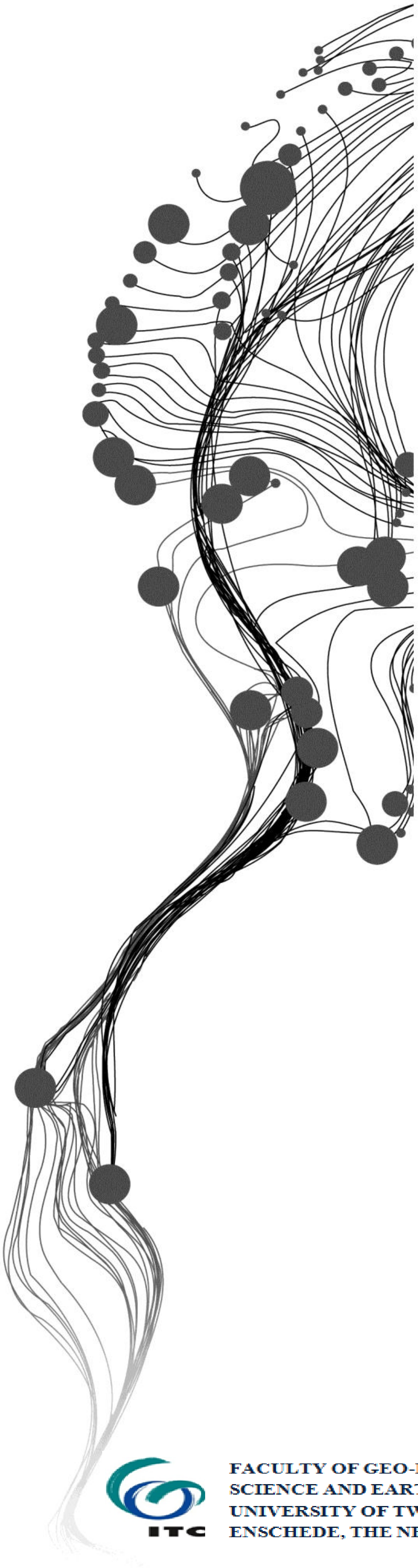
AUTOMATED GENERALIZATION OF LAND-USE/LAND-COVER

PRATIK YADAV
March, 2015

IIRS SUPERVISOR
PRASUN KUMAR GUPTA
DR. S. K. SRIVASTAV

IIRS SUPERVISOR
PROF.DR.IR. ALFRED STEIN



AUTOMATED GENERALIZATION OF LAND-USE/LAND-COVER

PRATIK YADAV
Enschede, The Netherlands [March, 2015]

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: Geoinformatics

THESIS ASSESSMENT BOARD:

Chairperson	:	
External Examiner	:	
ITC Supervisor	:	Prof. Dr. Ir. Alfred Stein
IIRS Supervisor	:	Prasun Kumar Gupta
IIRS Supervisor	:	Dr. S. K. Srivastav

OBSERVERS:

ITC Observer	:	Dr. Nicholas Hamm
IIRS Observer	:	Dr. S. K. Srivastav

ITC
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION, UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS

इसरो isro
INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
Department of Space, Government of India

DISCLAIMER

"Yesterday is history,
tomorrow is mystery,
but today is a gift.
That is why it is called the present."

-Master Oogway, Kung Fu Panda (2008),
Original Author- Bill Keane.

ABSTRACT

Automated generalization is a viable replacement for traditional, manual knowledge-based method with the capability to produce more accurate maps of coarser scales. Although, the basic concept behind all generalization remains the same, which is to remove details for better representation of information at coarser scale, the approach for generalization of thematic maps differs from topographic maps. This is due to the type of features present in the data, the rules on which generalization operators are based and the intended use. The current study is based on generalization of a finer scale land-use/land-cover dataset of scale 1:10k to produce coarser scale land-use/land-cover maps at the scale of 1:25k and 1:50k using *star* approach. National Urban Information System (NUIS) classification scheme, a hierarchal urban classification scheme is used in the present study. For this purpose, three operators are identified, namely elimination, reclassify and smoothening. The elimination operator is constructed using a modified version of polygon similarity model (Gao et al. 2013) which uses the sematic and geometric information of the polygons. The weights assigned to these operators for controlling their effect on the model was not previously reported. In the present study, the values of these weights are calibrated by considering a standard case for elimination and assigning variable value depending on the case of nearby polygons of the identified small polygons. Further using these three operators, eight sequences are identified to be used for producing maps at same classification level for 1:25k scale and level up classification scheme maps for 1:25k and 1:50k. The results of these sequences are compared on the grounds of least change caused in the percentage of class distribution as the main priority for land-use/land-cover generalization was to maintain the area of individual classes. Finally, the results of generalization are compared with the maps prepared by visual image interpretation for overall and individual class accuracy. The comparison reveals that not only the identified sequences produce maps with minimum change in class area, they also produce more accurate maps than the current approach of visual image interpretation used for producing these coarser scale land-use/land-cover maps. The current framework could serve as a solution for the production of land-use/land-cover maps at coarser scale map from finer scale maps, while providing more accurate results, maintaining the class distribution and benefiting in terms of time and cost.

Keywords: *Automated generalization, land-use/land-cover, NUIS urban classification scheme, polygon similarity model.*

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank the three most important persons for their contribution in this work, my supervisors. I am grateful to Prof. dr. ir. Alfred Stein for his constant support and valuable feedback. I would like to thank Dr. S. K. Srivastav (Head, GID) for his dual role in this research as a guiding supervisor and always encouraging course director. Countless thanks to Mr Prasun Kumar Gupta for being an inspiring teacher and great mentor. Without them this work would not have been possible.

I would like to thanks all the Teachers/Faculty members who guided me at IIRS and ITC. My deepest gratitude to Dr. Y. V. N. Krishna Murthy (Director, IIRS) and Dr. P. L. N Raju (Group Head, Geoinformatics) for the facilities provided at IIRS. Special thanks to Dr. Nicholas Hamm for his support and guidance during the course.

Now, a word of appreciation for my "*special five*", Abhishek Saikia, Kiledar Singh Tomar, Neeraj Agrawal, Vanya Jha and Akshara P. Byju. Thank you for staying throughout the journey even during the tough time. Abhishek Das, Pratiman, Vikrant, Ram, Surya, Raja, Raunak and Sanjay- my M.Tech friends, thank you for the wonderful memories. My ITC/UT colleagues- Jothi, Sneha, Abhijit, Riddhi, Mandy, Pascal, Anna, Dianna and Marisol – thank you for making the ITC stay memorable. I would like to thank my PG Diploma and ITEC classmates- Saikat, Ravi, Jagdeesh, Sanjeev, Amit sir, Shambhu sir and Wing Commander I. Malik. I am also thankful to Kanishk sir, Guru Sir, Ishan sir, Abhishek sir, Shishant sir, Danish, Amit sir, Kavisha, Amresh, Antra, Pooja, Pranay, Panini, Thapa Uncle, Bose Uncle, Verma *Bhaiya* and IIRS security staff, for being an important part of this wonderful journey.

Lastly my gratitude to my family- my parents, Priya and Aishwarya, thank you for your encouragement. Once Albus Dumbledore said "*Happiness can be found*, *even in the darkest of times*, *if one only remembers to turn on the light*", thank you all for being my *light*.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVATION

IHS - Intensity, Hue and Saturation

kml - Keyhole Mark-up Language

LISS - Linear Imaging Self-Scanning

LULC - Land-Use/Land-Cover

mmu - Minimum Mappable Unit

NMA - National Mapping Agency

NRSC - National Remote Sensing Centre

NUIS - National Urban Information System

PA - Producer's Accuracy

RF - Representative Factor

SDQ - Spatial Data Quality

UA - User's Accuracy

1 INTRODUCTION

1.1 Background

A major contributor in deciding the scale at which the map is produced is determined by its purpose and intended use. The purpose and usage of maps are so many that it becomes difficult to cater to the multiple requirements of map users. Therefore, scale plays a crucial role in map making. A map made at a very fine scale might be used for a purpose that is intended for a very large area and thus the high details are not required by the user (JoÃo 1998). To make such a map fit for use, it needs to be up-scaled from a finer resolution to a coarser resolution. Here, up-scaling is aggregating fine-scale information to a coarser scale. Up-scaling, however, introduces the problem of details that are abundantly present in a small area of map, hence reducing its legibility. Also the storage size of the data remains huge due to those details. Therefore, both the up-scaled map and the database benefit from generalization so as to make them fit for use as seen in Figure 1-1. Jenerette & Wu (2000, p. 104) defines generalization as *"Creating a legible map at a given scale from a more detailed geographical map"*.

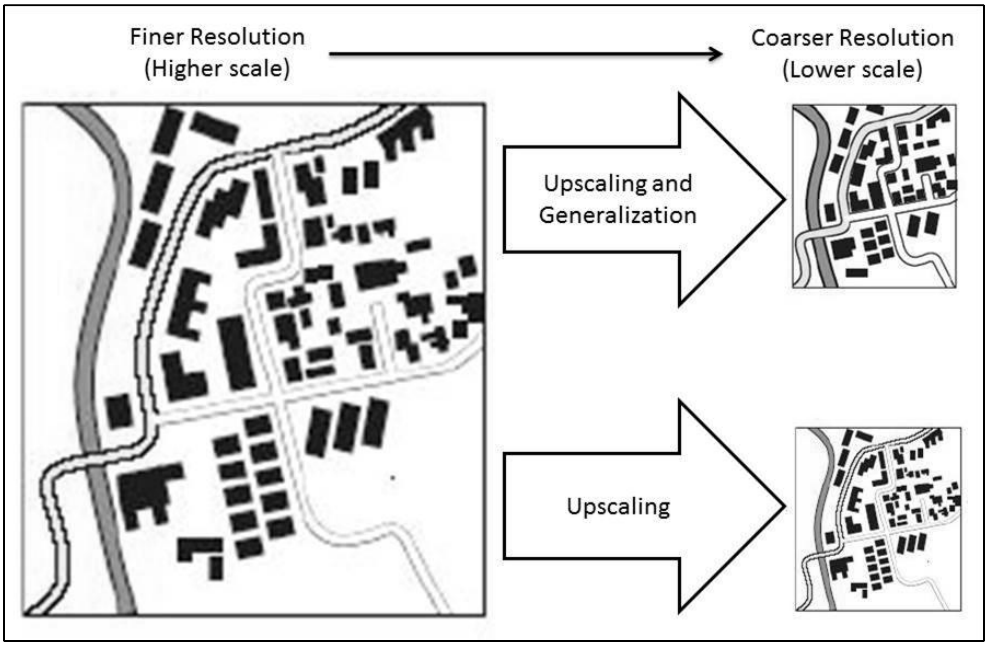


Figure 1-1. Up-scaling and generalization compared for a same map. While up-scaling (bottom) restores the same detail at the coarser scale, generalization (top) reduces detail for better representation while maintaining the core essence of the map.

Generalization is a scientific process that includes cartographer's understanding and knowledge. Generalization, however, is rather subjective and there is an absence of a formal structure. Traditional, knowledge based approach of generalization is complicated and results for a single data can vary as per different approaches of cartographers. Availability of large datasets have also introduced computational problems as currently the focus is more towards its automation. With new datasets every year and the requirement to have smart storage, the need for automated generalization has become increasingly important.

Generalization comprises of applying various operators on the database. Due to its non-structured framework, there is a variation in the naming of the operators in various texts, reports from agencies and in commercial software. For example, the basic concept behind smoothing and enhancement is the same, namely to reduce the number of nodes in a line or a polygon. The sequence of these operators, however, has a major influence on the results (Harrie & Sarjakoski 2002). The primary focus of most previous research was on generalization of topographic maps. Although the basic concept behind the generalization of all maps remains same, topographic maps differ from monothematic maps like soil maps, land-use/land-cover maps and road maps. This research focuses on generalization of land-use/land cover maps. Land-use/land-cover maps serve as a backbone for political development making. Urban development bodies use these map to analyse the growth and patterns related to any region. Due to their vast use and importance, land-use/land-cover maps are needed for various purposes where they are often required on different scales. The current practice by National Remote Sensing Centre (NRSC), India is to provide land-use/land-cover maps on three different scales which are 1:10k, 1:50k and 1:250k, which are prepared independently from each other. Preparation of these coarser scale maps from a single database using automated generalization will be beneficial in terms of both cost and time.

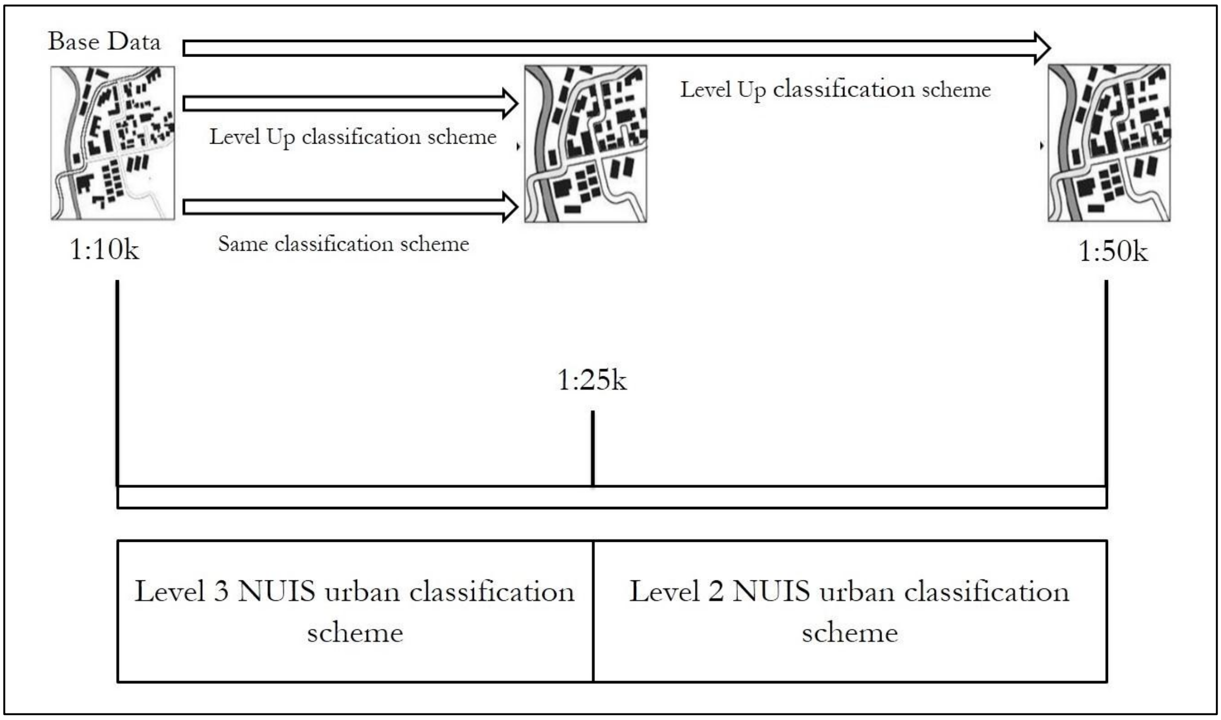


Figure 1-2. Proposed generalization framework using 1:10k as base data which will be used to produce maps of 1:25k and 1:50k.

The current study is based on finding a solution for automated generalization of land-use/land-cover maps. The land-use/land-cover map at a scale of 1:10k is used as base data (input) for generalization. This is up-scaled and generalized to two scale levels 1:25k and 1:50k. The 1:25k scale has two outputs: one with the same classification scheme as the 1:10k map (Level-III NUIS urban classification) and the other with the level up 1:50k classification scheme (Level-III NUIS urban classification) (Figure 1-2). This helps in finding the best NUIS urban land-use/land-cover classification scheme, in terms of accuracy, which can be employed at intermediate scales.

1.2 Previous Related Work

In previous years, various attempts have been made by researchers in the field of automated generalization using methods like model generalization, system based approach, agent based modelling, modular operator services, grid computing among many others (Basaraner 2002; Yang & Gold 1997; Lamy et al. 1999; Neun et al. 2009; Foerster et al. 2009; Chaudhry et al. 2009). But still a formal structure for generalization operators is absent. An attempt was made in previous research to formally classify these operators (Foerster et al. 2007). They identified five operators that are relevant to cartographic generalization.

In previous study by Gao et al. (2013), a framework was designed by means of basic generalization for improving representation of image-oriented classification map. They formulated a polygon similarity model that uses spectral, semantic and geometric information of polygons to eliminate small and unclassified polygons. The current study uses only semantic and geometric information of the polygons. The model helps to quantify this information to be used by the operators. Since these operators are based on different characteristics of the polygons, weights are assigned to these parameters (semantic and geometric). The optimal weights assigned to the amalgamation operator are derived by calibrating the model, since the rules defined are not exhaustive.

Till date the sequence of these operators is debatable as it depends on the knowledge of the cartographer and purpose of the map (Neun et al. 2009). The sequence in which these operators are applied play a major role in the generalized output and its quality. The current study also includes finding optimum sequence for generalization using the selected operators that have least effect on the area of land-use/land-cover classes.

Most of the research in the field of automated generalization has been done by the National Mapping Agencies in the European region (Stoter et al. 2011). Currently there are two approaches for generalization: *ladder* approach and *star* approach as shown in Figure 1-3 (Foerster et al. 2010). Since the star approach does not require preparation of intermediate scale maps, they are more efficient. Thus, the *star* approach is utilized in the present research, where 1:10k scale map serves as the base map and the primary input in the generalization process.

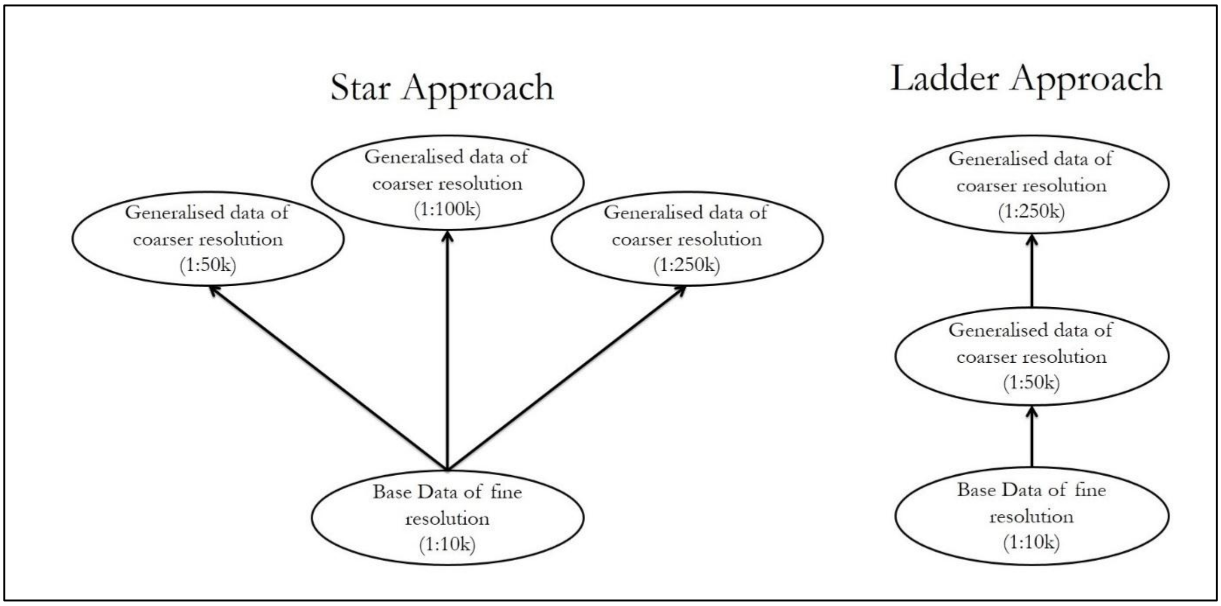


Figure 1-3. Star approach (left) and ladder approach (right) used by various NMAs in European region for generalization. While the later one is dependent of the intermediate results, star approach uses a single base data as input for all coarser level.

1.3 Motivation and Problem Statement

1.3.1 Research objectives

The main research objective is to develop automated generalization for land-use/land-cover maps in an urban environment.

1.3.2 Sub-objectives

- To develop generalization operators by integrating geometric and semantic information of polygons.
- To formulate a sequence of generalization operators that results in the smallest change in class distribution.
- To assess the accuracy of the modeled output by comparing with the corresponding map.
- To find the urban classification scheme that gives the highest accuracy at intermediate scales.

1.3.3 Research questions

- How can generalization be carried out using a model that integrates geometric and semantic information?
- What are the appropriate values for the parameters of such a model in the urban context?

- What is the optimal sequence of generalization operators, i.e. the sequence that results into the smallest change in class distribution?
- What is the accuracy of the modeled output?
- Which urban classification scheme gives the highest accuracy at the intermediate scale?

1.4 Innovation aimed at

Fully automated generalization is achieved only for topographic maps (Stoter et al. 2014). Automated Generalization for land-use/land-cover maps for urban environment has not been attempted previously. Sequencing of generalization operators and calibration of the weights assigned to their parameters in the method proposed by Gao et al. (2013) has also not been studied/reported before.

1.5 Thesis Structure

The thesis work is organized as follows-

Chapter 1: Introduction- The concept of Generalization is introduced and the thesis's motivation is stated. The research objectives and the research questions that are to be answered are presented along with previously carried out work.

Chapter 2: Literature Review- It deals with the detailed description of generalization concepts with reference to scale and purpose, and its effects on quality of data. Furthermore, details regarding various generalization operators and NUIS classification schemes are provided.

Chapter 3: Study Area and Data Preparation- Location and importance of the study area and the satellite images used to create the LULC data with details of interpretation techniques for preparing these maps are provided in this chapter.

Chapter 4: Methodology- I provides detailed description on the construction of the three identified operators and the calibration of associated weights; finding the sequence that results in least change of class distribution; and finding the accuracy of the modelled output.

Chapter 5: Results and Discussion- Effect of the various sequences on the output and finding the one that results in the smallest change in the area of classes are provided in this chapter.

Chapter 6: Conclusion and Recommendation- Final conclusion on the research and results with individual answers to research question and scope for future work as recommendation is highlighted here.

2 LITERATURE REVIEW

2.1 Scale and Generalization

Maps commonly represent a smaller scale version of the environment on which they are based on as they tend to represent the area on a smaller surface. It can be inferred that all maps are actually a generalized representation of their corresponding environment. This representation depends upon a large number of factors among which the two most governing are scale and intended purpose of the map.

Scale is defined as the ratio of the distance on a map to the corresponding distance on the surface the map represents. Most commonly it is expressed in terms of a ratio such as 1:1000 which signifies that one unit on the map will be equal to 1000 units on the actual ground. This representation of scale is called Representative Fraction (*RF*) or Natural Scale. When we describe a scale as a large scale maps, this means that the *RF's* denominator is small. Thus 1:1,000,000 maps are small scale maps whereas 1:1000 are large scale maps. The selection of scale for any map is based on the use of the map. While a town planner requires a map to be made at a scale of approximately 1:10,000 scales, so that he/she can easily identify the streets and building, a tourist might be comfortable with a 1:250,000 scale a map that shows a whole region at once. Thus most of the time a map made at a very detailed scale is required to be converted so that it can be used for a smaller scale purpose. This brings the role of generalization which reduce the details of map for its better representation at the changed scale.

According to Shekhar (2008, p. 955), map generalization is defined as:

"Map generalization is the name of the process that simplifies the representation of geographical data to produce a map at a certain scale with a defined and readable legend. To be readable at a smaller scale, some objects are removed; others are enlarged, aggregated and displaced one to another, and all objects are simplified. During the process, the information is globally simplified but stays readable and understandable."

The main goal of generalization is to maintain the essence of the map while reducing the unwanted information so that it contains the basic representation requirement. For this purpose, various operators are required which perform individually or collectively to do the desired.

2.2 Generalization Operators

A generalization operator can be defined as a set of rules to reduce the detail of a spatial data for better representation. One of the key issues faced in the field of generalization is the unstructured classification of operators. Often, two different operator will ultimately do the same changes in a data. For example, the key concept behind selection and elimination is same which is to reduce the complex data and make it more representative. Previous attempts have been made to formally classify these operators so as to have a structure among these operators (as shown in Table 2-1).

Table 2-1. Previous attempts to classify various generalization operators by McMaster & Shea (1992), Cecconi (2003), Yaolin et al. (2001) and Foerster (2007).

McMaster & Shea	
spatial transformations	Simplification
	Amalgamation
	Refinement
	Displacement
	smoothing
	Merging
	Exaggeration
	Aggregation
	Collapse
	Enhancement
Attribute transformations	Symbolization
	Classification

Cecconi	
<unspecified	Thematic selection
	Thematic aggregation
	Weeding
	Unrestricted simplification
Individual objects	Enlargement
	Exaggeration
	Fractalization
	smoothing
	Rectification
Individual or groups of object	Selection
	Elimination
groups of object	Displacement
	Amalgamation
	Combine
	Typification

Liu et al.
Simplification
Merge
Amalgamation
Aggregation
Classification
Selection

Foerster	
Model generalization	**Cartographic generalization**
Class Selection	Enhancement
Reclassification	Displacement
Collapse	Elimination
Combine	Typification
Simplification	
Amalgamation	

The current study involves systematic manipulation of data for reducing scale in a land-use/land-cover data. The dataset in this study is a vector layer of polygons with topological relationship i.e. cannot overlap or have space between boundaries. Thus, the following three operators are identified and used for the current study-

- **Elimination** reduces the complexity of the data by removing the features which are less visible (smaller in area).

- **Reclassify** changes the attribute of the features so that they are representative of the map at the new scale. In the current study, reclassify operator is used to change the classification scheme of land-use/land-cover from Level-III to Level-II.
- **Smoothening** transforms the objects to lesser complex features so that the visual appearance becomes less complicated with change in scale.

2.3 NUIS Land-use/Land-Cover Classification

The study of land-use/land-cover map generalization is selected because of its wide use in the field of development projects/policies in India. The development policies in India are based on five levels of planning, which are- (1) National Level-sectored cum inter-state/inter-regional planning; (2) State Level-sectored cum inter-district/inter-regional planning; (3) District/Metropolitan Level-regional planning; (4) Block Level-area planning; and (5) Panchayat Level-village planning (Raja 2012). These different levels of planning require specific scale maps, starting from 1:1k scale for utility mapping, 1:2k scale map for zonal planning, 1:10k scale for city level Master planning, 1:50k scale map for regional and State level planning (for small States) and the 1:250k scale for State level (for large scale) and country level development planning.

National Urban Information System (NUIS) defines urban land-use/land-cover classification scheme at four levels and is hierarchal in nature. For example – Level-I classification defines Built-Up which is sub-divided as Built-Up – Rural, Urban and Mining in Level-II. The level-II Urban Built-Up is further subdivided into 14 classes in level-III such as residential, commercial, industrial etc. (NUIS design and Standards 2008). The NUIS urban classification scheme is a hybrid scheme that merges both land-use and land-cover as shown in Table 2-2. The scheme is designed such that it is indicative and flexible. The concern is more towards the type of activity the land is engaged in rather that the ground cover on it. Thus, it can create confusion when compared with land cover as a forest inside an institute will be categorized as "Public and semi-public" rather than "Forest" class. It is designed for visual interpretation purpose and is not suitable for image derived land-use/land-cover maps.

Table 2-2. The hierarchical scheme relating Level-I, Level-II and Level-III for NUIS urban classification (NUIS design and Standards 2008).

Level-I (code)	Level-II (code)	Level-III (code)
Built Up (a)	Built Up (Urban) (aa)	Residential (aaa)
		Industrial (aab)
		Mixed Built Up area (aac)
		Recreational (aad)
		Public and semi-public (aae)
		Communication (aaf)
		Public utility and facility (aag)
		Commercial (aah)
		Transportation (aai)
		Reclaimed land vacant land (aaj)
		Vegetation area Trees (aak)
	Built Up (Rural) (ab)	
Agriculture (b)	Cropland (ba)	
	Fallow land (bb)	
	Plantation/ Orchards (bc)	
Forest (c)	Dense Forest (ca)	
	Open Forest (cb)	
	Plantation (cc)	
	Mangroves (cd)	
Grazing land Wastelands (d)	Salt-Affected (da)	
	Gullied /Ravenous (db)	
	Land with /without scrub (dc)	
	Barren /Rocky (dd)	
	Sandy area (de)	
Wetlands (e)	Marshy /Swampy (ea)	
	Mudflats (eb)	
	Waterlogged Salt pans (ec)	
Water bodies (f)	River/Streams (fa)	
	Canal (fb)	
	Lakes/ Ponds (fc)	
	Reservoirs (fd)	
	Tanks (fe)	
	Cooling Pond/Cooling Reservoir (ff)	
	Abandoned quarries with water (fg)	
Others (g)	Quarry / Brick Kilns (ga)	
	Dam / Barrage (gb)	
	Coral reef / Atoll (gc)	

2.4 Accuracy Estimation

Since generalization involves organized manipulation of data and reducing the level of details, it surely affects the spatial data quality (SDQ). Previous studies undertaken to assess the SDQ of the generalized data uses the semantic accuracy and geometric accuracy (Haunert & Sester 2008; Skopeliti 1997). Among various elements of spatial data quality, semantic accuracy and overall distribution of classes are important in this particular research. It is important to maintain the distribution of classes in terms of area as far as possible at various scales. This is because of the intended use of the map where if a map with a large number of small polygons of vegetation class when generalized will cause elimination of small polygons and will affect the overall percentage of vegetation in the map.

The accuracy assessment of the produced map will be a very critical part which will define whether the product is fit for use or not (Stoter et al. 2009). The current study uses generalization as specified by National Remote Sensing Agency, 2006. This estimates the accuracy of the map produced by visual interpretation. For estimating the interpretation quality, the vector layer is superimposed on the corresponding satellite data and checks are made for overall interpretation quality of major land cover types. Using stratified random sampling, random points are generated to compute user's accuracy (UA) and producer's accuracy (PA) of individual classes by compiling an error matrix. The ground truth for these points is based on high resolution satellite imagery (Google Earth) of the same period. Error matrix is further used to compute Cohen's Kappa coefficient *(κ)* which indicates how better the classification is as compared to randomly assigned value. Kappa *(κ)* is defines as-

$$Kappa\ (\kappa) = \frac{Observed\ accuracy - Chance\ agreement}{1 - Chance\ agreement} \qquad \ldots(2.1),$$

where *observed accuracy* (or *overall accuracy* in present study) is determined as the ratio of sum of diagonal elements of error matrix to total number of elements, while *chance agreement* is determined as the ratio of sum of diagonal elements (product of row and column for each class from error matrix) to the square of total number of elements.

The comparison of assessed accuracy of both generalization and image interpretation will help to understand weather the map made by automation provides more accurate maps. For this purpose, user's accuracy (UA), producer's accuracy (PA), overall accuracy and Kappa (κ) are used as the key components. Although Kappa (κ) is widely criticised, it is the most commonly used method to find accuracy of classified images.

3 STUDY AREA AND DATA PREPARATION

3.1 Study area

The city of Dehradun has been selected as the study area, which is located on the foothills of The Himalayas as shown in Figure 3-1. It is the capital and the biggest city of Uttarakhand State. Going through the phase of rapid growth and expansion, the city is now crawling towards the nearby sub-urban area. The increasing expansion will soon take over the nearby area into the city limits. This brings the need for understanding and analysing the land-use/land-cover pattern of the city and the surrounding areas to ensure sustainable growth. There is a large presence of sub-urban and rural areas near Dehradun which mostly depend on farming. This makes the site suitable for the research as it provides a variety of land-use/land cover classes.

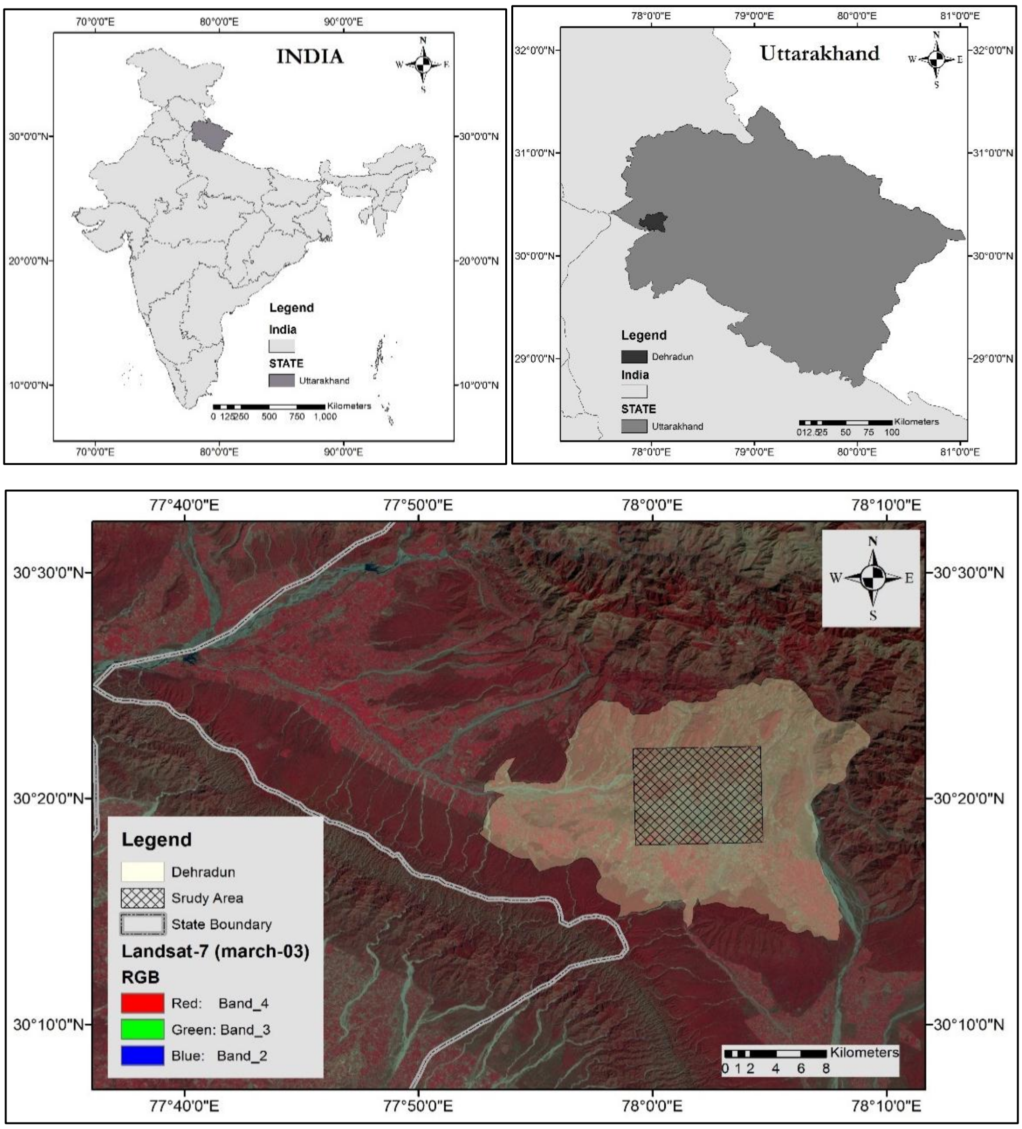


Figure 3-1. Location map of study area.

Since the whole administrative boundary was too big for the study, a part of the inner region of the city was selected having an area of approximately 68.68 km^2 (latitude 30^0 18’ 00” to 30^0 22’ 45” N and longitude 77^0 59’ 15” to 77^0 04’ 30” E). The area comprises of various land-use/land-cover activities and thus depicts a variety of classes for land-use/land-cover map.

3.2 Data Used and Pre-processing

Table 3-1. Description of satellite images used for preparing maps.

Satellite	Sensor	Date	Spatial Resolution
Cartosat-1	PAN	22 March 2013	2.5 m
Resourcesat 2	LISS-IV	7 March 2013	5.8 m
Resourcesat 2	LISS-III	31 March 2013	23.5 m

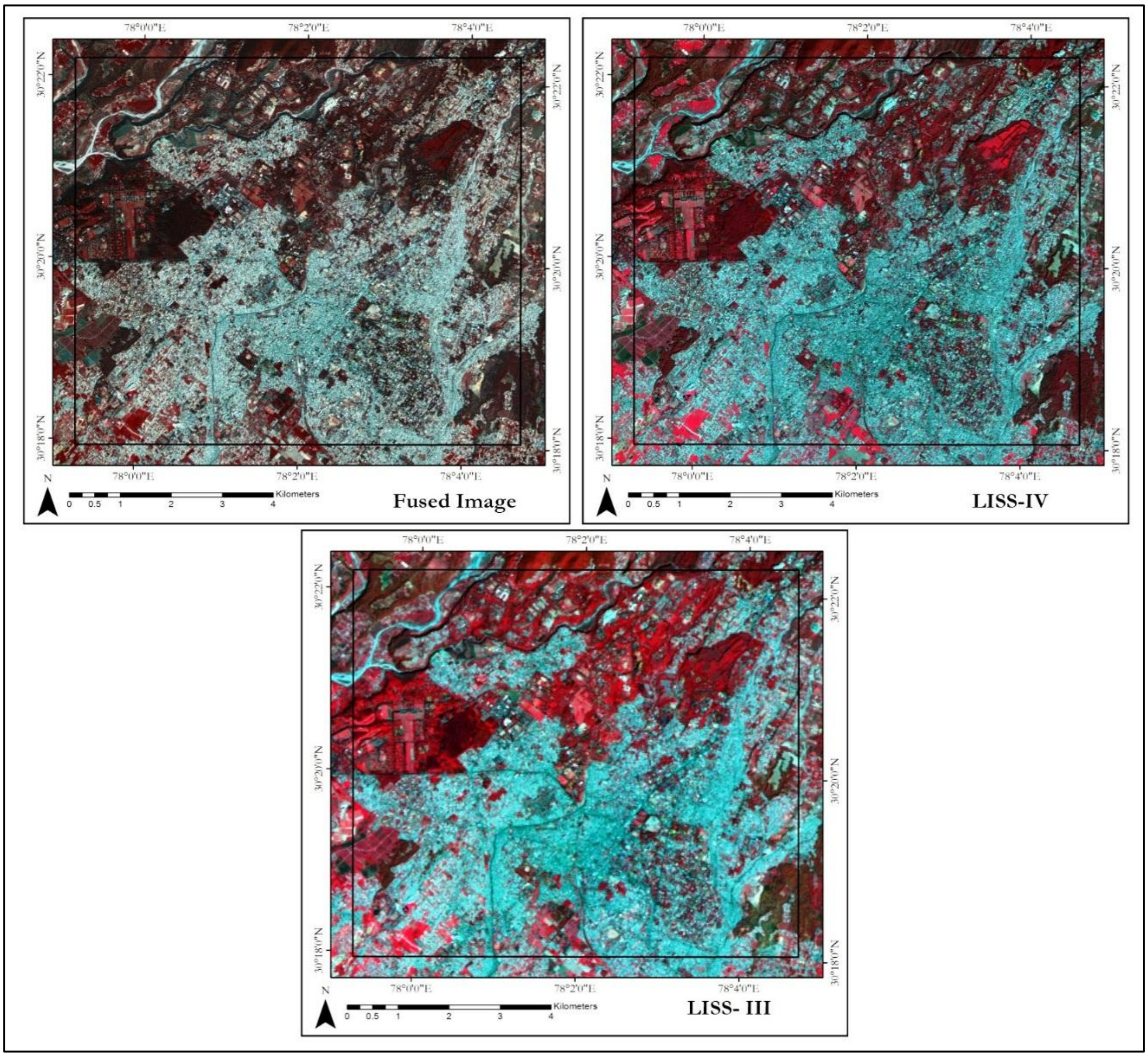


Figure 3-2. Satellite images used. Top Left- Fused image of resolution 2.5 m for preparing 1:10k map. Top right- LISS-IV image of resolution 5.8 m used for preparing 1:25k maps. Bottom- LISS-III image of resolution 23.5 m used for preparing 1:50k map.

Table 3-1 shows the description three satellite image used for preparing the visual interpreted maps. LISS-III and LISS-IV images are used for image interpretation at 1:50,000 and 1:25,000 scales, respectively. For 1:10,000 scale, fused image of PAN and LISS-IV is used as shown in Figure 3-2. The image fusion is done by IHS wavelet transformation producing a multi-spectral image of resolution 2.5 m.

3.3 Image Interpretation

Data preparation – The initial data preparation includes creating maps at three different scales using on-screen visual interpretation, which are-

- 1:10,000 scale – to be used as an input (base data) for generalization. This map is created using on-screen image interpretation of fused image (Cartosat-1 Panchromatic and Resourcesat-1 LISS-IV) with information from ground data. The map is based on NUIS Level-III urban land-use/land-cover classification scheme.
- 1:25,000 scale- two maps were prepared on this scale using LISS-IV data based on Level-II and Level-III NUIS urban land-use/land-cover classification scheme.
- 1:50,000 scale- Using LISS-III data, land-use/land-cover map for 1:50k scale is prepared. The classification scheme is based on Level-II NUIS urban land-use/land-cover classification scheme.

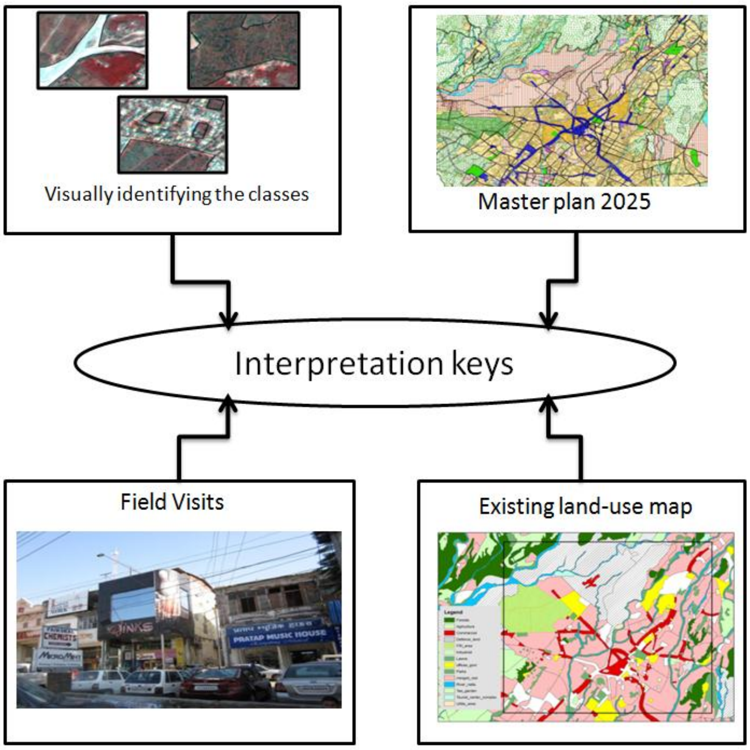


Figure 3-3. The interpretation key during preparation of land-use/land-cover maps. These keys helped to identify the appropriate class for delineated area.

For interpretation, the raster satellite images were displayed in ArcGIS 10.1 software with vector layer of study area boundary overlaid on top. The scale for the display of raster image is adjusted as per the required map scale (1:10k, 1:25k and 1:50k). The procedure followed then is to identify, delineate the area and assigning the appropriate class to it using interpretation keys (Figure 3-3). This delineation is based on image features such as tone, pattern, texture as well as the ground information from field and other source such as previous LULC map and master plan of Dehradun (2025). Among the prepared for maps, 1:10k Level-III classified maps serves the purpose of input data for generalization process (Figure3-4).

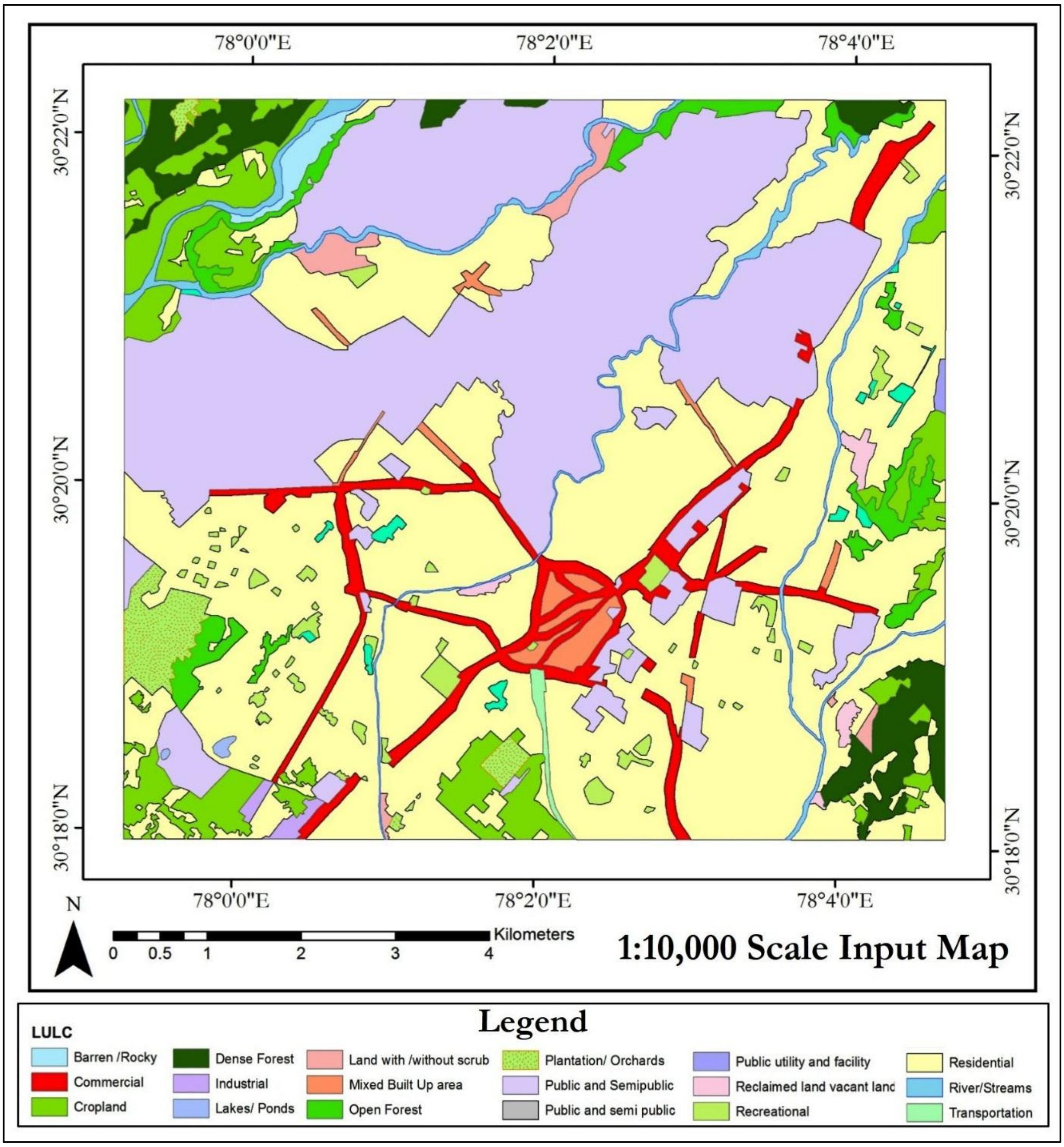


Figure 3-4. 1:10,000 scale maps prepared by visual interpretation with Level-III NUIS classification scheme. This map is used as an input data for other scales as per the star approach.

The accuracy assessment of these maps is based on ground truth. Using stratified random sampling, random points are generated and are used to generate error matrix for estimating user's accuracy (UA), producer's accuracy (PA), overall accuracy and Kappa (κ) as described in Section 2.4 (error matrix for maps shown in Appendix A.1). The overall accuracy as well as class accuracy for the four produced maps was computed using 450 random points (25 per class) as shown in Table 3-2. The ground truth for these points were taken from high resolution Google Earth image of 14 February 2013. For this purpose, the point layer was converted to kml file format, overlaid on Google earth and the land-use/land-cover class for the point was recorded. The results show the highest overall accuracy and Kappa *(κ)* value for 1:10k map and reduced values for coarser scales. This is because same 450 points are used for the accuracy estimation of all maps, created by random stratified sampling based on 1:10k scale map. Thus despite the up-scaling, the accuracy doesn't improves.

Table 3-2. Overall and Kappa *(κ)* accuracy of the prepared maps.

	1:10,000	**1:25,000**	**1:25,000**	**1:50,000**
NUIS Classification Level	III	III	II	II
Overall Accuracy	96.22%	77.56%	86.00%	76.67%
Kappa *(κ)*	96.00%	76.22%	80.99%	66.95%

4 METHODOLOGY AND IMPLEMENTATION

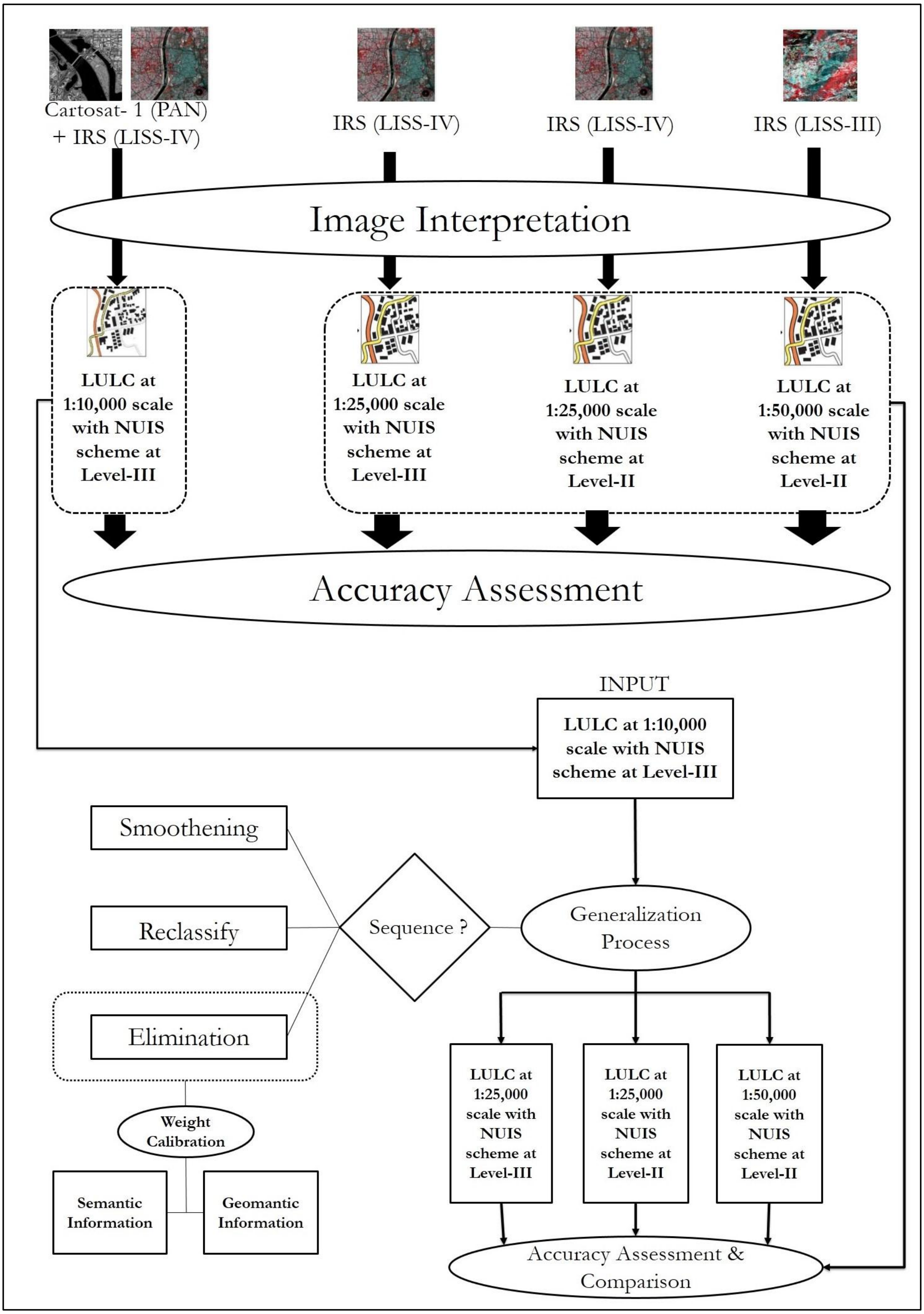


Figure 4-1. Research methodology.

4.1 Operators Construction

4.1.1 Polygon Similarity Model

Gao et al. (2013, p. 390) explains polygon similarity model as-

"The degree of similarity of two polygons depending on their contextual characteristics."

The model is designed for image-derived land-use/land-cover maps which uses spectral, semantic and geometric characteristics of polygons and quantify the similarity between two polygons as below:

$$S_{ik} = \omega_1 \cdot SE_{ik} + \omega_2 . GE_{ik} + \omega_3 \cdot (1 - SP_{ik}) \qquad \ldots(4.1),$$

where S_{ik} defines the similarity between the i^{th} and k^{th} polygons, SE_{ik}, GE_{ik} and SP_{ik}, represent similarity among the two polygons on semantic, geometric and spectral characteristics, respectively, and ω_1 , ω_2 and ω_3 are their weights (Gao et al. 2013).

In the present study, the map used is not image-derived and is based on visual interpretation. Also, the classification scheme used is a hybrid scheme using both land-use and land-cover, thus the similarity model will only use semantic and geometric characteristics of polygons and not the spectral characteristics. Thus, the reframed polygon similarity model is-

$$S_{ik} = \omega_1 \cdot SE_{ik} + \omega_2 . GE_{ik} \qquad \ldots (4.2).$$

Geometric Similarity (GE)

Geometric similarity (*GE*) is the ratio of the length of the shared boundaries of a small polygon with its neighbour polygon to the overall perimeter. The purpose is to reduce the possibility of generating new narrow-corridor conflicts due to elimination of the small polygon (Gao et al. 2013). *GE* is define as:

$$GE_{ik} = \frac{s_{ik}}{p_i} \qquad \ldots (4.3),$$

where S_{ik} is the shared boundary between polygon *i* and *k* and p_i is the perimeter of the small polygon *i*.

Semantic Similarity (SE)

Semantic similarity (*SE*) quantifies equivalence between land-use/land-cover classes of two polygons on the basis of a hierarchical system of land-use/land-cover classification. The relationship between polygons of two land-use/land-cover classes is given by:

$$SE_{ik} = \sum_{l=0}^{n} \frac{l.V_l}{n} \quad \ldots (4.4),$$

where *n* signifies the class levels described in land-use/land-cover classification scheme and *l* refers to the l^{th} level, $l = 1 \ldots n$. V_l is set as follows:

$$V_l = \begin{cases} 1, if\ two\ polygons\ belong\ to\ the\ same\ class\ at\ the\ \ lth\ level \\ 0, if\ two\ polygons\ belong\ to\ different\ classes\ at\ the\ lth\ level \end{cases}$$

SE value depends largely on the classification system. A three-level classification, such as the one used in present study, will yield the following four values of *SE*:

- *SE*=2 when two polygons have identical classes at Level-III.
- *SE*=1 when two polygons have identical classes at Level-II but not at Level-III.
- *SE*=1/3 when two polygons have identical classes at Level-I but not at Level-II and Level-III.
- *SE*=0 when two polygons have no identical classes at any level of classification.

4.1.2 Elimination

The traditional elimination operator is based on merging a polygon with either the largest nearby polygon or the polygon sharing the largest boundary. Such an operation ignores the semantics of the polygon. Consider the example in Figure 4-2 representing four polygons. While only using area as factor to decide a polygon to be selected to merge small polygon "FID-2", polygon "FID-1" is the ideal contender. But semantically, polygon "FID-3" is more closely related to "FID-2" as they belong to the same super class. This brings the role of polygon similarity model into view, as it quantifies the semantic and geometric similarity and selects the best possible solution.

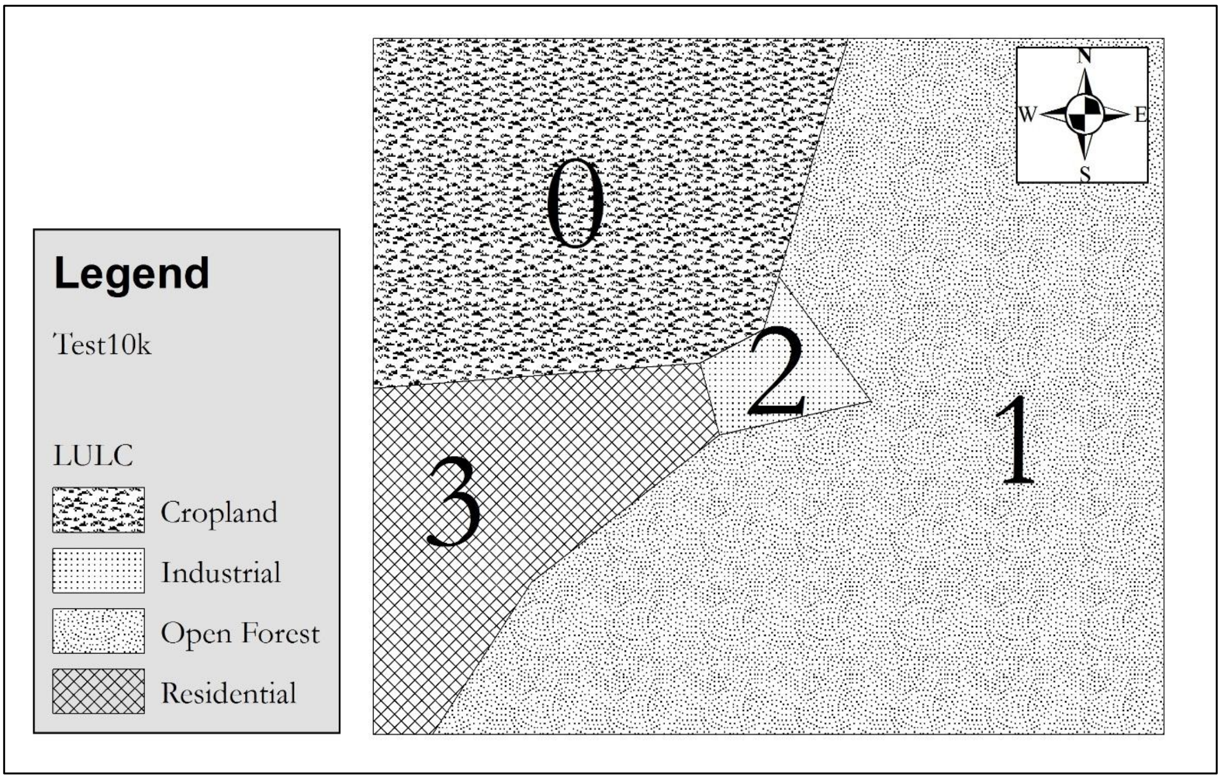


Figure 4-2. Sample data used to show elimination workflow. This dataset contains four polygons depicting four different classes where polygon "FID-2" represents a small polygons need to be eliminated.

The threshold limit to categorize a polygon as 'small' is derived from the scale at which the map is to be generalized. The minimum mappable unit (mmu, which in the present study is 3mm x 3mm) will be the limit and will be as follows-

1:10,000 – 900 m^2.

1:25,000 – 5625 m^2.

1:50,000 – 22500 m^2.

The identified small polygon is merged to a nearby polygon based on the polygon similarity model which uses Geometric Similarity (*GE*) and Semantic Similarity (*SE*).

There may be some classes which need to be kept in a restricted section so that a small polygon has least chance of getting merged with them due to their unique status. For example- small polygon merged with nearby 'River/streams' class polygon will create different boundaries for rivers and thus must not be allowed.

Considering all these factors, following rules have been identified for elimination operators-

- A small polygon surrounded from all sides with one larger polygon will be merged to it.
- If a small polygon is surrounded by one restricted and one unrestricted class polygon, then it will be merged in the unrestricted class polygon.
- If a small polygon is surrounded by more than one unrestricted class polygon, then it will be merged with the polygon with the highest similarity value as per the polygon similarity model.

The workflow of the elimination operator's algorithm (taking Figure 4-2 as case example) (Appendix A.4) is as follows-
Consider shape file "Test10k.shp" as the land-use/land cover map input file, with the attribute details shown in Table 4-1.

Table 4-1. Attribute table for shape file "Test10k.shp".

FID	Shape	LULC	Shape_Length	Shape_Area
0	Polygon	Cropland	4789.921	1397467
1	Polygon	Open Forest	8539.382	3161723
2	Polygon	Industrial	1605.671	130419.5
3	Polygon	Residential	3848.685	724810.3

Step-1: Using *arcpy* library, intersect "test10k.shp" with itself with the output as polylines ("shared_boundary.shp"). This produces a polyline shape file with lines representing shared boundaries between polygons.
Step-2: Using *dbfpy* module, load "test10k.dbf" into a variable. Then identify the smallest polygon and store its FID value, land-use/land-cover class and shape length (perimeter).
Step-3: Using the "FID" value, identify the polygons with shared boundary from "shared_boundary.dbf", storing their Land-use/land-cover type and shared boundary length.
Step-4: Based on the number of neighbouring polygons and their classes, identify the rule that needs to be applied.
Step-5a: If a single neighbouring polygon is *present*, then change the class of the small polygon to the neighbouring polygon's class.
Step-5b: If a single neighbouring polygon is *absent* then calculate the similarity value for each of the neighbouring polygon using previously stored variables using polygon similarity model. And change the land-use/land-cover class to the one having the highest value of similarity (Table 4-2). If an unrestricted class polygon is present in the neighbouring polygons, assign similarity value equal to 0.

Table 4-2. Computed Geometric Similarity (*GE*), Semantic Similarity (*SE*) and Overall similarity (*S*) for the three nearby polygons of small polygon "FID-2".

FID	LULC	Shared Boundary Length(m)	*GE*	*SE*	*S**
0	Cropland	398	0.248	0.000	0.124
1	Open Forest	975	0.607	0.000	0.304
3	Residential	233	0.145	1.000	0.573

*note that the weights assigned to both *GE* and *SE* are 0.5.

Step-7: Dissolve the layer "Test10k.shp" using dissolve tool from *arcpy* module taking "LULC" field as parameter with single parts allowed.

Thus the output of the given process produces a map with identified small polygons eliminated and merged with the nearby polygon of highest similarity *(S)* value based on the polygon similarity model.

4.1.3 Reclassify

The reclassify operator is used when the data base is subjected to a level change in the classification system. The operator uses the hierarchal relationship of the classes and change the LULC field of the polygon as per the super-class in which it falls (Figure 4-3). The operator uses *dbfpy* module to access the attributes and makes the changes according to the classification scheme and level (Appendix A.5). Further, using the *dissolve* tool from *arcpy* module, the polygons are merged within the same nearby classes.

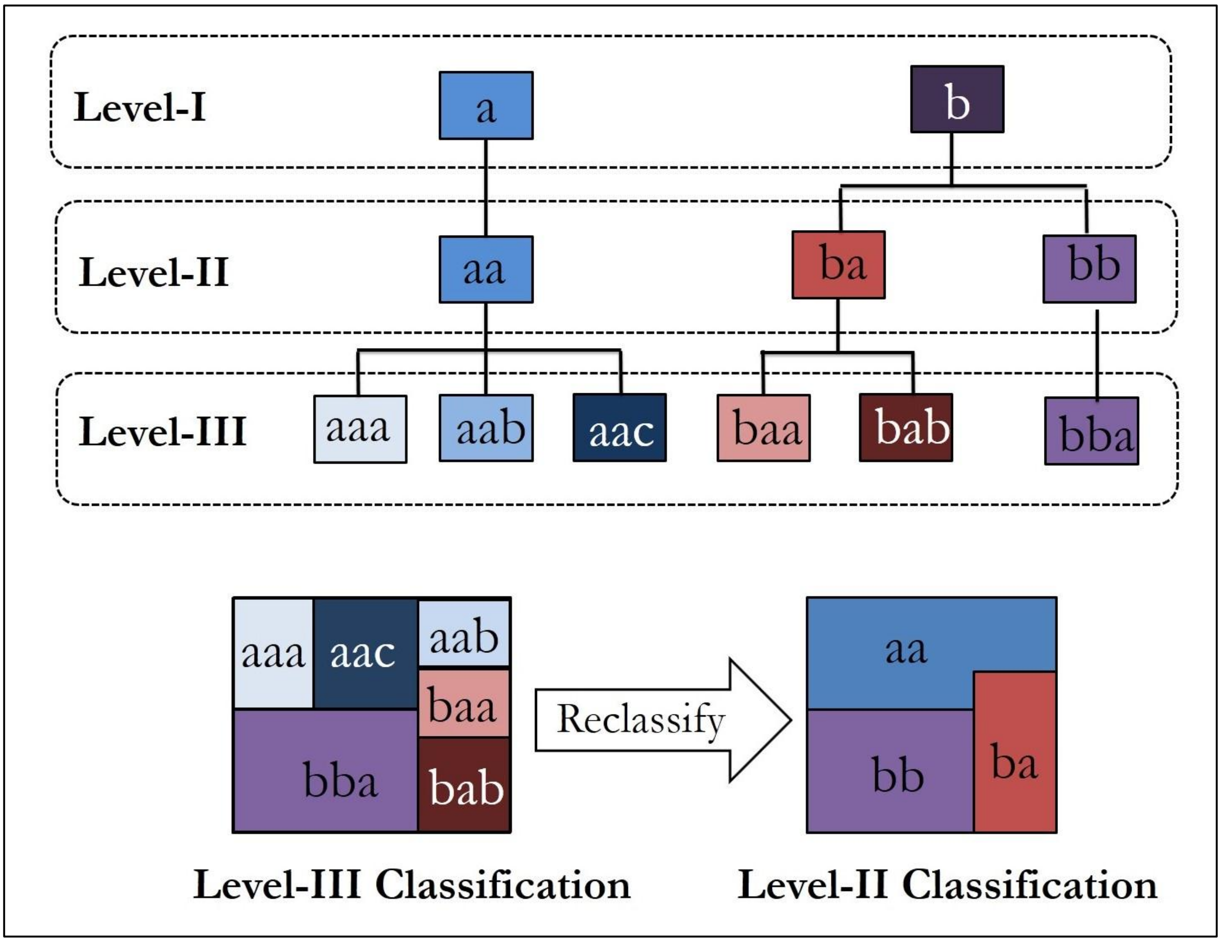


Figure 4-3. An illustration to show the change in polygons class after applying the reclassify operators. The changes in features are based as per the hierarchal relationships of classes in classification scheme (NUIS urban classification scheme in current study).

4.1.4 Smoothening

The task of a smoothening operator is to reduce the complexity of the features on a map so that they are visually more relevant as per the scale. It is also important to maintain the topology of the features: the current research demands an operator that produces features that neither overlap nor have gaps. An inbuilt tool is available in ArcGIS 10.1 software as Cartographic Tools> Generalization> Simplify Polygons which can be operated in two mode i.e. point removal or bend simplify. The bend simplify mode maintains the shape of the polygon and removes the extraneous bends in the boundary. This requires a tolerance value which is derived from the scale as minimum mappable length on map (3mm), which is-

1:25,000 scale – 75m

1:50,000 scale – 150m

4.1.5 Weight Calibration

The influence of *SE* and *GE* on the polygon similarity is controlled by weights ω_1 and ω_2 (Equation 4.2). The sum of these weights is unit, i.e. $\omega_1 + \omega_2$=1. Previous study by Gao et al. (2013, p. 393) states that-

> "*For a small polygon, GE does not influence the final similarity if the shared boundaries with its neighbours have nearly equal length. For such a case,* ω_1 *can be small so that SE makes a stronger difference. Otherwise, the importance of GE should be stressed and a larger weight assigned in order to avoid generating new conflicts after eliminating the small polygon.*"

To find the optimum value of these weights a standard situation is taken. Consider a small polygon "c" surrounded by two large polygon "a" and "b", where "a" is semantically more similar than "b" in Figure 4-4. The purpose of using semantic similarity in polygon similarity model is to merge the small polygon with the one which is semantically closer, while the inclusion of geometric similarity is to remove the creation of narrow corridors. Thus, the ideal weight combination for semantic and geometric similarity will be the one which tends to merge the small polygon to the most semantically closer polygon while reducing the chances of creation of small corridors.

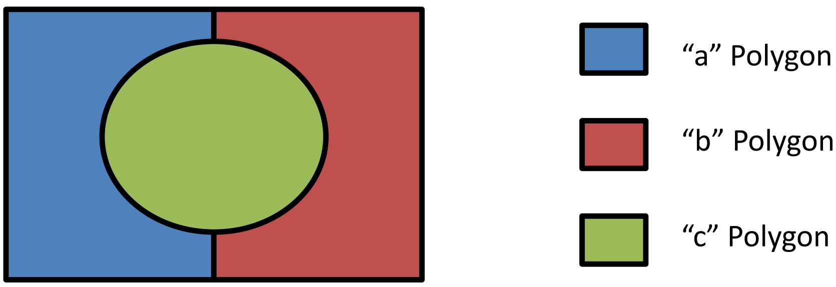


Figure 4-4. A standard case where a small polygon is surrounded by two polygons of different classes.

There are four values for Semantic similarity i.e. 2, 1, 0.33 and 0. Since the prepared data is made such that any two neighbouring similar land-use/land-cover polygons are merged together by using the *dissolve* tool, the remaining possible values for Semantic similarity (*SE*) is 1, 0.33 and 0.

Thus the possible combinations of semantic similarity (*SE*) among "c-a" and "c-b" are-

1. c-a=1, c-b=0.33
2. c-a=1, c-b=0.
3. c-a=0.33, c-b=0

Geometric similarity (*GE*) adopts the ratio of the length of the shared boundaries with its neighbouring polygon to its perimeter. Thus, its value can be between 0 and 1. In the given case i.e. when there are only two polygons, the possible combination for *GE* for c-a and c-b are-

1. c-a=0.1, c-b=0.9
2. c-a=0.2, c-b=0.8
3. c-a=0.3, c-b=0.7
4. c-a=0.4, c-b=0.6
5. c-a=0.5, c-b=0.5
6. c-a=0.6, c-b=0.4
7. c-a=0.7, c-b=0.3
8. c-a=0.8, c-b=0.2
9. c-a=0.9, c-b=0.1

To find the ideal weight, it is important to identify a value of *GE* which will serve to define the narrow-corridor. For this purpose, the value of *GE* taken as a threshold to define narrow corridor is 0.3 (Figure 4-5).

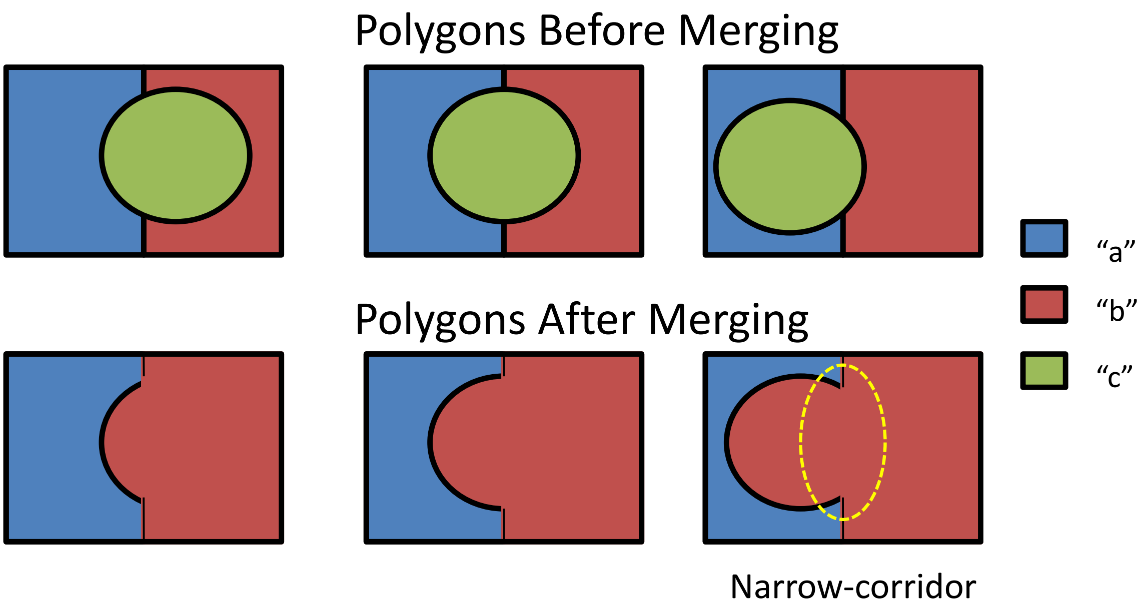


Figure 4-5. The creation of narrow corridor when a small polygon is merged with neighbour polygon base on highest similarity value *(S)*. *GE* in polygon similarity model is used as a measure to reduce the chances of creating such corridors.

Thus taking 0.3 as the value for threshold of creating a narrow corridor, similarity value *(S)* for polygon a and b based on the variation in weights value and geometric similarity value was computed for all the possible cases (Appendix A.2)

Based upon the computed value, it was observed that one universal value for these weights cannot serve the purpose for the model. Thus, the solution was to have multiple weights which will be based on semantic similarity of the two polygons sharing the largest boundary.

Table 4-3. Values for ω_1 and ω_2 based on different cases as per semantic value of two polygons sharing the largest boundaries.

Case	Semantic similarity *(SE)* of polygon "a"	Semantic similarity *(SE)* of polygon "b"	ω_1 Weight for *SE*	ω_2 Weight for *GE*
1	1	0.33	0.3	0.7
2	1	0	0.2	0.8
3	0.33	0	0.5	0.5
4	1	1	0.5	0.5
5	0.33	0.33	0.5	0.5
6	0	0	0.5	0.5

These value shown in Table 4-3 are based on the below given hypothesis.

- *GE* smaller than 0.3 will result in a narrow corridor.
- Small polygon "c" is surrounded by only two polygons. In case there are more than two polygons surrounding a small polygon, the two sharing the longest boundary are considered for the model.

4.2 Sequence of Operators

The eight identified sequences (two for Level-III generalization and six for Level-II generalization) using the three operators (Elimination, Reclassify and Smoothening) are applied on the 1:10,000 scale Level-III input data (Table 4-4). The intention is to find the changes caused by the order of these sequences on the output. The results are compared on the basis of change caused by these sequences on the classes.

Table 4-4. Sequences identified as per scales and level of classification.

Scale and Classification Level	Sequences (code)
1:25,000	Elimination > Smoothening (ES)
Level-III Classification	Smoothening > Elimination (SE)
1:25,000	Reclassify > Smoothening > Elimination (RSE)
Level-II Classification	Reclassify > Elimination > Smoothening (RES)
	Elimination > Reclassify > Smoothening (ERS)
	Elimination > Smoothening > Reclassify (ESR)
	Smoothening > Elimination > Reclassify (SER)
	Smoothening > Reclassify > Elimination (SRE)
1:50,000	Reclassify > Smoothening > Elimination (RSE)
Level-II Classification	Reclassify > Elimination > Smoothening (RES)
	Elimination > Reclassify > Smoothening (ERS)
	Elimination > Smoothening > Reclassify (ESR)
	Smoothening > Elimination > Reclassify (SER)
	Smoothening > Reclassify > Elimination (SRE)

5 RESULTS

5.1 Effect of Sequences on Output

A closer look on the outputs reveals the effect of these sequences on the map and the variation caused (Table 5-1, 5-2 and 5-3). The variation in outputs is due to the fact that the first operator used affects the geometric or semantic characteristic, or both, of polygons. This is later used by the second operator. Some operators produce similar results to each other while others produce largely varying output in terms of change in area of classes. A key factor to distinguish them is the change caused on individual classes (Figure 5-1, 5-2 and 5-3). This change (positive or negative) helps to define which sequence has resulted into the smallest change in the class distribution.

Table 5-1. Change caused by the applied sequences in terms of area and percentage change for 1:25,000 scale with Level-III NUIS classification scheme.

		1:10k Input	ES		SE		Image Interpreted	
Sr.	**LULC**	**Area (m²)** *(a)*	**Area (m²)** *(a_1)*	**Change (%)** ($\vert a_1-a\vert / a$)	**Area (m²)** (a_2)	**Change (%)** ($\vert a_2-a\vert / a$)	**Area (m²)** (a_3)	**Change (%)** ($\vert a_3-a\vert / a$)
1	Barren /Rocky	312415	314850	0.779	314850	0.779	364175	15.788
2	Commercial	2881241	2881949	0.025	2881949	0.025	2857395	0.852
3	Cropland	4356940	4373949	0.390	4371916	0.344	4416417	1.021
4	Dense Forest	2282868	2282663	0.009	2281267	0.070	2287198	0.260
5	Industrial	162809	162769	0.025	162769	0.025	177772	9.215
6	Lakes/ Ponds	42233	41652	1.376	41652	1.376	39614	4.826
7	Land with /without scrub	524464	522351	0.403	522351	0.403	391698	24.912
8	Mixed Built Up area	765428	763423	0.262	764959	0.061	876888	14.623
9	Open Forest	1691749	1685716	0.357	1686044	0.337	1735248	2.908
10	Plantation	277152	277557	0.146	277557	0.146	164477	40.801
11	Plantation/ Orchards	985350	991802	0.655	991802	0.655	902935	9.019
12	Public and Semipublic	20444120	20434571	0.047	20434599	0.047	20300665	0.655
13	Public utility and facility	54122	54893	1.425	54893	1.425	44433	19.327
14	Reclaimed land vacant land	256918	261848	1.919	261848	1.919	226757	13.658
15	Recreational	934986	844872	9.638	835112	10.682	563782	29.020
16	Residential	31326752	31405694	0.252	31412102	0.272	31891335	1.530
17	River/Streams	1251322	1249148	0.174	1254038	0.217	1344256	7.210
18	Transportation	129931	131090	0.892	131090	0.892	95753	27.197

Table 5-2. Change caused by the applied sequences in terms of area and percentage change for 1:25,000 scale with Level-II NUIS classification scheme.

			RSE		RES		ERS		ESR		SER		SRE		Image Interpreted	
Sr.	LULC	1:10k Area (m^2) (a)	Area (m^2) (a_1)	Change (%) ($\|a_1-a\|/a$)	Area (m^2) (a_2)	Change (%) ($\|a_2-a\|/a$)	Area (m^2) (a_3)	Change (%) ($\|a_3-a\|/a$)	Area (m^2) (a_4)	Change (%) ($\|a_4-a\|/a$)	Area (m^2) (a_5)	Change (%) ($\|a_5-a\|/a$)	Area m^2) (a_6)	Change (%) ($\|a_6-a\|/a$)	Area (m^2) (a_7)	Change (%) ($\|a_7-a\|/a$)
1	Barren /Rocky	312415	314850	0.779	314850	0.779	314850	0.779	314850	0.779	314850	0.779	314850	0.779	364175	16.568
2	Built Up (Urban)	56956307	56922482	0.059	56927430	0.051	56927430	0.051	56941109	0.027	56939321	0.030	56939321	0.030	57045913	0.157
3	Cropland	4356940	4375307	0.422	4376611	0.451	4383826	0.617	4373949	0.390	4371916	0.344	4363405	0.148	4393595	0.841
4	Dense Forest	2282868	2291310	0.370	2290005	0.313	2282791	0.003	2282663	0.009	2281267	0.070	2289778	0.303	2297681	0.649
5	Lakes/ Ponds	42233	42284	0.121	42284	0.121	42284	0.121	41652	1.376	41652	1.376	41652	1.376	39614	6.201
6	Land with /without scrub	524464	522885	0.301	522885	0.301	522885	0.301	522351	0.403	522351	0.403	522351	0.403	391698	25.315
7	Open Forest	1691749	1684746	0.414	1685074	0.395	1684746	0.414	1685716	0.357	1686044	0.337	1686044	0.337	1723270	1.863
8	Plantation	277152	277969	0.295	277969	0.295	277969	0.295	277557	0.146	277557	0.146	277557	0.146	164477	40.655
9	Plantation/ Orchards	985350	991610	0.635	991610	0.635	991610	0.635	991802	0.655	991802	0.655	991802	0.655	913622	7.279
10	River/Streams	1251322	1257355	0.482	1252080	0.061	1252408	0.087	1249148	0.174	1254038	0.217	1254038	0.217	1346755	7.627

Table 5-3. Change caused by the applied sequences in terms of area and percentage change for 1:50,000 scale with Level-II NUIS classification scheme.

			RSE		RES		ERS		ESR		SER		SRE		Image Interpreted	
Sr.	LULC	1:10k Area (m²) (*a*)	Area (m²) (a_1)	Change (%) (|a_1-a|/a)	Area (m²) (a_2)	Change (%) (|a_2-a|/a)	Area (m²) (a_3)	Change (%) (|a_3-a|/a)	Area (m²) (a_4)	Change (%) (|a_4-a|/a)	Area (m²) (a_5)	Change (%) (|a_5-a|/a)	Area m²) (a_6)	Change (%) (|a_6-a|/a)	Area (m²) (a_7)	Change (%) (|a_7-a|/a)
1	Barren /Rocky	312415	320051	2.444	320051	2.444	320051	2.444	320051	2.444	320051	2.444	320051	2.444	318426	1.924
2	Built Up (Urban)	56956307	56803472	0.268	56821927	0.236	56931019	0.044	56975632	0.034	56843863	0.197	56799604	0.275	57665254	1.245
3	Cropland	4356940	4414906	1.330	4354096	0.065	4378514	0.495	4345396	0.265	4411904	1.262	4419855	1.444	3687255	15.371
4	Dense Forest	2282868	2318370	1.555	2343417	2.652	2302577	0.863	2312805	1.311	2292629	0.428	2328654	2.006	2356569	3.228
5	Lakes/ Ponds	42233	38051	9.902	38051	9.902	38051	9.902	37736	10.648	45762	8.356	37736	10.648	33345	21.045
6	Land with /without scrub	524464	550034	4.875	552616	5.368	525651	0.226	521859	0.497	524370	0.018	536028	2.205	492468	6.101
7	Open Forest	1691749	1756289	3.815	1741567	2.945	1732497	2.409	1747048	3.269	1718125	1.559	1775886	4.973	1450512	14.260
8	Plantation	277152	177501	35.955	196516	29.095	196516	29.095	192912	30.395	268930	2.967	197495	28.741	92100	66.769
9	Plantation/ Orchards	985350	990806	0.554	988042	0.273	973680	1.184	973680	1.184	996882	1.170	990998	0.573	1003234	1.815
10	River/Streams	1251322	1305787	4.353	1324516	5.849	1282243	2.471	1253680	0.188	1252881	0.125	1268959	1.409	1581637	26.397

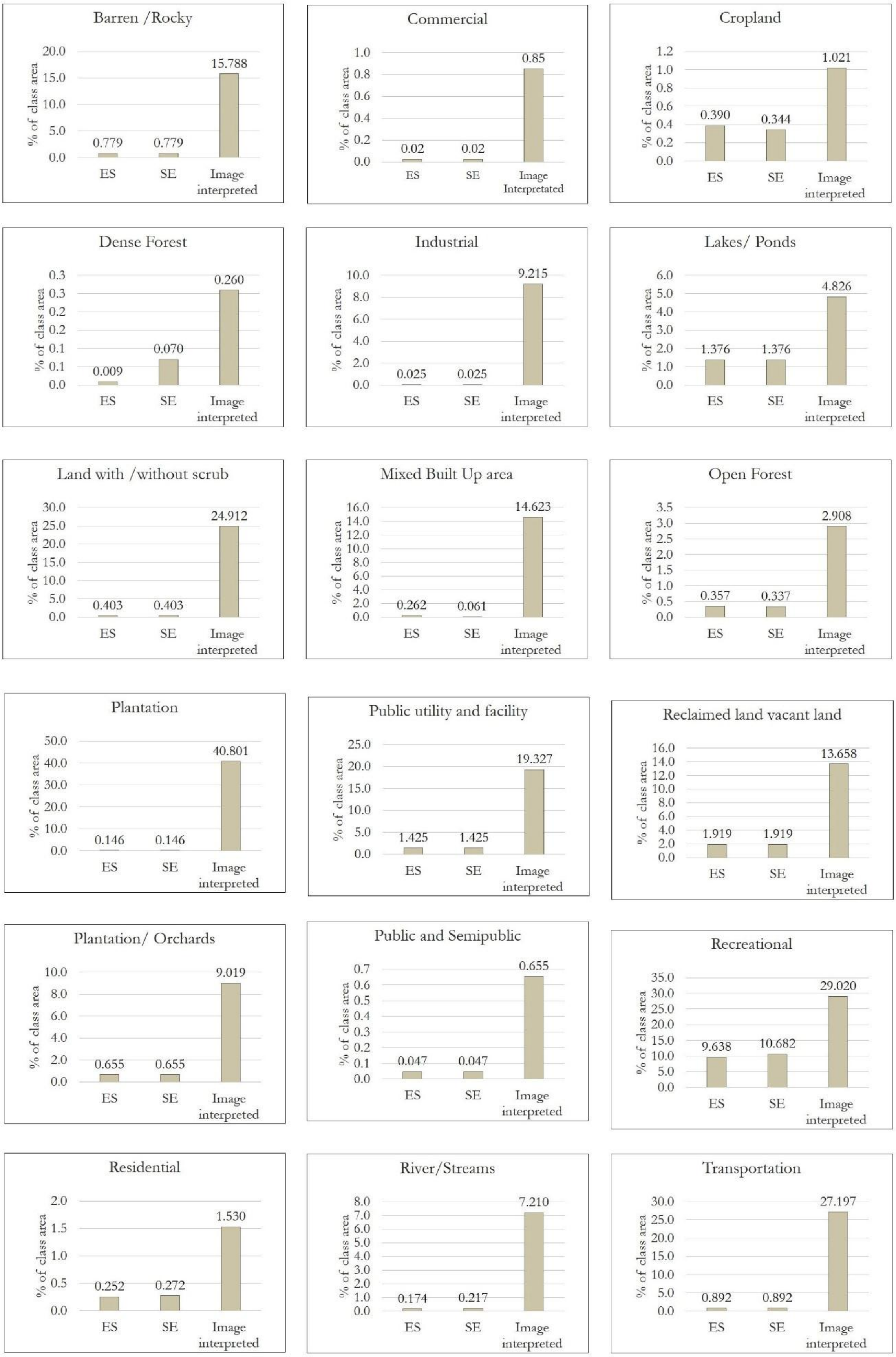


Figure 5-1. Graphs showing change in area caused by the two sequences in individual classes and its comparison with image interpreted maps for 1:25,000 scale at Level-III classification.

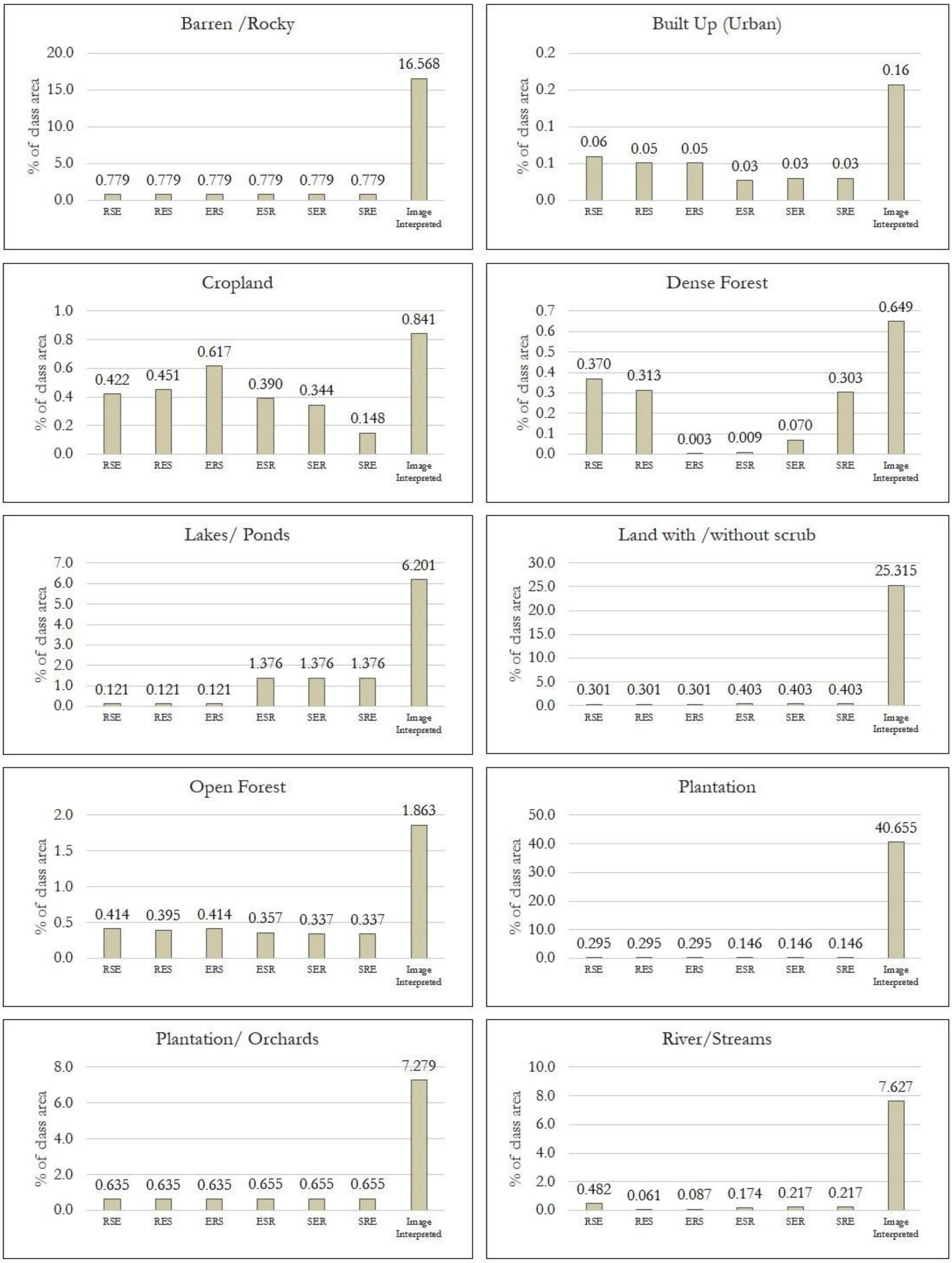


Figure 5-2. Graphs showing change in area caused by the six sequences in individual classes and its comparison with image interpreted maps at the 1:25,000 scale at Level-II classification.

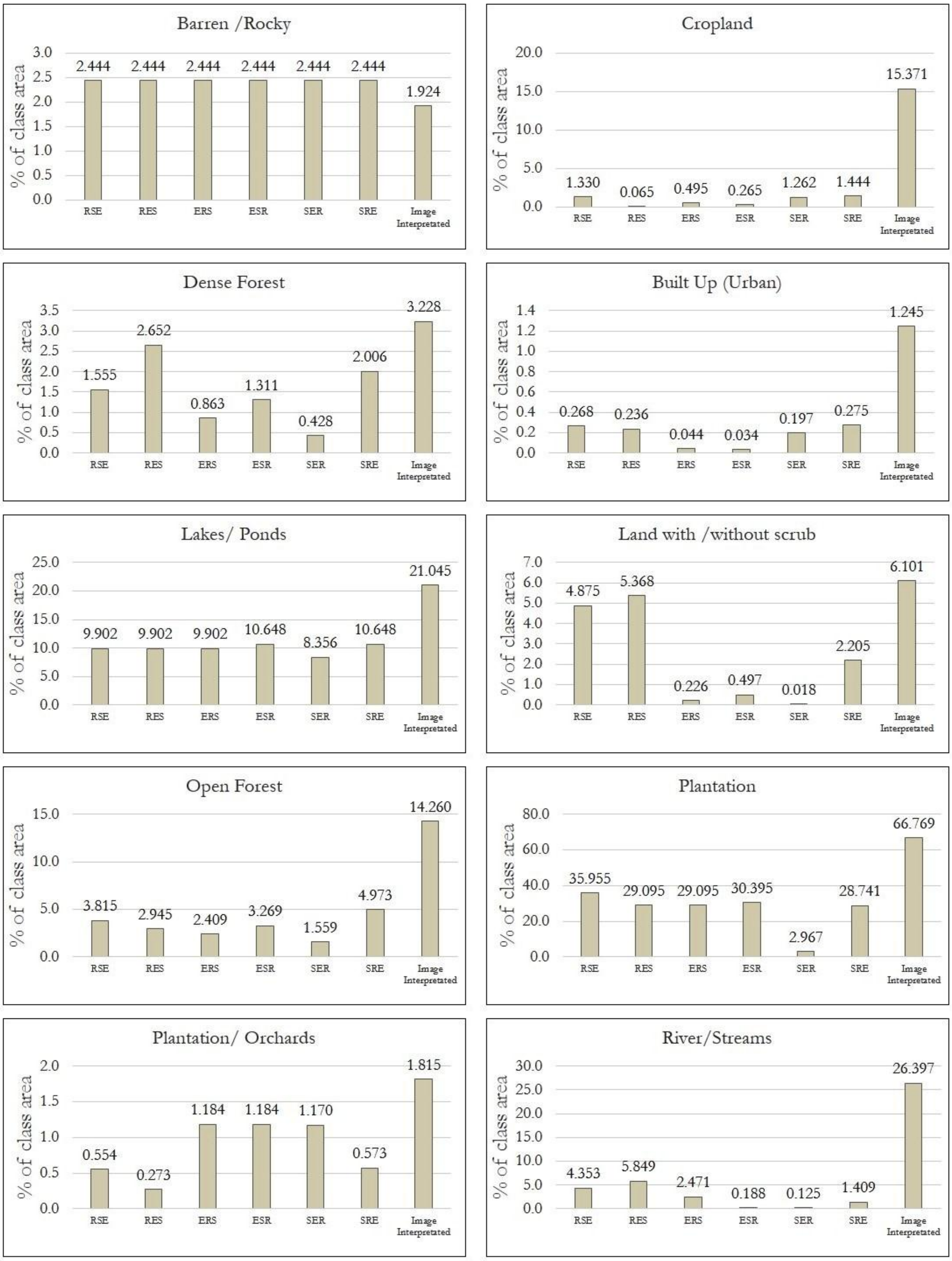


Figure 5-3. Graphs showing change in area caused by the six sequences in individual classes and its comparison with image interpreted maps at the 1:50,000 scale at Level-II classification.

Sequence	Change in area (m^2)	Percentage of total area
ES	225625	0.329
SE	239806	0.349
Image interpreted	1690688	2.462

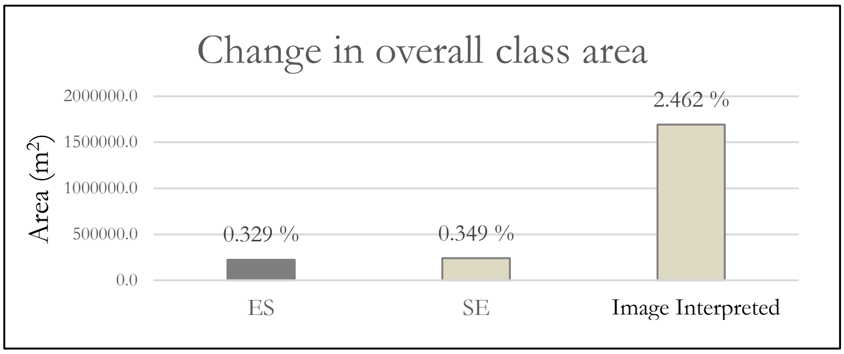


Figure 5-4. Comparison of sequences at the 1:25,000 scale with Level-III classification on the basis of overall change in class area.

Sequence	Change in area (m^2)	Percentage of total area
RSE	84812	0.123
RES	74260	0.108
ERS	75071	0.109
ESR	52605	0.077
SER	53970	0.079
SRE	50768	0.074
Image Interpreted	639576	0.931

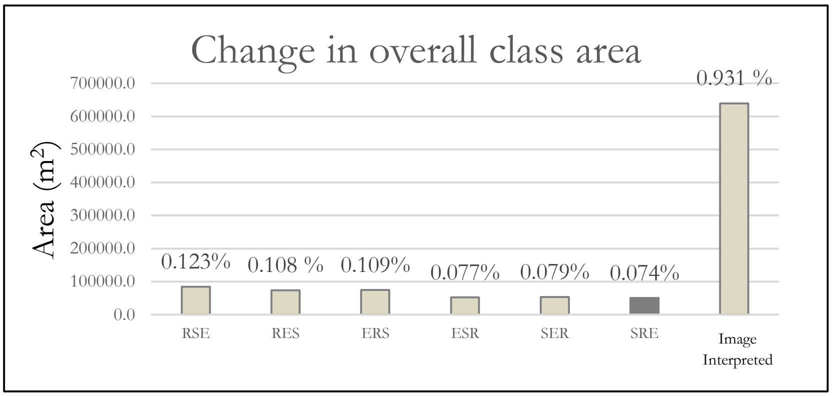


Figure 5-5. Comparison of sequences at the 1:25,000 scale with Level-II classification on the basis of overall change in class area.

Sequence	Change in area (m^2)	Percentage of total area
RSE	507803	0.739
RES	444083	0.647
ERS	243551	0.355
ESR	229111	0.334
SER	236117	0.344
SRE	476180	0.693
Image Interpreted	2273716	3.311

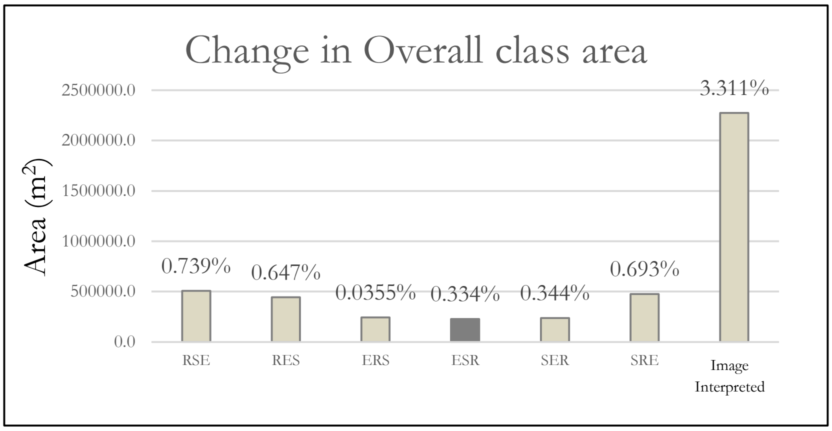


Figure 5-6. Comparison of sequences at the 1:50,000 scale with Level-II classification on the basis of overall change in class area.

The comparison of various sequences with image interpretation shows a better performance of generalization over image interpretation (Figure 5-4, 5-5 and 5-6). This applies to both individual and overall changes in area of classes. The reason is that generalization takes 1:10,000 Level-III classified map as an input, which is more detailed in terms of resolution and better in terms of accuracy. When the input is subjected to only two operators at the 1:25,000 scale map of Level-III classification, less/no changes are observed in most of the classes (Figure 5-1). This is due to the low value of parameters used for generalization and removing only minor level of details from map. As more operators are introduced, more variation in the results can be seen. The traditional approach of generalization of land-use/land-cover data comprises of using these operator in the following sequence, reclassify>elimination>smoothening. However, the current results reveals that this sequences causes the highest degree of change in the classes at both 1:25,000 and 1:50,000 scale. Thus. following (Table 5-4) are the identified sequences that results in the smallest change in the class distribution at the different levels (Second sub-objective).

Table 5-4. The sequences of operators that results in smallest change in area after generalization.

Generalization	Sequences (code)
1:25,000 Level-III	Elimination > Smoothening (ES)
1:25,000 Level-II	Smoothening > Reclassify > Elimination (SRE)*
1:50,000 Level-II	Elimination > Smoothening > Reclassify (ESR)**

*ESR and SER sequences also provide results close to SRE.
** SER and ERS sequences also provide results close to ESR.

The distribution of land-user/land-cover in a map is a very vital information that helps the policy makers to accurately assess a region. Keeping the class area close to the original during generalization will help to reduce the inconsistency of the output and make it more useful.

From the results, it can be observed that the classes having less, or no, presence of small polygons shows low affects of the sequences on final results. This is due to the fact that all operators affect either the sematic (land-use/land-cover class) or geometric (area) parameter of the polygons and thus give a different intermediate results that is used by the later operator. Since the operators uses the 1:10,000 scale dataset as input, which is made at a higher detail, they tend to produce less change in class area as compared to other coarser scale maps.

It can be observed that in all the three cases, the generalization output preserves the essence of the map far better than the maps made by visual image interpretation.

5.2 Comaprison of Modelled Output with Visual Interpretation Map

The accuracy of modelled outputs are compared with the maps prepared by visual image interpretation. For this purpose, the same 450 random points (25 per class at the 1:10k scale with Level-III), which were used to assess the accuracy of visual image interpreted maps, have been used to make the results comparable on same level (third sub-objective). Error matrix for individual maps is formulated to compute user's accuracy, producer's accuracy, overall accuracy and Kappa *(κ)* (Appendix A.3).

Table 5-5. Comparison of modelled output with corresponding maps made by visual interpretation.

Map scale	NUIS classification level		Image interpretation	Modelled Output
1:10,000	Level-III	Overall accuracy	96.22%	-
		κ	96.00%	-
1:25,000	Level-III	Overall accuracy	77.56%	94.44%
		κ	76.22%	94.12%
1:25,000	Level-II	Overall accuracy	88.44%	96.67%
		κ	83.27%	95.37%
1:50,000	Level-II	Overall accuracy	81.77%	90.89%
		κ	72.88%	86.94%

Accuracy assessment reveals that the maps produced by the generalization operators are more accurate than the maps prepared by visual interpretation at all three levels (Table 5-5). The two methods are also compared on the grounds of user's accuracy and producer's accuracy for individual classes in the corresponding maps (Figure 5-7, Figure 5-8 and Figure 5-9).

In terms of both user's accuracy (UA) and producer's accuracy (PA), the performance of generalization is better than the maps made by visual image interpretation over a very large margin. The intermediate scale of 1:25,000 scale has better accuracy at Level-II classification scheme as most of the classes are merged in the superclass and thus reduced the chances of misclassification. It is important to note that while 1:25,000 at Level-III also have a comparable accuracy value but with a higher level of sematic details due to the presence of detailed classes. Thus in terms of only accuracy, Level-II classification is a better option while in terms of semantic detail Level-III proves to be a better contender with a slightly low accuracy.

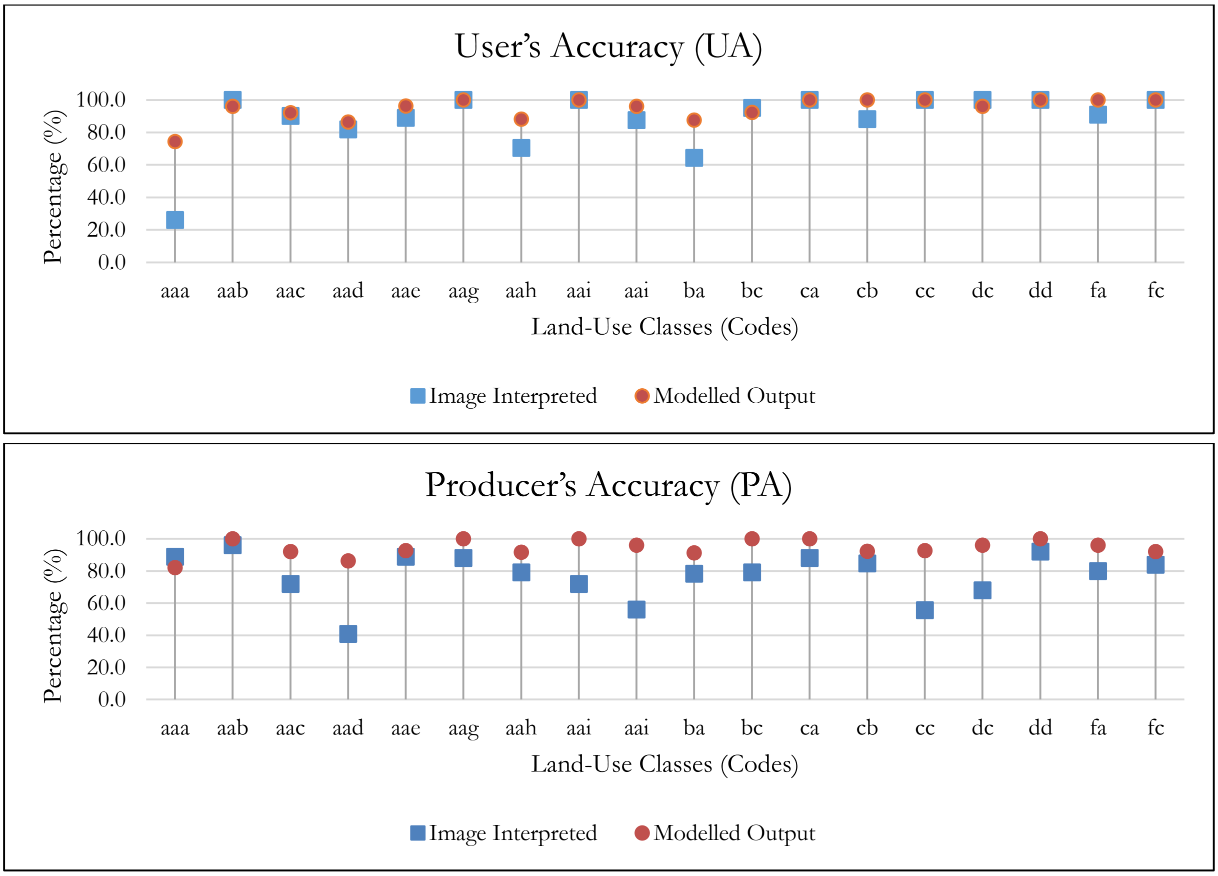


Figure 5-7. User's accuracy (UA) and producer's accuracy (PA) of the two corresponding maps made by image interpretation and by generalization at the 1:25,000 scale with Level-III NUIS classification scheme.

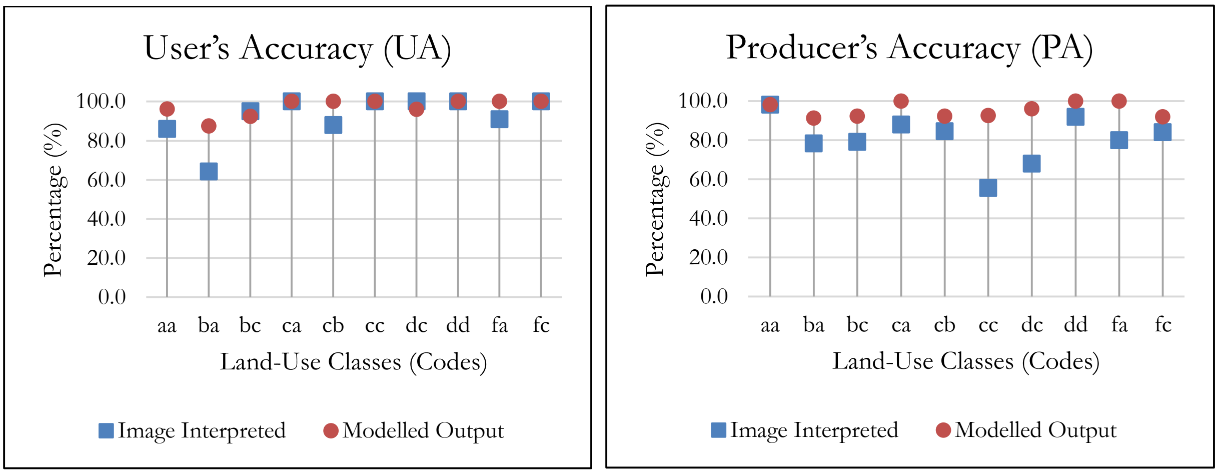


Figure 5-8. User's accuracy (UA) and producer's accuracy (PA) of the two corresponding maps made by image interpretation and generalization at the 1:25,000 scale with Level-II NUIS classification scheme.

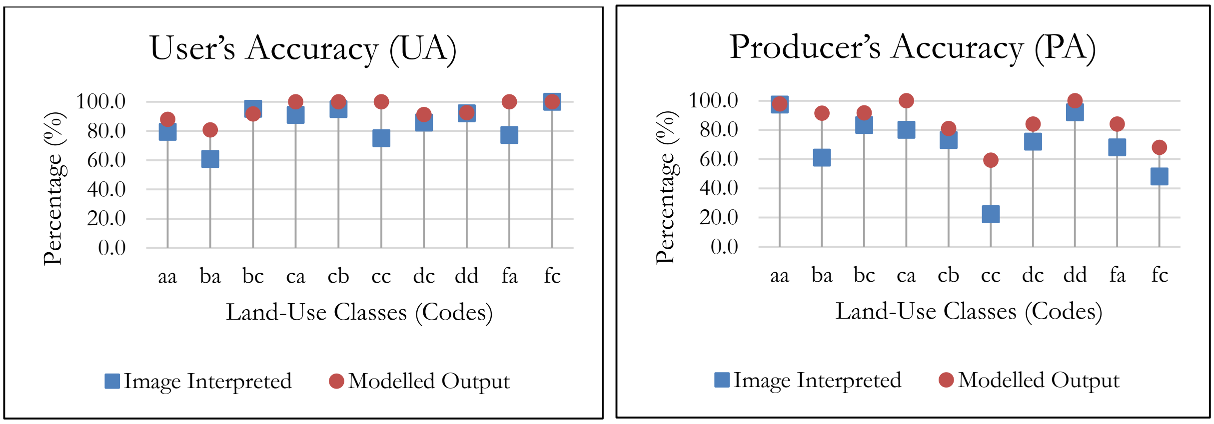


Figure 5-9. User's accuracy (UA) and producer's accuracy (PA) of the two corresponding maps made by image interpretation and generalization at the 1:50,000 scale with Level-II NUIS classification scheme.

The visual comparison of the two techniques reveals that the modelled outputs tend to maintain the essence of details even after the application of generalization operators. They stand comparable in front of maps made by visual image interpretation and contains more detail as shown in Figure 5-10, 5-11 and 5-12.

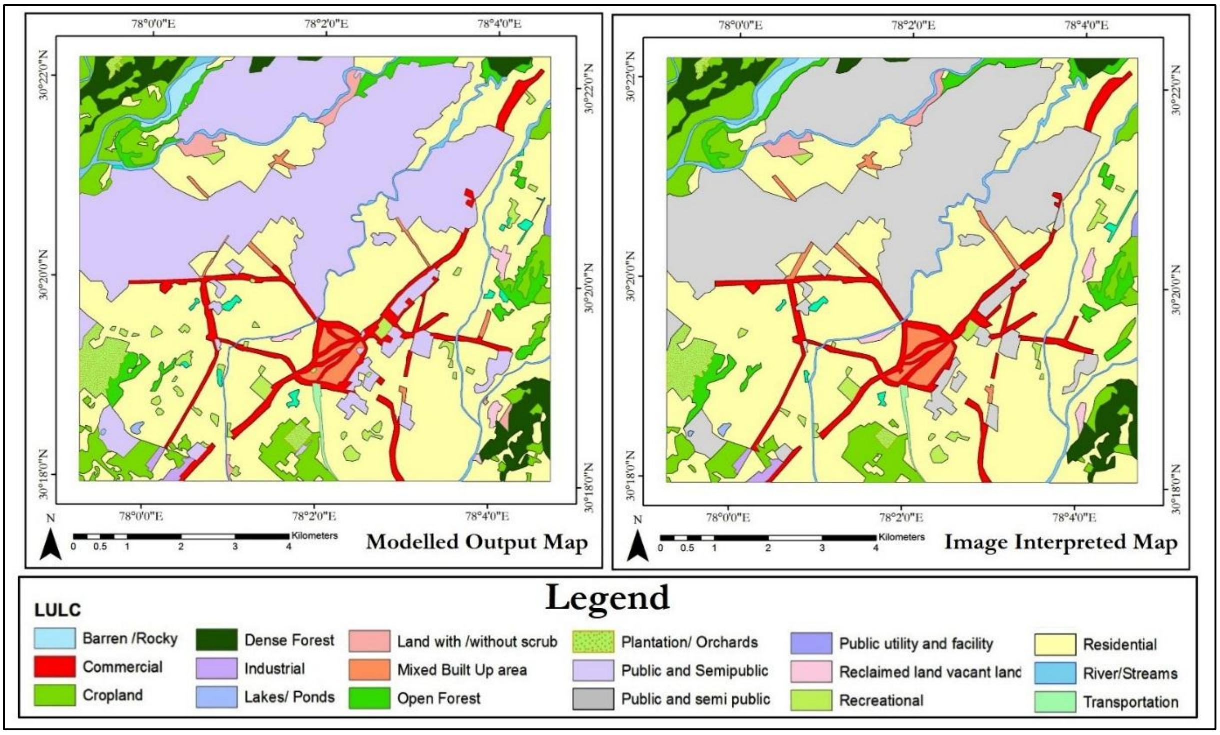


Figure 5-10. Visual comparison of modelled output map (left) with map made by visual image interpretation at the 1:25,000 scale with Level-III classification.

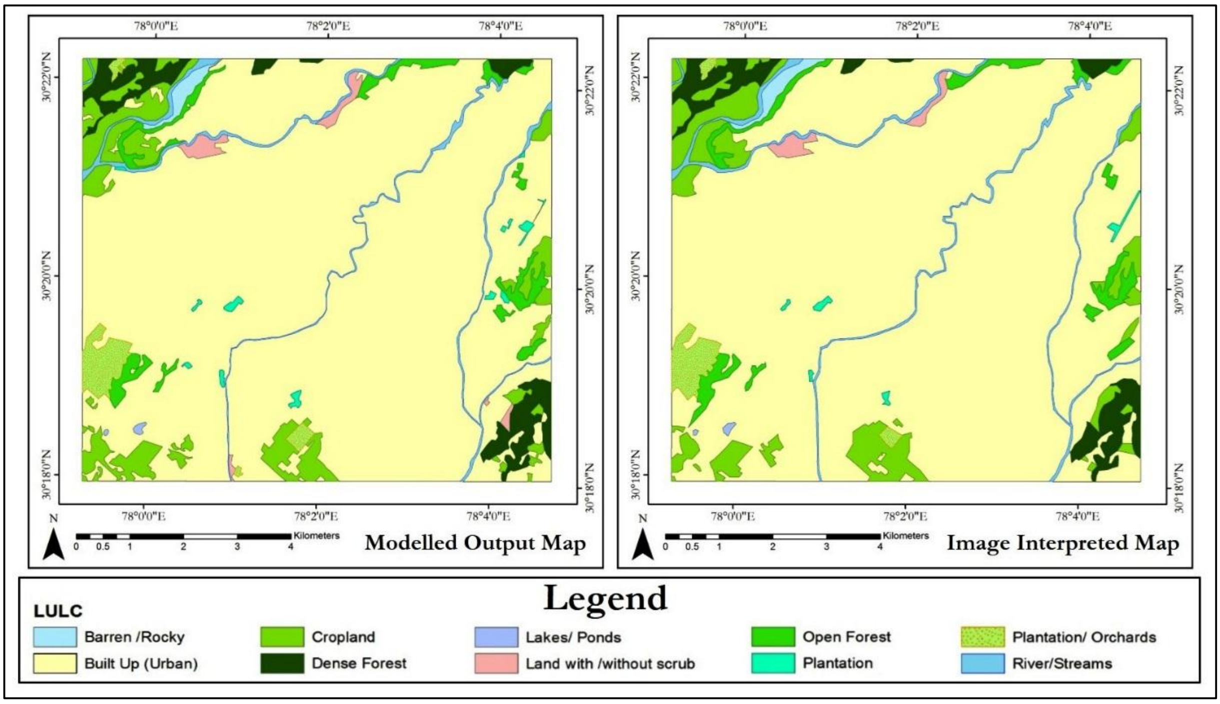


Figure 5-11. Visual comparison of modelled output map (left) with map made by visual image interpretation at the 1:25,000 scale with Level-II classification.

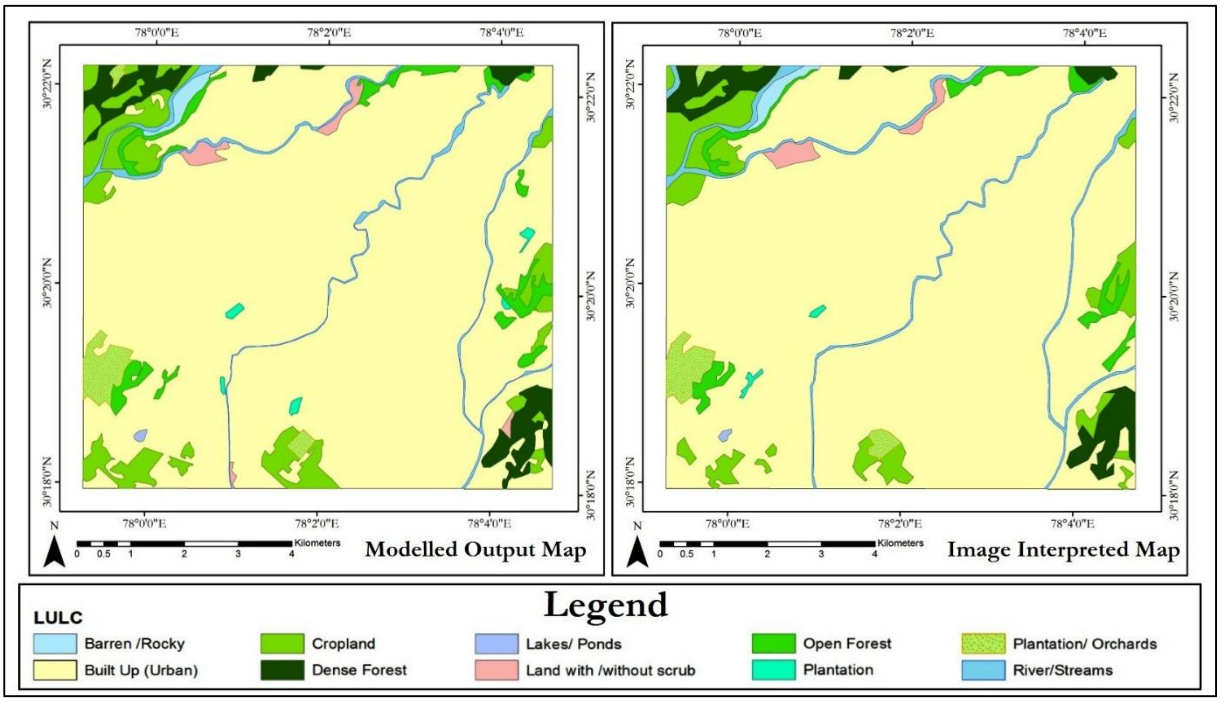


Figure 5-12. Visual comparison of modelled output map (left) with map made by visual image interpretation at the 1:50,000 scale with Level-II classification.

6 DISCUSSION

Most of land-use/land-cover maps are derived from remote sensing image by image classification. Those maps, however, do not comply with the actual activity within an area, as they represent only the classified land-cover derived from spectral value. For this purpose we require a hybrid classification containing both the essence of land-use and land cover such as NUIS urban classification scheme, which is also hierarchal in nature so as to have a relationship among various levels of classification. The creation of land-use/land-cover map by visual image interpretation for a large area is a time consuming process, as such maps cannot be made by image classification because the classification demands the type of activity on the ground and requires a manual interpretation. Creating such maps at different scales makes this process time and cost consuming. Automated generalization could be promising answer to this problem as it requires only creation of a base data whereas the coarser scale maps can be produced by an automated framework. As the input data used here will be based on high resolution image, it tends to be more accurate and better representation of ground reality in terms of classes. Mapping agencies could use this approach so as to be benefited in terms of resources utilized in creating maps at coarser scales by traditional visual image interpretation. Also, a formal structure in the generalization process will reduce the ambiguity caused by the traditional manual knowledge based approach of generalization which depends on the cartographer's skills. Among the various operators, the three selected in the present case are most reverent for the generalization of land-use/land-cover dataset.

In-built elimination tools currently available in commercial GIS software uses only the geometric information (area or shared boundary) of the nearby polygons which does not serve the purpose in case of land-use/land-cover polygons. The operator analysed in the present study use both the semantic and geometric information and thus are a better contender for elimination of land-use/land-cover polygons. *Elimination* operator requires a hierarchal classification scheme which in current study is NUIS urban classification scheme. This operator can be optimized to be used in other thematic environment such as soil maps or other classification scheme. As the operator derives its threshold of small polygon from the scale, it can also serve the purpose for coarser scale other than specified here.

The weights assigned to *SE* and *GE* are based upon a standard situation assuming that only two polygons are surrounding the small polygon as described in Section 4.1.5. This is a limitation of the present study as more cases are possible. Further study would be able to identify the weights for more possible combinations for small polygons and its nearby neighbours; more than two.

The *reclassify* operator is based on NUIS urban classification scheme which is hierarchal in nature. This is a major requirements for the present study as there should be a relationship among classes at various levels. With slight alterations, the operator can also be applied for other classification scheme.

The in-built tool available in ArcGIS 10.1 software is used for smoothening as the third operator. It is a major requirement for smoothening operator to reduce the complexity of the feature whereas at the same time to preserve the topology of the dataset. The used mode, "bend simplify", preserves the original shape of the feature while maintaining the topology of the layer and thus best suits as a smoothening operator. Similar as for elimination, this operator derives its input parameter form the scale and thus can be utilized at other scale levels.

Another aspect of the present research is the study of effect of various sequences of the generalisation operators and identifying the one that has the least effect on class distribution in terms of areal extent. The goal is to have a systematic manipulation while preserving the original essence of the dataset as much as possible. As these operators use the geometric information (area, perimeter, shared boundary length) and semantic information (land-use/land-cover class), their sequences cause change in the final output. This is due to the fact that the change caused by an operator on a polygon is to be used that are produced later by other operators. This could be easily observed at the places where there is a presence of small polygons, especially in classes such as "Recreational", "Dense Forest" and "Residential" (Figure 5.1). At the same time, some classes have almost no effect of sequence of operators such as "Barren Land" in Figure 5-2, which is because there are no small polygons in the class and nearby polygons. It can be also observed that as the scale increases, the overall change in class area also increases because the value of parameter in the operator increases. Table 5-5 and 5-6 also reveal that the largest change in area among these sequence occurs if the dataset is subjected to a level change (reclassify operator) first.

The comparison of generalized modelled output maps with visual interpretation, as shown in Table 5-10, confirms that generalization produces accurate results. the explanation is that the input used for generalization is a 1:10k scale map which has the highest accuracy as it is made by interpretation of 2.5m fused image. Figure 5-7, 5-8 and 5-9 also confirms the better accuracy of modelled output for most of the classes by comparing user's and producer's accuracy. Visual comparison in Figure 5-10, 5-11 and 5-12 signifies that the produced maps are legiable/clear, more detailed and fit for purpose of studying land-use/land-cover of an area.

Among the two results at the intermediate 1:25k scale, the Level-II classification yields a slightly higher overall accuracy and a higher Kappa *(κ)* value than the Level-III classification. Since most of the urban classes are diluted to "Built Up (Urban)", Level-II classification does not contain the details that might be required for the study of an urban area. While Level-III classification has a slightly lower accuracy, it is far more detailed than Level-II classification map. Thus, Level-II classification is a better choice if accuracy is the deciding factor, while Level-III classification brings the best of both detail and a slightly lower accuracy.

The present research provides a solution of systematic manipulation of data so as to obtain visually better represented maps for different scale while reducing the overall change caused by generalization process. The identified sequence of the operators reduces the change caused in the output and maintains the balance between data reduction while preserving the class distribution at the acceptable level of accuracy.

7 CONCLUSION AND RECOMMENDATION

7.1 Conclusion

This research aimed to formulate automated generalization of land-use/land-cover maps. The constructed operators use the semantic and geometric characteristic of the polygons with optimized weights so as to merge small polygons with most similar polygon based on polygon similarity model, with least chances of creating narrow corridors. Also, the identified sequences of generalization operators produce maps at given scale with smallest change in area of classes and far more better accuracy that the map made by tradition way of visual image interpretation technique. Such a process not only enables creating maps at different scale faster but also give more accurate representation of the ground reality.

7.1.1 Answers of Research Questions

1) How can generalization be carried out using a model that integrates geometric and semantic information?

Using the polygon similarity model (Gao et al. 2013), the semantic and geometric information of polygons can be utilized and quantified so that they can be used in a generalization operator. The elimination operator presented in this research takes into account the semantic similarity (*SE*) and geometric similarity (*GE*) values to merge a small polygon to the nearby polygon. The weights associated with *SE* and *GE* tend to control the effect of these characteristics on polygon similarity model. The reclassify operator uses only the semantic information of the polygon and assigns new classes based on the hierarchal relationship among classes.

2) What are the appropriate values for the parameters of such a model in the urban context?

Rather than assigning a universal value to the two weights ω_1 and ω_2 which control the effect of *SE* and *GE* in polygon similarity model, a case specific approach is used where two largest boundary sharing nearby polygons are considered and there semantic similarity (*SE*) with the small polygon is used to determine the weights. These value are based on the assumption that a narrow corridor is created where ever *GE* is smaller than 0.3.

3) What is the optimal sequence of generalization operators, i.e. the sequence that results into the smallest change in of class distribution?

Among the eight identified sequences, two were applicable when the input is subjected to generalization at the same level of classification, whereas six were applicable for a change in level of classification. The purpose of research was to identify the sequence that results in the least change in the areal extent of different classes which are-

- Elimination > Smoothening; at the 1:25,000 scale with the Level-III NUIS classification.
- Smoothening > Reclassify > Elimination; at the 1:25,000 scale with the Level-II NUIS classification.
- Elimination > Smoothening > Reclassify at the 1:50.000 scale with the Level-II NUIS classification.

4) What is the accuracy of the modeled output?

The accuracy of both generalized output and visual interpreted map are compared on the grounds of user's accuracy, producers' accuracy, overall accuracy and Kappa *(κ)*. The detailed comparison (Section 5.2) reveals that the accuracy of the generalization modelled output maps is higher than that of the maps prepared by visual image interpretation based on both the overall accuracy and the Kappa *(κ)* statistic. The performance of the modelled output is also better in user's accuracy and producer's accuracy for most classes. Therefore the generalization as used in the present research produces more accurate maps than the visually interpreted maps.

5) Which urban classification scheme gives the highest accuracy at the intermediate scale?

Among the two classification levels possible at the intermediate scale of 1:25,000, Level-II classification scheme provides better result in terms of overall and Kappa *(κ)* accuracy. The explanation is that most of the classes depicted in Level-III are merged in a Level-II superclass and thus creates a lower chance of having an error. In terms of level of details, the modelled output at Level-III classification is far more detailed but less accurate. Using the 1:25,000 map with the Level-III classification will only be recommended if high classification details are required.

7.2 Recommendations

Most of the current practices of making maps are still traditional where they are made separately at different scales. It is important to have a relationship among these scales as they all depict the same ground. A unified classification system for thematic maps must be followed so as to have a hierarchical relationship among these maps (and classes) at different scales. Making coarser scale maps by a standard generalization process needs to have a formalised structure. Such a process should be able to produce modelled maps that are more accurate.

The elimination operator used in the current study is based on polygon similarity model, where weights are calibrated under the hypostasis of two surrounding polygons. Further research can be done to optimise these weights for more number of surrounding polygons. Also, spectral information is not used in the present context as the maps are prepared by visual interpretation. For image classified maps, a similar approach can be utilized with further research towards the weights assigned to the three characteristics. It is also important to take into account the intended purpose of map to be prepared, and it must play a crucial role in defining the rule on which generalization is based. An identical approach for generalization of all kind of maps would not serve the purpose as each one require certain variation. Mapping organisation can incorporate these suggestion which will benefit them in terms of resources and time, with more accurate outputs. Commercial and open source GIS software can also incorporate tools which allow user to generalization data based on required theme.

REFRENCES

Chaudhry, OZ, Mackaness, WA & Regnauld, N 2009, 'A functional perspective on map generalisation'. *Computers, Environment and Urban Systems*, vol. *33*, no. 5, pp. 349–362, DOI: 10.1016/j.compenvurbsys.2009.07.002.

Foerster, T, Stoter, J & Kobben, B 2007, 'Towards a formal classification of generalization operators',*In Proceedings of the 23nd international cartographic conference ICC: Cartography for everyone and for you, International Cartographic Association (ICA), Moscow*. Available at: <http://kartoweb.itc.nl/kobben/publications/icc07_foerster_stoter_kobben.pdf> [Accessed September 15, 2014].

Foerster, T, Stoter, J & Morales, J 2009, 'Enhancing Cartographic Generalization Processing with Grid Computing Power',*In Proceedings of Grid Technologies for Geospatial Applications – pre-conference workshop AGILE 2000*. Available at: <http://ifgi.uni-muenster.de/archives/agile/submissions/AGILE-GridWorkshop-Foerster.pdf> [Accessed September 15, 2014].

Foerster, T, Stoter, JE, & Kraak, M 2010, 'Challenges for Automated Generalisation at European Mapping Agencies – A Qualitative and Quantitative Analysis', *Cartographic Journal*, vol. 47, no. 1, pp. 41-54. DOI: 10.1179/000870409X12525737905123.

Gao, W, Stein, A, Yang, L, Wang, Y, & Fang, H 2013, 'Improving Representation of Land-use Maps Derived from Object-oriented Image Classification', *Transactions in GIS*, vol. *17, no.* 3, pp. 387-405. DOI: 10.1111/j.1467-9671.2012.01368.x

Government of India, Town and Country Planning Organization 2008, *NUIS Design and Standards*. Available at :< http://tcp.cg.gov.in/nuis/Design_Standards.pdf> [Accessed September 15, 2014].

Harrie, L, & Sarjakoski, T 2002, 'Simultaneous graphic generalization of vector data set', *GeoInformatica*, vol. *6*, no. 3, pp. 233-261.DOI: 10.1023/A:1019765902987

Haunert, JH, & Sester, M 2008, 'Assuring logical consistency and semantic accuracy in map generalization', *Photogrammetrie Fernerkundung Geoinformation*, no. 3, pp. 165-173.Available at: <http://www.dgpf.de/neu/pfg-digital/2008/Heft_3.pdf#page=17>[Accessed September 15, 2014].

Jenerette, GD & Wu, J 2000, 'On the definitions of scale'. *Bulletin of the Ecological Society of America*, no. 81, pp. 104–105. Available at: <http://leml.asu.edu/jingle/Web_Pages/Wu_Pubs/.../Jenerett_Wu_2000.pdf> [Accessed September 15, 2014].

JoÃo, EM 1998, *Causes and Consequences of Map Generalisation*, CRC Press,pp. 1-27, ISBN: 0748407774.

Müller, JC 1991, 'Generalisation of spatial databases', In Maguire D J, Goodchild M F, Rhind D W (eds) *Geographical information systems: principles and applications,* pp. 457–75, ISBN: 0471321826.

Neun, M, Burghardt, D, & Weibel, R 2009, 'Automated processing for map generalization using web services', *GeoInformatica*, vol. 13, no. 4, pp. 425-52. DOI: 10.1007/s10707-008-0054-3.

National Remote Sensing Agency 2006, *National Land Use Land Cover Mapping using multi-temporal Satellite Data,* NRSA/RSGIS-A/NRC/NLULC-L3/TECHMAN/R02/May-06.

Raja, K 2012, *Five stages of Multi-level Planning in India.* Available at: <http://www.preservearticles.com/2012020322525/five-stages-of-multi-level-planning-in-india.html> [Accessed September 15, 2014].

Shekhar, Shashi, and Hui Xiong, eds. *Encyclopedia of GIS*. Springer, 2008.

Stoter, J, Nijhuis, R, Post, M, Van Altena, V, Bulder, J, Bruns, B, & Van Smaalen, J 2011, 'Feasibility study on an automated generalisation production line for multiscale topographic products'. Available at: <http://www.gdmc.nl/publications/2011/Automated_generalisation_production_line.pdf> [Accessed September 15, 2014].

Stoter, J, Post, M, van Altena, V, Nijhuis, R, & Bruns, B 2014, 'Fully automated generalization of a 1: 50k map from 1: 10k data', *Cartography and Geographic Information Science*, vol. 41, no. 1, pp. 1-13,DOI:10.1080/15230406.2013.824637.

Stoter, J, Burghardt, D, Duchêne, C, Baella, B, Bakker, N, Blok, C, & Schmid, S 2009, 'Methodology for evaluating automated map generalization in commercial software', *Computers, Environment and Urban Systems*, vol. *33*, no. 5, pp. 311-324, DOI: 10.1016/j.compenvurbsys.2009.06.002.

A. APPENDIX

A.1 Error matrix for Visually Interpreted Maps.

Table A-1. Error matrix for 1:50,000 scale map with Level-II classification.

LULC	Barren /Rocky	Commercial	Cropland	Dense Forest	Industrial	Lakes/ Ponds	Land with /without scrub	Mixed Built Up area	Open Forest	Plantation	Plantation/ Orchards	Public and Semipublic	Public utility and facility	Reclaimed land vacant land	Recreational	Residential	River/Streams	Transportation	Total	User's accuracy
Barren /Rocky	25																		25	100.00
Commercial		21						2			1					1			25	84.00
Cropland			22				1							1	1				25	88.00
Dense Forest				25															25	100.00
Industrial					24											1			25	96.00
Lakes/ Ponds						25													25	100.00
Land with /without scrub							24					1							25	96.00
Mixed Built Up area		2						23											25	92.00
Open Forest									25										25	100.00
Plantation										25									25	100.00
Plantation/ Orchards											24	1							25	96.00
Public and Semipublic												25							25	100.00
Public utility and facility													25						25	100.00
Reclaimed land vacant land			1											24					25	96.00
Recreational										1					23	1			25	92.00
Residential									1	1						23			25	92.00
River/Streams																	25		25	100.00
Transportation																		25	25	100.00
Total	25	23	23	25	24	25	25	25	26	27	25	27	25	25	24	26	25	25	433	96.22
Producer's accuracy	100.00	91.30	95.65	100.00	100.00	100.00	96.00	92.00	96.15	92.59	96.00	92.59	100.00	96.00	95.83	88.46	100.00	100.00	96.26	

Table A-2. Error matrix for 1:25,000 scale map with Level-III classification.

LULC	Barren /Rocky	Commercial	Cropland	Dense Forest	Industrial	Lakes/ Ponds	Land with /without scrub	Mixed Built Up area	Open Forest	Plantation	Plantation/ Orchards	Public and Semipublic	Public utility and facility	Reclaimed land vacant land	Recreational	Residential	River/Streams	Transportation	Total	User's accuracy (%)
Barren /Rocky	23																		23	100.00
Commercial		19						7								1			27	70.37
Cropland			18	2	1		2				2						2	1	28	64.29
Dense Forest				22															22	100.00
Industrial					24														24	100.00
Lakes/ Ponds						21													21	100.00
Land with /without scrub							17												17	100.00
Mixed Built Up area		2						18											20	90.00
Open Forest	1						1		22		1								25	88.00
Plantation										15									15	100.00
Plantation/ Orchards											19	1							20	95.00
Public and Semipublic		1		1					1			24							27	88.89
Public utility and facility													22						22	100.00
Reclaimed land vacant land			1											14			1		16	87.50
Recreational										1					9	1			11	81.82
Residential		2	4			4	5		3	11	2	2	3	11	13	24	2	6	92	26.09
River/Streams	1															1	20		22	90.91
Transportation																		18	18	100.00
Total	25	24	23	25	25	25	25	25	26	27	24	27	25	25	22	27	25	25	349	87.94
Producer's accuracy (%)	92.00	79.17	78.26	88.00	96.00	84.00	68.00	72.00	84.62	55.56	79.17	88.89	88.00	56.00	40.91	88.89	80.00	72.00	77.30	

Table A-3. Error matrix for 1:25,000 scale map with Level-II classification.

	Barren /Rocky	Built Up (urban)	Cropland	Dense Forest	Lakes/ Ponds	Land with /without scrub	Open Forest	Plantation	Plantation/ Orchards	River/Streams		User's accuracy (%)
Barren /Rocky	23										23	100
Built Up (urban)		221	5	1	4	5	4	12	2	3	257	85.9922179
Cropland		2	18	2		2			2	2	28	64.28571429
Dense Forest				22							22	100
Lakes/ Ponds					21						21	100
Land with /without scrub						17					17	100
Open Forest	1					1	22		1		25	88
Plantation								15			15	100
Plantation/ Orchards		1							19		20	95
River/Streams	1	1								20	22	90.90909091
	25	225	23	25	25	25	26	27	24	25	398	92.41870231
Producer's accuracy (%)	92	98.22222	78.26087	88	84	68	84.61538	55.55556	79.16667	80	80.78207	

Table A-4. Error matrix for 1:50,000 scale map with Level-II classification.

	Barren /Rocky	Built Up (urban)	Cropland	Dense Forest	Lakes / Ponds	Land with /without scrub	Open Forest	Plantation	Plantation / Orchards	River/Streams	Total	User's accuracy (%)
Barren /Rocky	23						2				25	92
Built Up (urban)		219	5	3	13	4	3	21	2	6	276	79.34782609
Cropland		3	14	2		2			1	1	23	60.86956522
Dense Forest			1	20					1		22	90.90909091
Lakes/ Ponds					12						12	100
Land with /without scrub		2				18				1	21	85.71428571
Open Forest			1				19				20	95
Plantation							2	6			8	75
Plantation/ Orchards			1						20		21	95.23809524
River/Streams	2	1	1			1				17	22	77.27272727
Total	25	225	23	25	25	25	26	27	24	25	368	85.13515904
Producer's accuracy (%)	92	97.33333333	60.86956522	80	48	72	73.07692308	22.22222222	83.33333333	68	69.68353772	

A.2 Similarity *(S)* for nearby polygons based on values of *SE* and *GE*

Table A-5. Similarity value for polygon "a" and "b" based on the variation in weights value and geometric similarity value. The standard value for Semantic similarity for "a" is 1 and "b" is 0.33.

	SE_a=1, GE_a=0.9 SE_b=0.33, GE_b=0.1	SE_a=1, GE_a=0.8 SE_b=0.33, GE_b=0.2	SE_a=1, GE_a=0.7 SE_b=0.33, GE_b=0.3	SE_a=1, GE_a=0.6 SE_b=0.33, GE_b=0.4	SE_a=1, GE_a=0.5 SE_b=0.33, GE_b=0.5	SE_a=1, GE_a=0.4 SE_b=0.33, GE_b=0.6	SE_a=1, GE_a=0.3 SE_b=0.33, GE_b=0.7	SE_a=1, GE_a=0.2 SE_b=0.33, GE_b=0.8	SE_a=1, GE_a=0.1 SE_b=0.33, GE_b=0.9
ω_1 =0.1	a 0.91	a 0.82	a 0.73	a 0.64	a 0.55	a 0.46	a 0.37	a 0.28	a 0.19
ω_2 =0.9	b 0.123	b 0.213	b 0.303	b 0.393	b 0.483	b 0.573	b 0.663	b 0.753	b 0.753
ω_1 =0.2	a 0.92	a 0.84	a 0.76	a 0.68	a 0.6	a 0.52	a 0.44	a 0.36	a 0.28
ω_2 =0.8	b 0.146	b 0.226	b 0.306	b 0.386	b 0.466	b 0.546	b 0.626	b 0.706	b 0.706
ω_1 =0.3	a 0.93	a 0.86	a 0.79	a 0.72	a 0.65	a 0.58	a 0.51	a 0.44	a 0.37
ω_2 =0.7	b 0.169	b 0.239	b 0.309	b 0.379	b 0.449	b 0.519	b 0.589	b 0.659	b 0.659
ω_1 =0.4	a 0.94	a 0.88	a 0.82	a 0.76	a 0.7	a 0.64	a 0.58	a 0.52	a 0.46
ω_2 =0.6	b 0.192	b 0.252	b 0.312	b 0.372	b 0.432	b 0.492	b 0.552	b 0.612	b 0.612
ω_1 =0.5	a 0.95	a 0.9	a 0.85	a 0.8	a 0.75	a 0.7	a 0.65	a 0.6	a 0.55
ω_2 =0.5	b 0.215	b 0.265	b 0.315	b 0.365	b 0.415	b 0.465	b 0.515	b 0.565	b 0.565
ω_1 =0.6	a 0.96	a 0.92	a 0.88	a 0.84	a 0.8	a 0.76	a 0.72	a 0.68	a 0.64
ω_2 =0.4	b 0.238	b 0.278	b 0.318	b 0.358	b 0.398	b 0.438	b 0.478	b 0.518	b 0.518
ω_1 =0.7	a 0.97	a 0.94	a 0.91	a 0.88	a 0.85	a 0.82	a 0.79	a 0.76	a 0.73
ω_2 =0.3	b 0.261	b 0.291	b 0.321	b 0.351	b 0.381	b 0.411	b 0.441	b 0.471	b 0.471
ω_1 =0.8	a 0.98	a 0.96	a 0.94	a 0.92	a 0.9	a 0.88	a 0.86	a 0.84	a 0.82
ω_2 =0.2	b 0.284	b 0.304	b 0.324	b 0.344	b 0.364	b 0.384	b 0.404	b 0.424	b 0.424
ω_1 =0.9	a 0.99	a 0.98	a 0.97	a 0.96	a 0.95	a 0.94	a 0.93	a 0.92	a 0.91
ω_2 =0.1	b 0.307	b 0.317	b 0.327	b 0.337	b 0.347	b 0.357	b 0.367	b 0.377	b 0.377

Table A-6. Similarity value for polygon "a" and "b" based on the variation in weights value and geometric similarity value. The standard value for Semantic similarity for "a" is 1 and "b" is 0.

	SE_a=1,GE_a=0.9 SE_b=0,GE_b=0.1	SE_a=1,GE_a=0.8 SE_b=0,GE_b=0.2	SE_a=1,GE_a=0.7 SE_b=0,GE_b=0.3	SE_a=1,GE_a=0.6 SE_b=0,GE_b=0.4	SE_a=1,GE_a=0.5 SE_b=0,GE_b=0.5	SE_a=1,GE_a=0.4 SE_b=0,GE_b=0.6	SE_a=1,GE_a=0.3 SE_b=0,GE_b=0.7	SE_a=1,GE_a=0.2 SE_b=0,GE_b=0.8	SE_a=1,GE_a=0.1 SE_b=0,GE_b=0.9
ω_1 =0.1	a 0.91	a 0.82	a 0.73	a 0.64	a 0.55	a 0.46	a 0.37	a 0.28	a 0.19
ω_2 =0.9	b 0.09	b 0.18	b 0.27	b 0.36	b 0.45	b 0.54	b 0.63	b 0.72	b 0.81
ω_1 =0.2	a 0.92	a 0.84	a 0.76	a 0.68	a 0.6	a 0.52	a 0.44	a 0.36	a 0.28
ω_2 =0.8	b 0.08	b 0.16	b 0.24	b 0.32	b 0.4	b 0.48	b 0.56	b 0.64	b 0.72
ω_1 =0.3	a 0.93	a 0.86	a 0.79	a 0.72	a 0.65	a 0.58	a 0.51	a 0.44	a 0.37
ω_2 =0.7	b 0.07	b 0.14	b 0.21	b 0.28	b 0.35	b 0.42	b 0.49	b 0.56	b 0.63
ω_1 =0.4	a 0.94	a 0.88	a 0.82	a 0.76	a 0.7	a 0.64	a 0.58	a 0.52	a 0.46
ω_2 =0.6	b 0.06	b 0.12	b 0.18	b 0.24	b 0.3	b 0.36	b 0.42	b 0.48	b 0.54
ω_1 =0.5	a 0.95	a 0.9	a 0.85	a 0.8	a 0.75	a 0.7	a 0.65	a 0.6	a 0.55
ω_2 =0.5	b 0.05	b 0.1	b 0.15	b 0.2	b 0.25	b 0.3	b 0.35	b 0.4	b 0.45
ω_1 =0.6	a 0.96	a 0.92	a 0.88	a 0.84	a 0.8	a 0.76	a 0.72	a 0.68	a 0.64
ω_2 =0.4	b 0.04	b 0.08	b 0.12	b 0.16	b 0.2	b 0.24	b 0.28	b 0.32	b 0.36
ω_1 =0.7	a 0.97	a 0.94	a 0.91	a 0.88	a 0.85	a 0.82	a 0.79	a 0.76	a 0.73
ω_2 =0.3	b 0.03	b 0.06	b 0.09	b 0.12	b 0.15	b 0.18	b 0.21	b 0.24	b 0.27
ω_1 =0.8	a 0.98	a 0.96	a 0.94	a 0.92	a 0.9	a 0.88	a 0.86	a 0.84	a 0.82
ω_2 =0.2	b 0.02	b 0.04	b 0.06	b 0.08	b 0.1	b 0.12	b 0.14	b 0.16	b 0.18
ω_1 =0.9	a 0.99	a 0.98	a 0.97	a 0.96	a 0.95	a 0.94	a 0.93	a 0.92	a 0.91
ω_2 =0.1	b 0.01	b 0.02	b 0.03	b 0.04	b 0.05	b 0.06	b 0.07	b 0.08	b 0.09

Table A-7. Similarity value for polygon "a" and "b" based on the variation in weights value and geometric similarity value. The standard value for Semantic similarity for "a" is 0.33 and "b" is 0.

	SE_a=0.33, GE_a=0.9 SE_b=0, GE_b=0.1	SE_a=0.33, GE_a=0.8 SE_b=0, GE_b=0.2	SE_a=0.33, GE_a=0.7 SE_b=0, GE_b=0.3	SE_a=0.33, GE_a=0.6 SE_b=0, GE_b=0.4	SE_a=0.33, GE_a=0.5 SE_b=0, GE_b=0.5	SE_a=0.33, GE_a=0.4 SE_b=0, GE_b=0.6	SE_a=0.33, GE_a=0.3 SE_b=0, GE_b=0.7	SE_a=0.33, GE_a=0.2 SE_b=0, GE_b=0.8	SE_a=0.33, GE_a=0.1 SE_b=0, GE_b=0.9
ω_1 =0.1	a 0.843	a 0.753	a 0.663	a 0.573	a 0.483	a 0.393	a 0.303	a 0.213	a 0.123
ω_2 =0.9	b 0.09	b 0.18	b 0.27	b 0.36	b 0.45	b 0.54	b 0.63	b 0.72	b 0.81
ω_1 =0.2	a 0.786	a 0.706	a 0.626	a 0.546	a 0.466	a 0.386	a 0.306	a 0.226	a 0.146
ω_2 =0.8	b 0.08	b 0.16	b 0.24	b 0.32	b 0.4	b 0.48	b 0.56	b 0.64	b 0.72
ω_1 =0.3	a 0.729	a 0.659	a 0.589	a 0.519	a 0.449	a 0.379	a 0.309	a 0.239	a 0.169
ω_2 =0.7	b 0.07	b 0.14	b 0.21	b 0.28	b 0.35	b 0.42	b 0.49	b 0.56	b 0.63
ω_1 =0.4	a 0.672	a 0.612	a 0.552	a 0.492	a 0.432	a 0.372	a 0.312	a 0.252	a 0.192
ω_2 =0.6	b 0.06	b 0.12	b 0.18	b 0.24	b 0.3	b 0.36	b 0.42	b 0.48	b 0.54
ω_1 =0.5	a 0.615	a 0.565	a 0.515	a 0.465	a 0.415	a 0.365	a 0.315	a 0.265	a 0.215
ω_2 =0.5	b 0.05	b 0.1	b 0.15	b 0.2	b 0.25	b 0.3	b 0.35	b 0.4	b 0.45
ω_1 =0.6	a 0.558	a 0.518	a 0.478	a 0.438	a 0.398	a 0.358	a 0.318	a 0.278	a 0.238
ω_2 =0.4	b 0.04	b 0.08	b 0.12	b 0.16	b 0.2	b 0.24	b 0.28	b 0.32	b 0.36
ω_1 =0.7	a 0.501	a 0.471	a 0.441	a 0.411	a 0.381	a 0.351	a 0.321	a 0.291	a 0.261
ω_2 =0.3	b 0.03	b 0.06	b 0.09	b 0.12	b 0.15	b 0.18	b 0.21	b 0.24	b 0.27
ω_1 =0.8	a 0.444	a 0.424	a 0.404	a 0.384	a 0.364	a 0.344	a 0.324	a 0.304	a 0.284
ω_2 =0.2	b 0.02	b 0.04	b 0.06	b 0.08	b 0.1	b 0.12	b 0.14	b 0.16	b 0.18
ω_1 =0.9	a 0.387	a 0.377	a 0.367	a 0.357	a 0.347	a 0.337	a 0.327	a 0.317	a 0.307
ω_2 =0.1	b 0.01	b 0.02	b 0.03	b 0.04	b 0.05	b 0.06	b 0.07	b 0.08	b 0.09

A.3 Error matrix for Generalization Modelled Output Maps.

Table A-8. Error matrix for 1:25,000 scale map with Level-III classification.

	Barren /Rocky	Commercial	Cropland	Dense Forest	Industrial	Lakes/ Ponds	Land with /without scrub	Mixed Built Up area	Open Forest	Plantation	Plantation/ Orchards	Public and Semipublic	Public utility and facility	Reclaimed land vacant land	Recreational	Residential	River/Streams	Transportation		User's accuracy (%)
Barren /Rocky	25	0																	25	100.00
Commercial		22						2			0					1			25	88.00
Cropland			21	0	0		1				0			1	0	1	0	0	24	87.50
Dense Forest				25															25	100.00
Industrial					24											1			25	96.00
Lakes/ Ponds						23													23	100.00
Land with /without scrub							24					1							25	96.00
Mixed Built Up area		2						23											25	92.00
Open Forest	0						0		24		0								24	100.00
Plantation										25									25	100.00
Plantation/ Orchards									1		24	1							26	92.31
Public and Semipublic		0	0	0		1			0	0	0	25							26	96.15
Public utility and facility													25						25	100.00
Reclaimed land vacant land			1											24			0		25	96.00
Recreational										1					19	2			22	86.36
Residential		0	1			1	0		1	1	0	0	0	0	3	23	1	0	31	74.19
River/Streams	0															0	24		24	100.00
Transportation																		25	25	100.00
	25	24	23	25	24	25	25	25	26	27	24	27	25	25	22	28	25	25	425	94.70
Producer's accurecy (%)	100.00	91.67	91.30	100.00	100.00	92.00	96.00	92.00	92.31	92.59	100.00	92.59	100.00	96.00	86.36	82.14	96.00	100.00	94.50	

Table A-9. Error matrix for 1:25,000 scale map with Level-II classification.

	Barren /Rocky	Built Up (urban)	Cropland	Dense Forest	Lakes/ Ponds	Land with /without scrub	Open Forest	Plantation	Plantation/ Orchards	River/Streams		User's accuracy (%)
Barren /Rocky	25										25	100.00
Built Up (urban)		220	2	0	2	0	1	2	2	0	229	96.07
Cropland		2	21	0		1			0	0	24	87.50
Dense Forest				25							25	100.00
Lakes/ Ponds					23						23	100.00
Land with /without scrub		1				24					25	96.00
Open Forest	0					0	24		0		24	100.00
Plantation								25			25	100.00
Plantation/ Orchards		1					1		24		26	92.31
River/Streams	0	0								24	24	100.00
	25	224	23	25	25	25	26	27	26	24	435	88.35
Producer's accuracy (%)	100.00	98.21	91.30	100.00	92.00	96.00	92.31	92.59	92.31	100.00	86.79	

Table A-10. Error matrix for 1:50,000 scale map with Level-II classification

	Barren /Rocky	Built Up (urban)	Cropland	Dense Forest	Lakes/ Ponds	Land with /without scrub	Open Forest	Plantation	Plantation / Orchards	River/Streams	Total	User's accuracy (%)
Barren /Rocky	25						2				27	92.59
Built Up (urban)		220	2	0	8	3	1	11	2	3	250	88.00
Cropland		2	21	0		1	1		0	1	26	80.77
Dense Forest			0	25					0		25	100.00
Lakes/ Ponds					17						17	100.00
Land with /without scrub		2				21				0	23	91.30
Open Forest			0				21				21	100.00
Plantation							0	16			16	100.00
Plantation/ Orchards		1	0				1		22		24	91.67
River/Streams	0	0	0			0				21	21	100.00
Total	25	225	23	25	25	25	26	27	24	25	409	94.43
Producer's accuracy (%)	100.00	97.78	91.30	100.00	68.00	84.00	80.77	59.26	91.67	84.00	85.68	

A.4 Python code for Elimination.

```
def sa(x,y):
    code= {'Built Up': '010000',
            'Built Up (Urban)':'010100' ,
                'Residential': '010101',
                'Industrial': '010102',
                'Mixed Built Up area' :'010103',
                'Recreational':'010104',
                'Public and Semipublic': '010105',
                'Communication' :'010106',
                'Public utility and facility':'010107',
                'Commercial': '010108',
                'Transportation': '010109',
                'Reclaimed land vacant land' :'010110',
                'Vegetation area Trees' :'010111',
            'Built Up (Rural)': '010200',
        'Agriculture' :'020000',
            'Cropland' :'020100',
            'Fallow land': '020200',
            'Plantation/ Orchards' :'0020300',
        'Forest' :'030000' ,
            'Dense Forest' :'030100'            ,
            'Open Forest' :'030200',
            'Plantation' :'030300',
            'Mangroves': '030400',
        'Grazing land Wastelands': '040000',
            'Salt-Affected' :'040100',
            'Gullied /Ravinous' :'040200',
            'Land with /without scrub': '040300',
            'Barren /Rocky' :'040400',
            'Sandy area': '040500',
        'Wetlands' :'050000',
            'Marshy /Swampy': '050100',
            'Mudflats': '050200',
            'Waterlogged Salt pans': '050300',
        'Water bodies' :'060000',
```

```
        'River/Streams' :'060100',
        'Canal': '060200',
        'Lakes/ Ponds' :'060300'          ,
        'Reservoirs' :'060400',
        'Tanks' :'060500',
        'Cooling Pond/Cooling Reservoir': '060600',
        'Abandoned quarries with water' :'060700',
      'Others' :'070000',
        'Quarry / Brick Kilns': '070100',
        'Dam / Barrage' :'070200',
        'Coral reef / Atoll': '070300'     }
  a= code[x]
  b= code[y]
  if a==b:
    se=2
  elif a[:4]==b[:4]:
    se=1
  elif a[:2]==b[:2]:
    se=0.33
  else :
    se=0
  return se
def S(l1,l2,s1,s2,g1,g2):
  print s1,s2
  if s1==1 and s2==0:
    w1=.2
    w2=.8
  elif s1==0 and s2==1:
    w1=.2
    w2=.8
  elif s1==1 and s2==0.33:
    w1=.3
    w2=.7
  elif s1==0.33 and s2==1:
    w1=.3
    w2=.7
  else:
```

```
        w1=.5
        w2=.5
    v1=(w1*s1) + (w2*g1)
    v2=(w1*s2) + (w2*g2)
    print w1,w2
    print l1,s1,g1,v1
    print l2,s2,g2,v2
    if v1>v2:
        return l1
    else :
        return l2


from dbfpy import dbf

db = dbf.Dbf("E:/Thesis/Sequence testing/50k/8sre/10kmain_d.dbf")

db2= dbf.Dbf("E:/Thesis/Sequence testing/50k/8sre/shared boundary.dbf")
limit=22500
count = 0
for i in db:
    rec = db[count]
    if i[3]<limit:

        print count , i[0]
##        print "XXXX"
##        print count
        l1=[]
        l2=[]
        l3=[]
        for j in db2:
            if j[0]==count:
                if j[6]!=i[0]:
                    #print
                    SA= sa(j[1],j[6])
                    ge=j[3]/i[2]
                    #print j
```

```
        #print j[0],j[6],j[3],i[2], ge,SA
        l1.append(j[6])
        l2.append(SA)
        l3.append(ge)
        #print
print l1
print l2
print l3
m1 = max(l3)
print m1
no1=[k for k, z in enumerate(l3) if z == m1]
print no1
ll=[]
ls=[]
lg=[]
ll.append(l1[int(no1[0])])
ls.append(l2[int(no1[0])])
lg.append(l3[int(no1[0])])
l1.pop(int(no1[0]))
l2.pop(int(no1[0]))
l3.pop(int(no1[0]))
print ll,ls,lg
if len(l1)!=0:
    m1 = max(l3)
    no1=[k for k, z in enumerate(l3) if z == m1]

    ll.append(l1[int(no1[0])])
    ls.append(l2[int(no1[0])])
    lg.append(l3[int(no1[0])])
    l1.pop(int(no1[0]))
    l2.pop(int(no1[0]))
    l3.pop(int(no1[0]))

print ll
print ls
print lg
```

```
    if len(ll)>1:
        cou=0
        for o in l1:
            if o=="River/Streams":
                ll.pop(cou)
                ls.pop(cou)
                lg.pop(cou)
            cou=cou+1



    if len(ll)>1:
        lu=S(ll[0],ll[1],ls[0],ls[1],lg[0],lg[1])
    else :
        lu=ll[0]
    print lu

    print "XXXX"
    #print i
    #rec = db[count]
    i["LULC"] = lu
    i.store()
    #del rec
    print
    print
    print
count=count+1
```

A.5 Python code for Reclassify.

```
def reclassify(x):
    code= {
                'Residential': 'Built Up (Urban)',
                'Industrial': 'Built Up (Urban)',
                'Mixed Built Up area' :'Built Up (Urban)',
                'Recreational':'Built Up (Urban)',
                'Public and Semipublic': 'Built Up (Urban)',
                'Communication' :'Built Up (Urban)',
                'Public utility and facility':'Built Up (Urban)',
                'Commercial': 'Built Up (Urban)',
                'Transportation': 'Built Up (Urban)',
                'Reclaimed land vacant land' :'Built Up (Urban)',
                'Vegetation area Trees' :'Built Up (Urban)',
            }
    a= code[x]
    return a
l=['Residential',
                'Industrial',
                'Mixed Built Up area' ,
                'Recreational',
                'Public and Semipublic',
                'Communication' ,
                'Public utility and facility',
                'Commercial',
                'Transportation',
                'Reclaimed land vacant land' ,
                'Vegetation area Trees' ]
from dbfpy import dbf
db = dbf.Dbf("E:/Thesis/Sequence testing/10kmainCopy.dbf")
for i in db:
    if i[0] in l:
        lu=reclassify(i[0])
        i["LULC"] = lu
        i.store()
```