 1999 and 2010 separately to get land cover information in each ward. Then each land cover was assigned with different weight factors as shown in Table 3.6 to get the information of number of population living in certain land cover types per ward. This population map was then converted into raster with 30 m cell size. Instead of absolute population at ward level, relative population density was calculated per sq.km.

Proximate causes
Proximity was determined by distance to major roads, minor roads, water supply, central business district, urban centers, major rivers, health facilities and educational facilities. For each cell, proximity was calculated as the Euclidean distance to the specific target location.

Road were classified into major roads and minor roads. This classification was based upon road map shown in 'Long term development plan for Kathmandu valley, 2002'. Major roads and ring road were assumed to remain constant throughout the study time period. Minor roads had slight changes between the time periods.

Urban centers are the core areas of the valley that have been performing as centre of business, commerce, administrative and political centers. Many services are delivered through these centres and thus they are centres of attraction for people. So it is expected that these urban centers will have significant influence on urban growth of the valley. To prepare distance map of urban centres first of all, core urban centres were digitized from Kathmandu valley land use zoning map of scale 1:10000 and Euclidean distance to these centres was calculated for all cells.

Distance to Major River was also considered as an important factor since these are the source of drinking water and irrigation for many inhabitants. Classification of Major River was based on land use map of 1995 from the Department of Survey.

Many literature use access to facilities such as health centres and educational centres as key factors of urban growth including- Cheng (2003); Aguayo *et al.* (2007); Priyanto (2010). So, these factors were also included in the study.

To create the factor map showing the distance to the industrial zones, industrial areas were extracted from the land use map of 2008 assuming that there were no significant changes in industrial areas during 1999-2008. This is true in the context since industrial areas remain constant throughout the years. To create factor map showing distance to the CBD, first the location of central business district was determined from the risk-sensitive land use plan report (EMI, 2010) and then the distance from the CBD was calculated for all cells.

Proportion of urban area in the surrounding
From an urban development perspective, the spatial influence principally comes from spatial agglomeration (Cheng & Masser, 2003). This means proportion of urban areas around certain location influence the growth of that location. In order to incorporate this spatial interaction effects, the factor maps showing proportion of urban land in the surrounding area was prepared. These maps show for each cell the proportion of urban cell in a 7x7 rectangular neighborhood. The size of neighborhood is determined by using different alternatives (such as- 3x3, 5x5 and 7x7) as suggested by Verburg *et al.* (2004); Xie *et al.* (2005); Hu & Lo (2007); Dubovyk *et al.* (2011).

Economic factors
Economic factors are also main driving forces of urban growth (Cheng & Masser, 2003). This study incorporates the land value as an economic variable. Land value in Kathmandu valley is generally determined by accessibility to road. The areas which are at closer proximity to road have higher land value than those which are further away. So for each ward the land value was determined by road category (Highway, blacktop, gravel, earthen, stone paved, brick paved, trail, stair and others). Land value was assigned to each road types to get land value contour. The whole road network was then converted to raster using land value attribute. It means cell closer to the road would achieve higher value and cell farther away from the road would receive lower values.

3.6.3. Data normalization

All input data were created in raster format. These data are either dichotomous or continuous type. The continuous data were normalized into the range of 0 -1. Normalization was done by minimum maximum linear transformation of input raster to achieve similar data range. This method is important since in multivariate statistical analysis such as logistic regression, all continuous variables should have the same the scale to compare the value within the same range (Cheng & Masser, 2003; Huang *et al.*, 2009; Priyanto, 2010).

3.6.4. Multicollinearity analysis

Dependencies between the explanatory variables are an important issue to account for in all multivariate methods (Cheng, 2003; Lesschen *et al.*, 2005). Before any regression analysis, data has to be checked on multicollinearity. If there is any correlation among the independent variables then it is an indication of collinearity. As collinearity increases among the independent variables, the level of efficiency of the estimates may become poor (Lesschen *et al.*, 2005). Moreover, very high levels of collinearity may definitely result in coefficients that are not statistically significant. Therefore it is prerequisite to check for multicollinearity of independent factors prior to modeling in any regression analysis.

There are two ways to detect multicollinearity. One way is to compute correlations(r) between all pairs of predictors. If r is very close to -1 or +1, then one of the two correlated predictors should be removed from the model. Another way is to calculate variance inflation factor for each predictor. Variance Inflation Factor (VIF) provides a reasonable indication of multi-collinearity between the variables (O'brien, 2007). Mathematically,

$$VIFi = \frac{1}{1-\mathrm{R}_{\mathrm{i}}^{2}} \tag{2}$$

where, R_i^2 is the coefficient of determination of the model that includes all the predictors except the i^{th} predictor. The most commonly used rule of thumb for tolerance of VIF is the 'rule of 10'. When any independent variable exceeds VIF value 10, then it should be eliminated from the analysis (Field, 2009). Therefore only those variables with VIF<10 were participated in LR model. After removing one variable with VIF>10, it is necessary to calculate the VIF for the remaining variables. This process needs to be repeated until VIF is lower than 10 for all the variables.

3.6.5. Spatial autocorrelation and sampling scheme

Spatial autocorrelation is designed to measure correlations within variables across space. It may be defined as "the property of random variables to take values over distance that are more similar or less similar than expected for randomly associated pairs of observations, due to geographic proximity" (Lesschen *et al.*, 2005). It is a method to measure how clustered/dispersed points are in space with respect to their attribute values(Priyanto, 2010).

Spatial autocorrelation can obscure the results of a regression analysis (Kok & Veldkamp, 2001). If autocorrelation is detected on the regression residuals, this can imply that non-linear relationships between the dependent and the independent variables are present or that one or more important predictor variables are missing (Overmars *et al.*, 2003). Many studies have examined spatial autocorrelation including- Kok & Veldkamp (2001); Cheng (2003); Overmars *et al.* (2003) which have found that the regression coefficient and significance of the contribution of individual variables are sensitive for the presence of autocorrelation. So, if autocorrelation is ignored it will lead to poor estimation of parameters and leads to false conclusion.

In logistic regression modeling, one of the most common methods to reduce spatial autocorrelation is by designing spatial sampling scheme. The most frequently used sampling scheme according to literature are either stratified random sampling and systematic sampling (Cheng, 2003). "Systematic sampling is effective to reduce spatial dependence but may lose some important information like relatively isolated sites when population is not spatially homogeneous. Conversely, stratified random sampling is efficient in representing population but low in efficiency in reducing spatial dependence especially local spatial dependence" (Cheng, 2003, pg 213). In this study systematic sampling scheme was applied to incorporate the effect of spatial autocorrelation. In other urban studies with similar areal extent and resolution, sampling window of 5x5 and 7x7 is proved to be effective to reduce spatial autocorrelation (Aguayo *et al.*, 2007; Wu *et al.*, 2009). In this study, different samples such as- 3x3, 5x5, 7x7, 9x9 and 11x11 were tested and final sampling was performed by applying 7x7 window size in Change analyst software as it gave significant p-values for most of the driving factors (see Table 4.11 in section 4.2.3).

One statistic for detecting spatial autocorrelation in LR model is Moran' I which measures spatial dependency. The value of Moran's I varies between +/-1, positive autocorrelation in the data produces positive values of I indicating tendency of clustering; negative autocorrelation produces negative values indicating tendency towards dispersion and no autocorrelation results in a value close to zero (Overmars *et al.*, 2003; Lesschen *et al.*, 2005). To examine the presence of spatial autocorrelation in the model, Moran's I index was calculated on the map of models residuals. In logistic regression analysis residual is defined as the difference between observed Y values and predicted probability.

3.6.6. Logistic regression model

LR has been widely applied to examine the influence of different factors on the pattern of land cover/use change. It predicts the probability of land cover change from non-built-up to built-up and estimates the empirical relationship between urban growth and driving factors that assumed to follow the logistic curve (Huang & Sin, 2010). The dependent variables (urban growth) in LR model are dichotomous and the independent variables (driving factors) are predictors of the dependent variable which can be measured on a nominal, ordinal, interval, or ratio scale. The general form of logistic regression is expressed as:

$$y = (b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n) \tag{3}$$

$$\mathrm{y} = \log\left(\frac{P}{1-P}\right) \tag{4}$$

$$\log\left(\frac{P}{1-P}\right) = (b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n) \tag{5}$$

Where, $x_1, x_2 \ldots x_n$ are driving factors or explanatory variables, y is a linear combination function of the explanatory variables representing a linear relationship (eq. (3)), b_0 is an intercept of the model, $b_1, b_2, \ldots b_n$ are regression coefficient or model parameters which are to be estimated, P is probability of transition of cell from non-urban to urban (Verburg *et al.*, 2002; Cheng & Masser, 2003). Function y is represented as the log of the odds or likelihood ratio that the dependent variable is 1 (eq. (4)). The value from equation 3 ranges from 0 to 1. If the value is close to 1, the probability of transition of cell from non urban to urban is very high and if it is close to 0, the probability of transition is very low. The predicted outcome is based on the predicted probabilities in which probability value greater and equal to 0.5 is classified as 1 and probability value less than 0.5 is classified as 0. Regression coefficients $b_1, b_2 \ldots b_n$ indicate the contribution of each driving factors on probability value P. A positive sign indicates that the particular factor will help to increase the probability of change and negative sign indicates that the particular factor will decrease the probability of change (Cheng & Masser, 2003). For e.g. if distance to major road has negative coefficient, then it indicates that further away from the major road *i.e.* high distance value, lower the probability of urban growth *i.e.* low probability value. Thus as a multivariate estimation method this technique helps to examine the direction of relationship and significance of the driving factors.

In this study binomial/binary logistic regression model is selected as it focuses on dichotomous result of 'urban' or 'non urban'. A value of 1 corresponds to the change of land cover from non urban to urban and a value of 0 corresponds to no change from non-urban to urban. The dependent variable is constructed from observed growth pattern using 1999 and 2010 land cover maps. The unknown parameters b_1, b,....., bn are estimated by using maximum likelihood algorithm which fits the greatest probability of observed growth (Czepiel, 2005; Huang *et al.*, 2009; Wu *et al.*, 2009). The modeling was performed in Change analyst extension tool in ArcGIS 10.

To interpret estimated parameters, Odds ratio (O.R) was used. Odd ratio is an indicator of change in odds resulting from a unit change in the predictor (Field, 2009). It is simply calculated as follows:

$$\Delta odds = \frac{odds\ after\ a\ unit\ change\ in\ the\ predictor}{original\ odds} \tag{6}$$

This proportionate change in odds is called odd ratio. If the value is greater than 1 then it indicates that as the value of predictor increase, the odds of outcome occurring increase and if the value is smaller than 1 it indicates that as predictor increases the odds of the outcome occurring decrease(Field, 2009).

The goodness of fit of LR model was based on the Chi-square test (Eq. 7). Chi-square is a statistical test that is commonly used to measure the fit of the observed values to the expected values. It tests the null hypothesis which states that independent variables have no influence on the model outcome (Christensen, 1997; Huang *et al.*, 2009). The formula for calculating Chi square is:

$$Chi-square(\chi 2) = \sum \frac{(observed\ value - expected\ value)^2}{expected\ value} \tag{7}$$

In order to accept or reject the null hypothesis, p-value is estimated and compared with significant level 0.05. If the estimated p-value is less than 0.05, then null hypothesis is rejected which proves that there is significant influence of independent variables on the model output.

To determine whether a variable is significant predictor of the model, T-Wald statistics (z-value) was used (Eq.8). Wald statistics tells whether the estimated coefficients for the predictor variables are significantly different from zero. If the coefficient is significantly different from zero then it can be assumed that the predictor is making a significant contribution to the prediction of the outcome (Cheng & Masser, 2003; Field, 2009). Wald statistic is calculated as follows:

$$Wald(z) = \frac{b}{SE_b} \tag{8}$$

Where, b is a coefficient of independent variable and SE_b is a standard error.

Two methods exist for the selection of significant variables- backward elimination and forward inclusion (Lesschen *et al.*, 2005). In forward inclusion procedure the independent variables are included in the process one by one until all the variables are involved in the iteration process and all significant variables are automatically selected. Whereas in backward elimination procedure all independent variables are included in the process at first and the insignificant variables are removed one by one until all the remaining variables are significant. Usually both methods produce the same result (Lesschen *et al.*, 2005). However, according to Dallal (2007), backward elimination has an advantage over forward elimination because in backward elimination joint predictive capability of the variables can be noticed as it starts with all variables in the model whereas forward elimination fails to identify this effect. So, a backward elimination procedure was applied to obtain the LR model with significant variables.

On the basis of Wald statistics the significance of each predictor variable was determined. If the p-value of the variable was greater than the confidence interval then the variable with highest significant value was removed and the model parameters were re-estimated. This process was repeated until all the variables had lower p-values than that of confidence interval.

Assumptions of Regression

There are number of assumptions that need to be satisfied to use the logistic regression model (Lesschen *et al.*, 2005) as listed below:

1. All independent variables are interval, ratio or dichotomous, and the all dependent variables are continuous.
2. Specification: a) all relevant predictors of the dependent variables are included in the analysis; b) no irrelevant predictors are included in the analysis; and c) the form of the relationship among the independent and dependent variables is linear.
3. Expected value of error is zero.
4. Errors are normally distributed for each set of values of the independent variables.
5. There is no autocorrelation among the error terms produced by different values of independent variables.
6. None of independent variables have perfect collinearity between each other.

3.6.7. Model evaluation

The evaluation of model was performed for the time span of 1999-2010. First, the development status of sample cells for 1999 and 2010 was determined. Then for each cell, the probability of change was computed with the estimated parameters. The model was then evaluated on the basis of percentage of correct prediction (PCP) and Kappa statistics.

Percentage of correct prediction (PCP)

PCP is one of the efficient way to assess the goodness of fit of the LR model that cross tabulate prediction with the observations (Xie *et al.*, 2005). It is defined as the percentage of observations that are correctly predicted i.e.

$$PCP = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \text{ x } 100 \tag{9}$$

Although PCP is an efficient way to measure model fit, it has an issue of classification rule. It arbitrarily use the classification rule that $P_i \geq 0.5 \rightarrow 1$ and $P_i < 0.5 \rightarrow 0$, this cut-off value treats an observation with $P_i=0.51$ as same as an observation with $P_i=0.99$ despite the fact that the former value explain much less than the latter. This can overstates the precision of PCP (Matt, 2012). So it is advisable to use alternative measure to evaluate the goodness of model fit (Dubovyk *et al.*, 2011).

Kappa statistic is an alternative method to measure the accuracy of the model prediction which ranges between -1 and 1 (Lesschen *et al.*, 2005). -1 is completely disagreement between observed and expected agreement, 1 is prefect agreement and 0 is exactly what would be expected by chance. The calculation is based on the difference between how much agreement is actually present (observed agreement) compared to how much agreement would be expected to be present by chance alone (expected agreement) (Viera & Garrett, 2005). Formula for Kappa statistics is:

$$Kappa\ (K) = \frac{P_o - P_e}{1 - P_e} \tag{10}$$

where, P_o stands for observed agreement and P_e for expected agreement. In addition to the standards kappa index, Lesschen *et al.* (2005) defines three variations of Kappa: Kappa for no information (K), Kappa for location (K_{loc}), and Kappa for quantity (K_{Histo}). K is an overall index of agreement, K_{loc} is an index that measures the agreement in terms of location only and K_{Histo} is the agreement in terms of quantity. According to Pontius (2000) a Kappa value higher than 0.5 can be considered as satisfactory for land use change modelling.

3.6.8. Predicting future urban growth

The probability map of urban growth for 2021 and 2032 were produced by fitting estimated model parameters and using factors of 2010 in equation 5.

3.7. Possibility of errors

The main probable source of error is due to the assumption that urban growth of the study area is explained only by the factors that are considered in the modeling. Others possible influences such as personal preference for household location, political factor, safety factor, land use policies etc. are not considered. Another source is due to accumulated error during GIS operations such as- calculating Euclidean distance, threshold value for determining window size, generalization of population data and land value. Another possible source is lack of recent data to validate the predictions of urban growth.

3.8. Software used

The software used in this study was Erdas Imagine 11, ArcGIS 10, Change analyst software (Huang *et al.*, 2009), Map comparison Kit 3, IBM SPSS statistics 20 and Microsoft office packages.

4. RESULTS AND DISCUSSION

This chapter deals with the main findings of the research including urban growth analysis, analysing driving factors of urban growth using a logistic regression (LR) model and projecting future urban growth pattern. The overall result of the analysis are presented in five sections: the first section explains about identification and quantification of spatio-temporal pattern of urban growth; the second section demonstrates the use of LR modeling to examine the relationship between urban growth and its driving factors; the third section presents evaluation of LR models; the fourth section demonstrates forecasting of future urban growth; and the last section presents discussion and possible implication of the result.

4.1. Identifying spatio-temporal pattern of urban growth

In order to examine the spatio-temporal change in urban growth, different land cover maps and land cover statistics were produced which are presented in the following sub-sections. The identified spatio-temporal pattern of urban growth is dependent on scale of analysis (resolution) and data availability. It would be more reliable if Landsat imageries were available for same season at equal time interval for the exact quantification of growth. But because of data constraints this study has used images from different season and time interval which introduce certain degree of uncertainty due to seasonal variation and temporal inconsistency.

4.1.1. Accuracy assessement of land cover classification

For the accuracy assessment of the land cover classifications 100 random reference points were generated from the high resolution imagery (see section 3.5c). Ismail & Jusoff (2008) indicate that an accuracy of 85% is acceptable level of digital image classification for Landsat imageries. However, in this study the waterbodies attained a very low accuracy level because most of the rivers have less than 30m width *i.e.* less than a pixel size of the image and comprise of shallow water. In addition sandy river beds have very similar spectral characteristics to that of built-up area. These facts made water bodies very difficult to detect and thus needed post classification refinement. Post classification refinement was done in two steps. First majority filter with a 3x3 window was used to eliminate misclassified pixels. Second the river network and other waterbodies were overlaid with the filtered image to retain the waterbodies in land cover maps. These features were extracted from the land use map of 1995. Before post classification refinement the overall accuracy of the classified images for the different years ranged between 82%-85% which increased to 89-94% after post processing (see Table 4.1 and Table 4.2). So overall the obtained land cover classification result is sufficiently accurate for the further analysis.

Table 4.1 Landsat image classification accuracies (%) before post processing

Land cover class	1989(%)		1999(%)		2010(%)	
	Producer's	User's	Producer's	User's	Producer's	User's
Urban	93	93	90	77	90	77
Water	10	50	10	100	10	100
Forest	100	100	100	100	90	100
Arable land	93	77	88	83	98	91
Overall accuracy	85		82		85	
Kappa statistics	0.76		0.72		0.76	

Table 4.2 Landsat image classification accuracies (%) after post processing

Land cover class	1989(%)		1999(%)		2010(%)	
	Producer's	User's	Producer's	User's	Producer's	User's
Urban	93	100	93	84	90	90
Water	80	89	80	80	80	89
Forest	100	100	100	100	100	100
Arable land	98	89	85	94	93	90
Overall accuracy	94		89		91	
Kappa statistic	0.91		0.83		0.86	

The highest producer's accuracy which corresponds to error of commission (see section 2.2.2) was obtained for forest class followed by arable land and urban class; water bodies attained only 10% producer's accuracy. After post-classification refinement, the producer's accuracy for water increased to 80% (Table 4.2). The result shows that there is still 20% inaccuracy in the classification which is probably due to conflict between urban class and bare class (included in arable land) which have similar spectral characteristics. The highest value of user's accuracy *i.e.* error of omission is attained by forest with 100% accuracy in all three years which is followed by urban class with accuracy level of 100%, 84% and 90% in 1989, 1999 and 2010 respectively after post classification refinement (see Table 4.2). This variation is probably caused by seasonal effect on land cover.

The Kappa statistics is automatically generated from the result of the land cover classification. The Kappa value of 0.76 and 0.72 in table 6 represents a probable 76% and 72% better accuracy than if classification is based on random unsupervised classification instead of the employed maximum likelihood classification. Landis & Koch (1977) indicate that the agreement is poor when Kappa<0.4, good when 0.4<Kappa<0.7 and excellent when Kappa>0.75. After refinement the Kappa value for all classified images is greater than 0.75 (see Table 4.2). Thus the overall accuracy is considered acceptable for this study.

4.1.2. Trend of land cover change in 1989, 1999 and 2010

After classification and accuracy assessment the trend of urban growth over time was analysed on the basis of the land cover maps generated. Figure 4.1 shows the land cover maps of the years 1989, 1999 and 2010. The map shows that significant urban expansion has occurred during this period. Most of the arable land and forest areas were transformed into urban lands. Table 4.3 demonstrate that the urban area of Kathmandu valley increased by 3% (19.2 km^2) in 1989 and 11% (71.6 km^2) in 2010 at an average annual rate of 0.4 km^2. Figure 4.3 is the bar chart which visually quantifies the area of land cover classes in different years. According to the chart, the study area is dominated by arable land and forest areas.

From Table 4.6, two main land cover transitions can be observed in between 1989-2010. Firstly during the period of 1989-1999, agricultural areas near the major roads and existing urban areas began to transform into urban land with an annual rate of 0.2 km^2. In this period a significant amount of agricultural land was converted into urban areas with the urbanization pattern following by the road networks and the existing built-up peripheries. The annual urban gain of 0.2% came at the expense of arable land. Secondly during 1999-2010 a different phenomenon of land cover transformation is observed compared to earlier period. In this period the urbanization continued at the proximity to major roads and existing built up areas but urban land gained its size at the expense of forest area. Annually 0.5 km² of forest area is lost due to conversion of urban built up area in this period.

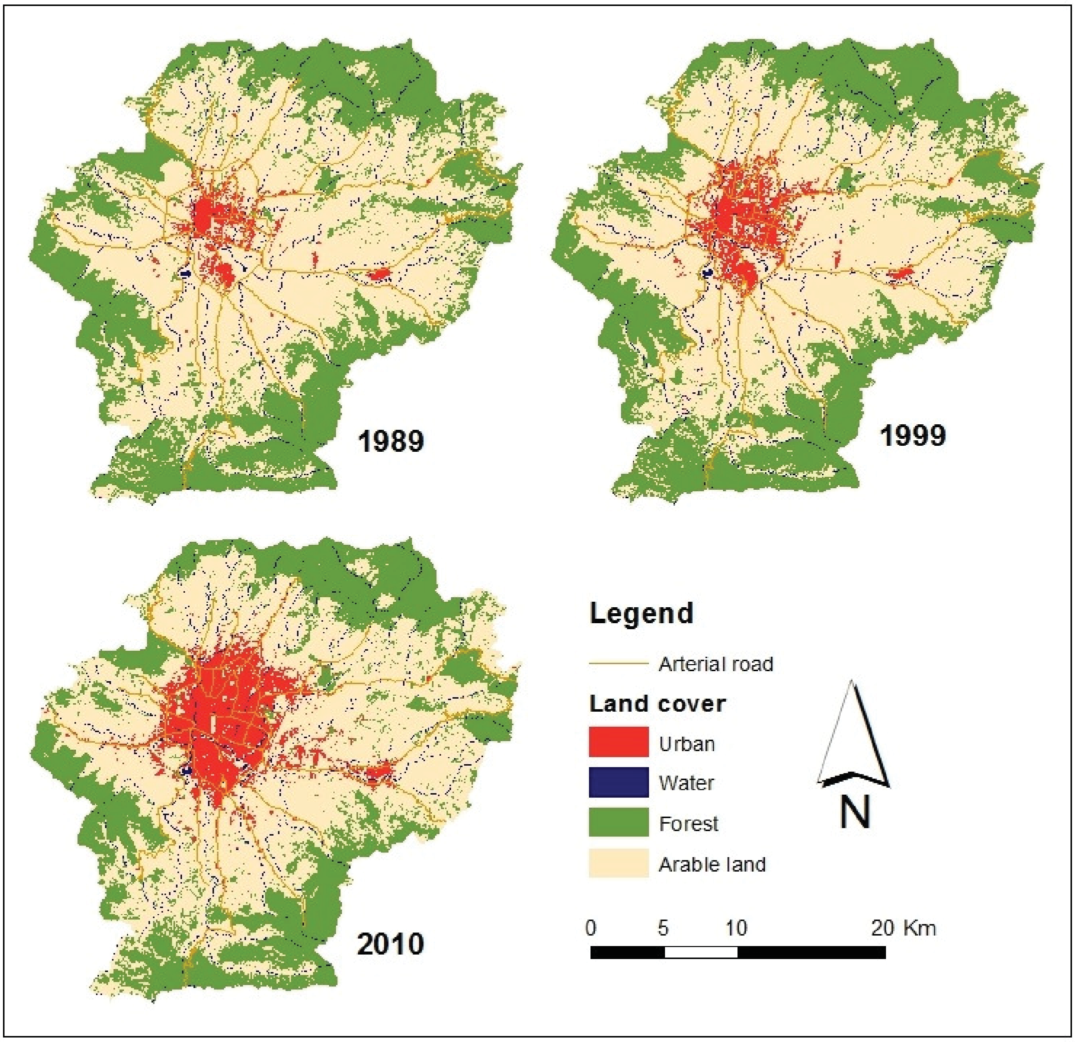


Figure 4.1 Land cover classification maps of Kathmandu valley

The land cover changes from one state to another can also be observed from Table 4.5 and Table 4.6. During 1989-1999, a significant amount of agricultural lands i.e. 16.5 km² was transformed into urban area and about 25.7km² of forest area is transformed into arable land (Table 4.4.) After 2000s, the urban area continued to expand around the existing urban periphery and along the major roads. This is the reason for the decline in arable land during 1999-2010 when 35.4 km² of arable land was transformed into urban area (Table 4.5). At the same time, forest area also faced a significant loss because of agricultural transformation. Around 56.5 km² of forest land was converted into arable land. By 1989-2010, around 58.3 km² of forest land has been converted into arable land and around 51.0 km² of arable land has been converted into urban area (Table 4.6). Of 71.6 km² of urban growth in 1989-2010, 97.4% was converted from arable land and 2.6% from forest area. This shows that the trend of conversion of forest land into arable land and arable land into built up area in the valley is noticeable.

Figure 4.2 visualize the spatial pattern of urban growth during 1989, 1999 and 2010. The total urban area of the Kathmandu valley increased from 19.3 km² in 1989 to 35.9 km² in 1999 with an annual expansion rate of 0.2%, which also means that the expanded area is almost two times the original urban area in 1989. The rate of expansion is not homogenous temporally. During 1999-2010 the speed of annual urban expansion accelerated to 0.5% and the expanded area almost doubled the original urban area in 1999.

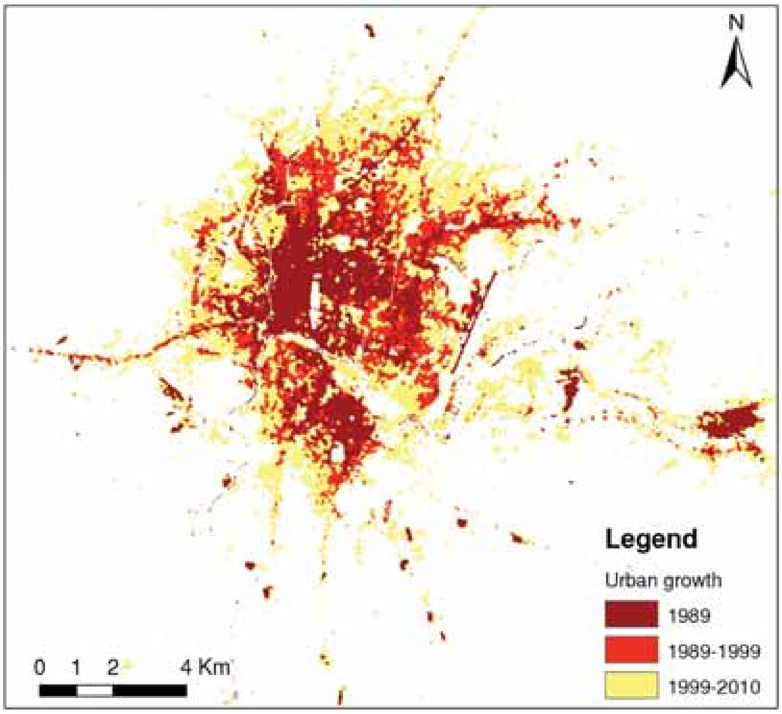
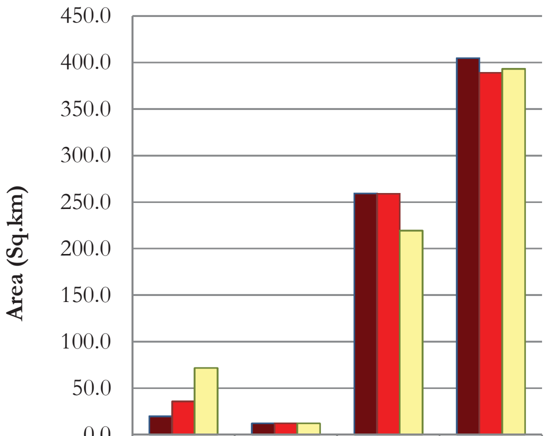


Figure 4.2 Spatial pattern of urban growth during 1989-2010



	Urban	Water	Forest	Arable land
1989 Area	19.8	12.2	259.4	404.8
1999 Area	35.9	12.4	258.9	389.0
2010 Area	71.6	12.2	219.4	393.0

Figure 4.3 Area of land cover classes in km² at the time of analysis

Table 4.3 Annual rate of land cover change (unit: km²/year)

	Area (km²)						Rate of change in %		
Classes	1989	%	1999	%	2010	%	1989-1999	1999-2010	1989-2010
Urban	19.8	3	35.9	5	71.6	10	0.2	0.5	0.4
Water	12.2	2	12.4	2	12.3	2	0	0	0
Forest	259.4	37	259.0	37	219.4	32	0	-0.5	-0.3
Arable land	404.8	58	389.0	56	393.0	56	-0.2	0	-0.1
Total area	696.2	100	696.2	100	696.2	100			

Table 4.4 The conversion matrix of land cover change from 1989 to 1999 (unit: km²)

1989	1999				
	Urban	Water	Forest	Arable land	Total
Urban	19.3	0.0	0.0	0.6	19.8
Water	0.0	12.2	0.0	0.0	12.2
Forest	0.1	0.1	233.4	25.8	259.4
Arable land	16.5	0.1	25.5	362.7	404.8
Total	35.9	12.4	258.9	389.0	696.2

Table 4.5 The conversion matrix of land cover change from 1999 to 2010 (unit: km²)

1999	2010				
	Urban	Water	Forest	Arable land	Total
Urban	35.6	0.0	0.0	0.2	35.9
Water	0.0	12.2	0.1	0.1	12.4
Forest	0.5	0.0	201.9	56.5	258.9
Arable land	35.4	0.0	17.5	336.1	389.0
Total	71.6	12.2	219.4	393.0	696.2

Table 4.6 The conversion Matrix of land cover change from 1989 to 2010 (unit: km²)

1989	2010				
	Urban	Water	Forest	Arable land	Total
Urban	19.2	0.0	0.0	0.6	19.8
Water	0.0	12.2	0.0	0.0	12.2
Forest	1.3	0.0	199.7	58.3	259.4
Arable land	51.0	0.0	19.7	334.0	404.8
Total	71.6	12.2	219.4	393.0	696.2

4.1.3. Pattern of urban growth

As mentioned earlier in section 2.1, spatial pattern of urban growth can be distinguished into three categories infill, expansion and scattered growth. In general infilling occurs in non built-up or vacant areas which is surrounded by existing built-up area and expansion occurs in the adjacent urban fringe spreading away from existing built-up patches. The third type of growth is scattered growth which means the formation of new urban patches without any direct spatial connection with existing built-up patches. In order to obtain infill growth and expansion areas of 1989-1999 and 1999-2010, a contiguous polygon of 1989 and 1999 built-up area were constructed (Figure 4.5). It is constructed by converting built-up area of 1989 and 1999 into polygon and aggregating all polygons which are within the distance of 200m. The threshold distance was adopted from the N.U.R.E.C. (2013) in which one of the criteria to define a contiguous built-up area is distance to building should be less than 200m. This polygon is taken as a reference to classify built-up area into infill type and outward expansion type. An assumption for determining infill and outward expansion is as follows- if urban growth is occurring inside the contiguous built-up area then it is infill type and if the growth is occurring outside the contiguous built-up area then it is expansion. Figure 4.4 shows the method that was adopted to determine infill growth and expansion.

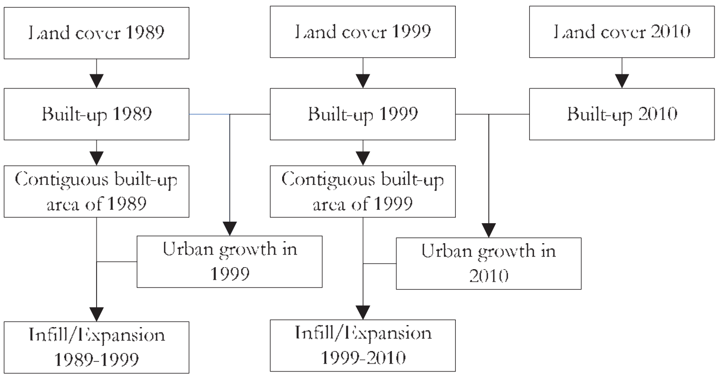


Figure 4.4 Method for determining infill and expansion

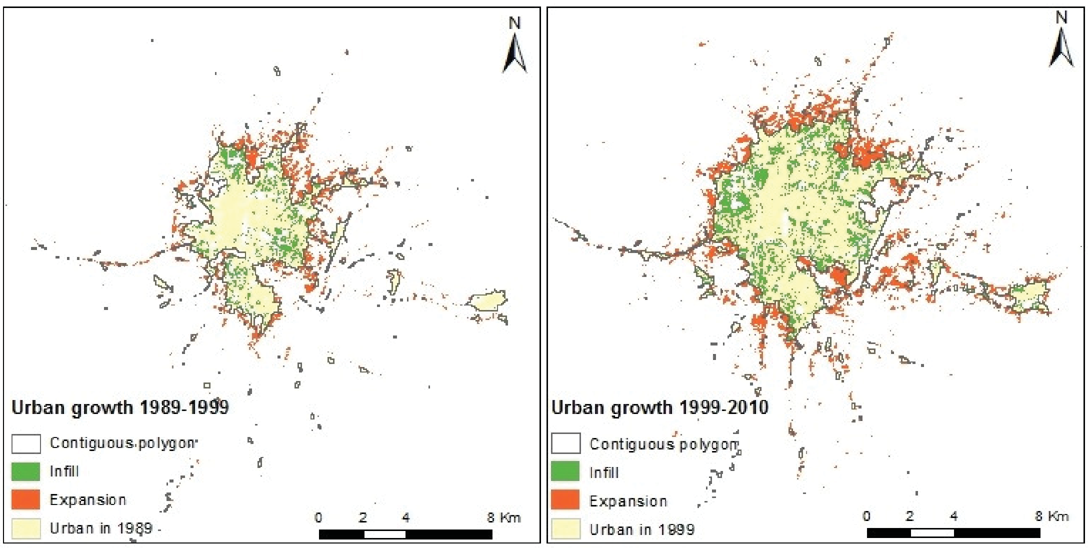


Figure 4.5 Infill development and outward expansion between 1989-1999 (left) and 1999-2010(right)

Figure 4.5 shows the spatial pattern of infilling and expansion of the valley during 1989-1999 and 1999-2010. During 1989-1999 the infill development was relatively low. Only around 5.7 km^2 of the vacant land inside the city were in-filled which is only 35% of total urban growth (see Table 4.7). Conversely, 65% of urban growth was occurring around the city fringes as outward expansion mainly along the four major transportation corridors-inner ringroad around the city periphery and arterial road in west, north and south. It should be noted that there are some cities like Bhaktapur, Thimi and Kirtipur located at the eastern and western part of the valley which is not the part of urban expansion since they are the traditional settlement of the valley. During 1999-2010 considerable areas have been developed along the major routes and around existing urban peripheries. Infill development was increased to 14.2 km^2 which is about 40% of total urban growth while 60% of land around city periphery was urbanized as an outward expansion.

Table 4.7 Types of urban growth (unit: km^2)

Type of growth	Area (km^2)			
	1989-1999	**%**	**1999-2010**	**%**
Expansion	10.9	65	21.7	60
Infill	5.7	35	14.2	40
Total	16.6	100	35.9	100

Figure 4.5 (right) shows that apart from infill and expansion, many new settlements are emerging in the valley in a form of scattered development. But this type of growth is not very prominent according to local experts. Moreover, it is very difficult to model scattered growth in logistic regression since its predictive power is poor for dispersed type of urban growth. Because of this reason scattered growth is not modelled separately, rather it is included in outward expansion. So the following sections will deal with LR modeling for overall urban growth, infilling and expansion only.

4.2. Identifying driving factors of urban growth

This section presents the main findings from LR modeling to examine the relationship between urban growths and driving factors in the study area during 1999-2010. The analysis of 1989-1999 driving forces was not performed due to data unavailability for 1989 driving forces. Section 3.6 already discussed about the preliminary assessment of driving factors which were based upon the literature review. This section will elaborate the results after applying LR model by using the methods explained in section 3.6.

4.2.1. Independent variable for urban growth

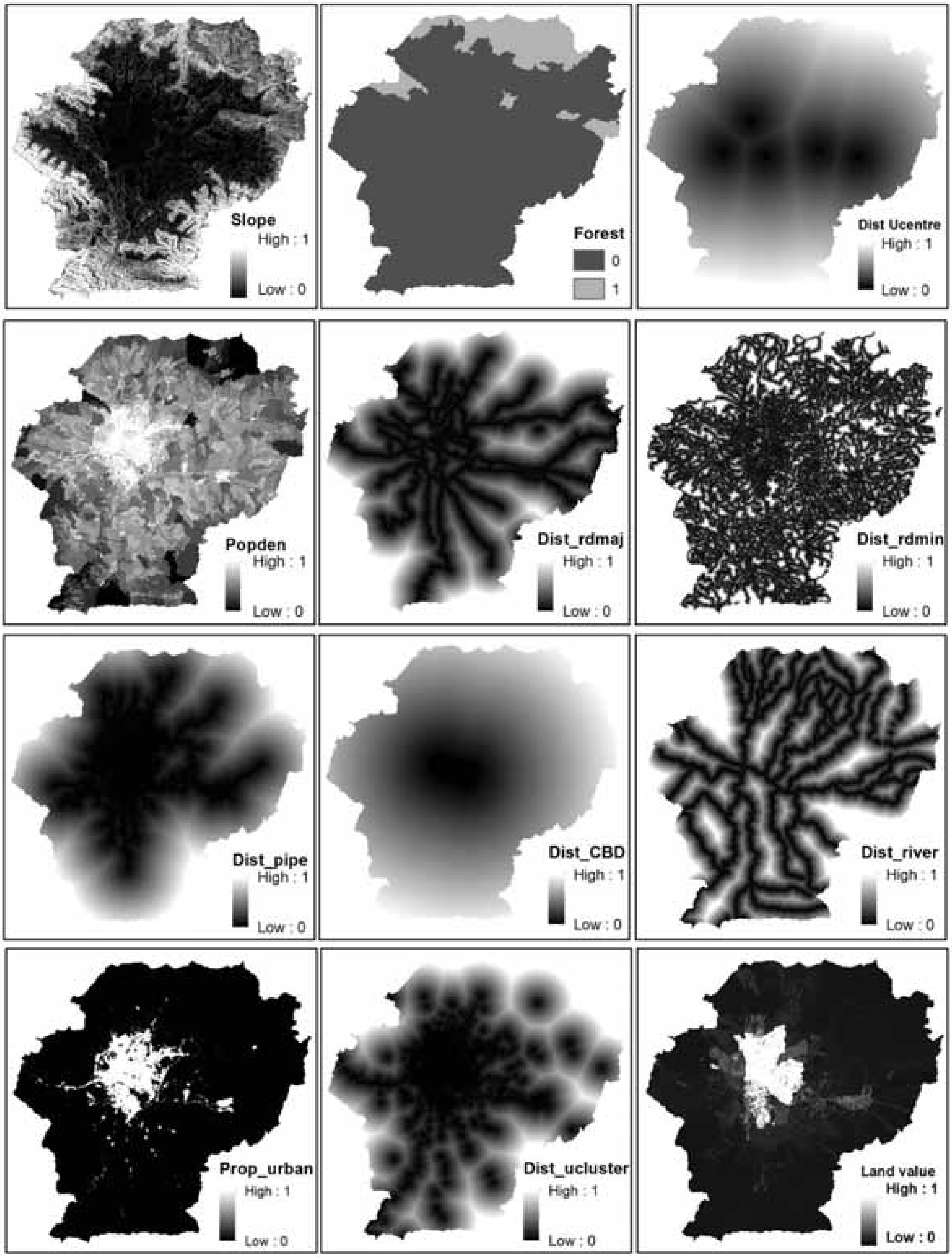
The LR model for overall urban growth, infill development and outward expansion were constructed on the basis of factors derived from the literature. Table 4.8 shows the list of independent factors that are derived from literatures and included in the LR model. Altogether 15 variables were entered in the LR model at first and modeling was performed for 1999-2010. Later variables were removed according to their collinearity value and significance level. A detail procedure for removal of variables is explained in section 4.2.2 and 4.2.3.

Figure 4.6 and Figure 4.7 are the factor maps of 1999 and 2010 used in this study. Only 8 factor maps for 2010 are shown in Figure 4.7 since other factors are same as that of 1999 (see section 3.6.2).

Table 4.8 List of variables and descriptions

Factor	**Variable in LRM**	**Description**	**Nature of variable**
Dependent	y	1-urban growth 0- no urban growth	dichotomous
Independent	*x*	-	-
Biophysical characteristics	Slope	Degree of slope	continuous
Zoning	Forest	1- forest; 0-not forest	dichotomous
Population density	Popden	Population density(person/sq.km)	continuous
Proximate causes	Dist_rdmaj	Distance to major roads(m)	continuous
	Dist_rdmin	Distance to minor roads(m)	continuous
	Dist_pipe	Distance to water supply lines(m)	continuous
	Dist_CBD	Distance to CBD(m)	continuous
	Dist_ind	Distance to industrial area(m)	continuous
	Dist_river	Distance to major rivers(m)	continuous
	Dist_Ucentre	Distance to urban centres(m)	continuous
	Dist_health	Distance to health facilities(m)	continuous
	Dist_edu	Distance to educational facilities(m)	continuous
Existing urban cluster	Prop_urban	Proportion of urban area in the surrounding (a rectangular neighbourhood of 7x7)	continuous
	Dist_ucluster	Distance to existing urban cluster(m)	continuous
Economic factor	Land_value	Value of land in each ward	continuous

Figure 4.6 Factor maps of 1999



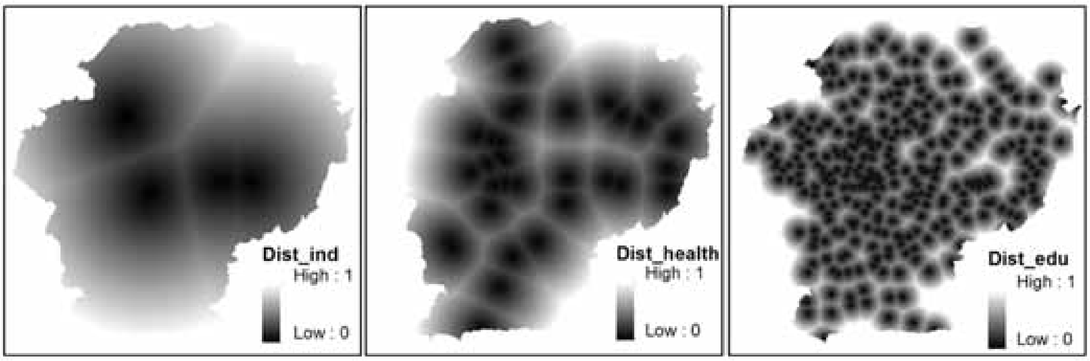
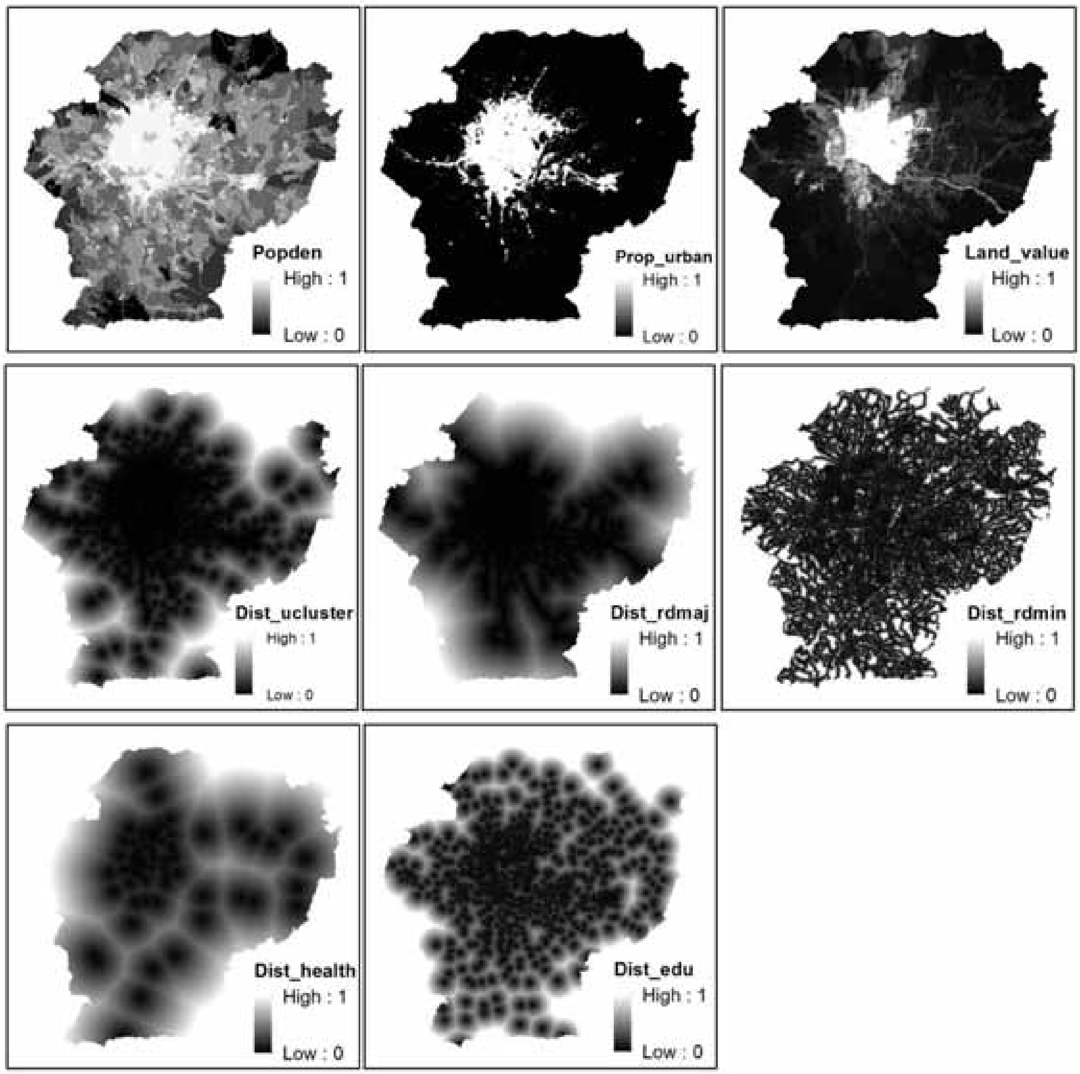


Figure 4.7 Factor maps of 2010



4.2.2. Multicollinearity diagnostics

Presence of collinearity among the variables can results in poor estimates of the model parameters (Lesschen *et al.*, 2005). So before any regression analysis can be performed it is pre-requisite to check multicollinearity among the independent variables. VIF is one of the indicative methods that help to check the collinearity among the variables (see section 3.6.4). According to Field (2009), VIF>10 should be eliminated from the analysis as it indicates the presence of very high collinearity. To overcome the problem of multicollinearity among the different driving factors, VIF test was performed for 1999-2010 driving factors. The test showed that distance to industries and distance to urban centres are highly correlated since they score VIF value 9.8 and 7.6 respectively (Table 4.9).

Table 4.9 Multicollinearity result for independent variables (Step 1)

S.N	Variables	Description	VIF
1	Slope	Degree of slope	1.9
2	Forest	1- forest; 0-not forest	1.7
3	Popden	Population density(person/sq.km)	2.2
4	dist_rdmaj	Distance to major roads(m)	1.4
5	Dist_rdmin	Distance to minor roads(m)	1.2
6	Dist_pipe	Distance to water supply lines(m)	3.6
7	Dist_CBD	Distance to CBD(m)	5.8
8	Dist_river	Distance to major rivers(m)	1.3
9	Dist_Ucentre	Distance to urban centres(m)	7.6
10	Dist_health	Distance to health facilities(m)	1.9
11	Dist_edu	Distance to educational facilities(m)	1.8
12	Prop_urban	Proportion of urban area in the surrounding (a rectangular neighbourhood of 7x7)	1.8
13	Dist_ucluster	Distance to existing urban cluster(m)	3.0
14	Land_value	Value of land in each ward	2.1
15	Dist_ind	Distance to industrial area(m)	9.8 (eliminated)

Table 4.10 Multicollinearity result for independent variables (Step 2)

S.N	Variables	Description	VIF
1	Slope	Degree of slope	1.8
2	Forest	1- forest; 0-not forest	1.6
3	Popden	Population density(person/sq.km)	2.2
4	dist_rdmaj	Distance to major roads(m)	1.3
5	Dist_rdmin	Distance to minor roads(m)	1.2
6	Dist_pipe	Distance to water supply lines(m)	3.6
7	Dist_CBD	Distance to CBD(m)	5.3
8	Dist_river	Distance to major rivers(m)	1.3
9	Dist_Ucentre	Distance to urban centres(m)	4.4
10	Dist_health	Distance to health facilities(m)	1.7
11	Dist_edu	Distance to educational facilities(m)	1.7
12	Prop_urban	Proportion of urban area in the surrounding (a rectangular neighbourhood of 7x7)	1.7
13	Dist_ucluster	Distance to existing urban cluster(m)	2.9
14	Land_value	Value of land in each ward	2.0

Although VIF for distance to industries is less than 10 (Table 4.9), it was eliminated from the analysis since the value is very close to 10. So at first distance to industries was eliminated. And the test was repeated again. The result in the second step showed that there was no significant multicollinearity among the variables since the VIF value for remaining variables are well below 10 (see Table 4.10). Therefore it can be safely concluded that there is no collinearity within explanatory variables in the current model.

4.2.3. LR model for overall urban growth 1999-2010

Logistic regression model for overall urban growth was built by using variables that were selected after multicollinearity diagnosis. Before building the regression model it is pre-requisite to choose the best sampling window size because presence of autocorrelation among the residuals produced by independent variables is very sensitive to the sampling window size (see section 3.6.5). So in order to choose the best sampling window size, different models were built for different sampling window varying from 3x3 to 11x11. At first these models were created by using all 14 factors. Then the final models were created by using backward stepwise procedure (see section 3.6.6 for backward elimination) which was used to select only those factors which have significant influence on the model estimation power (see Table A.1 to Table A.5 in Annexes I). The estimated parameters of the models were tested by using T-Wald statistics that determine whether the estimated coefficient of independent variable makes significant contribution to the result of model.

For each model, number of significant driving factors, percentage of correct prediction (PCP) and Moran's I were tested (Table 4.11). PCP was checked only for built-up area since the main concentration of the study is urban growth. In order to test the presence of spatial autocorrelation Moran's I was calculated on the residuals of estimated models (see section 3.6.5 for Moran's I). A residual in logistic regression model is defined as the difference between observed Y values minus predicted probability. The statistics for each model showed that sampling window size of 7x7 and 11x11 scored lower Moran's I with z-score within the critical range. As 7x7 has higher number of significant driving factors in comparison to 11x11, the model with 7x7 sampling size was selected as the final model for further analysis.

Table 4.11 Statistical test of LRMs of different sampling window size

S. No.	Criteria		Sampling window size				
			3x3	5x5	7x7	9x9	11x11
1	Significant Driving Factors		11	12	10	10	9
2	Percentage of correct prediction PCP		73.5	72.6	72.5	72.7	72.7
3	Spatial autocorrelation	Moran's Index	0.048	0.493	0.002	0.107	0.003
		z-score (critical values +/-1.96)	24.368	127.106	1.736	15.125	1.095

The final model of 7x7 sampling window was obtained on the fifth backward step after the elimination of factors like- distance to urban centres, distance to health, distance to river and forest since their T-wald statistics (p-values) were greater than assigned confidence level (0.05). The respective p-values for these variables are shown in Table A.6 in Annexes I. The model was significant with Chi-square value of 7465.285 and p-value less than 0.05. Table 4.12 shows the summary of model for overall urban growth. According to the table, different variable have different degree of influence on probability of urban growth. These effects are indicated by coefficient (*b*) which suggest that *b* variance in urban growth is explained by corresponding independent variable. A large positive *b* value indicates the strong positive relationship between urban growth and the predictor variables and the large negative *b* value indicates the strong negative relationship with urban growth.

In addition, each coefficients value is associated with corresponding odds ratio (O.R) which indicates that which variable has the highest influence to urban growth. The O.R shows the effect of increasing one unit of independent variable to the probability of urban growth holding the other independent variables constant at a certain value. From table 17, it can be seen that population density has the biggest influence on urban growth compared to other factors. It has coefficient value of 2.27 and O.R value 9.69 which can be interpreted as the increase of 1 unit in population density will influence 2.27 unit of urban area to change with probability value of 9.69. Other factors can be interpreted in the similar way and they can be ranked according to the value of O.R. as shown in table 17.

All the variables that have positive relation to urban growth have odds ratio (O.R)>1 indicating that in their presence the probability of urban growth is high. This means population density is the major driving force of urban growth with highest O.R of 9.6 followed by proportion of urban area in a neighbourhood with O.R 3.9 and land value with O.R 2.7. All the remaining variables that are negatively associated with urban growth have O.R<1 indicating that probability of urban growth is low if they are further away. For e.g. distance to water supply line (Dist_pipe) has the highest negative association with urban growth which indicates that as further away from a water supply line (high distance value) the less probability of urban growth (low probability value).

Table 4.12 Model parameters for 1999-2010 overall urban growth

Variables	b (Coefficient)	S.E.	Wald (z-value)	T-Wald test (p-value)	O.R	95% C.I.for O.R Lower	95% C.I.for O.R Upper
Constant	3.404						
Popden	2.271	.650	12.201	.000	9.690	2.709	34.656
Prop_urban	1.363	.615	4.911	.027	3.908	1.171	13.050
Land_value	.999	.400	6.242	.012	2.716	1.240	5.946
dist_rdmaj	-1.102	.313	12.360	.000	.332	.180	.614
Slope	-1.319	.647	4.163	.041	.267	.075	.949
Dist_edu	-2.476	.670	13.655	.000	.084	.023	.313
Dist_CBD	-2.753	.487	31.985	.000	.064	.025	.165
Dist_rdmin	-5.113	1.003	25.977	.000	.006	.001	.043
Dist_ucluster	-6.472	.561	133.113	.000	.002	.001	.005
Dist_pipe	-10.262	1.223	70.400	.000	.000	.000	.000

The strong positive relationship of population density seems reasonable in case of Kathmandu valley because both population growth and urban growth is consistently increasing as shown in Figure 1.1 in section 1.1. The valley urban area is still dominated by residential area which tends to have a high number of populations residing over there. Moreover urban growth of the valley is highly influence by land value with odds ratio of 2.7. Area which is at closer proximity to city centre has high land values and high tendency of growth. Conversely, an area which is farther away from city centre has low land value with low probability of urban growth. Another finding from the research is a significant relationship between proportion of urban area in a neighborhood area and probability of urban growth. The sparsely developed built-up area in the urban fringe in the valley is mostly clustered around the existing urban areas. This means areas which have higher proportion of urban areas have higher potential of growth than areas which have less urban proportion in the surrounding.

Other distance related factors such as distance to major road, distance to educational facilities, distance to CBD, distances to minor road and distance to existing urban cluster and distance to water supply line have negative association with urban growth. The model shows that urban growth of the valley has been controlled by road accessibility. The O.R for distance to major roads is 0.332 or 1/3.01 which indicates that the odds of urban development in an area closer to major road is 3.02 times as large as the odds of urban growth in an area 1 km further away from the major roads. This has contributed to the spatial pattern of linear development along the main roads. Similarly the result reveals that area closer to CBD has high probability to transform into built-up with an O.R 0.064 or 1/15.62 which is the evidence of the centralized mono-centric urbanization trend of the valley. Furthermore, distance to existing urban cluster has also negative association with urban growth, which means higher the distance from the existing urban areas lower is the probability of urban growth and lower the distance from the existing urban cluster higher is the probability of urban growth.

The model also demonstrates that slope is an important factor of urban growth in the Kathmandu valley with an O.R 0.26 or 1/3.74. This means the odds of urban growth in an area with lower degree of slope is 3.74 times as large as in the area with 1 degree more slope. This is confirmed from the land cover map as most of the growth is concentrated in valley floor which is relatively flat.

Table 4.13 Degree of influence of driving factors on urban growth according to expert's interview

Factors of urban growth		**High influence(3)**	**Moderate influence(2)**	**No influence(1)**	**Weight**
1. Bio-physical character	-Slope orientation		5		2
	-Degree of slope		4	1	1.8
2. Population growth	-Population density	4	1		2.8
3. Plans and policies	- Zoning		3	2	1.6
4. Proximate causes	-Distance to major roads	5			3
	-Distance to minor roads	4	1		2.8
	-Distance to ring road	5			3
	-Distance to water lines		4	1	1.8
	-Distance to sewer line		1	4	1.2
	-Distance to electric lines	1	4		2.2
	-Distance to CBD	3	2		2.6
	-Distance to industrial area			5	1
	-Distance to major rivers		1	4	1.2
	-Distance to urban centres	3	2		2.6
	-Distance to health facilities		5		2
	-Distance to educational facilities		5		2
5. Existing conditions	-Existing urban cluster	4	1		2.8
6. Economic factors	-Land value	5			3
7. Safety	-Political Safety		3	2	1.6

Regarding the degree of influence of driving factors, the result obtained from LR model and the result obtained from expert's interview is somehow coming to the same conclusion about the major driving factors of urban growth in Kathmandu valley (Table 4.13). For instance- population density, proportion of urban area in surrounding and land value has high influence on urban growth from both model and expert's interview whereas slope, distance to educational facilities and distance to water supply lines have moderate influence. However distance to major road is which has highest influence on urban growth according to expert's interview is showing moderate influence according to the model which is because of the fact that probability of urban growth decrease when distance from major increase. Besides there are some factors whose significance level are totally different in two cases. For example distance to urban centre, distance to health facilities and zoning which are significant drivers according to expert's interview are not significant at all according to the model and are thus eliminated from the model.

4.2.4. LR model for Infilling and Expansion 1999-2010

4.2.4.1. LR model for infill growth

LRM for infilling was obtained on the fifth backward step after eliminating the factors like-forest, land value, distance to educational facilitates and distance to CBD since their p-value were greater than 0.05 (See Table A.7 in Annexes I). The model was significant with Chi-square value of 503.98 with p-value less than 0.05. The sampling window for this model was reduced to 3x3 from 7x7 since it could not gain significant p-value for most of the variables used in the model which might be because of less infill growth. The estimated Moran's I showed that there is no spatial autocorrelation in the residual of the model. The overall PCP was 76% while PCP for built-op area is 97% (see Table 4.14).

Table 4.14 Statistical test of Infill growth model

S.N	Criteria		3x3
1	Significant Driving Factors		10
2	Percentage of correct prediction PCP		76%
3	PCP for built-up area only		97%
4	Spatial autocorrelation	Moran's Index	-0.002
		z-score (critical values +/-1.96)	-0.79

Table 4.15 shows the estimated parameters for the infill growth model where 10 variables are significant at p-vlaue less than 0.05. Among these significant variables, distance to urban centre, proportion of built-up area in a neighbourhood, distance to major road and population density have positive association to infill growth while other remaining variables have negative association indicating that increase in the value of these factor will decrease the probability of infill growth. All factor with positive coefficient have O.R>1 indicating that if the value of these variables is increased then infill growth tends to increase. While O.R<1 indicates that is the value of these variables are increased then probability of infill growth will decrease. The highest negative coefficient of the factor indicates the lowest influence of this particular factor on infill development of the area.

Distance to urban centre plays a significant role in infilling of vacant land inside the existing urban area. The probability of infill growth is very high on areas that are farther from urban centres than other areas. The odds of infill growth would increase by 19.7 if distance to urban centre is increased by one unit. This relationship is reasonable in Kathmandu valley since urban areas around the urban centres have already been filled without any space left for further infill development. Similarly, distance to major road has also a big influence on the probability of infilling. The closer the area to major roads the lower is the probability of infill growth with an odds ratio of 1.9 which means if the area is one unit farther away from major road then probability of infilling is increased by 1.9. This seems true in case of valley since lots of infilling are occurring inside the city pockets which are farther away from the major roads and highway. In contrast, proportion of urban areas in a surrounding area contributes very significant role in transformation of urban non built-up pockets into built-up. This factor has an odds ratio of 9.3 which implies that the probability of a neighbourhood to be infilled is 9.3 times larger when one urban cell is added to this neighbourhood. Moreover population density is also playing a major role to convert the non-built-up area inside the city into built-up with an odd ratio of 1. The probability of developing a vacant land inside the city is increased by 1 when population density is increased by one unit. This means the probability of infilling is higher in those areas where population density in high.

Table 4.15 Model parameters for 1999-2010 Infill growth

Variables	Coefficient(b)	S.E.	Wald (z-value)	df	T-Wald (p-value)	Odds ratio (O.R)	95% C.I.for O.R	
							Lower	Upper
Dist_ucentre	2.984	.751	15.782	1	.000	19.757	4.534	86.096
Prop_urban	2.234	.272	67.327	1	.000	9.335	5.475	15.916
dist_rdmaj	.649	.331	3.856	1	.050	1.914	1.001	3.661
Popden	.000	.000	136.176	1	.000	1.000	1.000	1.000
Dist_ucluster	-.831	.249	11.157	1	.001	.435	.267	.709
Dist_river	-1.734	.493	12.386	1	.000	.177	.067	.464
Dist_health	-1.813	.607	8.918	1	.003	.163	.050	.536
Slope	-2.410	.991	5.921	1	.015	.090	.013	.626
Dist_rdmin	-4.180	1.377	9.207	1	.002	.015	.001	.228
Dist_pipe	-11.832	2.743	18.601	1	.000	.000	.000	.002
Constant	.681	.285	5.692	1	.017	1.976		

Other factors such as- distance to urban cluster, distance to river, distance to health facilities, slope, distance to minor road and distance to water supply lines has negative relationship to infill urban growth. Infill growth tends to increase if distance to existing urban cluster is low with an odds ratio of 0.43 or 1/2.32. This indicates that the probability of infill growth is 2.32 times as high as the probability of an area that is a unit distance away from existing urban cluster. Similarly, distance to health centre is also playing important role for infill growth of the valley. This factor scored 0.16 O.R which can be interpreted as 1/6.25. According to this figure, the probability of infilling is 6.25 as big as the probability of another area which is farther away from health centre by one unit.

Another interesting finding from the result is the negative relationship between slope and infill growth. An area tend to have higher probability of infill growth if it has low degree of slope which seems logical as urban growth in the valley is occurring in relatively flat area than in surrounding elevated areas. Distance to river has also influenced the infill development of urban areas in the valley. Many rivers of the valley have no proper demarcation due to which river banks are encroached for different purposes. This is the reason for the infill growth in many parts of the city near the river banks. Another two factors *i.e.* distance to minor roads and distance to water supply lines have relatively low influence on the infill development since they scored lowest O.R as shown in table 22.

4.2.4.2. LR model for Expansion

LR model for expansion was obtained on the eight backward steps after the elimination of distance to health, forest, distance to urban centres, proportion of urban in a neighbourhood, slope and distance to river since their p-value were greater than 0.05 (see Table A.8 in Annexes I). The model was significant with Chi-square value of 1136.46 at p-value less than 0.05 and the estimated Moran's I showed that there is no spatial autocorrelation in the model's residuals. The overall PCP was 97% while PCP for built-up was only 22% (see Table 4.16). This is because most of the urban expansions are occurring in scattered manner. This type of growth is poorly predicted by LR model as it is not strong enough to predict dispersed growth.

Table 4.16 Statistical test of Expansion model

S.N	Criteria		7x7
1	Significant Driving Factors		7
2	Percentage of correct prediction PCP		97%
3	PCP for built-up area only		22%
4	Spatial autocorrelation	Moran's Index	0.00016
		z-score (critical values +/-1.96)	1.92

Table 4.17 gives the summary of estimated parameters of expansion model in which 7 factors are significant at p-value less than 0.05. According to the table, proportion of urban area in a neighbourhood and land value are positively associated to urban expansion while distance to CBD, distance to health centres, distance to urban cluster, distance to minor road and distance to water supply lines have negative relationship with urban expansion. All the factors which have positive coefficient value have O.R >1 indicating that in their presence urban expansion tends to increase. Conversely, those factors which have negative coefficient have O.R<1 indicating that increase in value of these factors will decrease the probability of urban expansion. In addition factor with the highest negative coefficient will have lowest influence in urban expansion.

According to Table 4.17, proportion of urban area in the surrounding have the biggest influence on urban expansion of the valley with O.R 19.6. This means the probability of urban area to expand will be 19.6 times greater if there is a unit increase in proportion of urban area in certain neighbourhood. This seems true in case of Kathmandu valley since most of the growth is occurring around the urban fringe where urban proportion is higher. Another major driving force for urban expansion is land value which scored the O.R of 5.2. This indicates that the probability of urban expansion is increased by 5.2 times if the land value is increased by one unit. The result seems ambiguous to certain extent since in general people would rather prefer cheaper land to build their house. But it is possible in Kathmandu valley because land value in Kathmandu valley is determined by accessibility to road networks. If access to road network is high then land value tends to be higher. Moreover people would prefer those locations which have higher accessibility inspite of higher land value. So although the land value is very high urban expansion is generally seen around road networks due to high accessibility.

Table 4.17 Model parameters for 1999-2010 Outward expansion

Variables	Coefficient (b)	S.E.	Wald	df	T-Wald (p-value)	Odds ratio (O.R)	95% C.I.for O.R	
							Lower	Upper
Constant	3.972	.669	35.226	1	.000	53.072		
Prop_urban	2.979	1.124	7.020	1	.008	19.666	2.171	178.134
Land_value	1.657	.672	6.080	1	.014	5.242	1.405	19.562
Dist_cbd	-2.721	.814	11.190	1	.001	.066	.013	.324
Dist_health	-3.018	.908	11.044	1	.001	.049	.008	.290
Dist_ucluster	-7.011	.993	49.884	1	.000	.001	.000	.006
Dist_rdmin	-9.286	1.865	24.782	1	.000	.000	.000	.004
Dist_pipe	-11.384	2.261	25.347	1	.000	.000	.000	.001

All other distance related factor such as distance to CBD, distance to health centres, distance to urban cluster, distance to minor road and distance to water supply line are negatively associated with urban expansion. Distance to CBD is one of the major drivers of urban expansion in the valley. The closer the distance from the CBD the higher is the probability of urban expansion with an O.R of 0.06. Distance to health facilities also plays a significant role for urban expansion as it score O.R of 0.04 or 1/20.4 which indicates that for every unit increase in distance to CBD will fall the probability of urban expansion by 20.4 times. Other factors such as- distance to urban cluster, distance to minor road and distance to water supply line have very low influence on urban expansion since they posses very low O.R and high negative coefficient value.

4.3. Model evaluation

Any urban growth model has to be evaluated for its accuracy of prediction before using it for the further analysis. Without doing so, it might produce a doubtful result. Therefore model evaluation is one of the important stages in assessing the quality of the developed LR model. There are different ways to evaluate the performance of LR model including percentage of correct prediction (PCP) and the Kappa statistic. The performance of urban growth model was evaluated by comparing the probability map of 2010 with the actual map of urban growth in 2010. The following sections present the evaluation of overall urban growth model, infill growth model and expansion model.

4.3.1. Evaluation of overall urban growth model

The PCP value for the overall urban growth model was quantified from the classification table (
Table 4.18). Normally classification assumes that if the probability is greater or equal to 0.5, then urban growth is expected to occur *i.e.* probabilities≥0.5 are classified as 1 and probabilities < 0.5 are classified as 0 (Lesschen *et al.*, 2005). By using this standard threshold value, the model predicts 96.29% of *y* occurrence correctly. The model seems to have a high prediction power due to the fact that the study area consists of higher proportion of unchanged cells which are correctly predicted. As a result it scored higher value of PCP. Since the focus of this research is urban growth, the model was evaluated only for the urban area. From
Table 4.18 below, the correct prediction for urban area is 72.5%. Remaining 27.5% could not be predicted by the model. This, on the one hand, suggests that not all the driving forces are presented in the LR model and, on the other hand, the phenomenon of urban growth can not be fully foreseen. However PCP greater than 70% is acceptable for urban growth models according to the literature.

Table 4.18 Model evaluation for 1999-2010 overall urban growth

Observed	Predicted					
	Non urban (0)	%	**Urban (1)**	%	**Total**	%
Non urban (0)	672437	99.1	6344	0.9	678781	100
Urban (1)	21724	27.5	57405	72.5	79129	100
Total	694161		63749		757910	
Correct predictions: 729842 Wrong prediction: 28068 Percentage of correct prediction (PCP): 96.29% PCP for urban area only: 72.5%						

Kappa statistic was used as an alternative method to measure the agreement between actual and predicted urban growth which ranges between -1 and 1(see section 3.6.7). It compares the two maps cell-by-cell to see if each pair of cells is equal or not in term of location and quantity. In addition to the standards Kappa index, Kappa for location (K_{loc}) and Kappa for quantity (K_{Histo}) was also determined. Kappa for location (K_{loc}) is the agreement of each pair of cells in term of spatial position on the map and Kappa for quantity (K_{Histo}) is the agreement of each pair of cells in terms of value (refer section 3.6.7). According to Landis & Koch (1977) the agreement is poor when Kappa<0.4, good when 0.4<Kappa<0.7 and excellent when Kappa>0.75. The comparison of observed urban growth and predicted urban growth in 2010 resulted in Kappa value=0.783, K_{loc}= 0.889 and K_{Histo}=0.881 (see Figure F.1 in Annexes II). The overall kappa value of 0.783 means there is 78.3% agreement between observed and predicted urban growth which is categorized as an excellent agreement. Thus it can be concluded that the driving factors used in the model are capable to describe the urban growth and they can be used to predict future urban growth pattern.

Figure 4.8 (left part) shows observed versus simulated urban growth in 2010. Although most of the urban and non urban areas are correctly predicted, the model does not estimate urban expansion fully correct. This is logical since the model is a statistical simplification of a complex real world phenomenon. From the figure it can be observed that the model is over-predicting those areas which are located inside the contiguous built-up area. This indicates that the vacant pockets inside the existing urban area have higher probability to convert into urban. Conversely, the areas which are outside of and disconnected to contiguous built-up area has lower probability to convert into urban. The right part of figure 16 shows the range and distribution of probability values of urban growth in 2010. Lighter tones indicate the higher probabilities of urban growth and darker tones indicated lower probabilities. To know the behaviour of these probability values, an under-prediction map and over-prediction maps are prepared along with the histogram as shown in Figure 4.9 and Figure 4.10.

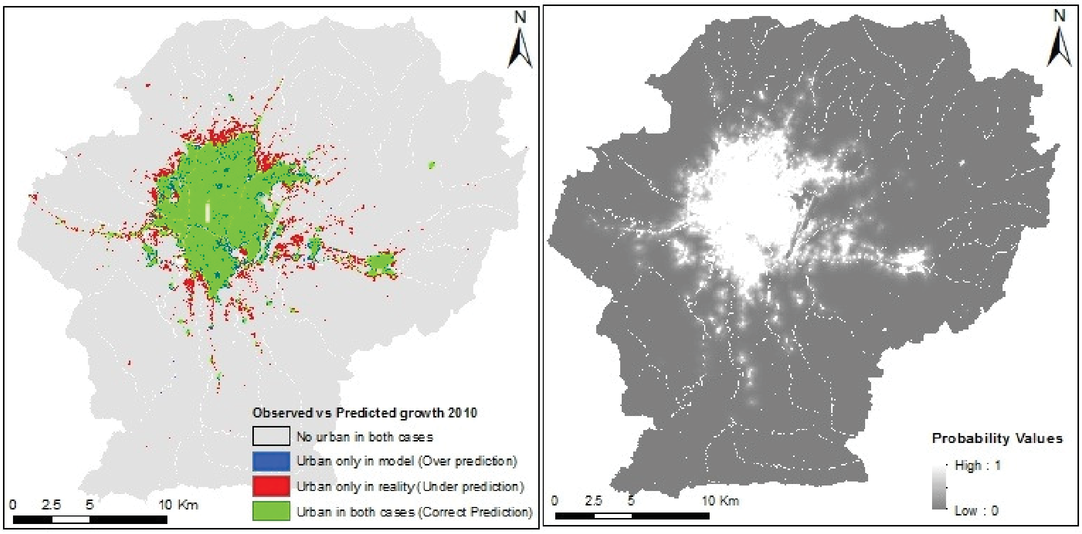


Figure 4.8 Observed and predicted urban growth at 2010 (left) and Probability map for 2010 (right)

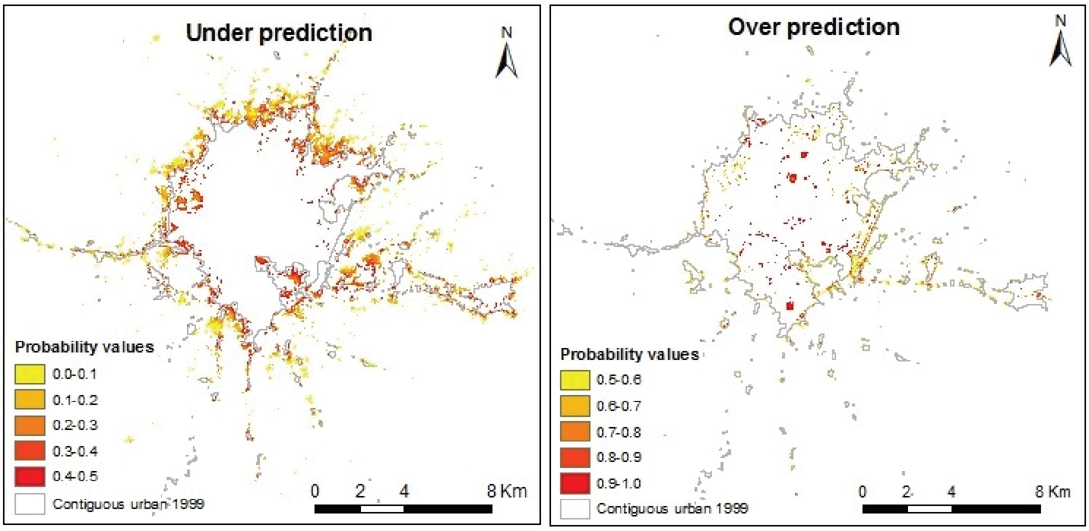


Figure 4.9 Range of probability values for under prediction (left) and over prediction (right) for 2010

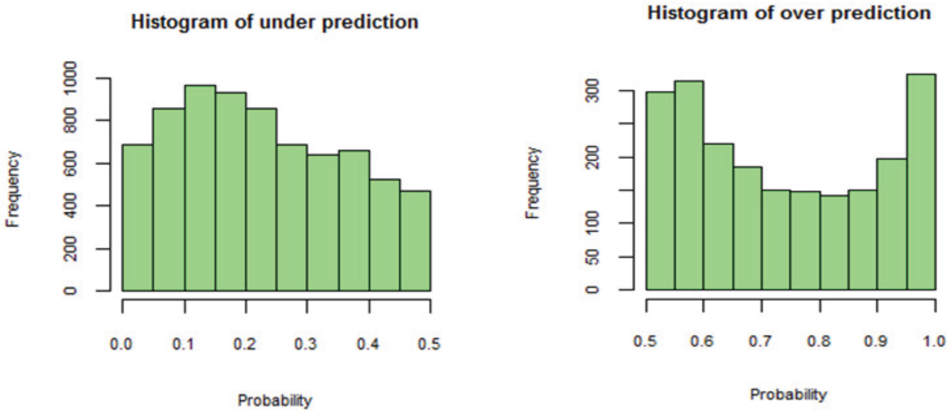


Figure 4.10 Histograms showing range of probability values for 2010

Figure 4.9 shows that most of the higher values of probability are concentrated near the city areas in the vacant pockets such as-parks, institutional areas and other cultivable and non-cultivable lands inside the city which are over-predicted by the model whereas lower probability values are distributed on the urban fringe areas that are under-predicted by the model. Figure 4.10 shows the number of under-predicted and over-predicted cells within different probability ranges. The left part of the figure shows that most of the under-predicted cells lie between probability values of 0.05-0.25 while the right part of the figure shows most of the over-prediction cells are in between the probability values range of 0.5-0.6 and 0.95-1. This indicates that infill development have higher probability values of urban growth than urban expansion.

4.3.1.1. Sensitivity analysis of overall urban growth model

To check the sensitivity of LR model for overall urban growth, different threshold values for classifying urban and non urban were tested by observing percentage of correct prediction (PCP) and its complement, the percentage of wrong prediction (PWP). From Table 4.19, it can be observed that although the threshold values can be between 0.6 to 0.15, but the value of PCP is always greater than 94% and the variation of PCP is only around 2%. So the main conclusion from the observation is that the model is not very sensitive to threshold values which might be due to three reasons. Firstly, the study area comprises of a large proportion of constant cells that did not changed from their original state (either 0 or 1). Because of this, model is correctly predicting all those constant cells and scoring higher PCP. Secondly, the driving factors which are used in the model are not entirely sufficient to explain the urban growth. Inspite of the elaborate literature review conducted (see Chapter 2) it is conceivable that one or more driving factors are not included in the current LR model. Thirdly, and probably most importantly, it must be mentioned that the LR model is a statistical simplification of reality whereas urban growth is a complex real world phenomenon.

Table 4.19 Model accuracy for whole study area at different classification thresholds

Threshold	**0.60**	**0.55**	**0.50**	**0.45**	**0.40**	**0.35**	**0.30**	**0.25**	**0.20**	**0.15**
PCP (%)	96.2	96.2	96.3	96.3	96.4	96.4	96.3	96.0	95.7	94.8
PWP (%)	3.8	3.8	3.7	3.7	3.6	3.7	3.7	4	4.4	5.2

4.3.2. Evaluation of Infill growth model

For infill growth model, the overall PCP was 76% which is high representational value for infill type of development (Table 4.20). However the model over-predicts around 97%, infilling almost all vacant pockets as shown in Figure 4.11. This indicates that in LR model, the probability of development of cells inside existing built-up area is very high. Because of this most of the vacant pockets are infilled even in the restricted areas such as-airport, institutional zones and public parks. These areas should have been excluded at the beginning so that model would not over estimate the growth.

Figure 4.11 shows the observed versus simulated infill growth in 2010. It shows that although the model correctly predicts a large amount of infill development, it also overestimates considerable amount of infill growth. The over-prediction is mainly occurring on those locations which have larger proportion of existing built-up area and closer proximity to ringroad. Figure 4.12 shows the range of probability values for under prediction and over prediction of infill growth in 2010. It shows that there is a few amount of under-prediction occurring mostly at the probability values 0.45 and 0.5. In contrast, a significant amount of over-prediction is occurring at the probability values 0.5 to 0.9. Figure 4.13 shows the quantity and distribution of under and over prediction at different ranges of probability values for infill growth.

The main conclusion from the analysis is that LR model for infill growth is not very accurate at prediction since it over-estimates significant amount of cells as infill. This is probably due to two reasons- first due absence of restricted areas as a driving factors which could have restrict growth and reduce over-prediction; and second due to high potential of vacant pockets to transform into built-up as they are closer to all facilities and have higher proportion of urban built-up in a surrounding.

Table 4.20 Model evaluation for 1999-2010 infill growth

Observed	Predicted					
	Non urban (0)	%	**Urban (1)**	%	**Total**	%
Non urban (0)	536	10	4657	90	5193	100
Urban (1)	405	3	15219	97	15219	100
Total	941		19876		20412	

Correct predictions: 15755

Wrong prediction: 5062

Percentage of correct prediction (PCP): 76%

PCP for built-up area only: 97%

Percentage of over-prediction: 90%

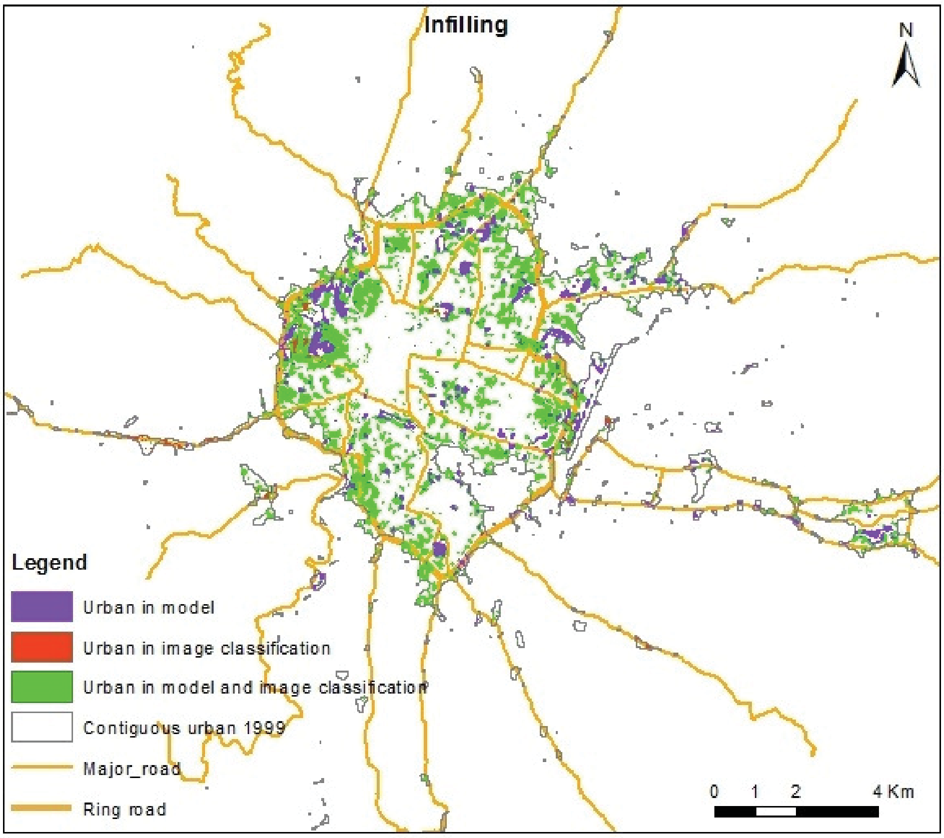


Figure 4.11 Comparison of observed and predicted infill growth in 1999-2010

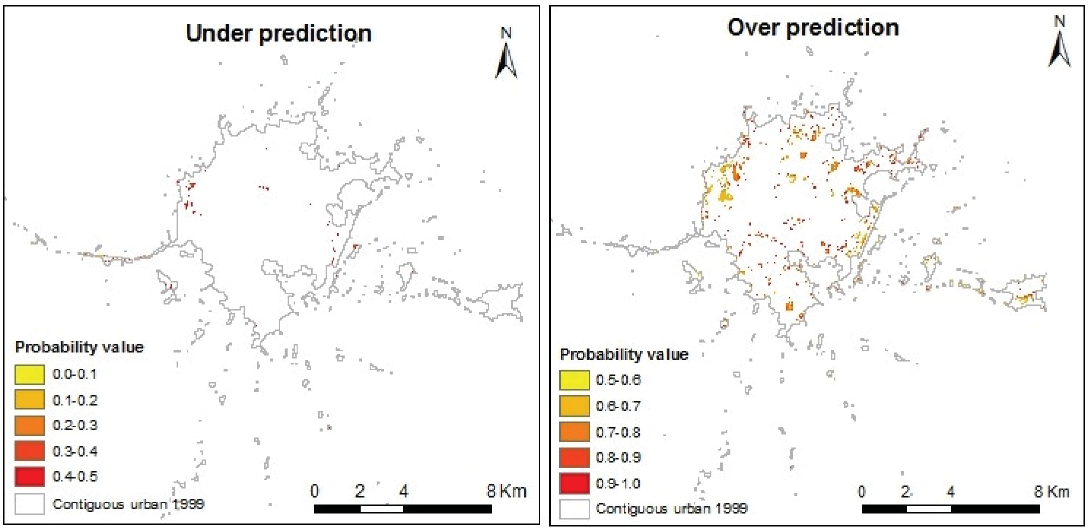


Figure 4.12 Range of probability values for under prediction (left) and over prediction (right) for 2010

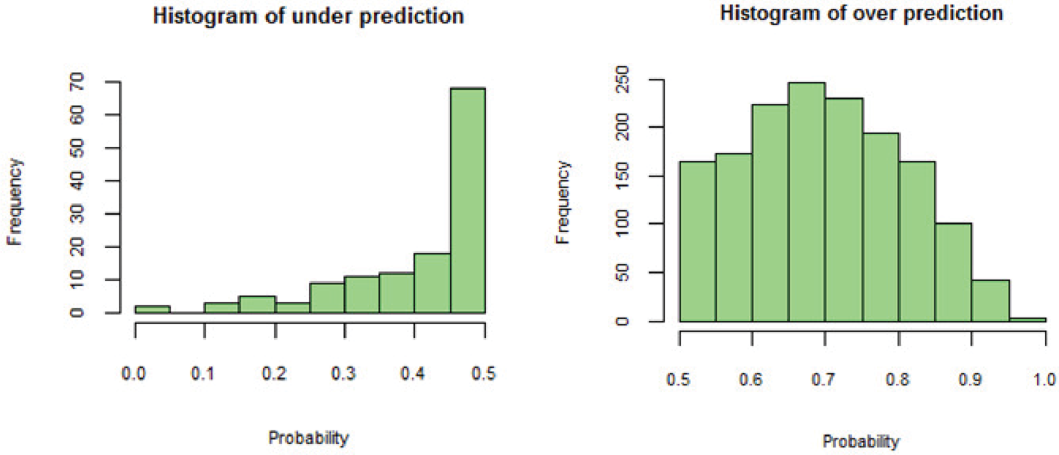


Figure 4.13 Histograms showing range of probability values for 2010

4.3.3. Expansion model

The overall accuracy for the expansion model was 97% with very low PCP for built-up area *i.e.* only 22% (Table 4.21). In Table 4.21, it can be observed that almost 99% of non-urban cells are correctly predicted by the model while 78% of urban cells are under-predicted. This is because of the fact that the study area comprised of large amount of non-urban cells than expansion cells. Most of the non-urban cells remain constant during the study period and are correctly predicted by the model. But expansion cells which are relatively very low could not be correctly predicted. This indicates that the expansion model is not very good in predicting outward expansion and scattered growth.

Figure 4.14 show that model is correctly predicting urban expansion predominantly on those locations which are attached to contiguous built-up area of 1999. Other scattered growth and dispersed growth which are disconnected to contiguous built-up area could not be fully predicted. Figure 4.15 show the spatial distribution of probability values for under-predicted and over-predicted cells. The frequency and distribution of probability values for under and over prediction can be seen in Figure 4.16. From both figures it can be observed that the amount of under-prediction is considerably higher than over-prediction. This indicates that expansion model is not very good at predicting scattered and disconnected growth.

Table 4.21 Model evaluation for 1999-2010 urban expansion

Observed	**Predicted**					
	Non urban (0)		**Urban (1)**		**Total**	%
Non urban (0)	669650	99%	3702	1%	673352	100
Urban (1)	18722	78%	5384	22%	24106	100
Total	688372		9086		697458	

Correct predictions: 675034

Wrong prediction: 22424

Percentage of correct prediction (PCP): 97%

PCP for built-up area only: 22%

From the evaluation result of expansion model, it can be conclude that LR model is not very good at predicting urban expansion in the region with predominant rural areas. If the region had higher proportion of urban area than rural area, it could have given better result for correct prediction of built-up. It might be also due to the driving factors which are not sufficient to explain the urban expansion.

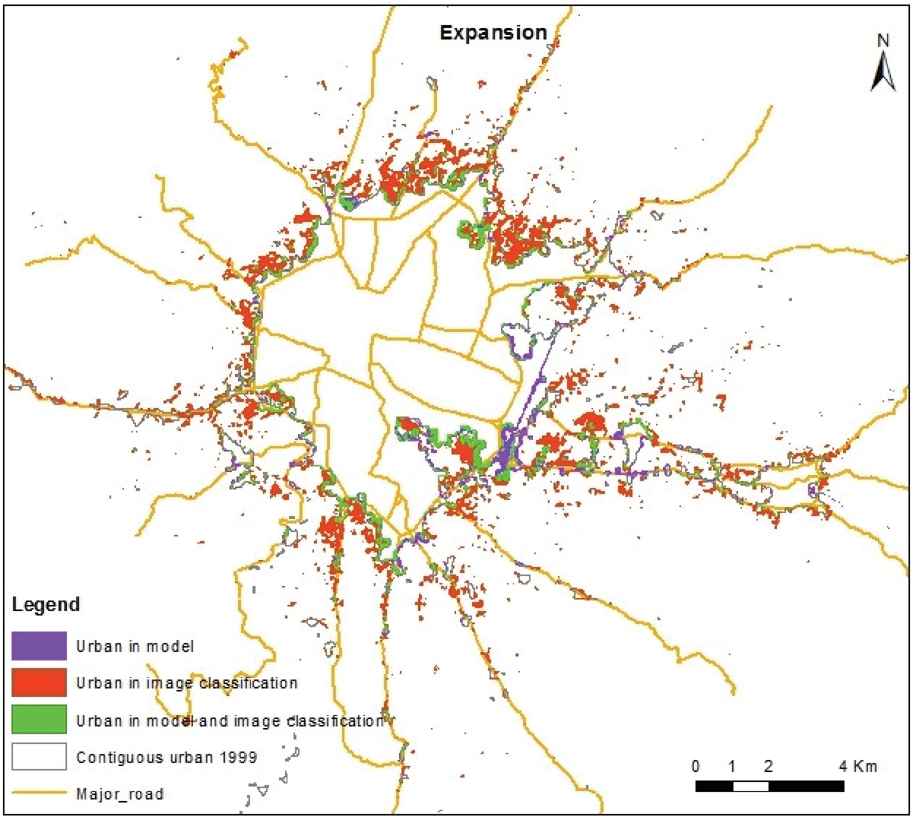


Figure 4.14 Comparison of observed and predicted urban expansion in 1999-2010

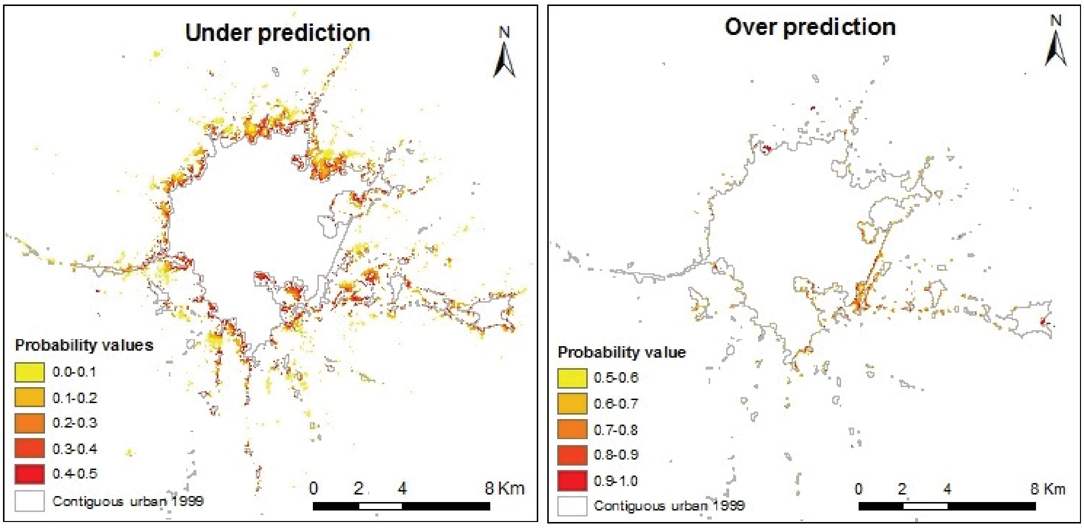


Figure 4.15 Range of probability values for under prediction (left) and over prediction (right) for 2010

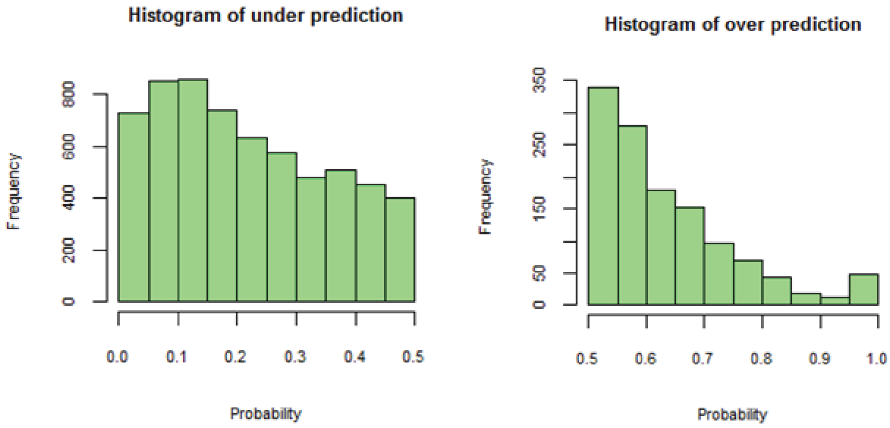


Figure 4.16 Histograms showing range of probability values for 2010

4.4. Probable areas of future urban growth

After evaluating the overall urban growth model for 1999-2010, the future urban growth was projected for the year 2021 and 2032 based upon the estimated parameters from LR model and driving factors of 2010. Figure 4.19 shows the spatial pattern of projected urban growth for 2021 and for 2032. According to Figure 4.17, in 2021 urban growth will be expanding along the major roads and highway corridors. As a result, most of the arable land in these areas is expected to be converted into built-up. Moreover for 2030, most of the areas in the southern and eastern part of the valley will be urbanized by eating up the arable land around the key urban centres. By this year according to the model, all municipalities except Bhaktapur in the east will be merged into greater Kathmandu city. Overtime, this outward expansion can threaten the reserved forest area in north part of the Kathmandu valley.

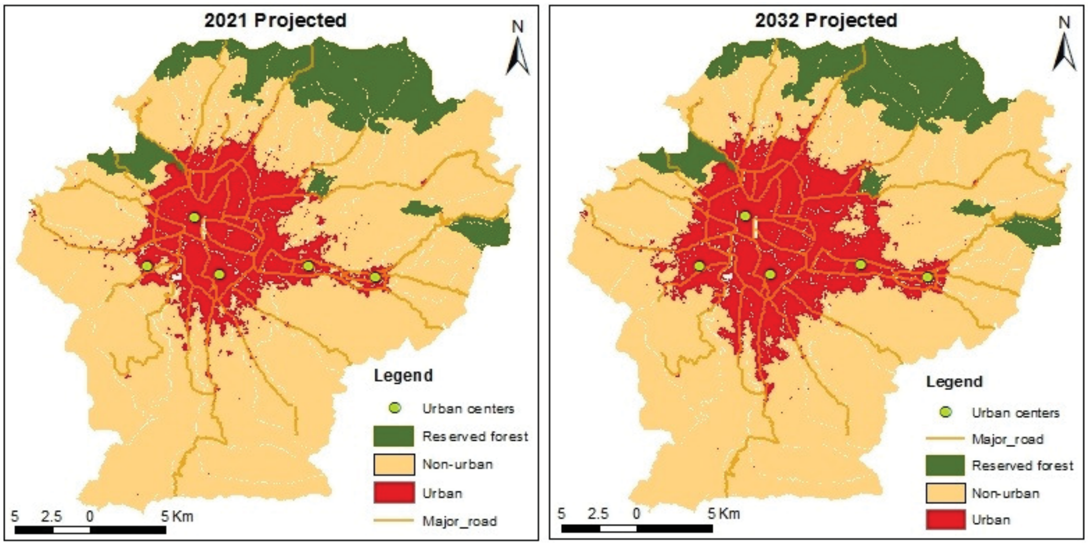
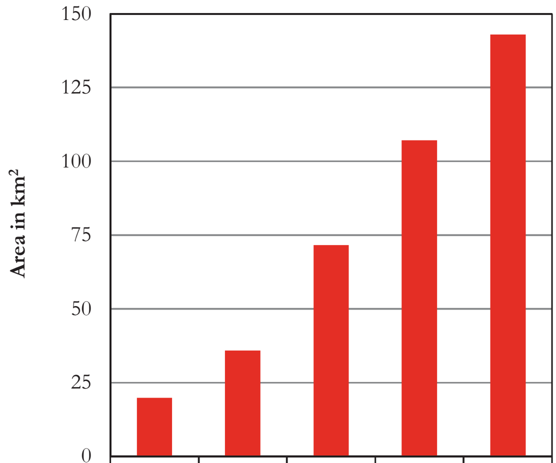


Figure 4.17 Projected urban growth for 2020 (Left) and 2030 (Right)

Figure 4.18 shows the trend of urban growth from 1989 to 2032. It shows that urban built-up areas which was around 19.8 km² in 1989 will increased by 5 fold in 2021 with an area of 107.2 and 7 fold in

2032 with an area of 143. This means by 2021 and 2032, the urban area is expected to become almost 1.5 and twice as much as the existing built-up area of 2010.

Figure 4.19 shows the expected spatial pattern of future urban growth in 2021 and 2032. From the figure it can be observed that most of the growth will occur in southeast and west direction. There is no prominent growth occurring in north direction in comparison to other sides probably due to topography *i.e.* higher degree of slope and greater distance to CBD and urban centres.



	1989	1999	2010	2020	2030
Urban	19.8	35.9	71.6	107.2	143.0
Non-urban	676.4	660.4	624.7	589.0	553.2

Figure 4.18 Trend of urban growth in km² (1989-2032)

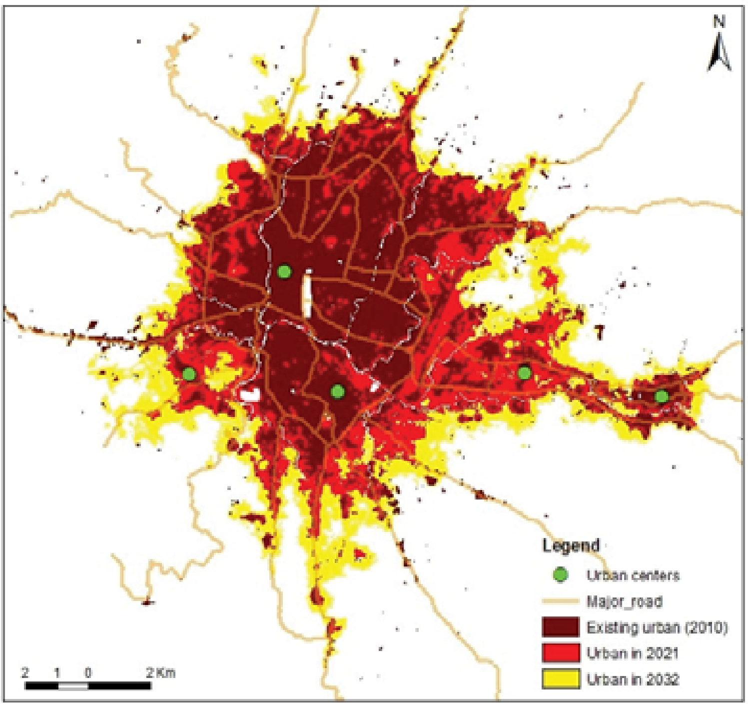


Figure 4.19 Projected urban growth pattern in 2021 and 2032

4.5. Discussion

This study has used the technologies of satellite remote sensing data combined with GIS to address spatio-temporal process of urban land cover change. The results of spatio-temporal analysis of land cover change indicate that there has been a notable change in urban land cover of Kathmandu valley during the entire study period. Two different study periods of time (1989-1999 and 1999-2010) had different mechanism of land cover change. First phase had urban development mainly at the expense of arable land and second phase had urban development mainly at the expense of forest area. In addition, different extent of infilling and expansion show that the trend of urban development was not regular spatially and temporally.

The use of Landsat data has proved to be successful to detect urban land cover change in the study area. The digital image classification along with GIS has demonstrated its ability to provide information on the rate, direction, and location of land cover change as a result of urbanization. However there has been an issue of uncertainness related to class inconsistency which has not been examined in this research. Although the overall accuracy of classified images was reasonably high, the accuracy of different land covers varies. For e.g. waterbodies attained significantly low accuracy in comparison to other classes which is due to site constraint such as- narrow width of the rivers which made it difficult to detect. Furthermore, sandy river beds were confused with urban class due to similar spectral properties. This confusion has hindered the attainment of accurate waterbodies. So to increase the accuracies of water bodies majority filter was performed at first to remove the misclassified pixels and then river networks were overlaid. Therefore the method has a limitation to improve accuracy of individual classes.

LR model was used to analyse urban growth and its driving forces which were constructed on the basis of driving factors derived from the literature and expert knowledge of planning professionals in Kathmandu valley. It was very challenging to acquire all these data. However it became true that even by using fewer data logistic regression can provide relevant information about spatial pattern of urban growth in future. In most of the urban studies, factors such as- population density, distance to major road and proportion of urban area in a neighbourhood are reported as major driving factors of urban growth (Cheng & Masser, 2003; Fang *et al.*, 2005; Hu & Lo, 2007; Huang & Sin, 2010). But the degree of influence of these variables differs from one case to another. In this study, the main driving factors for overall urban growth, infill development and outward expansion are similar but their degree of influence varies.

The evaluation of overall urban growth model was done on the basis of PCP and Kappa statistics which scored the PCP value of 72.5% and overall Kappa value of 0.783. The result from these evaluation methods indicates that the model is good at predicting overall urban growth. This was confirmed by comparing the result from other similar studies such as- Cheng & Masser (2003); and Huang *et al.* (2009). The future urban growth was predicted on the basis of driving factors of 2010 due to unavailability of data for future driving forces although many of these factors can subjected to change within few years. This may results in certain degree of uncertainity for the future urban growth projection. The evaluation results from infill growth model and expansion model shows that LR model is very good at predicting infill growth but at the same time it is much less accurate in predicting urban expansion. When it comes to forecasting urban expansion the model can much better predict the contiguous growth but has poor explanatory power when predicting dispersed or scatterd growth.

Besides, this research also indicates that LR is very sensitive for scale (resolution) of analysis. In this study the scale was chosen on the basis of resolution of Landsat imageries which has pixel size of 30x30 metre. Analysis conducted on higher spatial resolution is desirable since it may provide the detail information about causative effect of factors which are neglected in coarser scale. However, for city scale lower resolution is more reasonable because of the intense computational problem of finer resolution. Also there are some factors which can only be revealed by coarser scale of modeling especially for regional planning policies (Dubovyk *et al.*, 2011).

Spatial autocorrelation of the variables is an important issue which has to be taken care in LR model. According to Cheng & Masser (2003), the type of sampling scheme employed in the model can significantly affect the estimations of model parameters and model accuracy. This study incorporates systematic sampling scheme to reduce the spatial dependencies of model residuals. This method seemed to be effective for all three models to reduce autocorrelation. After testing different sampling window size varying from 3x3 to 11x11, a non overlapping 7x7 window was chosen for the overall growth model and expansion model as they provided with lowest Moran's I and the best parameter estimation. For infill growth model larger sampling window sizes resulted in insignificant p-values for the variables compared to the smaller sampling window. So 3x3 sampling window was selected for the further analysis since it resulted in better p-values for the variables in the model.

As illustrated in Table 2.3 in section 2.4, one of the disadvantages of LR model is lack of temporal dynamics in prediction of urban growth. Additionally it does not incorporate urban growth scenarios which can affect the pattern of future urban growth. LR model produce the probability map which indicates the possible locations of future urban growth but it does not indicate when it is going to happen. This is because the model does not proceed with time in a self-modifying way (Abebe, 2011). Furthermore, like CA model, LR model fail to incorporate the factors such as- people preference for household location, safety factors and political factors which might have significant influence on urban growth. For e.g. in Kathmandu valley one of the main drivers of urban growth is political and safety factor because there was a continuous inflow of rural-urban migration after Maoist insurgency which swell up the urban centres of the valley during mid 1990s (Timalsina, 2007).

Despite of all these limitations, the integration of remote sensing, GIS and LR model provide important information regarding the pattern and process of land cover change, relationship between urban growth and its driving factors and direction and magnitude of future urban growth. These results can be useful in urban planning domains to support policy development and decision making. The LR models for different types of growth provide different information for the decision makers. For instance- infill growth model informs the decision makers where are the probable areas of infilling and what are the major drivers of these types of growth while expansion model provides the information about what are the probable areas of urban expansion and which factor is driving urban expansion. These results would help decision makers to decide which areas need more interventions in order to manage the future urban growth. However there are some practical issues to implement these models. Abebe (2011) and Dubovyk *et al.* (2011) have noted that to use the LRM in urban planning practices, further improvement of the technique is required to make it more simple and user friendly. This tool has to incorporate different sampling schemes along with various statistical analyses for the LR modeling such as- spatial autocorrelation, multicollinearity test and model evaluation based on different parameters other than PCP alone.

5. CONCLUSIONS

This chapter discusses the main findings of this research based on the results as well as the methods and techniques applied to achieve answers to all research questions. It consists of two sections. The first section deals with the conclusive remarks on the main findings of the research. The second section points out possible directions for the further research based on findings and limitations of this study.

5.1. Conclusions

The main aim of the research was to determine driving factors of urban growth in Kathmandu valley and to forecast future urban growth in Kathmandu valley which was achieved by adopting several methods and techniques.

Regarding the method applied in this study, remote sensing and earth observation coupled with GIS has successfully identified the actual spatio-temporal process of the urban growth using Landsat data from different moments in time. Further, as a statistical approach, logistic regression analysis has made it possible to determine the significant driving forces of urban growth in the valley along with their degree of influence and direction of relationship with the urban growth. Moreover, data availability and data quality will have significant effect on the output of the LR modeling.

The following subsections draw the conclusions from the main findings of the research to address each objective that were set to meet the research aim.

5.1.1. Identification and quantification of urban growth pattern during 1989, 1999 and 2010

This section presents the main findings of the research that address the first objective. Three Landsat images of different years were classified using remote sensing and earth observation techniques for identifying and quantifying urban growth pattern of Kathmandu valley.

The digital image classification coupled with GIS has proved to be successful to detect urban land cover change in the study area. This technology has demonstrated its ability to provide information related to rate, pattern, direction and location of land cover change which is resulted from rapid urbanization of the valley. The result from this technique showed that the accuracy level was very poor for waterbodies (only 10%) before post-classification refinement. After post-classification refinement, the accuracy level was significantly increased for waterbodies. However, the issue of temporal and class inconsistency holds certain degree of uncertainty in the classification result which is propagated as an error factor to the further analysis.

Analysis of land cover maps prepared from classified images showed that there has been a significant change in urban built-up area during last 20 years at the expense of arable land and forest area. Based on these findings it can be concluded that all three land cover classes *i.e.* urban area, arable land and forest area has been subjected to prominent change during the entire study period.

Furthermore, the analysis of spatial pattern of urban growth was performed by defining two types of urban growth-infill and expansion. This definition is based on contiguous polygon of urban area with 200m threshold. But one of the limitations of using this boundary is that it classifies the urban growth based upon the position of delineated urban area which in fact is not always accurate. There is the possibility of infill occurring outside the boundary and expansion inside the boundary.

5.1.2. Identification of driving forces of urban growth

This section demonstrates the main findings for the second objective of the research which was concerned with determining the major driving forces of urban growth using LR model. To address the research questions of this objective, three models were prepared-overall urban growth model, infill growth model and expansion model. All the three models showed that the urban growth of Kathmandu valley is dependent on neighbourhood variables (proportion of urban area in a surrounding) and distance related factors (distance to minor road, distance to urban cluster, and distance to water supply line). The results suggest that city tends to grow in areas adjacent to already agglomerated area mitigating urban expansion. The driving factors of this type of growth are a good accessibility and connectivity. Nevertheless, not only these factors proved to be important but also population density, land value and slope.

Analysis of driving factor from overall urban growth model and experts interview show that some of the major driving such as- population density, proportion of urban area in a surrounding and land value are coming to the same conclusion while other factors such as-distance to urban centres, distance to health facilities and zoning are not coming to the consensus. This suggests that model and expert knowledge are two different entities which have different level of understanding. Model is simple statistical tool which operates at very local and detail level. In contrast, experts are the planning professional with a broad perception on urban growth phenomenon. Due to this, the result of major driving factors may not be always same.

Furthermore, from the model evaluation result, it was found that although the overall accuracies of all three LR models were very high but the PCP for built-up area were considerably low *i.e.* 72.5% for overall urban growth model and 22% for expansion model. Moreover the sensitivity analysis of overall urban growth model shows that it is not very sensitive to different threshold values. The results suggest that the proposed LR models are not very good to explain urban growth probably due to three reasons. First, the study area comprise of large proportion of unchanged cells which were correctly predicted by the model. Second, driving factors which were used in the model are not entirely sufficient to explain the urban growth. Third, the LR model is a statistical simplification of reality which could not fully explain the complex urban growth phenomenon.

5.1.3. Forecasting future urban growth

The third objective was to identify the probable areas of future urban growth. Overall urban growth was modelled for 2021 and 2032 on the basis of driving factors of 2010. Many of these factors are likely to change in future which may have different degree of influence on the urban growth. So the extrapolated urban growth might have certain degree of uncertainty regarding these factors. Furthermore, the probability maps are based on the extrapolation of historic urban growth process which is not certain to continue in the future. However it reflects the spatial pattern of urban growth, if historic process of growth continues in the future. If this is the case then the probability map generated for future urban growth will be helpful for planners and decision makers to make decision for guiding future urban growth of the valley.

5.2. Further research directions

For the further improvement in the findings of this study, some of the recommendations are listed below:

- This study has applied binomial LR model to analyse driving factors of urban growth. So effect of future urban growth on different land cover classes is not determined. Further research can be done using multinomial LR model to examine the effects of future urban growth on different land cover changes.

- Although LR model produces the probable areas of future urban growth, it does not specify when it is going to happen because it does not proceed with time in self-modifying way. This is one of the limitations of LR model. So further research can be done by using self-modifying approach so that factors in the model can be automatically updated with time.

- Due to unavailability of data for driving factors of urban growth during 1989-1999, LR model is constructed on the basis of 1999-2010 driving factors. So the model is not validated as it is constructed on the most recent data of 2010. This can be considered and extended in the future research.

- Earlier studies have already modelled urban growth of the valley by using Cellular automata. However, further studies can be done to investigate the differences between these two approaches- CA and LR modeling in the same study area and analyze both their advantages and disadvantages.

LIST OF REFERENCES

Abebe, F. K. (2011). *Modelling informal settlement growth in Dar es Salaam, Tanzania.* University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede. Retrieved from http://www.itc.nl/library/papers_2011/msc/upm/abebe.pdf

Adhikari, J. (2008). *Land reform in Nepal: problems and prospects*: NIDS.

Aguayo, M., Wiegand, T., Azócar, G. D., Wiegand, K., & Vega, C. E. (2007). Revealing the driving forces of mid-cities urban growth patterns using spatial modeling: a case study of Los Ángeles, Chile. *Ecology and Society, 12*(1), 13.

Allen, J., & Lu, K. S. (2002). Modeling and Predicting of Future Urban Growth in the Charleston, South Carolina Area. *Testimony presented to the United States Commission on Ocean Policy on January, 15*, 2002.

Anderson, J. (1982). Logistic regression. *Handbook of Statistics. North-Holland, New York*, 169-191.

Araya, Y. H., & Cabral, P. (2010). Analysis and modeling of urban land cover change in Setubal and Sesimbra, Portugal. *Remote Sensing, 2*(6), 1549-1563.

Arbury, J. (2005). *From urban sprawl to compact city: an analysis of urban growth management in Auckland.* Geography and Environmental Science)--University of Auckland.

Bagheri, M., Sulaiman, W. N. A., & Vaghef, N. (2012). Land Use Suitability Analysis Using Multi Criteria Decision Analysis Method for Coastal Management and Planning: A Case Study of Malaysia. *Journal of Environmental Science and Technology, 5*, 364-372.

Belal, A. A., & Moghanm, F. S. (2011). Detecting urban growth using remote sensing and GIS techniques in Al Gharbiya governorate, Egypt. *The Egyptian Journal of Remote Sensing and Space Science, 14*(2), 73-79.

Berling-Wolff, S., & Wu, J. (2004). Modeling urban landscape dynamics: A case study in Phoenix, USA. *Urban Ecosystems, 7*(3), 215-240.

Black, D., & Henderson, V. (1999). A theory of urban growth. *Journal of political economy, 107*(2), 252-284.

Booth, D., & Oldfield, R. (1989). A comparison of classification algorithms in terms of speed and accuracy after the application of a post-classification modal filter. *Remote Sensing, 10*(7), 1271-1276.

Camagni, R., Gibelli, M. C., & Rigamonti, P. (2002). Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion. *Ecological Economics, 40*(2), 199-216.

CBS. (2011). National population census Retrieved 23 August, 2012, from http://census.gov.np/index.php?option=com_content&view=article&id=22&Itemid=22

Cheng, J. (2003). *Modelling spatial and temporal urban growth*: Utrecht University.

Cheng, J., & Masser, I. (2003). Urban growth pattern modeling: a case study of Wuhan city, PR China. *Landscape and Urban Planning, 62*(4), 199-217.

Cheng, J., Masser, I., & Ottens, H. (2003). *Understanding urban growth system: Theories and methods*.

Christensen, R. (1997). Log-linear models and logistic regression.

Clarke, K. C., Hoppen, S., & Gaydos, L. (1996). *Methods and techniques for rigorous calibration of a cellular automaton model of urban growth.* Paper presented at the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment, 37*(1), 35-46.

Czepiel, S. A. (2005). Maximum likelihood estimation of logistic regression models: theory and implementation. *Published paper. Available on: http://czep. net/contact. html.(Assessed on April 5, 2011).*

Dallal, G. E. (2007). The Little Handbook of Statistical Practice, *Simplifying a Multiple Regression Equation* Retrieved 4th Feb, 2013, from http://www.jerrydallal.com/LHSP/simplify.htm

De Almeida, C. M., Vieira Monteiro, A. M., Camara, G., Soares-Filho, B. S., Coutinho Cerqueira, G., Lopes Pennachin, C., & Batty, M. (2005). GIS and remote sensing as tools for the simulation of urban land-use change. *International Journal of Remote Sensing, 26*(4), 759-774.

Dubovyk, O., Sliuzas, R., & Flacke, J. (2011). Spatio-temporal modelling of informal settlement development in Sancaktepe district, Istanbul, Turkey. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(2), 235-246.

EMI. (2010). Risk-Sensitive Land Use Plan Final Report, Kathmandu Metropolitan City, Nepal.

Fang, S., Gertner, G. Z., Sun, Z., & Anderson, A. A. (2005). The impact of interactions in spatial simulation of the dynamics of urban sprawl. *Landscape and Urban Planning, 73*(4), 294-306.

Field, A. (2009). *Discovering statistics using SPSS*: Sage Publications Limited.
Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote sensing of environment, 80*(1), 185-201.
Guo, H., Liu, L., Huang, G., Fuller, G., Zou, R., & Yin, Y. (2001). A system dynamics approach for regional environmental planning and management: A study for the Lake Erhai Basin. *Journal of Environmental Management, 61*(1), 93-111.
Haack, B. N., & Rafter, A. (2006). Urban growth analysis and modeling in the Kathmandu Valley, Nepal. *Habitat International, 30*(4), 1056-1065.
Hall, P. G., & Tewdwr-Jones, M. (2010). *Urban and regional planning*: Taylor & Francis.
He, C., Okada, N., Zhang, Q., Shi, P., & Zhang, J. (2006a). Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Applied Geography, 26*(3), 323-345.
He, C., Okada, N., Zhang, Q., Shi, P., & Zhang, J. (2006b). Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Applied Geography, 26*(3-4), 323-345.
Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote sensing of environment, 86*(3), 286-302.
Hu, Z., & Lo, C. P. (2007). Modeling urban growth in Atlanta using logistic regression. *Computers, Environment and Urban Systems, 31*(6), 667-688.
Huang, B., & Sin, H. L. (2010). Uncovering the Space–Time Patterns of Change with the Use of Change Analyst–Case Study of Hong Kong. *Advances in Earth Observation of Global Change*, 255.
Huang, B., Zhang, L., & Wu, B. (2009). Spatiotemporal analysis of rural–urban land conversion. *International Journal of Geographical Information Science, 23*(3), 379-398.
Ismail, M. H., & Jusoff, K. (2008). Satellite data classification accuracy assessment based from reference dataset. *International Journal of Computer and Information Science and Engineering, 2*(2), 96-102.
ITC. (2010). *GI Science and Earth Observation, a process-based approach*. Enschede.
Kivell, P. (1993). *Land and the city: patterns and processes of urban change*: Psychology Press.
Kok, K., & Veldkamp, A. (2001). Evaluating impact of spatial scales on land use pattern analysis in Central America. *Agriculture, Ecosystems & Environment, 85*(1–3), 205-221.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
Langford, M. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems, 30*(2), 161-180.
Lesschen, J. P., Verburg, P. H., & Staal, S. J. (2005). *Statistical methods for analysing the spatial dimension of changes in land use and farming systems*: Citeseer.
Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing, 28*(5), 823-870.
Maktav, D., & Erbek, F. S. (2005). Analysis of urban growth using multi-temporal satellite data in Istanbul, Turkey. *International Journal of Remote Sensing, 26*(4), 797-810.
Matt, G. (2012). Binary Response Models, *Percentage of correct prediction*. Retrieved from https://files.nyu.edu/mrg217/public/binaryresponse.pdf
McFadden, D. (1973). Conditional logit analysis of quantitative choice behavior In P. Zarembka F. i. econometrics (Ed.)
N.U.R.E.C. (2013). Network on Urban Research in the European Union, *Agglomeration Delimination* Retrieved 27th Jan, 2013, from http://www.oeaw.ac.at/isr/comet/documents/Final_Results/COMET_CTT/delimit/delimit-39.html
O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673-690.
Oluseyi, O. F. (2006). Urban land use change analysis of a traditional city from remote sensing data: The case of Ibadan metropolitan area, Nigeria. *Humanity and Social Sciences Journal, 1*, 42-64.
Overmars, K. P., de Koning, G. H. J., & Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling, 164*(2–3), 257-270.
Pontius, R. G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing, 66*(8), 1011-1016.
Priyanto, A. T. (2010). *The impact of human activities on coastal zones and strategies towards sustainable development : a case study in Pekalongan, Indonesia*. University of Twente, Faculty of Geo-information and Earth observation ITC, Enschede.

Rana, M. (2008). *Urban fringe Development pattern in kathmandu valley.* Master degree of science in Urban Planning, Institute of Engineering, Pulchowk.

Richards, J. A. (2012). *Remote sensing digital image analysis: an introduction*: Springer.

Shrestha, R. K. (2003). *Development Dynamics in Urban fringe, A Case study of Lalitpur SMC.* Master degree Thesis report, Institute of Engineering, Pulchowk.

Singh, A. (1989). Review Article Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing, 10*(6), 989-1003.

Sonar, S. G. (2008). Urban sprawl – A system dynamic approach

Thapa, R. B., & Murayama, Y. (2010). Drivers of urban growth in the Kathmandu valley, Nepal: Examining the efficacy of the analytic hierarchy process. *Applied Geography, 30*(1), 70-83.

Thapa, R. B., & Murayama, Y. (2011). Urban growth modeling of Kathmandu metropolitan region, Nepal. *Computers, Environment and Urban Systems, 35*(1), 25-34.

Timalsina, K. P. (2007). *Rural Urban Migration and Livelihood in the Informal Sector.* Department of Geography, Norwegian University of Science and Technology (NTNU).

USGS. (2012). U.S. Department of the Interior | U.S. Geological Survey Retrieved 09/09/2012, 2012, from <http://glovis.usgs.gov/> <http://earthexplorer.usgs.gov/>

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., & Mastura, S. S. A. (2002). Modeling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management, 30*(3), 391-405.

Verburg, P. H., van Eck, J. R. R., de Nijs, T. C. M., Dijst, M. J., & Schot, P. (2004). Determinants of land-use change patterns in the Netherlands. *Environment and Planning B, 31*(1), 125-150.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med, 37*(5), 360-363.

White, R., & Engelen, G. (2000). *High-resolution integrated modelling of the spatial dynamics of urban and regional systems.* Paper presented at the GeoComputation '99, July 25, 1999 - July 28, 1999, Fredericksburg, VA, USA.

Wilson, E., Hurd, J., & Civco, D. (2002). *Development of a model to quantify and map urban growth.*

Wu, B., Huang, B., & Fung, T. (2009). Projection of land use change patterns using kernel logistic regression. *Photogrammetric Engineering & Remote Sensing, 75*(8), 971-979.

Xiao, J., Shen, Y., Ge, J., Tateishi, R., Tang, C., Liang, Y., & Huang, Z. (2006). Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. *Landscape and Urban Planning, 75*(1-2), 69-80.

Xie, C., Huang, B., Claramunt, C., & Chandramouli, C. (2005). *Spatial logistic regression and GIS to model rural-urban land conversion.* Paper presented at the Proceedings of PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications.

Xu, C., Liu, M., Zhang, C., An, S., Yu, W., & Chen, J. M. (2007). The spatiotemporal dynamics of rapid urban growth in the Nanjing metropolitan region of China. *Landscape Ecology, 22*(6), 925-937.

Yuan, F., Sawaya, K. E., Loeffelholz, B. C., & Bauer, M. E. (2005). Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. *Remote sensing of environment, 98*(2–3), 317-328.

Zaki, R., Zaki, A., & Ahmed, S. (2011). Land Use and Land Cover Changes in Arid Region: The Case New Urbanized Zone, Northeast Cairo, Egypt. *Journal of Geographic Information System, 3*(3), 173-194.

Zeng, Y. N., Wu, G. P., Zhan, F. B., & Zhang, H. H. (2008). Modeling spatial land use pattern using autologistic regression. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVII*(B2), 115-118.

ANNEXES

I. Tables

Table A.1 Variables in the 3x3 LR model

		B	S.E.	Wald	df	Sig.	O.R	95% C.I.for O.R	
								Lower	Upper
Step 4[a]	Slope	-1.859	.309	36.235	1	.000	.156	.085	.285
	Forest	-3.411	1.005	11.518	1	.001	.033	.005	.237
	dist_rdmaj	-1.273	.139	84.529	1	.000	.280	.213	.367
	Dist_rdmin	-5.194	.443	137.527	1	.000	.006	.002	.013
	Dist_pipe	-10.808	.588	337.976	1	.000	.000	.000	.000
	Dist_CBD	-3.102	.217	203.777	1	.000	.045	.029	.069
	Dist_health	-1.162	.238	23.780	1	.000	.313	.196	.499
	Dist_edu	-1.676	.296	32.149	1	.000	.187	.105	.334
	Prop_urban	2.002	.250	64.084	1	.000	7.401	4.534	12.082
	Dist_ucluster	-5.469	.229	572.225	1	.000	.004	.003	.007
	Land_value	1.554	.174	80.021	1	.000	4.728	3.364	6.646
	Constant	3.804	.177	462.014	1	.000	44.886		

* Variables significant at 0.05.

Table A.2 Variables in the 5x5 LR model

		B	S.E.	Wald	df	Sig.	O.R	95% C.I.for O.R	
								Lower	Upper
Step 3[a]	Slope	-1.700	.499	11.589	1	.001	.183	.069	.486
	Forest	-2.453	1.013	5.868	1	.015	.086	.012	.626
	dist_rdmaj	-1.347	.231	34.015	1	.000	.260	.165	.409
	Dist_rdmin	-5.422	.726	55.717	1	.000	.004	.001	.018
	Dist_pipe	-11.234	.961	136.567	1	.000	.000	.000	.000
	Dist_CBD	-3.635	.444	67.020	1	.000	.026	.011	.063
	Dist_Ucentre	.699	.386	3.279	1	.070	2.012	.944	4.288
	Dist_health	-1.137	.401	8.018	1	.005	.321	.146	.705
	Dist_edu	-1.835	.485	14.311	1	.000	.160	.062	.413
	Prop_urban	2.135	.421	25.653	1	.000	8.455	3.701	19.315
	Dist_ucluster	-5.194	.376	191.199	1	.000	.006	.003	.012

Land_value	1.711	.294	33.860	1	.000	5.533	3.110	9.844
Constant	3.725	.293	162.093	1	.000	41.458		

* Variables significant at 0.05.

Table A.3 Variables in the 7x7 LR model

								95% C.I.for O.R	
		B	S.E.	Wald	df	Sig.	O.R	Lower	Upper
Step 5[a]	Slope	-1.319	.647	4.163	1	.041	.267	.075	.949
	Popden	2.271	.650	12.201	1	.000	9.690	2.709	34.656
	dist_rdmaj	-1.102	.313	12.360	1	.000	.332	.180	.614
	Dist_rdmin	-5.113	1.003	25.977	1	.000	.006	.001	.043
	Dist_pipe	-10.262	1.223	70.400	1	.000	.000	.000	.000
	Dist_CBD	-2.753	.487	31.985	1	.000	.064	.025	.165
	Dist_edu	-2.476	.670	13.655	1	.000	.084	.023	.313
	Prop_urban	1.363	.615	4.911	1	.027	3.908	1.171	13.050
	Dist_ucluster	-6.472	.561	133.113	1	.000	.002	.001	.005
	Land_value	.999	.400	6.242	1	.012	2.716	1.240	5.946
	Constant	3.404	.470	52.422	1	.000	30.086		

* Variables significant at 0.05.

Table A.4 Variables in the 9x9 LR model

								95% C.I.for O.R	
		B	S.E.	Wald	df	Sig.	O.R	Lower	Upper
Step 5[a]	Slope	-1.559	.830	3.525	1	.060	.210	.041	1.071
	dist_rdmaj	-1.152	.409	7.924	1	.005	.316	.142	.705
	Dist_rdmin	-3.230	1.202	7.215	1	.007	.040	.004	.418
	Dist_pipe	-9.889	1.581	39.130	1	.000	.000	.000	.001
	Dist_CBD	-3.509	.622	31.792	1	.000	.030	.009	.101
	Dist_river	1.104	.633	3.045	1	.081	3.017	.873	10.433
	Dist_health	-2.133	.694	9.441	1	.002	.119	.030	.462
	Prop_urban	2.477	.733	11.406	1	.001	11.906	2.828	50.129
	Dist_ucluster	-4.367	.631	47.908	1	.000	.013	.004	.044
	Land_value	1.809	.491	13.550	1	.000	6.106	2.330	16.000
	Constant	2.729	.496	30.325	1	.000	15.322		

* Variables significant at 0.05.

Table A.5 Variables in the 11x11 LR model

		B	S.E.	Wald	df	Sig.	O.R	95% C.I.for O.R	
								Lower	Upper
Step 6[a]	Slope	-1.681	1.020	2.719	1	.099	.186	.025	1.373
	Popden	2.181	.969	5.066	1	.024	8.859	1.326	59.207
	dist_rdmaj	-2.352	.517	20.682	1	.000	.095	.035	.262
	Dist_rdmin	-4.219	1.526	7.646	1	.006	.015	.001	.293
	Dist_pipe	-9.254	1.745	28.132	1	.000	.000	.000	.003
	Dist_CBD	-2.955	.707	17.448	1	.000	.052	.013	.208
	Dist_edu	-2.321	1.010	5.277	1	.022	.098	.014	.711
	Prop_urban	1.724	.876	3.868	1	.049	5.605	1.006	31.227
	Dist_ucluster	-5.977	.817	53.546	1	.000	.003	.001	.013
	Constant	4.205	.679	38.396	1	.000	67.034		

* Variables significant at 0.05.

Table A.6 Result of logistic regression for overall growth

Variables in the LR model for overall growth

		B	S.E.	Wald	df	Sig.	O.R	95% C.I.for O.R	
								Lower	Upper
Step 1[a]	Slope	-1.080	.659	2.683	1	.101	.340	.093	1.237
	Forest	-14.912	753.294	.000	1	.984	.000	0.000	
	Popden	2.008	.668	9.045	1	.003	7.451	2.013	27.583
	dist_rdmaj	-1.097	.320	11.783	1	.001	.334	.178	.624
	Dist_rdmin	-5.094	1.009	25.471	1	.000	.006	.001	.044
	Dist_pipe	-10.311	1.249	68.152	1	.000	.000	.000	.000
	Dist_CBD	-3.154	.605	27.194	1	.000	.043	.013	.140
	Dist_river	.334	.515	.420	1	.517	1.396	.509	3.830
	Dist_Ucentre	.635	.526	1.457	1	.227	1.886	.673	5.286
	Dist_health	-.375	.534	.494	1	.482	.687	.241	1.956
	Dist_edu	-2.278	.680	11.225	1	.001	.102	.027	.389
	Prop_urban	1.375	.620	4.911	1	.027	3.954	1.172	13.338
	Dist_ucluster	-6.486	.567	130.996	1	.000	.002	.001	.005
	Land_value	.942	.417	5.104	1	.024	2.566	1.133	5.810
	Constant	3.497	.476	53.916	1	.000	33.027		
Step 2[a]	Slope	-1.376	.652	4.456	1	.035	.252	.070	.906
	Popden	2.275	.654	12.084	1	.001	9.726	2.697	35.070
	dist_rdmaj	-1.109	.320	12.006	1	.001	.330	.176	.618
	Dist_rdmin	-5.055	1.007	25.198	1	.000	.006	.001	.046
	Dist_pipe	-10.261	1.253	67.087	1	.000	.000	.000	.000

	Dist_CBD	-3.138	.604	26.979	1	.000	.043	.013	.142
	Dist_river	.379	.513	.545	1	.460	1.460	.534	3.991
	Dist_Ucentre	.574	.526	1.190	1	.275	1.775	.633	4.980
	Dist_health	-.373	.534	.488	1	.485	.689	.242	1.961
	Dist_edu	-2.410	.677	12.655	1	.000	.090	.024	.339
	Prop_urban	1.328	.620	4.584	1	.032	3.774	1.119	12.731
	Dist_ucluster	-6.473	.568	129.805	1	.000	.002	.001	.005
	Land_value	.931	.416	5.000	1	.025	2.536	1.122	5.735
	Constant	3.420	.473	52.225	1	.000	30.573		
Step 3[a]	Slope	-1.400	.651	4.629	1	.031	.247	.069	.883
	Popden	2.316	.651	12.648	1	.000	10.134	2.828	36.313
	dist_rdmaj	-1.147	.315	13.256	1	.000	.318	.171	.589
	Dist_rdmin	-5.068	1.008	25.283	1	.000	.006	.001	.045
	Dist_pipe	-10.428	1.235	71.304	1	.000	.000	.000	.000
	Dist_CBD	-3.116	.603	26.676	1	.000	.044	.014	.145
	Dist_river	.411	.510	.649	1	.420	1.508	.555	4.097
	Dist_Ucentre	.502	.517	.946	1	.331	1.652	.600	4.548
	Dist_edu	-2.444	.676	13.072	1	.000	.087	.023	.327
	Prop_urban	1.310	.619	4.475	1	.034	3.707	1.101	12.477
	Dist_ucluster	-6.518	.564	133.601	1	.000	.001	.000	.004
	Land_value	.886	.412	4.640	1	.031	2.426	1.083	5.435
	Constant	3.416	.473	52.172	1	.000	30.439		
Step 4[a]	Slope	-1.359	.649	4.390	1	.036	.257	.072	.916
	Popden	2.287	.650	12.369	1	.000	9.845	2.752	35.212
	dist_rdmaj	-1.134	.315	12.988	1	.000	.322	.174	.596
	Dist_rdmin	-5.145	1.003	26.296	1	.000	.006	.001	.042
	Dist_pipe	-10.397	1.233	71.061	1	.000	.000	.000	.000
	Dist_CBD	-3.078	.600	26.349	1	.000	.046	.014	.149
	Dist_Ucentre	.485	.515	.886	1	.347	1.624	.592	4.458
	Dist_edu	-2.504	.671	13.910	1	.000	.082	.022	.305
	Prop_urban	1.366	.615	4.930	1	.026	3.919	1.174	13.088
	Dist_ucluster	-6.477	.560	133.715	1	.000	.002	.001	.005
	Land_value	.919	.409	5.034	1	.025	2.506	1.123	5.591
	Constant	3.438	.472	53.098	1	.000	31.122		
Step 5[a]	Slope	-1.319	.647	4.163	1	.041	.267	.075	.949
	Popden	2.271	.650	12.201	1	.000	9.690	2.709	34.656
	dist_rdmaj	-1.102	.313	12.360	1	.000	.332	.180	.614
	Dist_rdmin	-5.113	1.003	25.977	1	.000	.006	.001	.043
	Dist_pipe	-10.262	1.223	70.400	1	.000	.000	.000	.000
	Dist_CBD	-2.753	.487	31.985	1	.000	.064	.025	.165
	Dist_edu	-2.476	.670	13.655	1	.000	.084	.023	.313

Prop_urban	1.363	.615	4.911	1	.027	3.908	1.171	13.050
Dist_ucluster	-6.472	.561	133.113	1	.000	.002	.001	.005
Land_value	.999	.400	6.242	1	.012	2.716	1.240	5.946
Constant	3.404	.470	52.422	1	.000	30.086		

a. Variable(s) removed on step 2: Forest.
b. Variable(s) removed on step 3: Dist_health.
c. Variable(s) removed on step 4: Dist_river.
d. Variable(s) removed on step 5: Dist_Ucentre.

Table A.7 Result of logistic regression for Infill growth

Variables in the LR model for infill growth

		B	S.E.	Wald	df	Sig.	O.R	95% C.I.for O.R	
								Lower	Upper
Step 1[a]	Dist_cbd	-.344	.681	.256	1	.613	.709	.187	2.691
	Dist_edu	.168	.683	.061	1	.806	1.183	.310	4.512
	Dist_health	-1.996	.690	8.372	1	.004	.136	.035	.525
	Dist_pipe	-11.372	2.865	15.754	1	.000	.000	.000	.003
	dist_rdmaj	.651	.334	3.792	1	.052	1.917	.996	3.690
	Dist_rdmin	-4.165	1.398	8.873	1	.003	.016	.001	.241
	Dist_river	-1.745	.502	12.092	1	.001	.175	.065	.467
	Dist_ucentre	3.082	.814	14.330	1	.000	21.808	4.421	107.568
	Dist_ucluster	-.776	.269	8.323	1	.004	.460	.272	.780
	forest	-21.379	40192.970	.000	1	1.000	.000	0.000	
	Land_value	.034	.349	.010	1	.922	1.035	.522	2.051
	Popden	.000	.000	135.893	1	.000	1.000	1.000	1.000
	Prop_urban	2.249	.277	66.053	1	.000	9.483	5.512	16.313
	Slope	-2.391	.994	5.785	1	.016	.092	.013	.642
	Constant	.684	.357	3.673	1	.055	1.981		
Step 2[a]	Dist_cbd	-.373	.680	.302	1	.583	.688	.182	2.610
	Dist_edu	.164	.683	.058	1	.810	1.178	.309	4.492
	Dist_health	-2.024	.689	8.621	1	.003	.132	.034	.510
	Dist_pipe	-11.322	2.862	15.654	1	.000	.000	.000	.003
	dist_rdmaj	.646	.334	3.738	1	.053	1.908	.991	3.673
	Dist_rdmin	-4.175	1.398	8.915	1	.003	.015	.001	.238
	Dist_river	-1.756	.502	12.244	1	.000	.173	.065	.462

	Dist_ucentre	3.104	.814	14.544	1	.000	22.287	4.521	109.859
	Dist_ucluster	-.780	.269	8.419	1	.004	.458	.270	.776
	Land_value	.028	.349	.006	1	.937	1.028	.519	2.037
	Popden	.000	.000	135.915	1	.000	1.000	1.000	1.000
	Prop_urban	2.256	.277	66.447	1	.000	9.546	5.549	16.420
	Slope	-2.368	.993	5.680	1	.017	.094	.013	.657
	Constant	.695	.357	3.800	1	.051	2.004		
Step 3[a]	Dist_cbd	-.389	.650	.359	1	.549	.678	.190	2.420
	Dist_edu	.177	.664	.071	1	.790	1.193	.325	4.383
	Dist_health	-2.013	.675	8.892	1	.003	.134	.036	.502
	Dist_pipe	-11.354	2.834	16.051	1	.000	.000	.000	.003
	dist_rdmaj	.643	.332	3.755	1	.053	1.902	.993	3.644
	Dist_rdmin	-4.182	1.395	8.993	1	.003	.015	.001	.235
	Dist_river	-1.749	.495	12.497	1	.000	.174	.066	.459
	Dist_ucentre	3.113	.806	14.925	1	.000	22.491	4.635	109.125
	Dist_ucluster	-.786	.259	9.248	1	.002	.456	.274	.756
	Popden	.000	.000	136.379	1	.000	1.000	1.000	1.000
	Prop_urban	2.256	.277	66.438	1	.000	9.544	5.548	16.418
	Slope	-2.371	.993	5.702	1	.017	.093	.013	.654
	Constant	.710	.306	5.380	1	.020	2.033		
Step 4[a]	Dist_cbd	-.409	.645	.402	1	.526	.664	.188	2.353
	Dist_health	-1.982	.664	8.896	1	.003	.138	.037	.507
	Dist_pipe	-11.395	2.828	16.237	1	.000	.000	.000	.003
	dist_rdmaj	.638	.331	3.709	1	.054	1.893	.989	3.625
	Dist_rdmin	-4.130	1.380	8.951	1	.003	.016	.001	.241
	Dist_river	-1.738	.493	12.431	1	.000	.176	.067	.462
	Dist_ucentre	3.150	.794	15.735	1	.000	23.335	4.921	110.646
	Dist_ucluster	-.787	.258	9.274	1	.002	.455	.274	.755
	Popden	.000	.000	136.376	1	.000	1.000	1.000	1.000
	Prop_urban	2.244	.273	67.615	1	.000	9.430	5.523	16.098
	Slope	-2.378	.992	5.744	1	.017	.093	.013	.648
	Constant	.730	.296	6.075	1	.014	2.075		
Step 5[a]	Dist_health	-1.813	.607	8.918	1	.003	.163	.050	.536

	Dist_pipe	-11.832	2.743	18.601	1	.000	.000	.000	.002
	dist_rdmaj	.649	.331	3.856	1	.050	1.914	1.001	3.661
	Dist_rdmin	-4.180	1.377	9.207	1	.002	.015	.001	.228
	Dist_river	-1.734	.493	12.386	1	.000	.177	.067	.464
	Dist_ucentre	2.984	.751	15.782	1	.000	19.757	4.534	86.096
	Dist_ucluster	-.831	.249	11.157	1	.001	.435	.267	.709
	Popden	.000	.000	136.176	1	.000	1.000	1.000	1.000
	Prop_urban	2.234	.272	67.327	1	.000	9.335	5.475	15.916
	Slope	-2.410	.991	5.921	1	.015	.090	.013	.626
	Constant	.681	.285	5.692	1	.017	1.976		

a. Variable(s) removed on step 2: forest.

b. Variable(s) removed on step 3: Land_value.

c. Variable(s) removed on step 4: Dist_edu.

d. Variable(s) removed on step 5: Dist_cbd.

Table A.8 Result of logistic regression for Expansion

Variables in the LR model for Expansion

								95% C.I.for O.R	
		B	S.E.	Wald	df	Sig.	O.R	Lower	Upper
Step 1[a]	Dist_cbd	-2.220	1.102	4.056	1	.044	.109	.013	.942
	Dist_edu	.010	1.231	.000	1	.993	1.010	.090	11.290
	Dist_health	-3.026	1.047	8.361	1	.004	.049	.006	.377
	Dist_pipe	-10.151	2.256	20.244	1	.000	.000	.000	.003
	dist_rdmaj	-.221	.672	.108	1	.742	.802	.215	2.990
	Dist_rdmin	-8.912	1.960	20.675	1	.000	.000	.000	.006
	Dist_river	1.107	.936	1.399	1	.237	3.025	.483	18.936
	Dist_ucentre	-.378	.920	.169	1	.681	.685	.113	4.161
	Dist_ucluster	-9.074	1.165	60.659	1	.000	.000	.000	.001
	forest	-14.931	1136.108	.000	1	.990	.000	0.000	
	Land_value	1.569	.786	3.986	1	.046	4.801	1.029	22.394
	Popden	.000	.000	3.387	1	.066	1.000	1.000	1.000
	Prop_urban	.156	1.624	.009	1	.923	1.169	.048	28.211
	Slope	-.026	1.089	.001	1	.981	.974	.115	8.235
	Constant	5.208	.823	40.051	1	.000	182.813		
Step 2[a]	Dist_cbd	-2.219	1.099	4.078	1	.043	.109	.013	.937

	Dist_health	-3.025	1.036	8.517	1	.004	.049	.006	.370
	Dist_pipe	-10.151	2.256	20.246	1	.000	.000	.000	.003
	dist_rdmaj	-.220	.671	.108	1	.742	.802	.216	2.985
	Dist_rdmin	-8.911	1.952	20.838	1	.000	.000	.000	.006
	Dist_river	1.106	.931	1.412	1	.235	3.022	.488	18.736
	Dist_ucentre	-.379	.920	.169	1	.681	.685	.113	4.156
	Dist_ucluster	-9.072	1.152	62.003	1	.000	.000	.000	.001
	forest	-14.925	1133.111	.000	1	.989	.000	0.000	
	Land_value	1.570	.775	4.107	1	.043	4.806	1.053	21.933
	Popden	.000	.000	3.386	1	.066	1.000	1.000	1.000
	Prop_urban	.157	1.623	.009	1	.923	1.170	.049	28.151
	Slope	-.027	1.089	.001	1	.980	.974	.115	8.229
	Constant	5.208	.823	40.085	1	.000	182.776		
Step 3[a]	Dist_cbd	-2.194	1.096	4.009	1	.045	.111	.013	.955
	Dist_health	-3.087	1.038	8.845	1	.003	.046	.006	.349
	Dist_pipe	-10.080	2.264	19.818	1	.000	.000	.000	.004
	dist_rdmaj	-.191	.673	.080	1	.777	.826	.221	3.090
	Dist_rdmin	-8.957	1.952	21.055	1	.000	.000	.000	.006
	Dist_river	1.180	.926	1.624	1	.203	3.255	.530	19.996
	Dist_ucentre	-.463	.922	.252	1	.615	.629	.103	3.832
	Dist_ucluster	-9.095	1.155	62.005	1	.000	.000	.000	.001
	Land_value	1.618	.773	4.383	1	.036	5.042	1.109	22.924
	Popden	.000	.000	3.555	1	.059	1.000	1.000	1.000
	Prop_urban	.176	1.626	.012	1	.914	1.193	.049	28.864
	Slope	-.348	1.077	.104	1	.747	.706	.086	5.827
	Constant	5.206	.820	40.295	1	.000	182.331		
Step 4[a]	Dist_cbd	-2.185	1.093	4.000	1	.046	.112	.013	.957
	Dist_health	-3.078	1.034	8.856	1	.003	.046	.006	.350
	Dist_pipe	-10.051	2.246	20.022	1	.000	.000	.000	.004
	dist_rdmaj	-.194	.672	.083	1	.773	.824	.221	3.076
	Dist_rdmin	-8.943	1.947	21.099	1	.000	.000	.000	.006
	Dist_river	1.190	.921	1.668	1	.196	3.287	.540	19.999
	Dist_ucentre	-.469	.920	.260	1	.610	.626	.103	3.798

	Dist_ucluster	-9.167	.949	93.383	1	.000	.000	.000	.001
	Land_value	1.617	.772	4.387	1	.036	5.040	1.110	22.897
	Popden	.000	.000	3.578	1	.059	1.000	1.000	1.000
	Slope	-.349	1.077	.105	1	.746	.705	.085	5.820
	Constant	5.249	.718	53.458	1	.000	190.361		
Step 5[a]	Dist_cbd	-2.119	1.068	3.935	1	.047	.120	.015	.975
	Dist_health	-3.148	1.006	9.795	1	.002	.043	.006	.308
	Dist_pipe	-10.107	2.244	20.283	1	.000	.000	.000	.003
	Dist_rdmin	-8.918	1.944	21.038	1	.000	.000	.000	.006
	Dist_river	1.176	.920	1.635	1	.201	3.242	.534	19.672
	Dist_ucentre	-.501	.913	.301	1	.583	.606	.101	3.628
	Dist_ucluster	-9.240	.916	101.747	1	.000	.000	.000	.001
	Land_value	1.650	.764	4.666	1	.031	5.206	1.165	23.259
	Popden	.000	.000	3.640	1	.056	1.000	1.000	1.000
	Slope	-.348	1.078	.104	1	.747	.706	.085	5.837
	Constant	5.163	.654	62.376	1	.000	174.748		
Step 6[a]	Dist_cbd	-2.156	1.064	4.107	1	.043	.116	.014	.932
	Dist_health	-3.158	1.005	9.876	1	.002	.042	.006	.305
	Dist_pipe	-10.192	2.233	20.837	1	.000	.000	.000	.003
	Dist_rdmin	-8.898	1.944	20.955	1	.000	.000	.000	.006
	Dist_river	1.159	.917	1.595	1	.207	3.186	.528	19.236
	Dist_ucentre	-.529	.910	.339	1	.561	.589	.099	3.502
	Dist_ucluster	-9.264	.913	102.872	1	.000	.000	.000	.001
	Land_value	1.655	.764	4.689	1	.030	5.233	1.170	23.404
	Popden	.000	.000	3.596	1	.058	1.000	1.000	1.000
	Constant	5.177	.652	62.959	1	.000	177.077		
Step 7[a]	Dist_cbd	-2.525	.860	8.612	1	.003	.080	.015	.432
	Dist_health	-3.276	.987	11.028	1	.001	.038	.005	.261
	Dist_pipe	-10.334	2.220	21.664	1	.000	.000	.000	.003
	Dist_rdmin	-8.913	1.942	21.059	1	.000	.000	.000	.006
	Dist_river	1.174	.918	1.637	1	.201	3.236	.535	19.563
	Dist_ucluster	-9.265	.913	103.056	1	.000	.000	.000	.001

	Land_value	1.579	.755	4.373	1	.037	4.852	1.104	21.323
	Popden	.000	.000	3.551	1	.059	1.000	1.000	1.000
	Constant	5.208	.651	64.031	1	.000	182.638		
Step 8[a]	Dist_cbd	-2.458	.851	8.336	1	.004	.086	.016	.454
	Dist_health	-3.276	.983	11.112	1	.001	.038	.006	.259
	Dist_pipe	-10.216	2.207	21.423	1	.000	.000	.000	.003
	Dist_rdmin	-9.121	1.931	22.306	1	.000	.000	.000	.005
	Dist_ucluster	-9.146	.906	101.927	1	.000	.000	.000	.001
	Land_value	1.596	.753	4.488	1	.034	4.933	1.127	21.592
	Popden	.000	.000	4.128	1	.042	1.000	1.000	1.000
	Constant	5.263	.648	66.033	1	.000	193.046		

a. Variable(s) removed on step 2: Dist_edu.

b. Variable(s) removed on step 3: forest.

c. Variable(s) removed on step 4: Prop_urban.

d. Variable(s) removed on step 5: dist_rdmaj.

e. Variable(s) removed on step 6: Slope.

f. Variable(s) removed on step 7: Dist_ucentre.

g. Variable(s) removed on step 8: Dist_river.

II. Figures

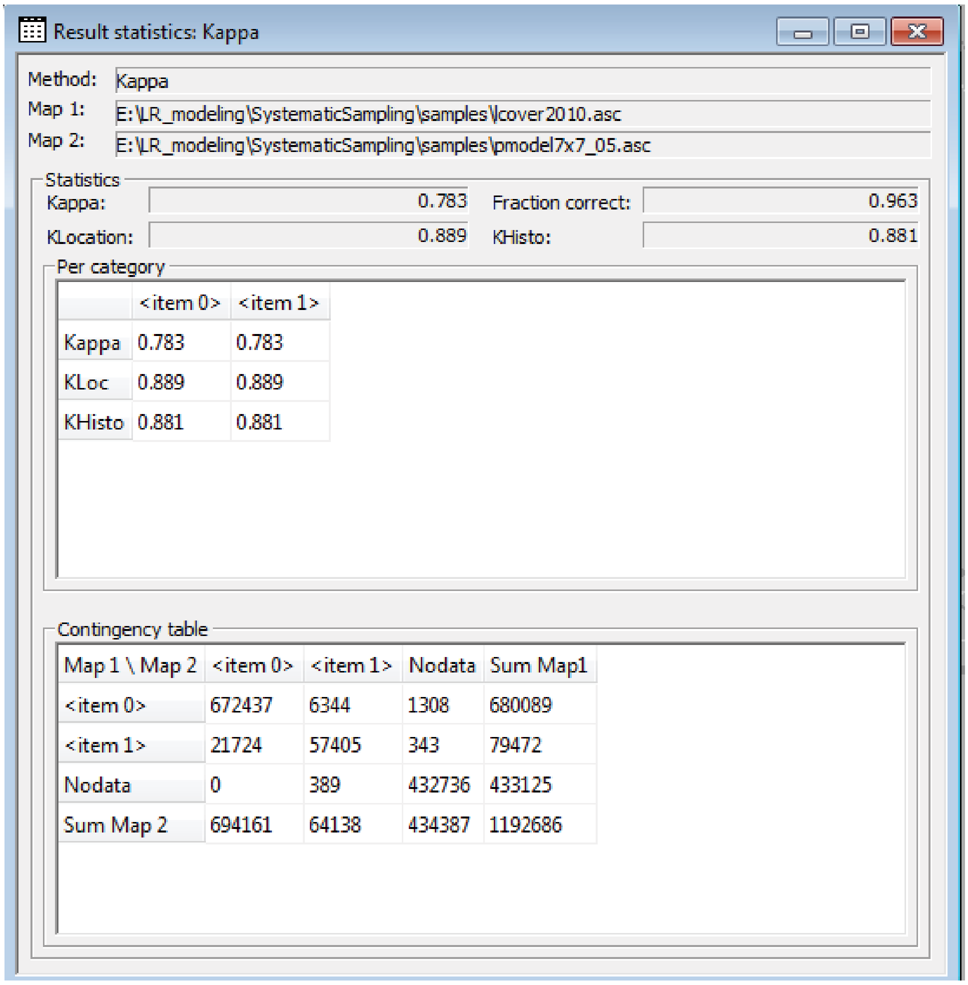


Figure F.1 Kappa statistics based on observed urban growth in 2010

III. Expert interview

Introduction

This research focuses on determining factors which drives the urban growth and forecasting future urban growth pattern in Kathmandu valley. In this regards, this interview will help to address expert opinion about the urban growth and its driving factors. In addition it will give the insight of urban development policies in the valley.

Name of interviewee: …………………………………….
Organization: …………………………………..
Post: …………………………………………..

Checklist for interview:

A. Understanding the concept of urban growth

a. According to the literature, there are three types of urban growth as shown in fig below. Could you please state where are these growths occurring in the valley?

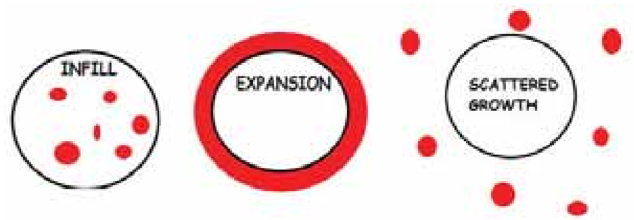


Figure F.2 Different types of urban growth

b. Which one do you think is the most prominent in Kathmandu valley? Why?

B. Understanding the drivers of urban growth

a. From literature review, following driving factors of urban growth have been listed. Could you please tick them according to their degree of influence?

b. In your view, what are the key factors of urban growth in Kathmandu valley?

c. What are the driving factors of infill growth, outward expansions and scattered growth?

Factors of urban growth		**High influence**	**Moderate influence**	**No influence**
1. Bio-physical character	-Slope orientation			
	-Degree of slope			
2. Population growth	-Population density			
3. Plans and policies	- Zoning			
4. Proximate causes	-Distance to major roads			
	-Distance to minor roads			
	-Distance to ring road			
	-Distance to water lines			
	-Distance to sewer line			
	-Distance to electric lines			
	-Distance to CBD			

	-Distance to industrial area			
	-Distance to major rivers			
	-Distance to urban centres			
	-Distance to health facilities			
	-Distance to educational facilities			
5. Existing conditions	-Existing urban cluster			
6. Economic factors	-Land value			
7. Safety	-Political Safety			
If other please specify	-			
	-			
	-			

C. Plan and policies of urban growth in Kathmandu valley

a. What types of plan and policies are there to control urban growth in Kathmandu valley?

b. One of the innovative parts of Long term development plan of Kathmandu valley is urban containment policy. What is the basis for delineating this boundary?

IV. Training Set Survey Sheet

Point ID	X	Y	Land cover	Roof Type*	Const. Year
1	346072.00	3062040.00			
2	345495.00	3062320.00			
3	344439.00	3061960.00			
4	334034.00	3065751.00			
5	333724.00	3065648.00			
6	333059.00	3065561.00			
7	333166.00	3065057.00			
8	333799.00	3066744.00			
9	334845.00	3062143.00			
10	334200.00	3062999.00			
11	333682.00	3062935.00			
12	334995.00	3061862.00			
13	330158.00	3062577.00			
14	329837.00	3062963.00			
15	333968.00	3066118.00			
16	335130.00	3065996.00			
17	336063.00	3063980.00			
18	340908.00	3062850.00			
19	340869.00	3063300.00			
20	344985.00	3061885.00			
21	343833.00	3061337.00			
22	343921.00	3062323.00			
23	334199.00	3065001.00			
24	335380.00	3061232.00			
25	334085.00	3063590.00			
26	335041.00	3064089.00			
27	336075.00	3065345.00			
28	337993.00	3064438.00			
29	330693.00	3064141.00			
30	330787.00	3062647.00			
31	333331.00	3065508.00			
32	334291.00	3065741.00			
33	339988.00	3063046.00			
34	338205.00	3062260.00			
35	334795.00	3060549.00			
36	336886.00	3068024.00			
37	336561.00	3070060.00			
38	333511.00	3068964.00			
39	330973.00	3067242.00			
40	332564.00	3065446.00			

1 Land cover Type: Building->B, Road-> R, Water->W, Forest->F, Vegetation->V, Bare->Ba
2 Roof Type: Tiles-> T, corrugated iron sheet->C, Reinforced cement concrete->RCC, Others->O
3 *If land cover is building or road