

GEORG-AUGUST UNIVERSITÄT GÖTTINGEN

FACULTY OF FOREST SCIENCES AND FOREST ECOLOGY

- DEPARTMENT OF FOREST INVENTORY AND REMOTE SENSING -



**DEVELOPMENT OF EFFICIENT FOREST INVENTORY TECHNIQUES
FOR FOREST RESOURCE ASSESSMENT IN SOUTH KOREA**

JONG-SU YIM

GÖTTINGEN, 2009

**DEVELOPMENT OF EFFICIENT FOREST INVENTORY TECHNIQUES
FOR FOREST RESOURCE ASSESSMENT IN SOUTH KOREA**

A dissertation to obtain the degree of Doctor
at the Faculty of Forest Sciences and Forest Ecology of
Georg-August University at Göttingen

By

Jong-Su Yim

born in Hwa-Seong, REPUBLIC OF KOREA

GÖTTINGEN, 2009

D7

Advisor: Prof. Dr. Christoph Kleinn

Referee: Prof. Dr. Joachim Saborowski

Date of oral examination: 12. 12. 2008

Diese Arbeit ist durch die Niedersächsische Staats- und Universitätsbibliothek, SUB-Göttingen, unter folgender Internetadresse elektronisch veröffentlicht:

<http://webdoc.sub.gwdg.de/diss/2009/yim/>

ACKNOWLEDGEMENTS

This work was funded by the German Research Foundation (DFG) and the Korea Science and Engineering Foundation (KOSEF). I thank for their financial support.

I would like to express my sincere thanks to my supervisor, Prof. Dr. Christoph Kleinn, who provided the opportunity for me to work at the University of Göttingen and obtain the financial support from the DFG. His personal and scientific advice has been invaluable. My special thanks go to Prof. Dr. Klaus von Gadow and Prof. Dr. Joachim Saborowski for their superior lectures in forest development and forest biometric information which were very helpful throughout my studies.

I thank to the my fellows, Dr. Lutz Fehrmann, Stefan Kunth, Dr. Hans Fuchs, Axel Buschmann, Paul Magdon for R-Programming, Haijun Yang, Hamid Reza Riyahi, and Dr. František Vilčko, for their valuable comments and assistances to my studies and personal matters, and my colleagues, Ulrike Docketor, Hendrik Heydecke, Reinhard Schlote, and Margret Krüger for technical and administrative matters in the Department of Forest Inventory and Remote Sensing at the University of Göttingen.

I am also grateful to Prof. Dr. Man-Yong Shin, dean of the College of Forest Science at the Kook-Min University, for his support and help. I also thank the researchers in the Division of Forest Resource Information, Korea Forest Research Institute, for their support and valuable comments, and my Korean fellows for assistance with field work. I also wish to thank Prof. Dr. Erkki Tomppo and his colleagues in Finland, who have given me help and advice to my studies.

Finally, I wish to express my gratitude to my wife, Song-Hui Lee, my daughter, Suh-Yoon, and my family in South Korea for their support, patience, and encouragement.

Sincerely thank to all,

Jong-Su Yim

LIST OF CONTENTS

List of Contents..... i

List of Tables.....iv

List of Figures v

1. INTRODUCTION 1

1.1 Forests in South Korea..... 1

1.2 National forest inventory..... 3

1.3 Satellite remote sensing 5

1.4 Problem statement and Research questions..... 7

1.5 Objectives..... 8

2. MAPPING OF FOREST COVER TYPE 9

2.1 Introduction..... 9

2.2 Materials and Methods 12

2.2.1 Study area 12

2.2.2 Field data 13

2.2.3 Satellite data 15

2.2.4 Map data 15

2.2.5 Topographic correction..... 16

2.2.6 Classification 17

2.2.7 Evaluation 19

2.3 Results..... 20

2.3.1 Topographic correction..... 20

2.3.2 Classification 22

2.3.3 Evaluation 23

2.6 Discussion and Conclusion 27

3. ESTIMATION OF FOREST ATTRIBUTES.....	31
3.1 Introduction.....	31
3.2 Basics of the <i>k</i> -Nearest Neighbor technique	32
3.2.1 General information.....	32
3.2.2 Characteristics	33
3.3 Materials and Methods	37
3.3.1 Study area	37
3.3.2 Field data	38
3.3.3 Satellite data	40
3.3.4 Map data	41
3.3.5 Application of the <i>k</i> -NN technique	41
3.3.6 Evaluation	44
3.4 Results.....	45
3.4.1 Satellite images.....	45
3.4.2 Characteristics in the <i>k</i> -NN process.....	47
3.4.3 Map production	54
3.5 Discussion and Conclusion	57
4. PLOT DESIGN OPTIMIZATION.....	61
4.1 Introduction.....	61
4.2 Materials and Methods	64
4.2.1 Sample cluster unit.....	64
4.2.2 Cluster configuration	65
4.2.3 Statistical Analysis.....	66
4.2.4 Cost analysis.....	69
4.3 Results.....	71
4.3.1 Statistical characteristics	71
4.3.2 Cost analysis.....	74
4.4 Discussion and Conclusion	77

5. SAMPLING DESIGN OPTIMIZATION	81
5.1 Introduction.....	81
5.2 Materials and Methods	83
5.2.1 Artificial forest population	83
5.2.2 Simulation of sampling designs.....	83
5.2.3 Comparison.....	87
5.3 Results.....	89
5.3.1 Artificial forest population	89
5.3.2 Sampling simulation	91
5.4 Discussion and Conclusion	95
6. OVERALL CONCLUSION	98
7. SUMMARY	99
8. REFERENCES	103
9. ANNEXES	111

List of Tables

Table 1. 1: History of the National Forest Inventory in South Korea	3
Table 1. 2: Examples of forest definition in some countries and FAO.....	4
Table 1. 3: Examples of current earth observation satellites by spatial resolution.....	5
Table 2. 1: Technical information for the satellite image used.....	15
Table 2. 2: Distribution of field points for each stratum	17
Table 2. 3: Estimated values of the Minnaert constant for each band	20
Table 2. 4: Error matrices for assessing the classification accuracy of the digital forest map and the classified images for both classifiers with field plot data.....	23
Table 2. 5: Error matrix for assessing the classification accuracy of the NNC classified image and digital forest map over the entire test area per pixel	25
Table 2. 6: The result of the chi-square test for field plot data with digital map and both classified images.....	25
Table 3. 1: Distribution of land use/cover classes for the study area (KFS, 2004a)	37
Table 3. 2: Summary statistics based on field plot data ($n=191$ sub-plots).....	39
Table 3. 3: Technical information of the satellite images.....	40
Table 3. 4: Estimated values of the Minnaert constant for the satellite images	45
Table 3. 5: Correlation coefficients between DN _s on the topographically normalized images and the growing stock based on field plot data ($p_{0.05} = 0.138$, $p_{0.01} = 0.181$).....	45
Table 3. 6: The radii of the horizontal reference areas and the minimum, maximum and mean number of the field plots as a reference plot.....	47
Table 3. 7: Weighting parameters for each band and reference window	53
Table 3. 8: Error matrix for assessing the classification accuracy of the NNC classified image and field plot data.....	54
Table 3. 9: Characteristics for growing stock and forest maps in the k -NN process.....	55
Table 4. 1: Variance table for cluster sampling.....	68
Table 4. 2: Summary of the time study for the pilot inventory	74
Table 5. 1: Summary of characteristics for different sampling designs	84
Table 5. 2: Mean estimators for the simulated systematic sampling designs	86
Table 5. 3: Summary statistics of growing stock (m^3) per pixel unit (0.0625 ha) for each stratum for the given artificial forest population	89
Table 5. 4: Summary of estimations by simulation for different sample sizes under stratified random sampling and simple random sampling	91
Table 5. 5: Summary of estimations by sample size for different sampling designs	93

List of Figures

Figure 1. 1: Distribution of forest proportion (left) and forest cover types (right) in South Korea. The forest cover types are divided into three types; coniferous forest (C), deciduous forest (H), and mixed forest (M) (KFS, 2004a).....	2
Figure 2. 1: Location of the study area (left) and distribution of field sample points for the 3 rd NFI on the Digital Elevation Model (right).	12
Figure 2. 2: Plot design for the 3 rd and 4 th NFI cycles in South Korea; each cluster plot consists of 4 sub-plots (right).	13
Figure 2. 3: Geo-referencing of field sample plots; coordinates of the permanent sub-plots measured on 1:50,000 topographic maps (a) and coordinates of those on 1:25,000 DEM data (b).....	14
Figure 2. 4: Representation of the incident solar angle (i) and the solar zenith angle (θ_0), where L_T is the observed value and L_H is the normalized value.	17
Figure 2. 5: Comparison of the minimum distance (left) and the nearest neighbor (right) classifiers: for instance, an un-classified pixel (+) belongs to class B by the NNC.....	18
Figure 2. 6: Comparison between raw (a) and topographically normalized (b) images (Landsat TM 4:3:2).	21
Figure 2. 7: The mean digital numbers for the different forest classes and bands (C: coniferous forest H: deciduous forest and M: mixed forest).	22
Figure 2. 8: Distribution of digital numbers by forest classes in bands 4 and 5.	22
Figure 2. 9: Comparison of the forest cover maps for different approaches in the study area ...	26
Figure 3. 1: Research flowchart for the k -NN technique.	33
Figure 3. 2: Location of the test area and distribution of field plots for the study area in the current NFI cycle (2006-2010), where the samples are established with a systematic square grid of 4km.....	38
Figure 3. 3: Field points on the DEM and the two cluster plot designs used.	39
Figure 3. 4: Original digital number (left) and mean digital number for the 3 x 3 window filtering (right) for each sub-plot centre within the plot design defined using Landsat ETM+.	43
Figure 3. 5: Comparison of relative RMSE for different images and different numbers of neighbors (k).	46
Figure 3. 6: Relative RMSE and bias for the horizontal reference areas and different numbers of neighbors (k).	48
Figure 3. 7: Distribution of the number of field plots for the altitude classes.....	49
Figure 3. 8: Estimated mean (a), RMSE (b), RMSE% (c), and bias (d) for the HRA-10km and stratifications by the VRA and forest cover types and different numbers of neighbors (k), where the “true” mean value was computed based on field plot data.	50

Figure 3. 9: Relative RMSE for the different neighbor weighting functions (a) and image filtering technique (b) with the HRA-10km and Landsat ETM+.	51
Figure 3. 10: Overall accuracy for different reference windows and different numbers of neighbors (k).	53
Figure 3. 11: Continuous thematic map of growing stock (a), forest and non-forest map by the NNC as mask map (b), and growing stock map within the forest (c) over the study area, which can be used as an artificial model forest in chapter 5.	56
Figure 3. 12: Distribution of frequencies for the growing stock classes and different numbers of neighbors with the HRA-10km.	59
Figure 4. 1: Pilot cluster plot design as used in this study comprising 10 circular sub-plots. Various standard cluster shapes can directly be formed from this design.	65
Figure 4. 2: Cluster configurations for different cluster sizes in this study.	66
Figure 4. 3: Distance distribution between pairs of points for the 25 clusters in the pilot study.	67
Figure 4. 4: Examples: walking distances for different travel routes and cluster configurations; the dotted lines are the return distances.	70
Figure 4. 5: Covariance functions for the key attributes by distance between pairs of sub-plots based on the given pilot clusters.	71
Figure 4. 6: Intra-cluster correlation coefficients for different key attributes and different cluster configurations (see Figure 4.2).	72
Figure 4. 7: Relative standard errors for each attribute by cluster configuration.	73
Figure 4. 8: Total walking distances for the two travel routes and different cluster configurations.	75
Figure 4. 9: Total times for the two travel routes and different cluster configurations.	75
Figure 4. 10: Example: forest strata per sub-plot within a cluster (C: coniferous forest, H: broad-leaved forest, and M: mixed forest).	78
Figure 5. 1: Thematic maps used as an artificial forest population in this study: forest cover types (top) and growing stock per hectare within the forest (bottom).	90
Figure 5. 2: Relative efficiency for sample sizes; sampling designs: SRS (simple random sampling), STR (stratified random sampling), SYS (systematic sampling), sys+pre (stratified systematic sampling), sys+post (systematic sampling with post-stratification), and sys+clu (systematic cluster sampling).	94

Abbreviations

BMVEL	Federal Ministry of Consumer Protection, Food, and Agriculture in Germany (Bundesministerium für Verbraucherschutz, Ernährung und Landwirtschaft)
CATIE	Tropical Agricultural Research and Higher Education Center
DEM	Digital Elevation Model
DTM	Digital Terrain Model
EC	European Commission
FAO	Food and Agriculture Organization of the United Nations
FGIS	Forest Geographic Information System
FRA	The FAO-led Global Forest Resources Assessment
GIS	Geographic Information System
GPS	Global Positioning System
KFRI	Korea Forest Research Institute
KFS	Korea Forest Service
NDVI	Normalized Difference Vegetation Index
NFI	National Forest Inventory
NGIS	National Geographic Information System
NWGS	Non Wood Goods and Services
PCA	Principle Component Analysis
SFM	Sustainable Forest Management
USDA	United States Department of Agriculture
VI	Vegetation Index

1. INTRODUCTION

1.1 FORESTS IN SOUTH KOREA

Forests in South Korea were mostly destroyed by over-cutting and illegal harvesting during the Colonial Period (1910-1945) and the Korean War (1950-1953). However, since the early 1960s, an organized investment has been made in the forestry sector as part of a National Economic Development Plan.

Additionally tree planting, forest protection, and nature conservation practices of the “*New Villages Movement*” played a major role in the restoration of forest land (FAO, 1997). Currently, the forest covers an area of about 64,063km², representing about 64% of the total land area (about 99,600km²), whereas the area for farming represents 20% and other uses total 16%. In recent years, the forest area has been decreasing with an average annual reduction of about 72 km² (KFS, 2004a).

Forest cover maps from interpretation of aerial photographs show that coniferous forests account for 41.9% of the total forest, while deciduous, mixed, and other forests comprise 28.9%, 25.5%, and 3.7%, respectively (Figure 1.1). In natural forests, however, the area of coniferous forest (mainly, *Pinus densiflora*) has been continuously diminishing, while that of deciduous forest has been increasing over time through ecological succession (Chung, 1996; Lee *et al.*, 2004b). The main tree species are Japanese red pine (*Pinus densiflora*), Korean white pine (*Pinus koraiensis*), Japanese larch (*Larix leptolepis*), and oak species (*Quercus mongolica*, *Quercus variables*, and *Quercus acutissima*) (KFS, 2004a).

Most of the forested areas were planted after the heavy forest depletion that followed the Korean War. Thus, almost 60% of the forest stands are less than 40 years old and the total growing stock is merely 468 million m³, with growing stock volume per hectare estimated to be approximately 73 m³/ha. The density of forest roads is less than 4 m/ha, although these roads are crucial for forest management as well as for reaching field sample plots for forest inventories (KFS, 2004a).

INTRODUCTION

The topography of Korea is dominated by low hills in the southern and the western regions, which grow gradually larger as one progresses toward the eastern and northern regions. On the whole, the western and southern slopes of the Korean Peninsula are wide, with some plains and basins along the rivers, while the eastern slope is very steep with high mountains that border the East Sea. The mean altitude is 482 meters above sea level, and forests over 1,000 m are mostly in the northeastern districts.

According to a forest soil report (KFRI, 2004), the areas suitable for intensive forest management are those with slopes up to 25 degrees (or 47%), which cover about 40% of the total forest area, whereas forest areas with a slope of greater than 31 degrees (or 60%) account for about 31% where it is difficult to practice silviculture.

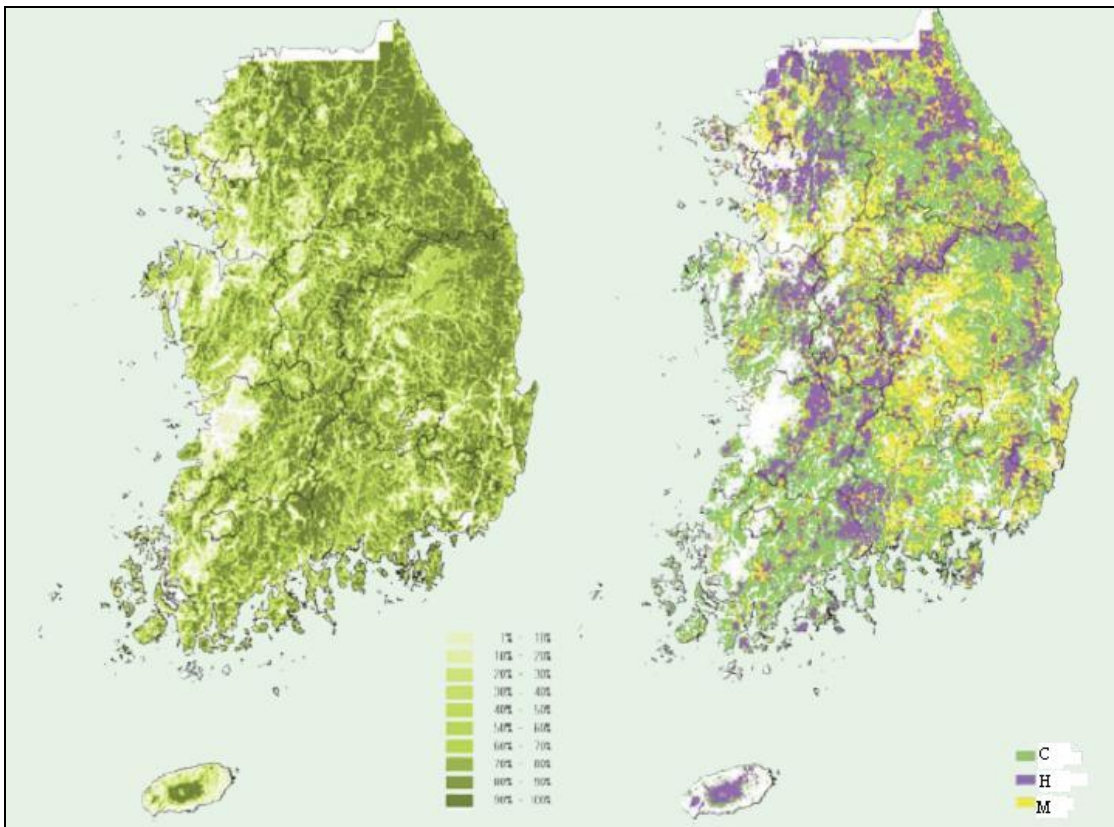


Figure 1. 1: Distribution of forest (left) and forest cover type (right) in South Korea (KFS, 2004a). The forest cover type is divided into three types; coniferous forest (C), deciduous forest (H), and mixed forest (M).

1.2 NATIONAL FOREST INVENTORY

The principal goal of National Forest Inventories (NFIs) is to produce statistically sound and reliable forest resource information of an entire country, which is relevant and required for forest policy decision-making and monitoring at the national level and forestry monitoring and related sectors such as environmental studies, land use management, biodiversity, and assessment of environment impact (KFS, 2004b).

There is a NFI system in South Korea, which has been developed in the 1960s and first implemented in 1972 (Table 1.1). Currently, the NFI is in its fifth cycle and has been reorganized and expanded to support sustainable forest management planning and to provide basic data and information for international processes and conventions such as Montreal Process, Conservation on Biological Diversity, and the UN Framework Convention on Climate Change (KFS, 2002).

Table 1. 1: History of the National Forest Inventory in South Korea

Survey period	Source	Content
1917-1961	Forest management document	
1961-1964	Field survey	Forest status inventory
1972-1975		The 1 st NFI, entire country
1978-1980		The 2 nd NFI, entire country
1986-1992	Aerial photograph + Field survey	The 3 rd NFI, national annually inventory
1996-2005		The 4 th NFI, national annually inventory
2006-2010		The 5 th NFI, national annually inventory

In the fourth NFI cycle, the entire country was divided into 10 sub-areas based on districts that comprise 9 provinces. The NFI was carried out in one sub-area per year over a 10-year period and the KFRI published a report for the implemented province annually. The main goal of the NFI was to provide information for the reforestation plan and to support in decision-making for the forest policy at a specific point in time (KFRI, 1996). Since the NFI was implemented through a rotation system by province, it was hardly possible to provide current information for the entire country at the same time. Critical information included only timber-oriented attributes: area and growing stock by forest cover types, age classes, dominant tree species, *etc.* for forest

INTRODUCTION

conditions, ownerships, and administrative units.

The NFI also provides data for sub-national geographical or administrative units and is a basis for global forest assessments and other international processes in the context of sustainable management of natural resources. With the increasing use of the forest resource for purposes other than timber production, the scope of the NFI has been expanded to provide more information about forest resources, including biodiversity, regeneration, soil, *etc.* (KFRI, 2006).

In the Korean Land Management Act of Forest, the forest is defined, for example, as collectively standing and growing trees and bamboos, and the land that contains them. Beyond those definitions, the forest is also defined according to land use. However, it is impossible to completely adapt this definition to the Korean NFI, as some ambiguities are generally involved. Whenever large area forest inventories are carried out, a clear and quantitative definition of the forest is required. The definition of that can vary with countries and purposes (FAO, 2004; EC, 1997; Lund, 2007). In the Korean NFI, its definition has changed to correspond to the FAO's definition. Its components are area, proportion of crown cover, tree height, and width (Table 1.2).

Table 1. 2: Examples of forest definition in some countries and the FAO.

Countries & FAO	Minimum threshold values				
	Area (ha)	Crown Cover (%)	Tree Height (m)	Width (m)	Number of trees per ha
Germany	0.1	50		10	
United States	0.4	10	4	36	
FRA 2000	0.5	10	5	20	
Korea (KFRI, 1996)	1	30			1,200 or 1,600 (dbh≥6 cm)
Korea (KFRI, 2006)	0.5	10	5	30	

Source: <http://home.comcast.net/~gyde/DEFpaper.htm#forest> (09/2007)

1.3 SATELLITE REMOTE SENSING

Forests cover large areas and are mostly located in mountain ranges and/or in out-of-the-way areas. Due to the high variability and complexity of the resources, the information about forest resources is extensive. Moreover, it is required that the information should be updated from time to time because of the changes in forest resources over time. In order to address these issues, satellite remote sensing that allows for frequent measurement and monitoring at lower cost has been widely used in the forestry sector.

In general, satellite-based remotely sensed data can be distinguished into *Low*-, *Medium*- and *High*-resolution satellite (Table 1.3). *Low-resolution satellite data* have proved useful for a continental view, generating small-scale maps of the forested landscape and detecting hot spots of severe deforestation within densely forested landscapes. *Medium-resolution satellite data* are relatively inexpensive, and are also suitable for larger areas. Much research has been performed, particularly for Landsat imagery, where customized approaches are available. *High-resolution satellite data* are suitable for interpreting and analyzing smaller sites. In particular, South Korea has launched KOMPSAT-1 and KOMPSAT-2.

Table 1. 3: Examples of current earth observation satellites by spatial resolution

Classification	Satellite	Launched	Bands	Spatial resolution (m)	Swath (km)
Low Resolution	AVHRR-3	05/1998	6	1,090	2,700
	MODIS	12/1999	2 / 5 / 29	250/500/1000	2,330
	OrbView-2	07/1997	8	1,000	2,800
Medium resolution	Landsat-5	03/1984	6 / 1	30/120	185
	Landsat-7	04/1999	1 / 6 / 1	15/30/60	185
	SPOT-4	03/1998	1 / 4	10/20	60
	ASTER	12/1999	4 / 6 / 5	15/30/90	60
High Resolution	IKONOS-2	09/1999	1 / 4	1/4	11
	Quickbird-2	10/2001	1 / 4	0.6/2.5	16
	SPOT -5	05/2002	1 / 4	2.5 or 5 /10	120
	KOMPSAT-2	07/2006	1 / 4	1/4	15

<http://www.asprs.org/news/satellites/index.html> (Stoney, 2006)

INTRODUCTION

For forest inventory applications, satellite data may be used in principle (i) for the entire population of interest (the entire country and its forests) or (ii) for sample regions only, where the sample regions (i.e., the “remote sensing imagery sample plots”) can be placed around the field sample plots. In the first case, *low-* or *medium-* resolution satellite imagery is commonly applied; with *medium-* resolution imagery, one can use multivariate statistical approaches to produce predictions of specific forest characteristics at any forest location. These approaches lead to a regionalization of estimations. Forest cover maps can then be produced for the entire inventory region. In the latter case, the imagery is essentially utilized to improve the precision of area estimation of forest and forest cover types.

From the early 1990s, *low-* and *medium* spatial resolution satellite image data have been employed to produce land cover/use type map by the KFRI and some other studies (Kim *et al.*, 1989; Kim, 1991; Rho and Lee, 1995; Lee *et al.*, 1994). Currently, the research on satellite remote sensing is increasing, particularly for the monitoring of the land/forest cover change over the Korean Peninsula including North Korea (Lee, 1994; Lee *et al.*, 1998; Kim and Park, 2000), as well as forest health and forest fire monitoring (Kim *et al.*, 2003). Furthermore, with the enhancement of remote sensors, the use of satellite imagery having *high* spatial resolutions is increasing to identify dominant tree species at a local or landscape unit (Lee and Kim, 2000; Cho, 2002; Chung *et al.*, 2001).

The implementation of remote sensing techniques is closely related to Geographic Information System (GIS) applications. For a variety of applications, remote sensing, as one source of potential input to a GIS, is very valuable. It represents a powerful technology for providing input data for measurement, mapping, monitoring, and modeling within a GIS context (Wilkinson, 1996). Since 1995, as part of the NGIS (National Geographic Information System) in South Korea, a Forest Geographic Information System (FGIS) has been implemented in the forestry sector. The principal objectives of the FGIS are digital mapping of thematic maps related to forestry sectors and developing application for forest and natural resources management. In recent years, several digital thematic maps relevant and related to the forest resource, including forest cover types, land use classes, forest soil types, and forest roads over the whole country have been digitalized for application in the FGIS (Kim, 2004; KFRI, 2004).

1.4 PROBLEM STATEMENT AND RESEARCH QUESTIONS

In South Korea, the NFI was implemented at a 10-year interval by province in its fourth cycle (1996-2005). The main goal of this NFI was to provide fundamental information for the reforestation plan and to support in decision-making for forest management and policy at the national level. In order to achieve these goals, forest inventory techniques were developed based on both field sampling and aerial photo interpretation. This system, however, leaves room for optimization as follows:

- i. Despite increasing information on forest resources to support sustainable forest management planning, the goal of this NFI was mainly to provide information on the timber-production function of forest resources.
- ii. Despite the development of remote sensors and their extended application techniques, there has so far not been an attempt to fully integrate satellite remote sensing into the NFI system.
- iii. The NFI provides information for national and regional geographical or political units, which is relevant and required for national-level decision-making and monitoring. Since the information for smaller area units is available only as mean values from this NFI, it is difficult to support management planning at a small-area unit.
- iv. The cluster plot was designed completely on empirical findings rather than scientific grounds. Thus, the most efficient cluster plot design for the Korean forest conditions is needed, and must be designed based on both statistical soundness and cost-effectiveness.
- v. Since this NFI was implemented using a rotation system, it is hardly possible to provide reliable information over the entire country at the same time.

In order to solve the first point on the list, the fifth NFI (2006-2010) has been expanded to provide more information about natural resources in the forest (KFRI, 2006). Therefore, this thesis addresses the other points that focus on the integration of satellite remotely sensed data with forest inventory data, and sampling and plot design optimization.

1.5 OBJECTIVES

The overall objective of this thesis is to develop efficient inventory techniques for large area forest resource assessment in South Korea, which essentially address the application of satellite remote sensing and sampling issues for field surveys, as well as their efficient integration and combination.

This entailed designing efficient forest inventory techniques, which allow the collection of the required forest resource information within a specific time and budget framework. In this thesis, the specific biophysical conditions of South Korea are taken into consideration. In order to address the overall objective, the technical aspects are organized into the following four chapters:

- I. Mapping of forest cover types: Forest cover maps are a key product of the NFI system in South Korea. The objective of this chapter is to delineate forest cover types by integrating forest inventory data and digital satellite imagery.
- II. Estimation of forest attributes: The main aim of this chapter is to estimate key forest attributes over a test area by integrating forest inventory data and digital satellite imagery using the *k*-Nearest Neighbor technique.
- III. Plot design optimization: This chapter is aimed at determining the most efficient cluster plot as sampling unit design considering both statistical soundness and cost-effectiveness for forest resource assessment in the Korean forest.
- IV. Sampling design optimization: By applying a simulation study using an artificial forest population, various sampling design options shall be simulated and compared to determine the most efficient sampling design, while integrating findings and results from the technical objectives (I), (II), and (III).

2. MAPPING OF FOREST COVER TYPE

2.1 INTRODUCTION

Forest cover maps are an important element in both forest resource management and scientific research. As different forest types are associated with different economic and environmental values, there is a need for detailed maps that provide the current status of forest types. Therefore, these maps provide an important baseline for forest managers and other policy makers.

Forest cover maps are a key product of the Korean NFI system. Since the early 1970s, aerial photographs have been used to identify forest cover types in the Korean NFI system. However, the use of aerial photography is waning because interpretation and processing is laborious and aerial photographs are often out-of-date (Kim *et al.*, 1989). Moreover, the forest cover classification using aerial photos has traditionally relied on subjective decisions through a visual interpretation process. This process is, therefore, neither “*transparent*” nor “*reproducible*” (Drăgut and Blaschke, 2006).

Since the launch of the Landsat-1, satellite remotely sensed data have been widely used to enhance natural resources information and to detect their change over time. Digital satellite data acquired from sensors with different characteristics (e.g., spatial, temporal, and spectral resolutions) have been used for land cover classification. The ability to repeatedly obtain digital satellite imagery, continuity of the obtained images, and their wide availability are some characteristics of satellite imagery that have contributed to the current development of remote sensing, image processing, and GIS technologies (Holmgren and Thuresson, 1998; Kleinn, 2002).

Besides, with the development of Global Positioning System (GPS), accurate information on the location of each field point can be acquired. Since this development, supervised classifiers have been widely proliferated for land cover/use classification by combining digital satellite data and ground truth data. Within forestry, GPS-based field plot data from forest inventories can serve as training data for forest cover type mapping (Tokola *et al.*, 1996; Haapanen *et al.*, 2004). There are numerous supervised classifiers, all of which can be grouped in a number of ways. Franklin *et al.* (2003) divided them into four groups: parametric, non-parametric, image segmentation, and spectral-temporal classification.

MAPPING OF FOREST COVER TYPE

In image classification process, while parametric classifiers are done dependent upon some assumptions that input data are drawn from a given probability distribution; non-parametric classifiers do not rely on the assumptions. Much research has indicated that non-parametric classifiers may provide better classification results than parametric classifiers, mostly Maximum Likelihood Classifier (MLC), for complex landscapes (Hardin, 1994; Paola and Schowengerdt, 1995).

The Nearest Neighbor Classifier (NNC), which represents one of non-parametric classifiers, has been mostly used for land cover classification due to its several advantages based both statistical and practical grounds (Zhang *et al.*, 2006):

- (i) it is simple and easy to categorize new observations based on distances in feature spaces (e.g., as high or infinitely dimensional);
- (ii) it is feasible to categorize a large number of classes that occur within the characteristic of interest; and
- (iii) from a statistical point of view, the error rate of the NNC is guaranteed to approach Bayes rule, as the sample size approaches infinity.

Ince (1987) found that the NNC produced higher accuracies than the MLC and was more robust. Hardin (1994) compared the performance of parametric and non-parametric classifiers, particularly the NNC, and concluded that the neighborhood-based classifiers (in particular a distance-weighted neighbor classifier) were superior to parametric classifiers, particularly when a training dataset was large and contained the same class proportions as the population to be classified. On the other hand, this classifier requires a relatively large amount training data and a high processing time (Hardin and Thomson, 1992).

In the application of pixel-wise classifiers, the following major error sources have been discussed (Chen and Stow, 2002; Maselli *et al.*, 2005): (i) spatial match between field plots and pixels on digital satellite images, (ii) quality of training samples, and (iii) mixed pixel problem. The first error relates to the fact that pixel-level applications are sensitive to field plot locations and rectification errors. Thus, in the selection of satellite data, the spatial resolution of the satellite data must be considered to correspond to an observation unit for field sampling: if the spatial resolution is much smaller or larger than a plot size defined, digital numbers per pixel cannot be representative of field plots. The second error source is the availability and variability

of training data used in a test area. Finally, the mixed pixel problem due to the heterogeneity of landscapes, the complexity of tree species composition, and the limitation in spatial resolution of satellite data is common when using *low*- and *medium* resolution data. Foody (2002) pointed out that the presence of mixed pixels is to be a major problem for land cover classification.

In South Korea, *Low*-resolution satellite data (e.g., AVHRR) was used to identify vegetation cover classes over the Korean Peninsula (Lee, 1994). Most studies on land use/cover classification at a regional level used *Medium*-resolution satellite data, in particular Landsat MSS and TM. Kim *et al.* (1989) compared the results of classification using Landsat TM with digital forest maps from aerial photographs, but the accuracy was modest. This result was also found in a case study by Kim (1991). In his study, the result with ancillary information (DTM) was more accurate than without. Cho (2002) compared different classifiers using Landsat TM and IKONOS, pixel-based and segment-based with the maximum likelihood and majority principle. His results showed that the best classifier differed according to satellite imagery; the majority principle was superior using Landsat TM while the segment-based with the MLC was superior using IKONOS. Park *et al.* (2001) used multi-temporal Landsat TM data to identify land cover categories. Rho and Lee (1995) used Landsat MSS and TM data to detect the change in forest cover over time. In recent years, the research using *High*-resolution satellite data is increasing, in particular IKONOS (Cho, 2002; Chung *et al.*, 2001) and KOMPSAT-1 (Lee and Kim, 2000).

However, satellite data have not been fully integrated into the Korean NFI. Moreover, although the NFI field data are able to serve as training data, these data have not yet directly contributed to the forest cover classification because they do not define the forest strata per field observation unit. The aim of this chapter is to evaluate the possibility of combining digital satellite data and forest inventory data from the Korean NFI, for forest cover classification.

2.2 MATERIALS AND METHODS

2.2.1 Study area

The study area, Pyeong-Chang County, is located in northeastern South Korea and covers an area of approximately 1,463 km². The county lies between 37°16'N and 37°49'N, and between 128°14'E and 128°46'E (Figure 2.1). Approximately 82% (or 1,199 km²) of the county area is covered by forests; the farming area is only 10% and land for other uses covers about 6% (KFS, 2004a). The county lies over a relatively hilly mountain range; the average slope is approximately 20 degrees (or 36 %). The altitude ranges from 210 to 1,570m and the average altitude is approximately 670m. The main tree species are Japanese red pine (*Pinus densiflora*), Korean pine (*Pinus koraiensis*), Japanese larch (*Larix leptolepis*), Mongolian oak (*Quercus mongolica*), and other deciduous tree species.

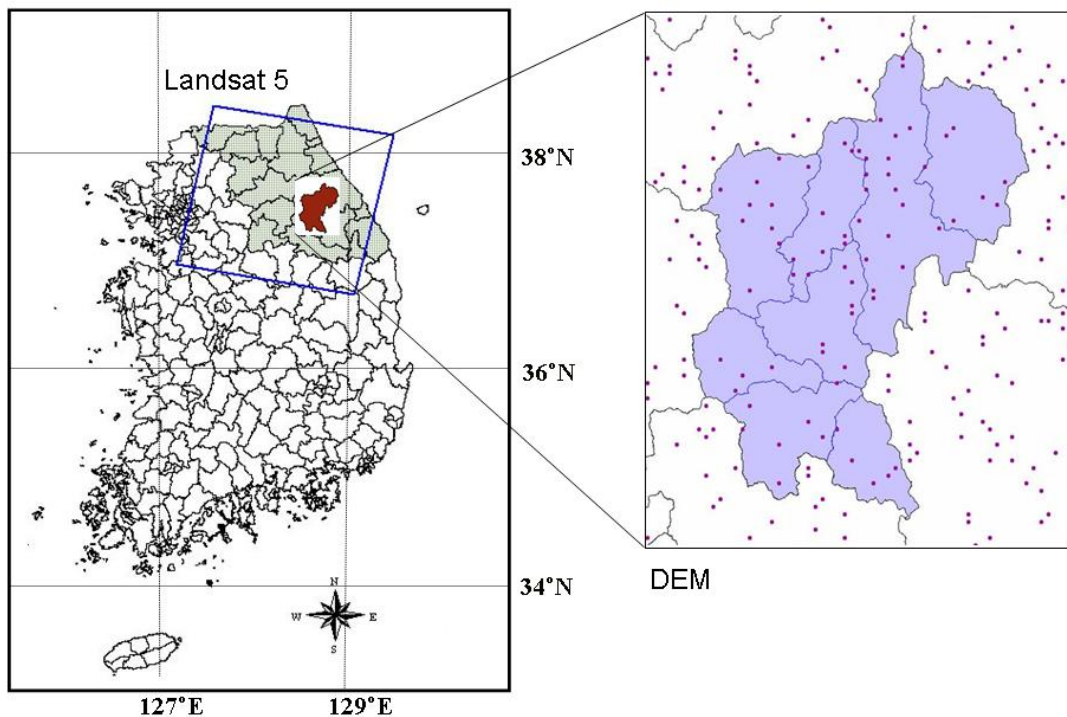


Figure 2. 1: Location of the study area (left) and distribution of field sample points for the 3rd NFI on the Digital Elevation Model (right).

2.2.2 Field data

The field plot data for the test area were extracted from the 3rd NFI database, conducted by the KFRI. This NFI employed a stratified systematic sampling with clusters consisting of four sub-plots. As shown in Figure 2.2, both permanent and temporary sub-plots, with a fixed-area of 0.05 ha, were used.

The clusters were systematically established at every intersection of 1 km x 1 km grids on 1:50,000 topographic maps. These clusters were located within the forest boundary by land cover/use situations. The temporary sub-plots were established on the northern, the eastern, and the southern aspects of a selected center sub-plot (permanent sub-plot) within a cluster, at a distance of 50 m.

For the study area, the field inventory was carried out in the 1986 year. A total of 227 clusters fell into the area for which DEM data were available (Figure 2.1). Since the coordinates of the center points of the field plots were recorded on 1:50,000 topographic maps without GPS recordings, the field plots needed to be geo-referenced so that they could be matched to the digital satellite data (Figure 2.3).

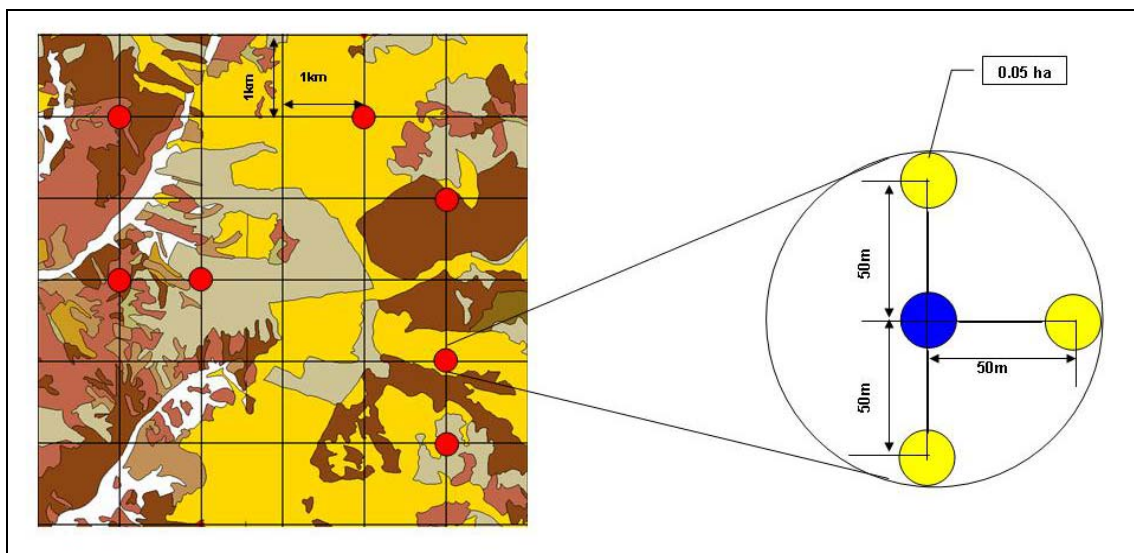
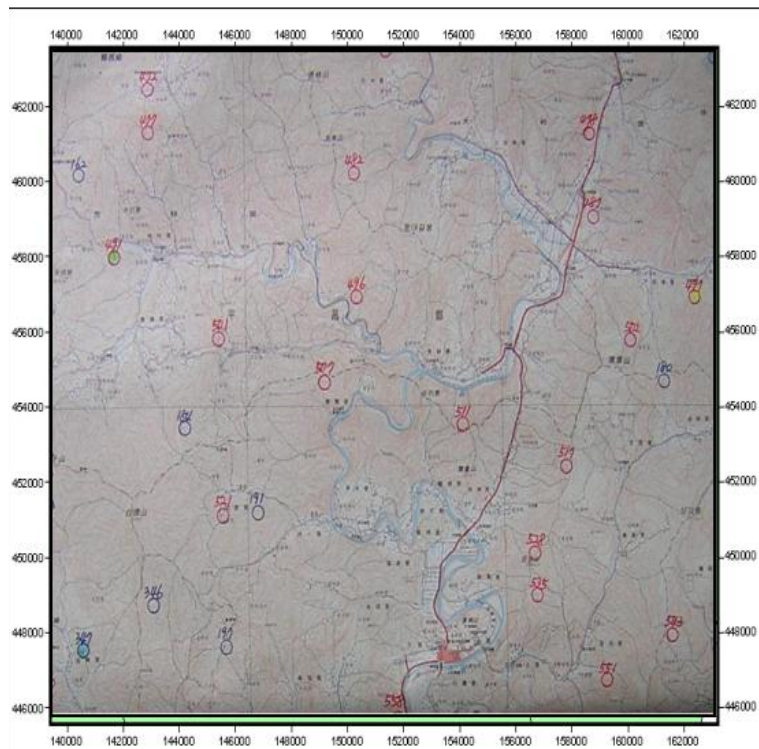
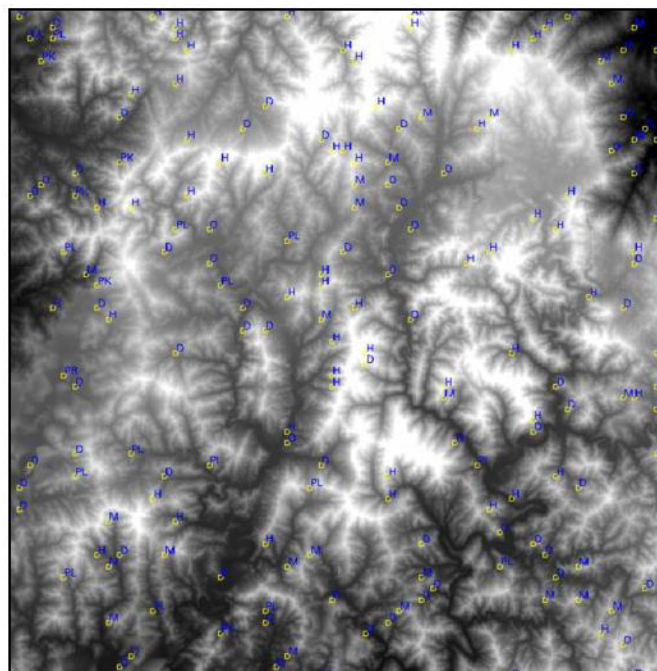


Figure 2. 2: Plot design for the 3rd and 4th NFI cycles in South Korea; each cluster plot consists of 4 sub-plots (right).

MAPPING OF FOREST COVER TYPE



(a) Location of field points on the topographic map (Source from KFRI)



(b) Geo-referenced field points on the Digital Elevation Model

Figure 2. 3: Geo-referencing of field sample plots; coordinates of the permanent sub-plots measured on 1:50,000 topographic maps (a) and coordinates of those on 1:25,000 DEM data (b).

2.2.3 Satellite data

Landsat TM imagery was selected for its spatial resolution and spatial scale during the field survey period. The dataset was stored in an image file format and processed by the KFRI. The thermal band (band 6) was not used because of its poor spatial resolution and low contrast in the forest area. The image was relatively cloud-free and geometrically corrected to an overall RMSE of one pixel (25 m). Other metadata for the satellite image are presented in Table 2.1.

Table 2. 1: Technical information for the satellite image used

Sensor	Landsat TM-5
Acquisition date	02. May 1989
Path / Row	115 / 34
Sun elevation / azimuth	57.45 / 126.01
Map projection / Datum	Transverse Mercator / Tokyo (Korea)
Spheroid	Bessel_1841
Latitude of origin	38
Central meridian	129° 00' 10.405 E
Re-sampling method	Nearest neighbor

2.2.3 Map data

The Digital Elevation Model (DEM) data were produced from 1:25,000 digital topographic maps. They were then used to extract topographic variables for reducing the topographic effects on satellite imagery. The spatial resolution of the DEM was 25m, corresponding to one pixel of the Landsat TM.

In order to compare the classification results from the combination of field plot data and satellite imagery with a “true map”, we took digital forest maps of the study area from the KFRI. They were assembled from the photo interpretation of 1:15,000 black and white aerial photographs, and the follow-up field checking. These maps were digitized and converted into a GIS layer with polygons that can directly be overlaid with the geo-coded image data (Kim, 2004).

2.2.5 Topographic correction

Digital satellite data for mountainous regions include radiometric distortions known as topographic effects. The image classification of multi-spectral data over mountainous terrain is often unsuccessful because surfaces of the same class but with different slope angles (s) and aspects reflect differently, as shown in Figure 2.4.

These effects have been seen to vary considerably with small changes in solar elevation and azimuth, and slope angle in terrain. To eliminate these effects, the Minnaert constant method has been frequently used. It was outlined by Smith *et al.* (1980), based on a principle developed by Minnaert (1941) that is given by the following equation:

$$L_H = \frac{L_T \cos(e)}{(\cos(i) \cos(e))^k} \quad (2-1)$$

where L_H = the normalized brightness value,
 L_T = the observed brightness value,
 $\cos(i)$ = the cosine of the incidence angle,
 $\cos(e)$ = the cosine of the exitance angle or slope angle, and
 k = the empirically derived Minnaert constant.

The Minnaert constant (k) can be estimated by the backward radiance correction transformation model (Colby, 1991). The k value is the slope of the regression line:

$$\log(L_T \cos(e)) = \log(L_H) + k \log(\cos(e) \cos(i)). \quad (2-2)$$

To estimate the Minnaert constants for each spectral feature (hereafter defined as band), the DNs for all bands from the Landsat TM imagery and topographic variables, such as elevation, slope, and aspect from the DEM data were extracted at centre points of the field plots within the forest area.

The value of the Minnaert constant lies usually between 0 and 1, which is used to describe the roughness of the surface. When phenomena on the surface of the earth reflect incident radiation equally in all directions, it is called the Lambertian behavior, and then the value of the Minnaert constant is equal to 1 (Tokola *et al.*, 2001).

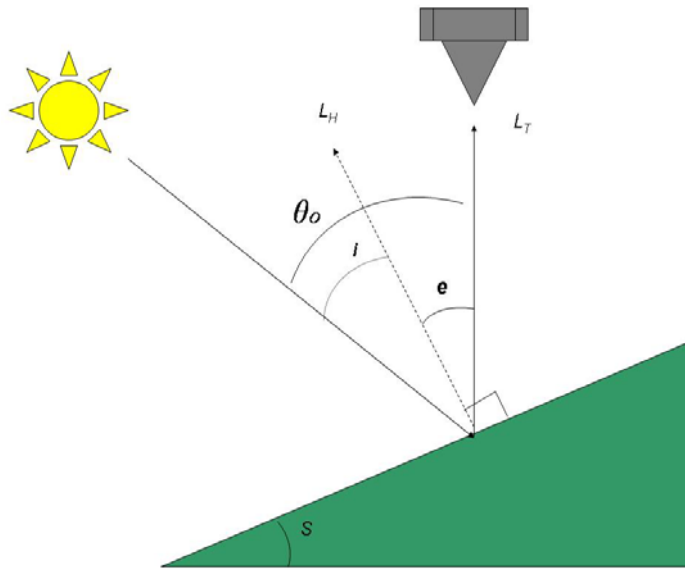


Figure 2. 4: Representation of the incident solar angle (i) and the solar zenith angle (θ_0), where L_T is the observed value and L_H is the normalized value (modified from Jensen, 1996).

2.2.6 Classification

Forest cover types must be clearly defined before they can be identified. In this study, the types per sub-plot were defined according to the definitions in the 4th NFI (see Annex 1). The available field plot data per forest stratum served as training data. Subsets of field plots were selected to train a classifier before classifying pixels on digital satellite data.

All 798 sub-plots from the 227 clusters were classified into three forest types. The types depend on the number of trees by dominant tree species per sub-plot. Then, 172 points within non-forest areas were randomly selected from digital topographic maps. Table 2.2 indicates that the distribution of field points per sub-plot unit that served as training data for each stratum.

Table 2. 2: Distribution of field points for each stratum

Stratum	Forest			Non-forest	Total
	Coniferous	Deciduous	Mixed*		
Number of points	366	327	105	172	970

Mixed forest*: 24-74 % of number of trees by deciduous tree species

The Maximum Likelihood Classifier (MLC) that is a parametric classifier considers not only the average numbers in assigning classification but also their variability in each class (Lillesand *et al.*, 2004). Despite the assumption of input data, it is mostly used as a baseline in land cover/use classification and forest cover classification as well as other remote sensing applications (Kim *et al.*, 1989; Cho, 2002).

However, since remotely sensed measurements of forest cover types mostly do not meet the assumptions of such as the MLC, it is useful to perform non-parametric classifiers. In this study, the Nearest Neighbor classifier (NNC) was used to identify the forest classes. This classifier is similar to the Minimum Distance Classifier (MDC) in the parametric approach. In the MDC, an un-classified pixel is assigned to a closest training class centroid, whereas the NNC requires distances between the un-classified pixel and every training pixel in feature spaces (Koukal, 2004), as shown in Figure 2.5. To select the nearest neighbor class at an un-classified pixel in the NNC, the distances in feature spaces were computed by the Euclidean distance metric. For a detailed algorithm of the NNC, see next chapter.

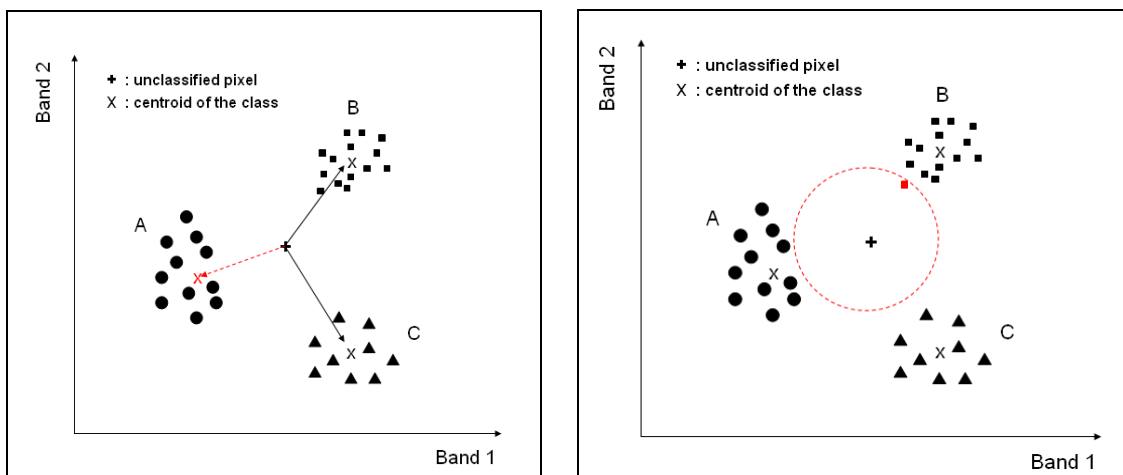


Figure 2. 5: Comparison of the minimum distance (left) and the nearest neighbor (right) classifiers: for instance, an un-classified pixel (+) belongs to class B by the NNC (modified from Koukal, 2004).

2.2.7 Evaluation

The “leave-one-out cross-validation” is enabling to assess accuracy even when limited ground truth samples are available for training and accuracy assessment. This analysis allows for an accuracy statement for the digital forest map and both classified images. The accuracy can be defined in terms of the degree of misclassification, which can be computed from the confusion or error matrix (Congalton, 1991).

Additionally, the Kappa statistic was also used as a measure of the classification accuracy. This statistic measures the strength of agreement of the row and column variables. The value of kappa (\hat{k}) is computed as:

$$\hat{k} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})} \quad (2-3)$$

- where r : the number of rows in the error matrix,
 x_{ii} : the number of observations in row i / column i ,
 x_{i+} : the total of observations in row i ,
 x_{+i} : the total of observations in column i , and
 N : the total number of observations included in matrix.

The distributions of classification results from different data sources; aerial photos and satellite images, and the two classifiers were also compared with the classification result from field plot data as an expected distribution. For this, the chi-square goodness-of-fit test was used (Rencher, 1993; Koukal, 2004). If a classification result significantly differs from an expected distribution, the classification result is biased. The test statistic is defined as:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \quad (2-4)$$

where the expected classes are the number of plots per stratum from field plot data, while the observed classes are extracted from the digital forest map and classified images.

2.3 RESULTS

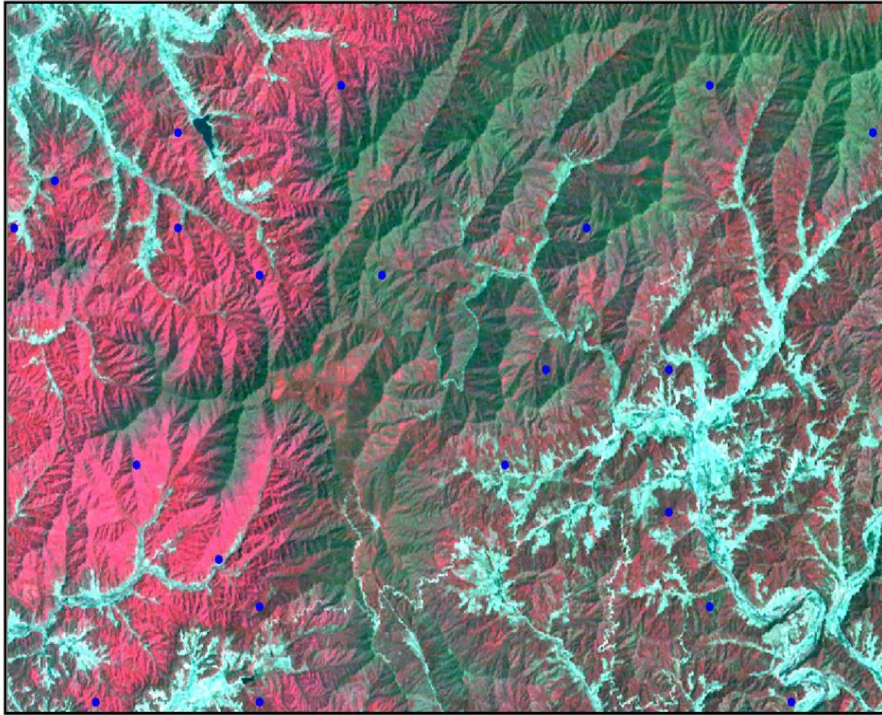
2.5.1 Topographic correction

Table 2.3 shows the empirically calculated k values for each band for the study area. The estimated values can be used to describe the roughness of the surface. In our case, the estimated values of k ranged from 0.2402 to 0.5237. The greatest range of difference was observed between bands 3 (red) and 4 (near infrared). When comparing different bands, the band 7 (thermal infrared) had the highest value of k .

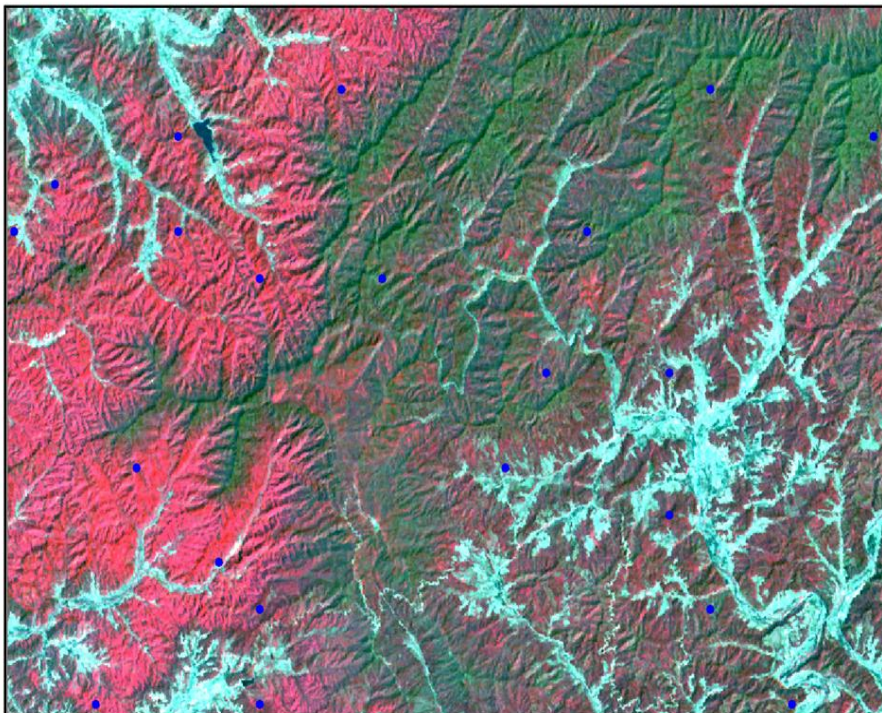
Table 2. 3: Estimated values of the Minnaert constant for each band

Band	1	2	3	4	5	7
Minnaert constant	0.2402	0.2587	0.3682	0.5035	0.5083	0.5237

After the Minnaert constant k was derived, the topographic correction was performed. A reduction of the topographic effects was visually apparent in the normalized image. The topographically normalized image shows that the dark sides (shadowed areas) on the raw image become brighter whereas the solar facing slopes appear in a rather darker tone on the normalized image (Figure 2.6b).



(a) Raw image



(b) Topographically normalized image

Figure 2. 6: Comparison between (a) raw and (b) topographically normalized images (Landsat TM 4:3:2).

2.3.2 Classification

Figure 2.7 depicts the mean digital numbers (DNs) for the different forest classes and bands. At bands 4 and 5, the spectral differences among forest classes could easily be identified, whereas the differences at the other bands were not so obvious. In the case of the deciduous forest (H), the mean value was lowest at band 4 and highest at band 5.

Figure 2.8 illustrates the distribution of DN by forest classes in bands 4 and 5. In the classification process, coniferous and deciduous forests can clearly be discriminated while the mixed forest (M) can hardly be distinguished from those.

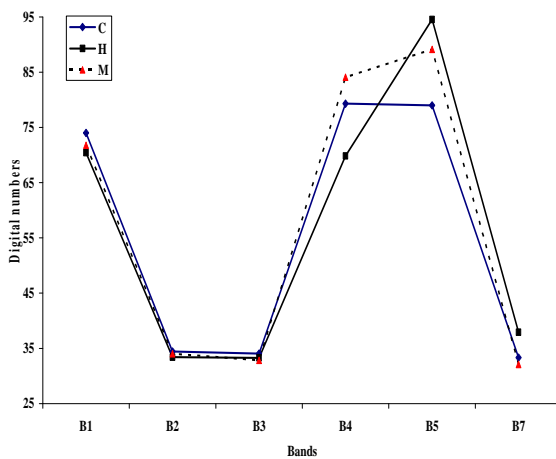


Figure 2. 7: The mean digital numbers for the different forest classes and bands (C: coniferous forest; H: deciduous forest; and M: mixed forest).

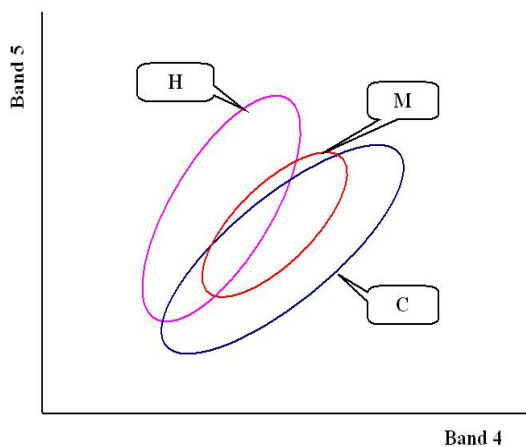


Figure 2. 8: Distribution of digital numbers by forest classes in bands 4 and 5.

2.3.3 Evaluation

Digital forest map

The digital forest map was derived from the interpretation of aerial photos in which the forest was divided into the three forest classes with a minimum area of 1 ha. Table 2.4 presents the error matrix for assessing the classification accuracy of the digital forest map with field plot data. In the accuracy analysis, the coniferous (C) and deciduous forests (H) had higher accuracy than the mixed forest (M). The user and producer accuracies of the mixed forest were 26% and 36%, respectively. The overall accuracy was about 70%. User and producer accuracies ranged from 26 to 80% and from 36 to 93%, respectively. The value of kappa was estimated to be 0.58.

Table 2. 4: Error matrices for assessing the classification accuracy of the digital forest map and the classified images for both classifiers with field plot data

Classification results		Field plot data				Total	User accuracy
		C	H	M	Non		
Digital forest map	C	280	43	37	9	366	76 %
	H	29	199	22	1	250	80 %
	M	37	70	48	2	149	26 %
	Non	21	15	8	160	205	78 %
	Total	366	327	105	172	970	Overall accuracy 70 %
Producer accuracy		77 %	61 %	36 %	93 %		
Kappa value = 0.58							
Classified image by MLC	C	196	123	45	27	391	50 %
	H	43	135	23	10	211	64 %
	M	98	59	33	15	205	16 %
	Non	29	10	4	120	163	74 %
	Total	366	327	105	172	970	Overall accuracy 50 %
Producer accuracy		54 %	41 %	31 %	70 %		
Kappa value = 0.31							
Classified image by NNC	C	288	32	21	23	364	79 %
	H	39	272	13	8	332	82 %
	M	27	13	63	5	108	58 %
	Non	12	10	8	136	166	82 %
	Total	366	327	105	172	970	Overall accuracy 78%
Producer accuracy		79 %	83 %	60 %	79 %		
Kappa value = 0.69							

Classified images

Classified images were produced using both classifiers (Figure 2.9). Classification results from the cross validation are presented in Table 2.4. The accuracy for the MLC was modest; the user and producer accuracies ranged between 16 and 74%, and between 31 and 70%, respectively. The accuracy for the NNC was greatly improved compared to that for the MLC; its user and producer accuracies ranged between 58 and 82%, and between 60 and 83%, respectively. Particularly in the case of the mixed forest (M), the accuracy for the NNC was appreciably improved. Overall accuracies for the MLC and the NNC were 50% and 78%, respectively. The estimated kappa value for the NNC (0.69) was about twice as large as for the MLC.

Digital map vs. classified image by the NNC

The accuracy of the classification result by the NNC was assessed using the digital forest map as a reference (Table 2.5). Here, the pixel size was a square grid of 25 m (0.0625 ha). Compared with other classes in the classification accuracy assessment, the accuracy of the mixed forest class was lowest. Within the classified image, most of the mixed forest class on the digital forest map was divided into the deciduous forest (about 43%) and the coniferous forest classes (about 36 %). On the contrary, the accuracy of the non-forest class was highest. As a result, the overall accuracy was modest (48%). User and producer accuracies ranged from 19 to 60% and from 12 to 63%, respectively. The estimated value of kappa was to be 0.28.

Figure 2.9 shows the digital forest map (a) and the classified images by the MLC (b) and the NNC (c) for the study area. Due to the relatively large minimum area that was defined, forest cover types on the digital forest map can more clearly be discerned than those within both classified images.

MAPPING OF FOREST COVER TYPE

Table 2. 5: Error matrix for assessing the classification accuracy of the NNC classified image and digital forest map over the entire test area per pixel

Classification		Digital forest cover map				Total	User accuracy
		C	H	M	Non		
Classified Image	C	401081	169179	135135	90358	795753	50%
	H	240009	411964	158982	52129	863084	48%
	M	86692	75142	42541	18889	223264	19%
	Non	106590	48389	31267	275584	461830	60%
Total		834372	704674	367925	436990	2343931	Overall accuracy 48%
Producer accuracy		48%	59%	12%	63%		

Kappa value = 0.28

The classification result of field plot data compared with the digital forest map and both classified images by the Chi-square goodness-of-fit test is presented in Table 2.6. The goodness-of-fit test indicates that the digital forest map and the MLC classified image differ significantly from the classification result of the field plot data, but there is no statistically significant difference found between the field plot data and the NNC classified image.

Table 2. 6: The result of the chi-square test for field plot data with digital map and both classified images

	Digital forest map	MLC	NNC
Field plot data	42.9*	138.5*	0.38 ^{NS}

* : significant at 5% level, ^{NS} : not significant, degree of freedom = 3

$$\chi^2_{0.05,3} = 7.815$$

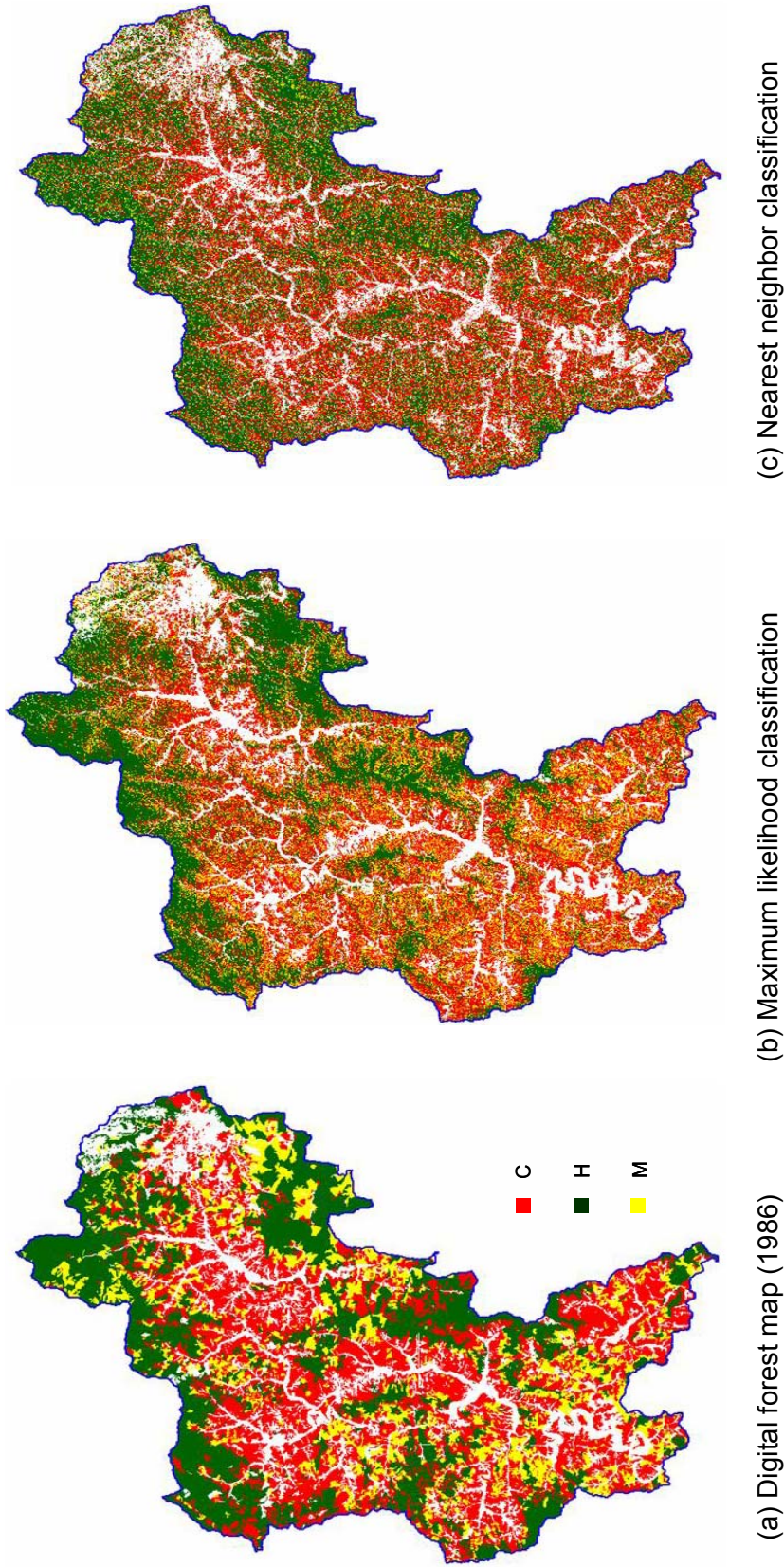


Figure 2.9: Comparison of the forest cover maps for different approaches in the study area; digital forest map (a), and classified images by the MLC (b) and the NNC (c), respectively. The observation units differ from depending on data source: digital forest map (1 ha) and classified images (0.0625 ha); thus, the forest cover types on the digital forest map are obviously identified, whereas they are highly fragmented within the classified images.

2.6 DISCUSSION AND CONCLUSION

The main objective of this chapter was to evaluate the potential of digital satellite imagery for forest cover mapping by combining them with field plot data from the NFI. In order to address this objective, field plot data available from the 3rd NFI and digital satellite imagery (Landsat TM-5) were combined through pixel-wise classifiers and then the results were compared with the existing digital forest map as a reference.

Topographic correction

Topographic characteristics, in particular over mountainous area, have an influence on the natural spectral variability, which varies by band within a single satellite image. Radiometric correction of topographic effects is required in the Korean Peninsula, where two-thirds of the territory is covered with forests on hilly terrain (KFRI, 2004).

The Minnaert constant method, which has produced the most reliable results for normalizing these effects in most studies for Korean forest conditions, was applied (Lee and Yoon, 1997; Cho, 2002). Generally, in order to reliably estimate the Minnaert constant (k), all pixels in a DEM-masked image are used. However, this requires a high processing time when the area of interest is too large and/or high spatial-resolution imagery is applied. In this study, since the test area is relatively large, a small number of pixels for forest inventory plots ($n=227$) were used. The estimated values of k ranged from 0.24 to 0.52 were fulfilled in the range expected (0.2-0.6) for that of Korean forest conditions (Lee and Yoon, 1997).

Although the effectiveness of applying the Minnaert constant has not been realized in all cases, it must be performed to reduce topographic effects on satellite imagery in the image pre-processing stage. Considering that the Minnaert constant is relevant to the surface roughness of the area of interest, the use of forest inventory points that may cover varying topographical characteristics is feasible to reliably estimate the Minnaert constant. Furthermore, to successfully reduce topographic effects, the Minnaert constants should be estimated per forest stratum since original DNs under the same topographic condition also reflect differently according to forest strata and therefore the topographic effects vary considerably with forest strata (Lee and Yoon, 1997). In this study, however, the sample size did not allow for such stratification.

Supervised classification

Despite the geo-referencing process for field points, their locations were still unreliable because this process was performed on the basis of marked points on topographic maps (Figure 2.3). With respect to spatial match, if a GPS receiver is used for locating the field points, the spatial matching error between field plots and pixels on the Landsat TM can be reduced. This is because the spatial resolution of Landsat TM, a square of 25 m, is nearly comparable to the defined plot size (500 m²).

The quality of the training data is related to the definition of the forest types of interest and their variability. In this study, the definition of the proportion of the number of trees by dominant tree species was taken to classify field plots per sub-plot as an observation unit into the forest cover types. However, this procedure is neither a laborious task nor does it produce dubious classification results, even under the complexity of the composition of tree species per sub-plot. Consequently, for field data from a forest inventory to be used directly, there is a need for clearly defined the forest strata of interest per field observation unit.

In order to successfully achieve the NNC, it is necessary to have a sufficient training dataset. This dataset must cover all variations of the strata of interest, and an equal number of available reference samples for each stratum (Davies, 1988). In the given training dataset, the number of reference samples varied with the forest cover types. In particular, the number of samples available for mixed forest was relatively small (Table 2.2), which caused the high classification errors (Table 2.4).

From an ecological point of view, most mixed forests in South Korea are composed of Japanese red pine and oak species. The Japanese red pine forests (coniferous forest) are mostly located in mountainous areas without any artificial disturbance, and have gradually changed into the mixed forests through ecological succession (Chung, 1996; Lee *et al.*, 2004b). Thus, the mixed forest may encroach on the range of the coniferous forest in the feature spaces, as shown in Figure 2.8.

Additionally, with respect to the of forest disturbance history in South Korea, most forests were planted after the Korean War. In the reforestation plan, coniferous tree species, like the Korean pine and Japanese larch, were mostly planted over large areas. Most of the mixed forest was generated by natural regeneration, without artificial disturbance due to its inaccessibility (Park 1986). Thus, the forest structure is

more complex and highly fragmented. In order to eliminate the uncertainty about mixed forests, satellite data showed that the mixed forest class for the forest cover classification belonged to other forest classes (Cho, 2002; Chung *et al.*, 2001). These different disturbances may also be a major obstacle in improving the accuracy of classification.

The accuracy of the NNC classified image was assessed using the existing digital forest map as a reference. The accuracy, however, was modest ($\hat{\kappa}=0.28$) and similar to the results of the case study by Kim *et al.* (1989). According to them, this may be related to (i) different definitions of the observation unit and (ii) the different dates for acquisition from remote sources and from field survey. From a spatial scale point of view, for aerial photo interpretation, the forest is defined as an area of 1 ha. This is much larger than the area used for satellite data classification (a square grid of 25 m in this study), which depends upon the spatial resolution of the satellite data.

Accordingly, each forest class on the digital forest map is split into different forest classes within the classified image. In other words, the classified images manifested “*salt-and-pepper*” effects due to the smaller observation unit, as shown in Figure 2.9. In this context, there is a need for suitable spatial resolution (i.e., definition of the observation unit). In addition, because of the large definition in the fifth NFI (KFRI 2006), 0.5 ha, there is a need for more research on post-classification processing to improve correspondence with the definition as well as the accuracy of the classifications.

Even though there are a variety of errors in the image classification process, digital satellite data can represent a reasonably useful and more cost-effective data source over a large area for forest cover classification. The classification accuracy can be improved with the application of new classification techniques, an understanding of historical disturbances for the area of interest, sufficient ground truth data for each forest stratum, and clear definitions of forest strata.

MAPPING OF FOREST COVER TYPE

3. ESTIMATION OF FOREST ATTRIBUTES

3.1 INTRODUCTION

There are two main types of forest inventory in South Korea: national and management forest inventories. The national forest inventory (NFI) provides data and information for national and regional or administrative units, and international processes, which is relevant and required for national-level decision making and monitoring (KFS, 2002). The latter inventory has been implemented to provide baseline information for forest management planning at a single stand unit.

However, these inventory types cannot easily be combined due to their different scales and objectives. In addition, since the information for regional and municipal units is only available as mean values from the NFI, it would require complex small-area estimation procedures to support management planning at a regional or municipal level (KFS, 2004a). The NFI field data at a municipality level is less useful for municipal purposes due to a small number of samples. Thus, to acquire information for forest management planning in a small-area unit, the conventional procedure is to independently carry out a field-based survey. Hence, the inventory cost will be comparatively high. Moreover, the last NFI implemented a rotation system by province so that current and reliable information cannot be provided over the entire country at the same time (KFS, 2002).

Field measurements in combination with remotely sensed data are of interest for forest inventory, offering the possibility to use accurate field data together with full image coverage from digital satellite and/or airborne sensors. Since the launch of Landsat TM in 1982, a correlation has been found between forest characteristics and the spectral response of multi-spectral satellite data based on:

- statistical regression approach (Ahern *et al.*, 1991; Lee *et al.*, 2004a) and
- *k*-Nearest-Neighbor (*k*-NN) technique (Tomppo, 1991; Tokola *et al.*, 1996).

This correlation between forest characteristics and spectral features on satellite imagery is not high in all cases, but can frequently be used to establish statistical models. A developed model for a forest characteristic can be used to estimate unobserved points.

On the other hand, the k -NN technique is to combine forest characteristics from forest inventories with remotely sensed data (Tomppo, 1991). In the k -NN technique, field plots observed across an entire area covered by satellite imagery serve as training data for any unobserved area, and that satellite imagery can be used as an indicator dataset for estimating the areas represented (Tokola *et al.*, 1996).

For the estimation of forest attributes, this technique has been operational in the Finnish NFI since 1990 and has been extensively used and advanced. In the last decade, it has successfully contributed to large area forest inventories, particularly in the Nordic countries and the United States (Nilsson, 1997; Katila and Tomppo, 2001; Franco-Lopez *et al.*, 2001; McRoberts *et al.*, 2002).

3.2 BASICS OF THE k -NEAREST NEIGHBOR TECHNIQUE

3.2.1 General information

The k -Nearest Neighbor technique represents one of the simplest techniques in statistical discrimination and is also used as an instance-based learning algorithm in pattern recognition. It is a non-parametric approach, wherein a new observation is identified as a pre-defined class of observations from a learning or training dataset that is closest to the new observation (Mitchell, 1997).

In order to achieve this technique, the similarity between previous observations and the new observation, which can usually be determined based on distances computed from ancillary information such as remotely sensed data, is required. Training data of the form $(X, f(x))$ are used to train a discrete-valued or real-valued target function $Y = f(x)$. There is no general description of the target function in the training stage, so that the training data are just stored regardless of any assumptions. Accordingly, in order to produce reliable estimates of new observations by this technique, a large and representative training dataset is required. The training dataset should cover “*the full range of variability*” that occurs within the class or characteristic of interest (Mitchell, 1997; Koukal, 2004).

3.2.2 Characteristics

In the estimation of forest characteristics, the major steps involved in performing the k -NN technique are shown in the flowchart (Figure 3.1). To do these steps, three data sources such as ground truth data, remotely sensed data, and mask map data are usually required. The key advantage of this technique is that all inventory variables can be estimated at the same time.

On the other hand, one disadvantage is the requirement for sufficient training data. Accordingly, to successfully achieve this technique, the key is to obtain a sufficient training dataset. It also requires a high processing time when a large number of training data and/or a complicated operational option are applied, as well as when several test runs are needed.

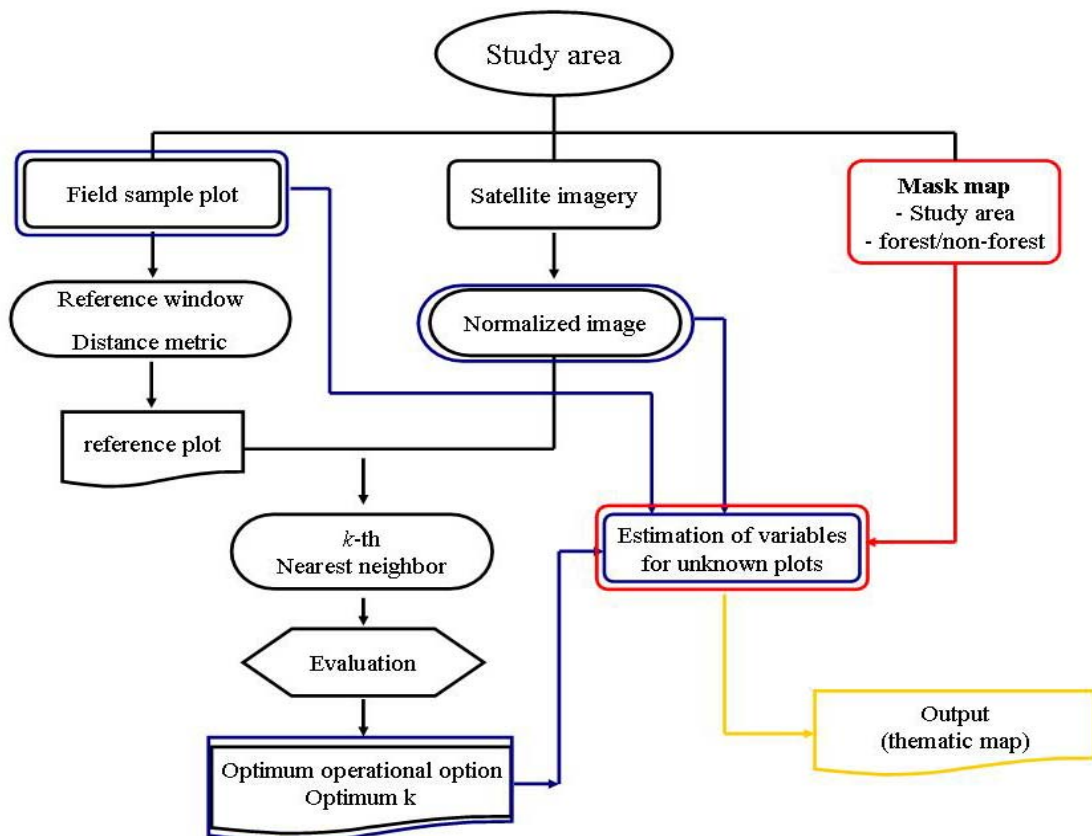


Figure 3.1: Research flowchart for the k -NN technique.

Some weaknesses of this technique have been discussed. Altman (1992) pointed out that it may give biased estimates with increasing the number of k . The biased estimates can also appear along the borders in feature spaces, because the nearest neighbor distances tend to be greater and the neighbors may be concentrated in one direction only (Katila, 2004). In the last decade, various alternative operational options have been discussed and evaluated to overcome the weaknesses and to improve the precision of estimates for the k -NN technique:

- Use of remotely sensed data: With respect to remotely sensed data as a data source, medium resolution imagery (e.g., Landsat TM and ETM+) is regarded as the most suitable satellite source for large area inventories due to the spatial and spectral resolution and spatial scale. In addition to the data source, different digital satellite data may be used not only individually, but several remotely sensed data can be combined as well, including digital aerial photographs, Lidar imagery, and radar imagery.
- Distance metrics: To compute the similarity among reference plots and target plots in feature spaces, different distance metrics have been applied (Nilsson, 1997; Franco-Lopez *et al.*, 2001). The results by Franco-Lopez *et al.* (2001) showed that the Mahalanobis distance did not improve the quality of the estimation as compared to the Euclidean distance; this result was contrary to the result reported by Nilsson (1997). Additionally, to reduce the processing time for searching large numbers of reference samples, Finley *et al.* (2006) used a squared mean distance. Since a complicated distance metric can require more processing time, the Euclidean distance metric is frequently preferred due to its time-efficiency in the k -NN process. The distance, $d_{t,r}$, is computed as follows:

$$\text{Euclidean distance} \quad d_{t,r} = \left(\sum_{i=1}^m (x_{i,r} - x_{i,t})^2 \right)^{\frac{1}{2}} \quad (3-1)$$

where $x_{i,r}$ and $x_{i,t}$ are digital numbers of target plot t and reference plot r for each band i , respectively and m is a number of bands within a single image.

- Distance-weighting for neighbors: Altman (1992) demonstrated that the bias in the estimates can be reduced with distance-weightings (a). For this, several weightings were tested in the k -NN estimator. In most studies, estimates with distance-weightings for neighbors were more precise (Katila and Tomppo, 2001). In contrast, some studies showed that non- or small weighting for neighbors gave better results (Franco-Lopez *et al.*, 2001; Tokola *et al.*, 1996).

$$\text{Distance-weighting} \quad w_{i,r} = \frac{\frac{1}{d_{i,r}^a}}{\sum_{r=1}^k \frac{1}{d_{i,r}^a}} \quad (3-2)$$

a = weighting parameter

- Weights for each band within a single image: The correlation between bands and forest characteristics may not be equal for all bands within a single image. Assuming that there exists a linear combination of bands that can provide better results, additional weights can be computed and applied to the original digital numbers. For this, a weighted Euclidean distance has been applied (Equation 3-3). A weighting parameter (b_i) for a band (i) can be derived from the downhill optimization method (Franco-Lopez *et al.*, 2001), genetic algorithm (Tomppo and Halme 2004), and empirical constants (Tokola *et al.*, 1996).

$$\text{Weighted-Euclidean distance} \quad d_{i,r} = \left(\sum_{i=1}^m b_i^2 (x_{i,r} - x_{i,t})^2 \right)^{\frac{1}{2}} \quad (3-3)$$

- Reference windows for stratification: Forest characteristics are sensitive to specific situations such as topographical or structural conditions. In other words, estimates based on the similarity only in feature spaces can be unreliability and highly biased due to heterogeneous conditions between the reference and target plots. To address this challenge, various ancillary information, such as site quality maps, land-use maps, and soil class maps, have been considered to search reference plots that are restricted within an identical class to a target plot.

Another approach restricts an area (or window) that is searched for neighbors within the geographical horizontal and vertical boundaries (Tokola, 2000; Katila and Tomppo, 2001; Koukal, 2004).

- Image enhancement technique: Spatial filtering techniques have been applied for improving the spatial match between field plots and pixels on satellite imagery (Halme and Tomppo, 2001; Franco-Lopez *et al.*, 2001). Additionally, an image enhancement technique, called transformation, is also frequently used to generate new features (e.g., NDVI, VIs and others using tasseled cap transformation and PCA) from a single imagery because new features may have a stronger correlation with forest characteristics (Franco-Lopez *et al.*, 2001; Tomppo and Halme, 2004).
- The value of k : A number of studies have proposed different optimal values of k ranged from 1 to 15 (Franco-Lopez *et al.*, 2001; Tokola *et al.*, 1996; Nilsson, 1997; Reese *et al.*, 2003). McRoberts *et al.* (2002) found that a small k ($k < 5$) may result in RMSE values larger than a standard deviation of observations when estimating forest proportion. The optimal value of k is affected by the number of available reference samples, their variation, satellite imagery and forest conditions at a specific area of interest. In addition to the optimum value of k , Franco-Lopez *et al.* (2001) contended that an optimal number of k was dependent upon user objectives: if the objective is to produce a thematic map, the use of only the nearest neighbor ($k=1$) should be preferred because the variability of samples can be retained, while if the objective is to provide estimates, the value of k is determined having the minimized estimation error.

The main aim of this chapter is to analyze the potential of the k -NN technique in the Korean NFI. To address this aim, a pilot field inventory based along the new NFI design is implemented in a test area. The k -NN technique is used to estimate growing stock per unit within the forest area for the entire test area. Moreover, to improve the precision of estimates, more alternative options in the k -NN process are tested, including stratification, image enhancement, and weighting for neighbors and spectral features.

3.3 MATERIALS AND METHODS

3.3.1 Study area

The study area, Yang-Pyeong County, is centrally located in the Korean Peninsula, covering about 87,446 ha. The county consists of about 72% (or 63,242 ha) of forest, whereas the area for farming is only 15% and that for other uses is about 13% (Table 3.1). The county lies between 37°22'N and 37°40'N, and 127°18'E and 127°51'E (Figure 3.2). The county is in the hilly region with a complex topography, as shown in Figure 3.3. The elevation ranges from 20 to 1,157 m above sea level. According to a forest soil inventory report (KFRI, 2004), the areas with slopes greater than 21° cover about 82% of the total forest area. The main tree species are Japanese red pine (*Pinus densiflora*), Korean pine (*Pinus koraiensis*), Japanese larch (*Larix leptolepis*), Mongolian oak (*Quercus mongolica*) and other deciduous tree species.

Table 3. 1: Distribution of land use/cover classes for the study area (KFS, 2004a)

Class		Land-use*	
		Area (ha)	Proportional (%)
Forest	Coniferous	42,025	48 (66.4)
	Deciduous	16,293	19 (25.8)
	Mixed	4,436	5 (7.0)
	Un-stocked	488	1 (0.8)
	Total	63,242	72 (100)
Non-forest		24,204	28
Total		87,446	100

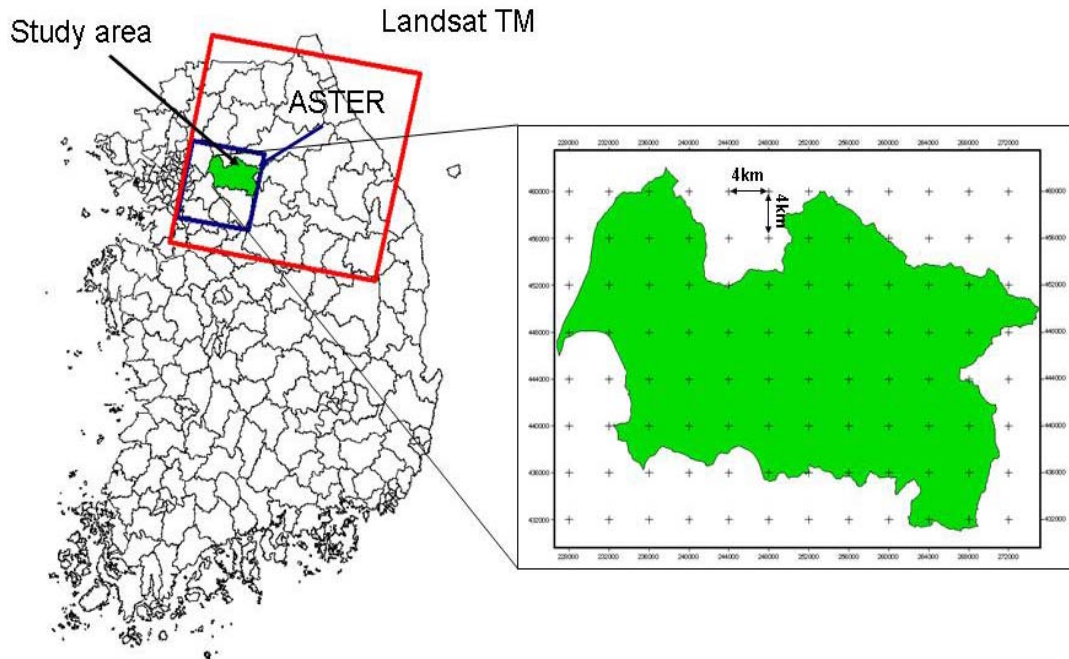


Figure 3. 2: Location of the test area and distribution of field plots for the study area in the current NFI cycle (2006-2010), where the samples are established with a systematic square grid of 4km.

3.3.2 Field data

In the fifth Korean NFI, a systematic cluster sampling design is applied, wherein the samples are established with a systematic square grid of 4km (Figure 3.2). This pilot study was carried out during 2005/2006. A total of 39 samples fall within forests in the test area. In this study, two cluster plots were used; one cluster having 4 sub-plots (30 clusters, plot design A) and the other cluster having 10 sub-plots (9 clusters, plot design B), as shown in Figure 3.3. A total of 191 sub-plots can serve as training data. Each sub-plot was composed of two concentric circular plots; all trees with diameter at breast height (DBH) of 6 to 20cm and with DBH above 20 cm were registered at 10m and 15m radii, respectively. The coordinates for each sub-plot centre were recorded by a GPS receiver (Garmin V DELUXE).

On each of the concentric circular plots, the basal area of individual trees was multiplied by the extension factors related to each radius for calculating the basal area per hectare. The volumes of individual trees were calculated using the stem volume functions for Japanese red pine, Korean pine, larch and oak (the oak function was used for all deciduous tree species) (see Annex 3). Table 3.2 presents summary statistics based on field plots.

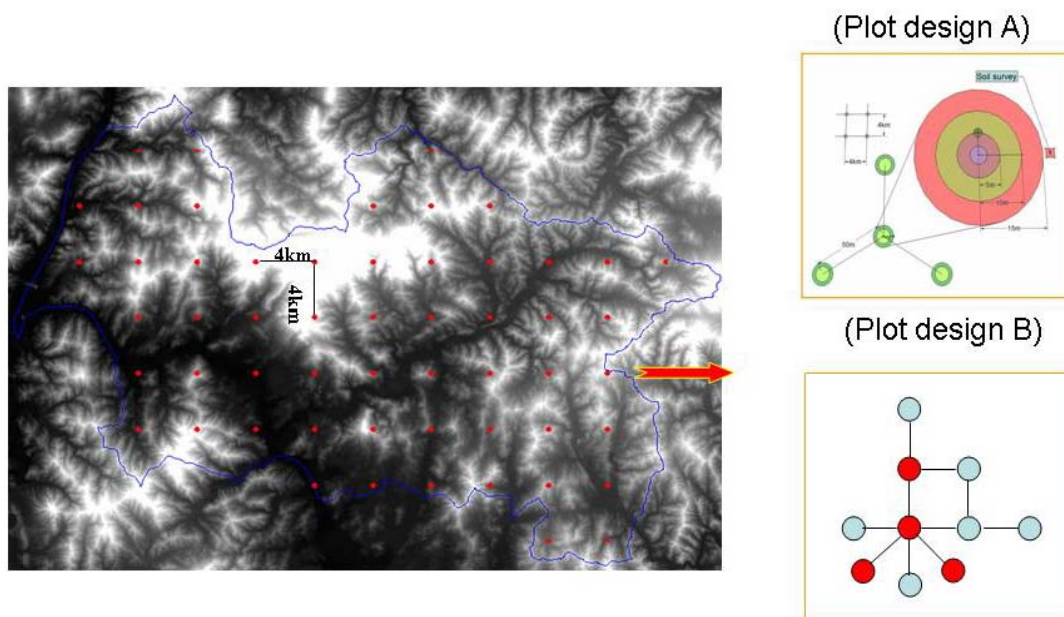


Figure 3. 3: Field points on the DEM and the two cluster plot designs used.

Table 3. 2: Summary statistics based on field plot data ($n=191$ sub-plots)

Variables	V (m ³) / ha	BA (m ²) /ha	N / ha	Number of observed tree species per plot
Mean	112.9	17.7	1,009	4.6
Min.	3.3	0.9	113	1.0
Max.	239.1	35.4	2,576	9.0
Standard deviation	47.2	6.7	529	2.1

ESTIMATION OF FOREST ATTRIBUTES

3.3.3 Satellite data

For this study, different digital satellite images were acquired from different satellite sensors, Landsat ETM+ and ASTER.

Landsat ETM+

Landsat ETM+ data, which are widely employed in the *k*-NN estimation, was acquired on 28 April 2002. The thermal infrared (Band 6) was not used because of its poorer spatial resolution (60m) and low contrast in the forest area. More detailed information is presented in Table 3.3. The imagery was relatively cloud-free and geometrically corrected with an overall RMSE of 1 pixel (25m).

Table 3.3: Technical information of the satellite images

Sensor		Landsat 7 ETM+	ASTER
Band (μm)	1	0.45-0.52 (B)	0.52-0.60
	2	0.52-0.60 (G)	0.63-0.69
	3	0.63-0.69 (R)	0.76-0.86
	4	0.76-0.90 (NIR)	1.60-1.70
	5	1.55-1.75 (MIR)	2.145-2.185
	6	10.4-12.50 (TIR-L) 10.4-12.50(TIR-H)	2.185-2.225
	7	2.08-2.35 (MIR)	2.235-2.285
	8	-	2.295-2.365
	9	-	2.360-2.430
Acquisition date		28. April 2002	22. September 2003
Path / Row		115 / 34	115 / 034
Sun elevation / azimuth		59.4 / 134.6	50.3 / 155.4
Bands		1 / 6 / 1	3 (1) / 6 / 5
Swath width (km)		185	60
Spatial resolution (m)		15 / 30 / 60	15 / 30 / 90
Map projection / Datum		Transverse Mercator / Tokyo (Korea)	
Spheroid		Bessel_1841	
Latitude of origin		38	
Central meridian		127° 00' 10.405 E	
Re-sampling method		Nearest neighbor	

ASTER

ASTER (Advanced Space borne Thermal Emission and Reflection Radiometer) is a medium spatial resolution multi-spectral imager onboard spacecraft Terra, a satellite launched in December 1999 (Abrams, 2000). ASTER has three subsystems operating in different feature spectral ranges, namely the visible and near infrared (VNIR), shortwave infrared (SWIR) and thermal infrared (TIR). The spatial resolution is 15, 30 and 90 m for VNIR, SWIR and TIR, respectively (Table 3.3).

An imagery (AST_L1B.003:2017513589) acquired on 22 September 2003 and processed into Level 2 was taken. In this study, TIR bands having a spatial resolution of 90m were not used (Table 3.3). The imagery was relatively cloud-free and geometrically corrected with an overall RMSE of 1 pixel (15m and 30m) for each subsystem.

3.3.4 Map data

Digital Elevation Model (DEM) was produced for topographic analysis of elevation, slope and aspect from 1:25,000 digital topographic maps. The spatial resolution of the DEM was 25 m corresponding to that of the Landsat ETM+. In order to generate realistic thematic maps of forest attributes as derived by the k -NN technique, a thematic map of forest/non-forest classes is required as a mask map at first. As mentioned in the previous chapter, even though a digital forest map for the study area exists, it has not yet been updated. Therefore, a thematic map of forest and non-forest classes was produced based on the previous chapter.

3.3.5 Application of k -NN technique

The similarity between reference and target plots in feature spaces can be computed using different distance metrics. In this study, the similarity was computed with the Euclidean distance metric (Equation 3-1). With respect to the distance-weighting approach, three weightings were tested, including no weighting ($a=0$), inversely proportional to the distance ($a=1$) and inversely proportional to the square of the distance ($a=2$) in equation 3-2. The sum of weights ($w_{t,r}$) is equal to 1.

Once the distances among neighbors and their weights were calculated, the k -NN technique was applied to each pixel. An estimated value \hat{v}_t for a target plot t can be expressed as the weighted sum of the observed values v_r at the selected reference plots. The value for the target plot t is computed as:

$$\hat{v}_t = \sum_{r=1}^k w_{t,r} \times v_r \quad (3-4)$$

Reference window

In South Korea, forests represent a highly variable and complex structure due to varying ecological and topographical conditions, cultivation practices and site quality. Therefore, reference windows that are sensitive to these changes have been applied to select reference samples, which have similar conditions to a target plot.

- Horizontal reference area: The relationship between growing stock and image features depends on site conditions. When a large HRA is applied, estimates may be biased because plots within the training dataset on different site conditions can be selected as reference plots, and it also requires a high processing time. In this sense, a minimum HRA that may cover all the local variations in the forest attributes of interest is needed. The study area is relatively small, and thus the HRAs were divided into five areas from 10km to 50km radius. In this case, a target plot close to the boundary of the study area can use only a small number of available reference plots. For this, the distances between the XY-coordinates of target and reference plots were calculated by the Euclidean distance metric.
- Vertical reference area: Over complex terrain, forest structures are affected by micro-climates related to local topographic variables such as altitude, slope, and aspect (Shin *et al.*, 2001). The altitude above sea level varies from 20 to 1,157m, which should influence forest structure and growth (Park, 1988).
- Stratification by forest cover types: Ancillary information based on existing digital maps or field plots can be used either pre- or post-stratification, respectively. In this study, the field data were classified into three forest strata in terms of the proportion of the number of trees by the dominant tree species per sub-plot; coniferous, deciduous, and mixed forests (see Annex 1).

Image filtering

An image enhancement technique called spatial filtering was applied. The goal of the spatial filtering is to improve spatial matching between field plots and pixels on the imagery. For this purpose, a 3 x 3 window mean filter, which corresponds to the plot design defined for each sub-plot point, was constructed, as shown in Figure 3.4.

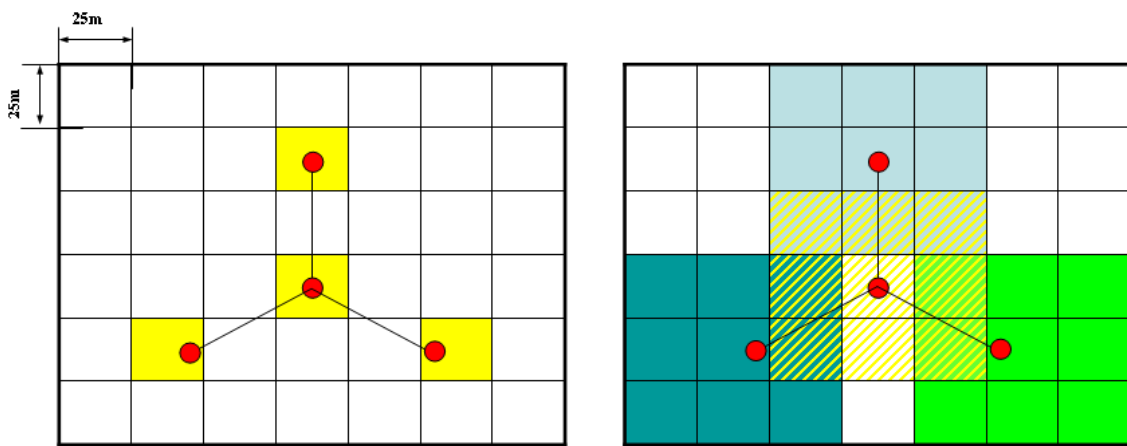


Figure 3.4: Original digital number (left) and mean digital number for the 3 x 3 window filtering (right) for each sub-plot centre within the plot design defined using Landsat ETM+.

Weighting parameters for each band

Not all of the bands in a single image share the same influence on the estimation of a forest attribute. The development of weighting parameters for each band was carried out by applying the downhill simplex optimization method developed by Nelder and Mead (1965). The weighting parameters (b_i) for each band (i) were computed to minimize RMSE with a value of k and the reference windows (Equation 3-3).

3.3.6 Evaluation

For determining an optimal operational option and for selecting an optimal value of k in the k -NN technique, two evaluation techniques were applied:

- Minimum RMSE (Root Mean Square Error) and bias by the leave-one-out cross-validation method and
- Confusion matrix by attributes' classes.

After obtaining an independent estimate from different operational options within the k -NN model for each of the pixels in the training dataset, the estimates were evaluated using an estimation error, which measures how well a model estimates the response value of a new observation. For every trial, the precision of estimates was examined using the RMSE, relative RMSE (RMSE%), and bias:

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}, \quad RMSE\% = \frac{RMSE}{\bar{\hat{y}}} \times 100 \quad (3-5)$$

$$bias(\bar{e}) = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (3-6)$$

where y_i and \hat{y}_i are the observed values and the estimated values on plot i , respectively, $\bar{\hat{y}}$ is the estimated mean of the estimates, and n is the number of plots. The RMSE% and bias for each operational option were compared to determine an optimal value of k and operational option.

In addition, the confusion matrix was also used to evaluate the accuracy of estimates for growing stock classes. The estimated and observed values were divided into four classes; 0-50, 50-100, 100-150, and above 150 m³/ha. The overall accuracy (OA) was compared for different reference windows and different numbers of neighbors.

3.4 RESULTS

3.4.1 Satellite images

The digital numbers (DNs) corresponding to each field point were extracted from both images, while topographic variables such as elevation, aspect and slope were extracted from the produced DEM data. They were used to calculate Minnaert constants for each band (Table 3.4). The estimated values of k ranged from 0.2731 to 0.5173 and from 0.2994 to 0.5181 for Landsat ETM+ and ASTER, respectively. With the estimated Minnaert constants, topographic effects on both images could be reduced.

Table 3. 4: Estimated values of the Minnaert constant for the satellite images

Constant (k)	Bands								
	1	2	3	4	5	6	7	8	9
Landsat ETM+	0.4299	0.4428	0.7473	0.2731	0.5173	-	0.4299	-	-
ASTER	0.2994	0.3813	0.5181	0.4531	0.3994	0.4573	0.4144	0.433	0.3349

With the two topographically normalized images, the relationship between the growing stock and their DN (Table 3.5) was analyzed. The correlation was very low; this result is similar to a study that used Landsat TM for the Korean forest (Lee *et al.*, 1994). When comparing the correlation by satellite data, Landsat ETM+ proves to have higher correlations than does ASTER, except for the band 3. In particular, the shortwave infrared (SWIR) bands for ASTER have the lowest correlation with the growing stock.

Table 3. 5: Correlation coefficients between DN on the topographically normalized images and the growing stock based on field plot data ($p_{0.05} = 0.138$, $p_{0.01} = 0.181$)

	Band								
	1	2	3	4	5	6	7	8	9
Landsat ETM+	0.20**	0.14*	0.16*	-0.18*	0.06	-	0.14*	-	-
ASTER	0.10	0.01	-0.19**	-0.08	0.01	-0.01	0.01	0.02	0.01

Notes: * and ** are significant at the 0.05 and 0.01 probability levels, respectively.

ESTIMATION OF FOREST ATTRIBUTES

Two different satellite data were employed independently and in combination as data source. Analyses were conducted using common operation options in the k -NN process, inverse distance weighting for neighbors and the non-reference window.

The precision using the combined data set could be slightly improved as compared to using the images separately when $k < 12$. Estimates using ASTER were more precise than those using Landsat ETM+ for any number of neighbors, as shown in Figure 3.5. The difference in RMSE% between the two images was approximately 10% ($10 \text{ m}^3/\text{ha}$) when $k = 1$, and it tended to decrease with increasing k .

Although the estimation that used ASTER was more precise, the identification of the optimal operational options in the k -NN technique was performed using Landsat ETM+ only. With regard to their spatial extends, Landsat imagery allows for estimating for a large area at a provincial level, while ASTER is only available for a smaller area at a municipal level (see Figure 3.2).

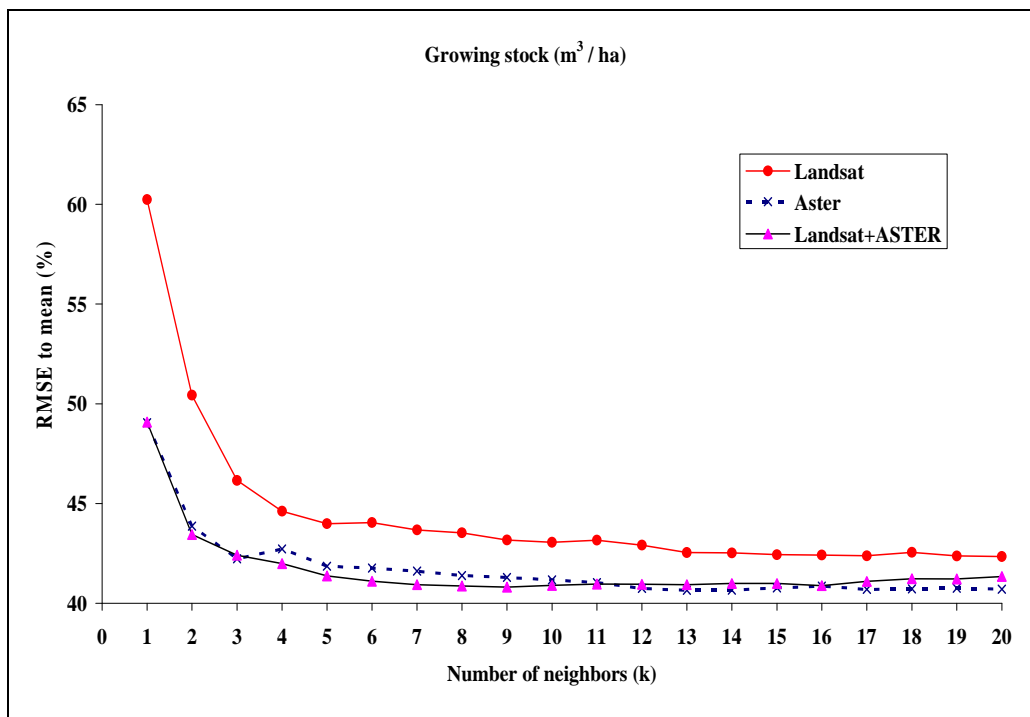


Figure 3. 5: Comparison of relative RMSE for different images and different numbers of neighbors (k).

3.4.2 Characteristics in the k -NN process

Selection of reference window

Horizontal reference area (HRA)

Table 3.6 shows the summary of the number of available reference plots by horizontal radius. When the minimum radius (10 km) was applied, the number of available reference plots at a target plot ranged from 26 to 78 plots.

Table 3. 6: The radii of the horizontal reference areas and the minimum, maximum and mean number of the field plots as a reference plot

Radius	Minimum number of reference plots	Maximum number of reference plots	Mean number of reference plots
10km	26	78	50
20km	70	184	117
30km	120	190	171
40km	177	190	189
50km	190	190	190

The variation in estimates depends more on the numbers of neighbors rather than on the HRAs, as shown in Figure 3.6. The RMSE% decreased rapidly as the value of k increased from 1 to 4, and decreased only slightly for larger k for all HRAs. Despite the small number of available reference plots at the 10 km radius of HRA (HRA-10km), the RMSE% was slightly smaller or similar to the other HRAs. The highest RMSE% was observed at the 40 km radius for any number of neighbors.

The biases of the estimates decreased with increasing number of available reference plots, i.e., larger reference radius when $3 < k < 15$. The biases with the HRA-10km were below $\pm 1 \text{ m}^3/\text{ha}$ when $k > 3$. This result was similar to the maximum HRA (50 km radius). The difference between the two HRAs was no more than $\pm 1 \text{ m}^3/\text{ha}$ when $3 < k < 12$. As a result, the HRA-10km was found to be the most efficient HRA in the given sampling intensity.

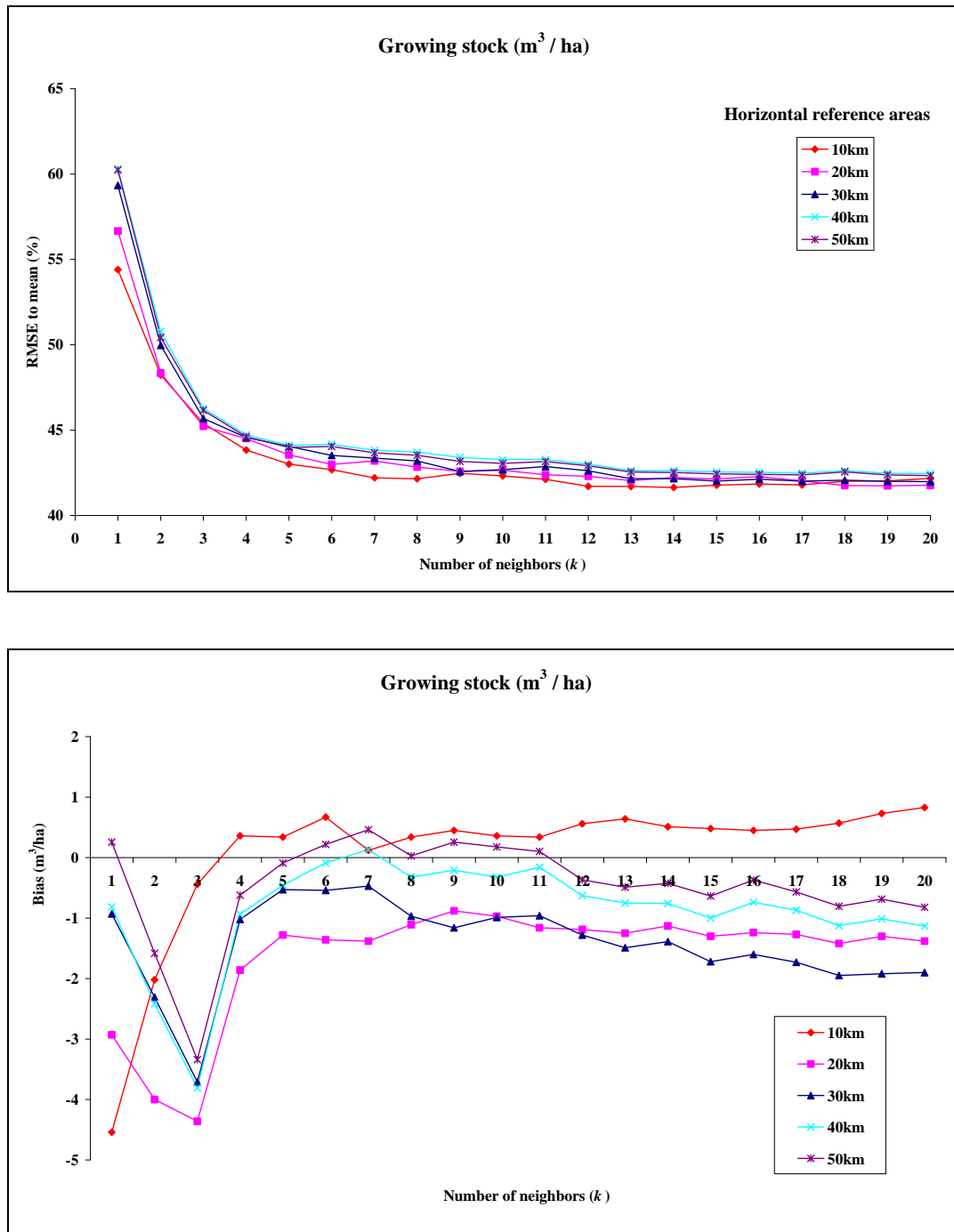


Figure 3. 6: Relative RMSE (RMSE%) and bias for the horizontal reference areas and different numbers of neighbors (k).

Stratification of field data by VRA and forest cover type

The vertical reference area (VRA) was divided into four altitude classes based on the number of available reference samples: 0-150m (44 sub-plots), 150-300m (85 sub-plots), 300-450m (47 sub-plots), and above 450m (15 sub-plots); thus the maximum number of reference plots at the above 450m altitude class was only 14 plots. Figure 3.7 shows the distribution of the number of field plots for the altitude classes. The reference samples for each forest cover type was also unequally used; coniferous (72 sub-plots), deciduous (66 sub-plots), and mixed forests (53 sub-plots).

The trend in the RMSE for both stratification windows was similar to that observed with the HRAs (Figure 3.8). The RMSE% for the stratification by forest cover types was lower than that for the VRA by approximately 2%. The biases of estimates for both approaches were lower than $\pm 2 \text{ m}^3/\text{ha}$ when $k > 1$. Accordingly, the stratification by forest cover types gave more precise results than the VRA.

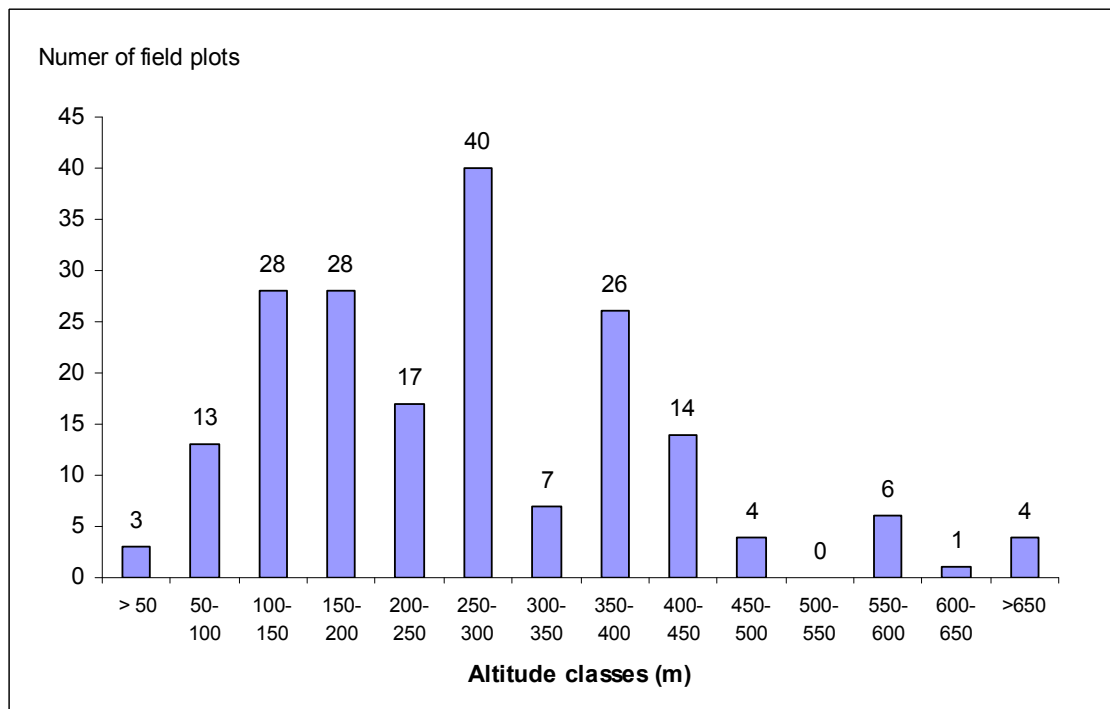


Figure 3. 7: Distribution of the number of field plots for the altitude classes.

ESTIMATION OF FOREST ATTRIBUTES

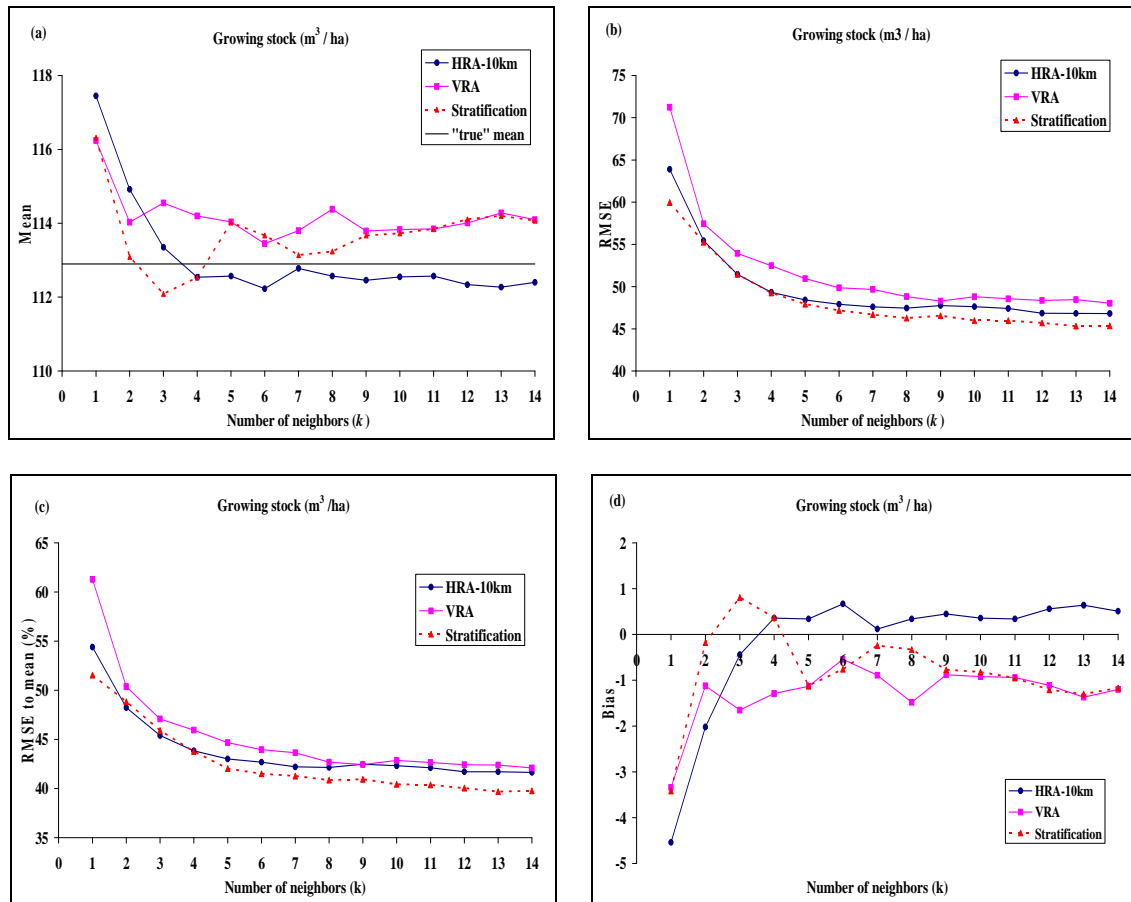


Figure 3.8: Estimated mean (a), RMSE (b), RMSE% (c), and bias (d) for the HRA-10km and stratifications by the VRA and forest cover types and different numbers of neighbors (k), where the “true” mean value was computed based on field plot data.

Comparison of stratification vs. HRA-10km

The means of the estimates for the HRA-10 km were underestimated when $k > 3$ and they slightly decreased with increasing value of k , whereas the means of the estimates for the VRA and stratification were overestimated, except for the $k = 3$ and 4 nearest neighbors. The difference in RMSE% between the HRA-10km and stratification increased with additional neighbors when $k > 4$.

The estimation for stratification was more precise than that of the HRA-10km. With regard to the bias of estimates, however, the estimation with the HRA-10km had a lower bias when $k > 2$. The bias of the estimates for the HRA-10km was slightly lower than that for stratification. The difference, however, was small.

Distance-weightings

The results of applying the three different distance-weighting functions for growing stock estimation are shown in Figure 3.9 (a). With weighting functions, estimates were more accurate than without the weighting function ($a = 0$). The weighting with the inversely proportional to the distance ($a = 1$) gave slightly better results when $k < 10$.

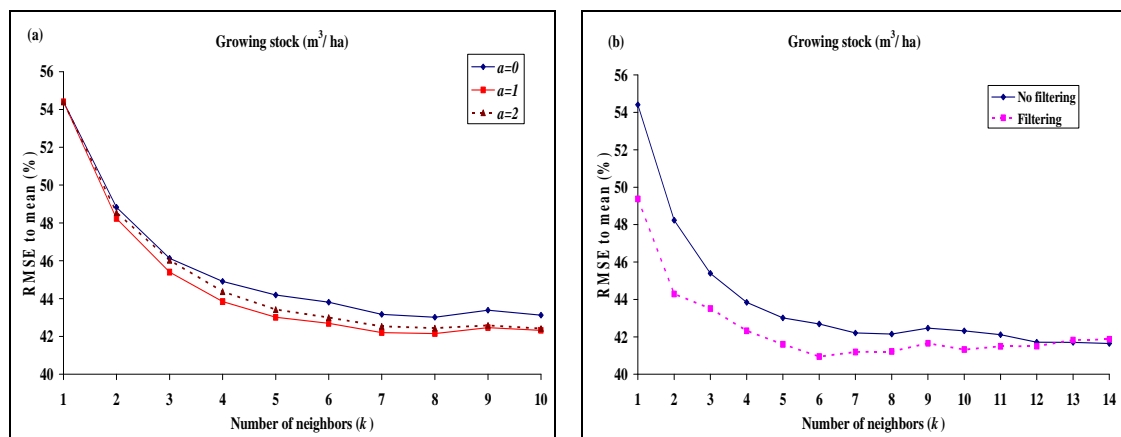


Figure 3. 9: Relative RMSE for the different neighbor weighting functions (a) and image filtering technique (b) with the HRA-10km and Landsat ETM+.

Image enhancement

The result of the spatial filtering technique is shown in Figure 3.9 (b). Slightly more precise results can be obtained by using the 3 x 3 pixel window mean filtering in the estimation. The precision increased as the number of neighbor plots decreased when neighbors were smaller than $k = 12$. There were no benefits when using more than $k = 12$ nearest neighbors as the RMSE% were similar from this point.

Selection of number of nearest neighbors

Different criteria were taken into account to select an optimal value of k : RMSE, RMSE%, bias and overall accuracy. The optimal value of k was determined such that the RMSE% and the bias of the estimates were minimized or stabilized, and the overall accuracy was high.

The RMSE decreased with increasing value of k until a minimum RMSE was reached with the HRA-10 km (Figure 3.6) and for the stratification (Figure 3.8). The minimum may not be reached before $k = 20$, depending on the size of field data as reference data. In particular, the RMSE% rapidly decreased when the value of k increased from 1 to 5 for both reference windows. The RMSE varied from 46.8 to 63.9 m³/ha and from 45.1 to 60 m³/ha, respectively. The minimum RMSE was 46.8 m³/ha when $k = 14$ with the HRA-10km, and 45.05 m³/ha when $k = 20$ for the stratification. However, at $k = 5$ nearest neighbors for both windows, the RMSE showed little tendency to level off, and the values were 48.4 and 47.9 m³/ha, respectively.

The variation in RMSE% at values of k larger than $k = 5$ were nearly stable. When the RMSE% for $k = 5$ was compared to those for $k = 14$ and $k = 20$ that gave the minimum RMSE, their difference was no more than 2 % for both windows. Therefore, for selecting the optimum value of k using the cross-validation, the relatively stable RMSE, which has leveled off at a value of k , could be a more efficient method than the minimum RMSE.

The growing stock in the confusion matrix consisted of four classes: 0-50, 50-100, 100-150 and above 150 m³/ha. The overall accuracy (OA) in classification comparing the HRA-10km and stratification is shown in Figure 3.10. The OA ranged from 0.35 to 0.41 and from 0.35 to 0.42, for HRA-10km and for stratification, respectively. The results of this study were modest because the study area and the training data set were relatively small. The estimation for the stratification also gave a higher accuracy than that with the HRA window when $k > 2$. The highest OA was observed for $k = 5$ and $k = 18$ for stratification, resulting in a value of 0.42.

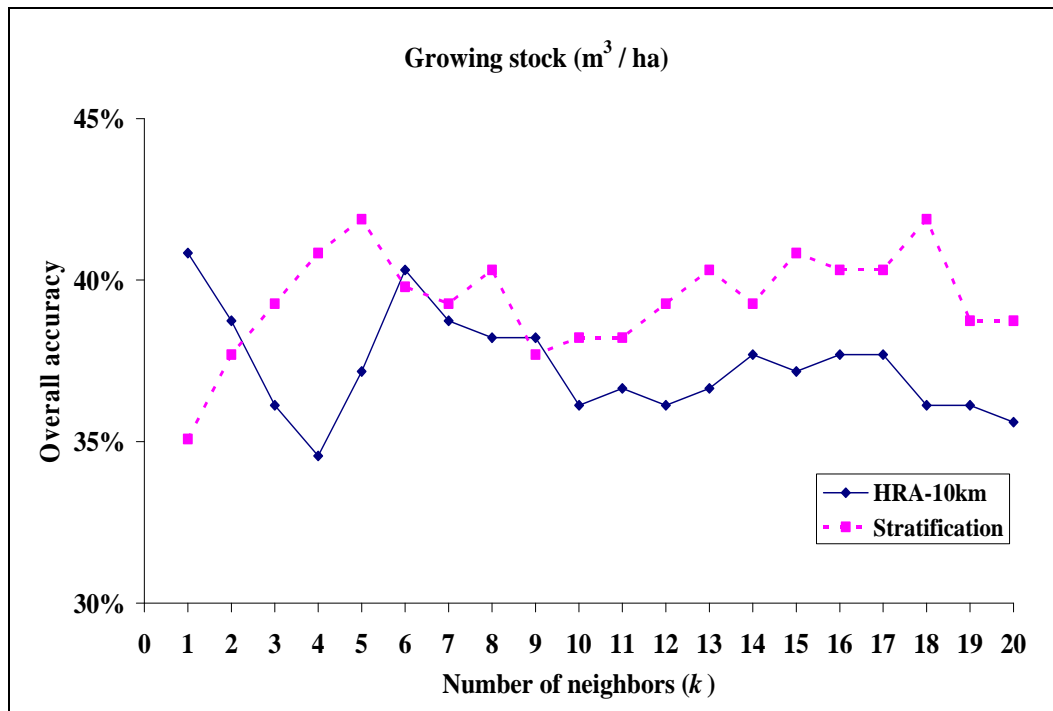


Figure 3.10: Overall accuracy for different reference windows and different numbers of neighbors (k).

Weighting parameters for each band

As presented in Table 3.5, the relationships to growing stock were different for the Landsat ETM+ bands. In order to find weighting parameters for each band, the downhill simplex optimization method was applied for both reference windows with the value of $k = 5$. The estimation errors for both windows decreased by less than 1% only through the weighting parameters, as presented in Table 3.7.

Table 3.7: Weighting parameters for each band and reference window

	Landsat ETM+ bands						RMSE
	1	2	3	4	5	7	(m ³ /ha)
HRA-10km ($k=5$)	1.0521	0.9355	1.0311	1.0253	1.0149	1.0309	48.42 (48.21)*
Stratification ($k=5$)	1.3659	1.4473	0.9463	1.0356	0.3602	0.6688	47.93 (45.41)*

*The values in parentheses are the RMSEs using equal weightings ($b=1$) for each band.

3.4.3 Map production

Forest map

The study area map was drawn from the existing digital forest map, which was produced from the 3rd NFI cycle, because changes have been small. The forest/non-forest map was produced based on the finding in chapter 2 by combining the Landsat imagery and field plot data used in this chapter.

The 191 sub-plots from 39 clusters were classified into three forest cover types and served as training data. The forest strata of each sub-plot were defined by the proportion of number of trees by dominant tree species (see Annex 1). Table 3.8 shows the number of field plots for each stratum: coniferous (72 sub-plots), deciduous (66 sub-plots) and mixed (53 sub-plots). In addition, 39 points within the non-forest area such as water, road and agricultural and residential land were extracted from the digital forest map. The pixel-wise classification by the NNC ($k = 1$) that gave the most accurate result in chapter 2 was applied. The Landsat image was classified into forest and non-forest and then the forest was subdivided into three forest cover types (Figure 3.11). Table 3.8 presents the result of the error matrix of the NNC classified image and field plot data. Overall accuracy reached 74% and the estimated value of kappa was 0.65. This result is similar to other studies that used Landsat TM for the Korean forest conditions (Lee, 1991; Cho, 2002).

Table 3. 8: Error matrix for assessing the classification accuracy of the NNC classified image and field plot data

Classification		Field plot data				Total	User accuracy
		C	H	M	Non		
Classified Image	C	55	9	5	0	69	80 %
	H	9	44	6	3	62	71 %
	M	5	7	36	0	51	75 %
	Non	3	6	6	36	51	71 %
Total		72	66	53	39	230	Overall accuracy 74 %
Producer accuracy		76 %	67 %	68 %	92 %		

Kappa value = 0.65

Growing stock map

In order to produce a thematic map of growing stock over the entire study area, coordinate information and digital numbers of center points for each pixel within the study area were extracted. Then, with the selected number of neighbors ($k = 5$) that were determined for the selected reference window (HRA-10km), estimates of the growing stock per pixel unit over the study area were computed with the operational options given in Table 3.9, producing a continuous digital layer (Figure 3.11).

Table 3.9: Characteristics for growing stock and forest maps in the k -NN process

Operational options	Growing stock map (continuous)	Mask map 2 classes: forest, non-forest
Satellite source	Landsat ETM+	
Distance metric	Euclidean distance metric	
Distance-weighting for neighbors	Inversely proportional to the distance ($a = 1$)	-
Value of k	5	1
Spatial filtering	No filtering	No filtering
Reference window	Horizontal Reference Area of 10km radius	-
Feature weighting	Equal ($a = 1$)	Equal ($a = 1$)

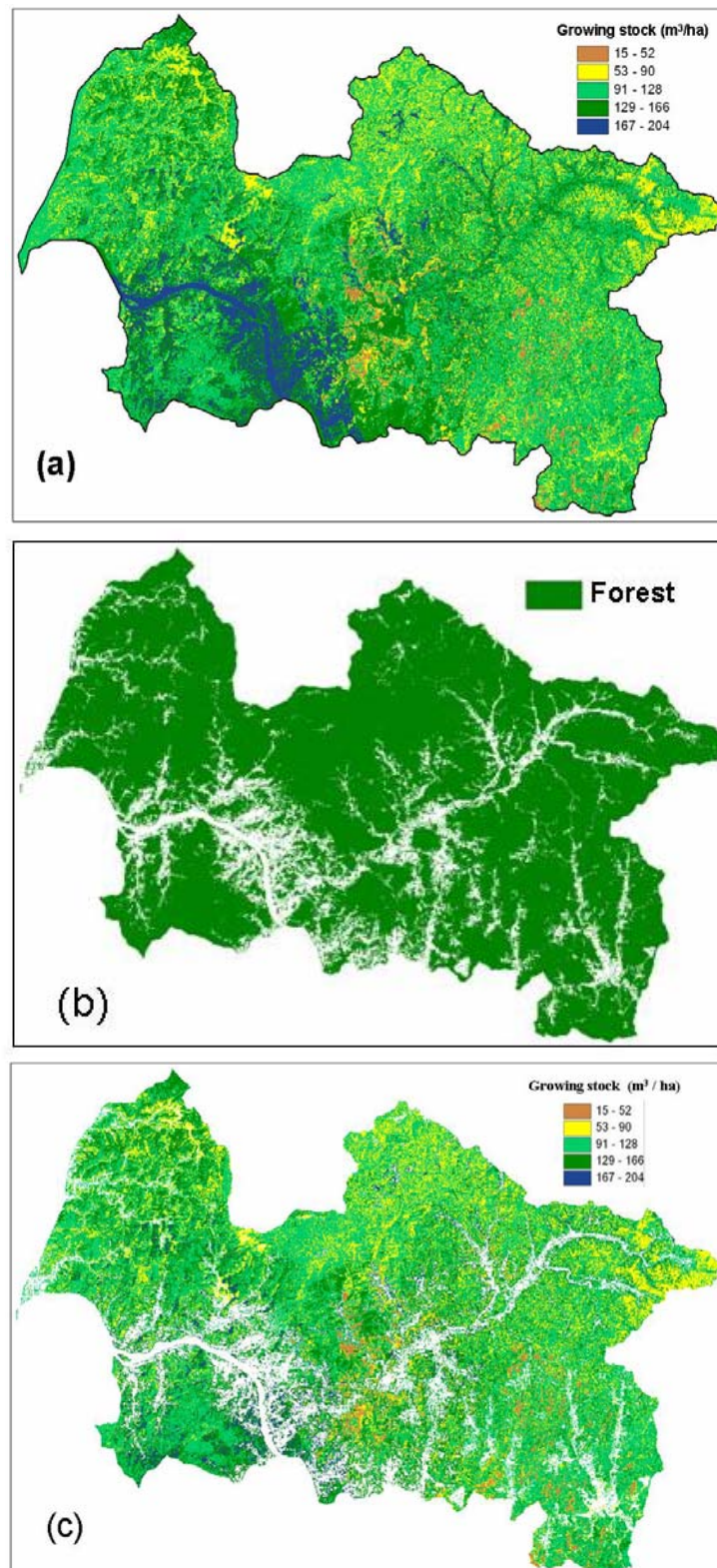


Figure 3. 11: Continuous thematic map of growing stock (a), forest and non-forest map as mask map (b), and growing stock map within the forest (c) over the study area, which can be used as an artificial model forest in chapter 5.

3.5 DISCUSSION AND CONCLUSION

Since 1990, the k -NN technique has been widely used in large area forest inventories to produce geo-referenced information, forest thematic maps, stratified area estimation, and small-area statistics by integrating forest inventory data and remotely sensed data. This chapter was conducted to analyze the possibility of the k -NN technique in the Korean NFI. To address this objective, a pilot study was implemented for a small area. Several operational options were tested, including different satellite sources, reference windows for the selection of reference samples, distance-weightings for neighbors, image filtering, and weightings for spectral features.

Data sources for the k -NN technique

In most previous studies, the estimation error measured with RMSE% was between 50-80% when estimating growing stocks. Despite a small number of reference plots ($n=191$) used in this study, the RMSE% for growing stock was relatively small, ranging from 40 to 60%. The estimation error may be sensitive to forest development; that is, the forest of the test area has a relatively small variation in growing stocks since it is relatively young, mostly less than 50 years old (KFS, 2004a).

In order to successfully achieve the k -NN technique, the key is to obtain a sufficient training data set that have to be involved in all variations that occur within a characteristic of interest. In the current Korean NFI, GPS-based field data have been collected every year in about 800 clusters (3200 sub-plots) over the entire country. If the field data serve as training data in the k -NN estimation, this may be sufficient.

Two satellite images (Landsat ETM+ and ASTER) were used independently as data source, as well as in combination. Since the combined data set allowed more variations in feature spaces over a single image, the precision with the combined data set was slightly better than using each independently. In addition to combined data set, as the spectral responses at a target point vary by season under natural conditions, if multi-seasonal satellite data are used as the data source, the precision should improve (Tokola *et al.*, 2001).

The comparison between the two images is difficult because the images were acquired during different seasons. When the two sensors were compared, the estimation using ASTER was more precise than that using Landsat ETM+. However, when the spatial scale is considered, the swath width of ASTER is approximately 60

km, which is smaller than that of Landsat ETM+ (185 km), and may be adequate to provide estimates for a small-area when sufficient field data is available. The use of ASTER imagery for large area inventories is expensive and there is a trade-off between precision and budget. When considering the spatial match between the field plot and its spatial resolution, the plot size (500m²) is nearly comparable to the spatial resolution of Landsat ETM+.

Characteristics in the k-NN process

The *k*-NN technique is very flexible for combining ancillary information. In this study, forest cover types and geographical variables were used as reference windows to search reference plots at a target plot. The RMSE% for the smallest HRA (a 10 km radius) was higher than that for the stratification by forest cover types, but the bias was small. This indicates that the forest attributes in a complex landscape may be affected more by geographically neighboring forests than by homogeneous conditions. The VRA for the altitude classes, however, did not improve the precision. It may be related to the availability and variability of reference samples for each class. In order to fully achieve stratification to the *k*-NN estimation, a training set includes equal numbers of training data sets for each stratum. The number of available reference samples for each class by the VRA, however, was different and relatively small so that the precision did not improve.

Nilsson (1997) pointed out that the estimates from the *k*-NN technique tend towards the mean with an increasing value of *k*. This can also be observed in this study, as shown in Figure 3.12; most of the estimates were aggregately classified into the 100-150 m³/ha class around the mean, by increasing of the value of *k*. This means that the variability of estimates is to be decreased with increasing *k*. This problem can be overcome by stratification, since the tendency towards the mean does not affect the estimates for each stratum equally. This result was shown in Figure 3.10. The overall accuracy for the HRA-10km tend to decrease with increasing number of neighbors when *k* > 5, whereas for the stratification, it oscillated more or less around a trend when *k* > 2, i.e., the variability of estimates was preserved.

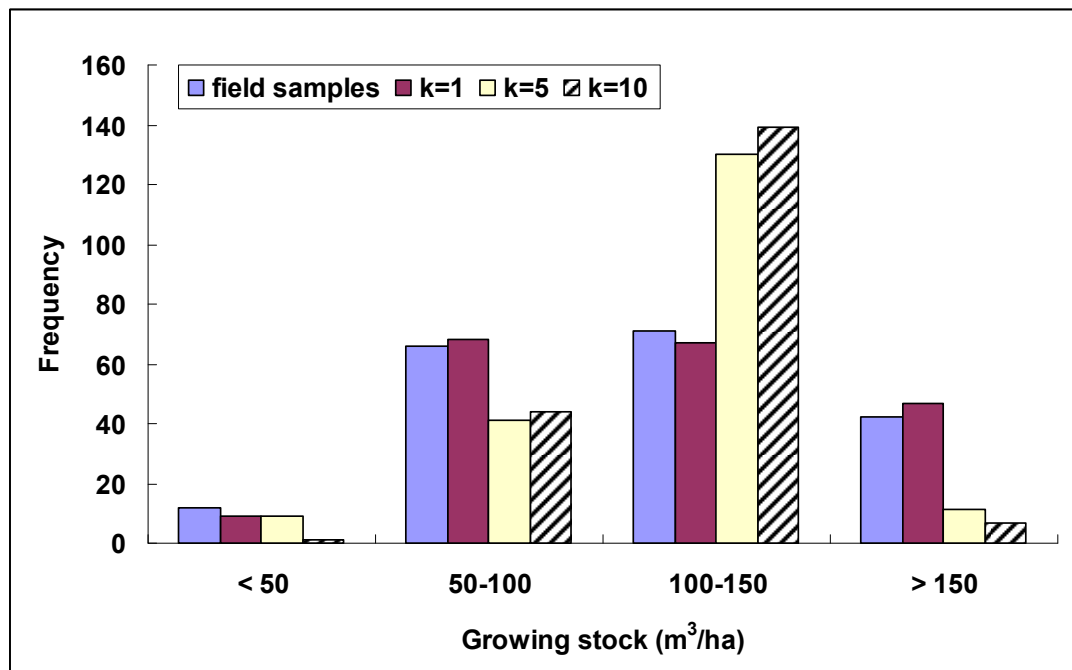


Figure 3. 12: Distribution of frequencies for the growing stock classes and different numbers of neighbors with the HRA-10km.

Stratification can be applied either individually or in combination with different stratification criteria. Katila and Tomppo (2001) applied a two-stage stratification by soil classes and horizontal boundaries as a reference window, and the precision of estimates improved. They found that stratification can also reduce the bias of estimates when reference samples for each stratum are sufficient. In this study, however, the training data did not allow for such stratification. Currently, digital maps via the Korean FGIS are increasing. If the current NFI data serve as training data and various digital maps can be used as a basis for stratification, more reliable information can be produced.

Although the locations of field plots were recorded by a GPS receiver, a spatial filtering technique for reducing spatial matching error was applied because it is difficult to obtain accurate location information in the Korean forest conditions. The estimation error for the 3 X 3 window mean filter improved. The spatial filtering, however, causes changes in original digital numbers of satellite images. This means that the filter's smoothing effect can mask small differences between digital numbers.

ESTIMATION OF FOREST ATTRIBUTES

Considering that the k -NN technique is sensitive to their differences, spatial filtering method is inappropriate for a pixel-based approach. Consequently, spatial filtering has a trade-off between spatial matching errors and a loss of original information in feature spaces.

The correlation between bands and forest attributes may not be equal for all bands within a single image. To address this concern, the weighted Euclidean distance was tested. The estimation error, however, was not affected by these weights in the downhill optimization method because the weighting parameters computed were quite similar or small. This result was also presented by Franco-Lopez *et al.* (2001). The weighting parameters for bands must be related to the relationship with the forest attributes of interest and the spectral variability in feature spaces.

Even though there are several errors and drawbacks involved in the k -NN technique, the k -NN maps, by integrating forest inventory data and satellite data, allow to support forest management planning and forest statistics reporting for small- and large area units. However, to successfully apply the k -NN technique in the Korean NFI, the following points must be considered:

- (i). Considering the complexity of landscape in the Korean forest conditions, location error of field plots is a major error source for matching them to satellite data. Above all, accurate locations of field plots have to be acquired;
- (ii). With regard to the number of reference samples, if field data from the current NFI serve as training data it may be sufficient. In addition, if management inventory data at a regional level are available, the precision for a small area unit can be improved; and
- (iii). Finally, the application of stratification for the selection of reference samples can improve the precision of estimates and can preserve the range of variability of estimates. In recent years, digital maps applicable for stratification in the forestry sector are increasing and topographic factors over mountainous area can also be considerable as ancillary information.

4. PLOT DESIGN OPTIMIZATION

4.1 INTRODUCTION

Forest management requires reliable current data on the forest resources. In forest inventories, sampling techniques are used for data collection and estimation because a complete census would require an enormous expense in terms of time and money (Johnson, 2000). It thus becomes necessary to develop the most efficient sampling strategy for a forest inventory, for which three basic design elements must be considered: (a) the overall design of selecting sample units (sampling design; see chapter 5), (b) the design of individual sample units themselves and the method of selecting individual trees within these units (plot design), and (c) the estimation design for estimating attributes of interest. In this context, the plot design defines how to select sample trees to be observed around the sample points, or variables to be observed at the sample sites selected by a given sampling design. Thus, plot design considerations refer to the characteristics of the observation unit under a sampling design (Kleinn *et al.*, 2002).

Clustering of samples is an appropriate tool for increasing the efficiency of a field data-collection since cluster plots can reduce travel time considerably by concentrating measurements, although setting up separate sub-plots can increase the costs. Cluster plots have been widely applied as sampling unit or plot designs for natural resources assessment, e.g. large area forest inventories (EC, 1997; BMVEL, 2001; Khunrattansiri, 2005).

This is especially true for large and/or inaccessible areas. In large area forest inventories, much of the expense (often 50%) is incurred by traveling to the clusters (Scott, 1991), so it is often more efficient to sample one large cluster than to sample two smaller clusters in one working day (Arvanitis and O'Regan, 1972). Scott (1998) demonstrated cluster plots can provide the opportunity to "spread out" the sampling unit; thus, more "independent" or "new" information can be collected at each plot as opposed to simply measuring one large plot. In most forest inventories, a cluster plot gives more accurate information per unit cost than any other plot designs (Tokola and Shrestha, 1999). Furthermore, it is also possible to research into spatial variability.

PLOT DESIGN OPTIMIZATION

When planning a cluster plot design, several questions are relevant to the statistical grounds (Scott, 1981; Kleinn, 1996):

- (i) what is the spacing between clusters?
- (ii) what is the shape of the cluster?
- (iii) what is the number of plots per cluster? and
- (iv) what is the sample plot configuration?

For determining an optimal solution, prior information about the spatial distribution and correlation of the variables of interest is needed. The covariance function can be found by simple transformations of variograms, such as are frequently used in geo-statistics (Kleinn and Jost, 1994). The function of distances between pairs of points can be used to compare the efficiencies of various sample designs (Korhonen and Maltamo, 1993; Scott, 1998).

In natural resource assessment, computer simulated designs have been applied to compare various alternative designs. They offer a cheaper and more flexible planning tool. Cluster plot simulation has been studied by Köhl (1986) and Päivinen (1987). Köhl (1986) used the intra-cluster correlation to evaluate the efficiency of sampling designs, while Päivinen (1987) evaluated the variation between repeated inventories and costs. Kleinn (1994) compared the performance of line sampling to other forms of cluster sampling based on a standard covariance function. Furthermore, he also compared the statistical and economic efficiency of various cluster shapes (Kleinn, 1996). Scott (1991) developed an optimal cluster plot that was derived from the relationship between variance and cost function for the Forest Health Monitoring Program of the United States. Tokola and Shrestha (1999) compared various cluster plot designs to determine the most efficient cluster plot by using a model forest population by combining field sample data and satellite image data. In that study, the design efficiency coefficient (DEFF), based on the intra-cluster correlation coefficient and cluster size, was used to compare different cluster configurations and the variation between repeated simulations under given sampling designs. Scheuber and Köhl (2003) demonstrated that cluster sampling is an appropriate tool for the assessment of Non-Wood-Goods and Services (NWGS), where an optimal cluster plot can be found by visualizing the variance structures by means of spatial statistics such as variograms.

Additionally, an appropriate cost function must be developed with regard to economic efficiency. With a cost function or cost map, the method would be more suitable for the planning of practical forest inventories. The primary cost of conducting an inventory is personnel time, e.g., planning, field work, supervision, and data synthesis and analysis. The costs associated with simple sampling can be divided into the following categories: total cost of the inventory (C_T), overhead or non-field costs (C_O), and field costs (C_F), which include the cost of travel from a base of operations to and from the field (Johnson, 2000). The fundamental cost function is then

$$C_T = C_O + C_F \cdot \quad (4-1)$$

The costs of interest to the person designing a sampling operation are those associated with the field operations. The simple cost function provided by Hansen *et al.* (1953) is as follows:

$$C_F = c_1 n + c_2 nm \quad (4-2)$$

where c_1 = sampling unit (cluster) cost, such as locating the cluster,
 c_2 = sub-sampling unit (i.e., sub-plot in cluster sampling) cost,
 n = number of clusters in the sample, and
 m = number of sub-plots within one cluster.

In this cost function, the first cost (c_1) is defined as those costs that vary in relation to the number of clusters (i.e., sampling intensity) taken in the sample. These costs include those of selecting, traveling to, and locating a cluster and designating the sub-plots within each cluster. The second cost (c_2) describes costs that also vary in relation to the number of sub-plots per cluster, so-called “cluster size”. These costs are dependent upon the inventory activities (such as measuring the trees, recording the information, *etc.*) per sub-plot, and the spatial arrangement of sub-plots within a cluster (Khunrattansiri, 2005). Scott (1981) demonstrated that in developing a cost function, some variables that are integers or continuous cannot be optimized; thus, prior practical information, such as the feasible cluster size (m), cluster shape, and distance between sub-plots, is necessary to find optimal solutions. In some studies, the cost function was used to determine an efficient cluster size, where the cluster size was derived from the relation between the cost ratio (c_1/c_2) and the intra-cluster correlation coefficient. However, if the ICC is negative, it cannot be used to compute the optimum cluster size (Scheuber and Köhl, 2003; Khunrattansiri, 2005).

There are limited studies on plot design optimization for forest resources assessment in South Korea. Furthermore, most studies have focused mainly on plot size and shape for a fixed-area plot at a specific forest stand (Kim, 1966; Lee and Han, 1984; Yim and Shin, 2006). Although a cluster plot was used in the 4th NFI cycle, the plot design was dependent on empirical findings, giving primary consideration to practical aspects in the field (KFRI, 1996). A small number of studies on plot design options, particularly the cluster plot, have been conducted, notwithstanding the need for an optimally sound and cost-effective plot for the Korean forest situation (KFS, 2005; Shin and Han, 2006).

The objective of this chapter is to assist in determining the most efficient cluster plot for the Korean forest conditions. To do this, a pilot cluster is empirically designed and then several cluster configurations which are applicable to the Korean NFI are simulated. With respect to statistical soundness, several statistical factors for key attributes are taken into consideration. Furthermore, a cost function in terms of units of time is also developed to measure the economic efficiency.

4.2 MATERIALS AND METHODS

4.2.1 Sample cluster unit

To determine an efficient cluster plot, a pilot cluster plot was taken to collect field data. Figure 4.1 illustrates the pilot cluster plot, which consists of 10 sub-plots and allows for the evaluation of different spatial arrangements of sub-plots, varying distances between pairs of sub-plots and different cluster sizes. The maximum distance from the center sub-plot was set at 100m; this distance can be considered a maximum feasible distance given the steep terrain conditions (KFS, 2004b). A total of 25 clusters were collected in the field. As the focus of this study was on plot design optimization, the clusters were randomly selected within one county.

On each sub-plot, all trees with at least a 6 cm diameter at breast height (DBH) were measured, within a circular plot with a fixed-area of 500 m² (Yim and Shin, 2006). The variables, including DBH, tree species, height, *etc.* were measured.

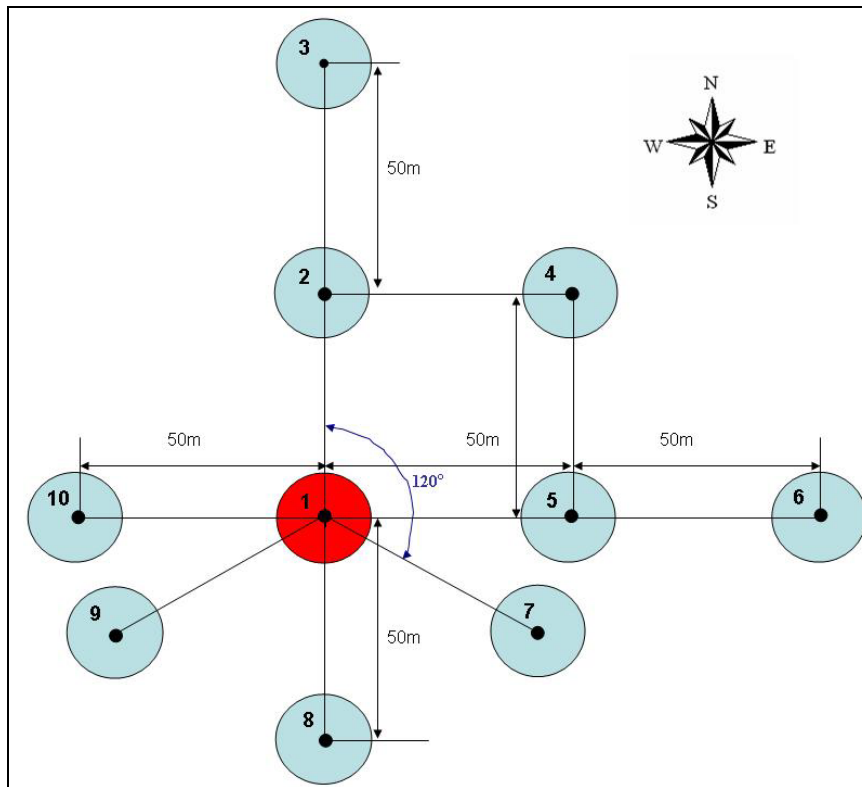


Figure 4. 1: Pilot cluster plot design as used in this study comprising 10 circular sub-plots. Various standard cluster shapes can directly be formed from this design.

4.2.2 Cluster configuration

In this study, several cluster configurations frequently used for large area forest inventories were established on the basis of the given pilot cluster plot, as represented in Figure 4.2. Considering that one cluster has to be carried out by a field-crew within one working day, the cluster size ranged between 3 and 5 sub-plots.

Among the given cluster configurations, the shape of cluster 3 was used in the 4th Korean NFI cycle (KFRI, 1996), while the modified triangular shape (like cluster 4) has been used in the current NFI cycle (2006-2010) and in the United States (USDA, 2005). The square and L-shaped clusters have been adopted, for example, in the German (BMVEL, 2001) and Finnish NFIs (Katila and Tomppo, 2001), respectively.

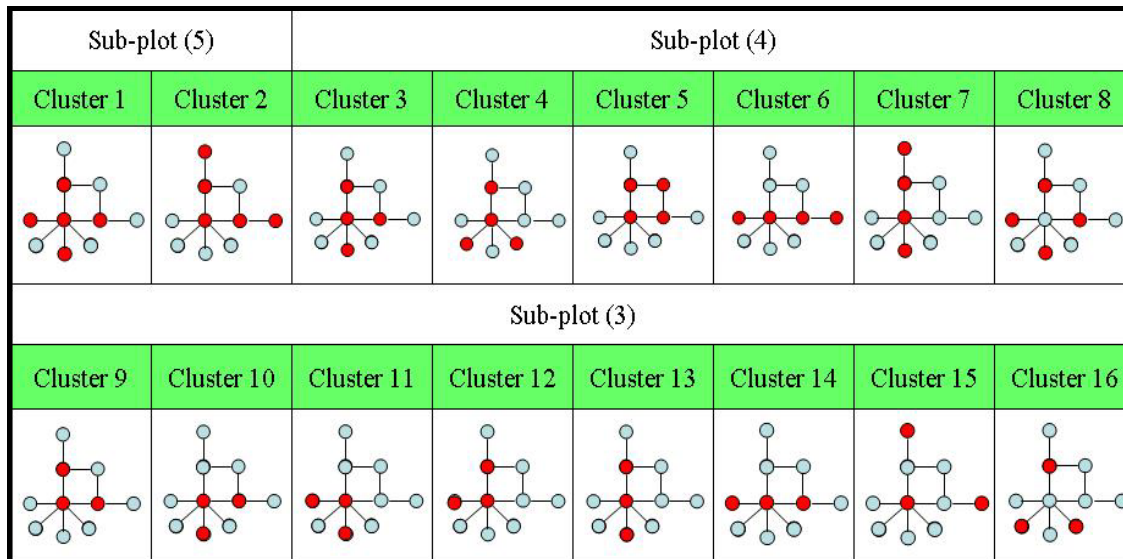


Figure 4. 2: Cluster configurations for different cluster sizes in this study.

4.2.3 Statistical Analysis

In order to evaluate and compare the given cluster configurations, the following characteristics are taken into consideration:

- covariance function for determining an efficient distance between sub-plots,
- intra-cluster correlation for evaluating geometric arrangements of sub-plots, and
- standard error for the cluster size.

Covariance function

The advantage of the cluster plots is that the required travel time per sub-plot is minimized. One disadvantage compared to a random layout of the sub-plots is that a possible spatial correlation of variables among the sub-plots, namely “Spatial auto-correlation”, may influence the precision of the sampling design (Kleinn and Morales, 2001). This disadvantage can be reduced by increasing the distance between pairs of sub-plots per cluster. In most studies, the spatial variation has been described with the average covariance function (Kleinn 1994; 1996). The most efficient distance can be determined by combining the covariance function between pairs of sub-plots with the travel cost.

The covariance is the measure of how two observed values at a pair of points vary in relation to each other. If there is little or no association between the two observations from a given pair of sub-plots, the covariance will be close to zero. In this case, each cluster provides more independent information. The covariance of the two observed values, y_1 and y_2 for a pair of sub-plots, is symbolized by $cov(y_1, y_2)$ and is defined as:

$$cov(y_1, y_2) = E[(y_1 - \mu_1)(y_2 - \mu_2)] \quad (4-3)$$

where μ_1 and μ_2 are the means of y_1 and y_2 , respectively. In this context, the two means are equal ($\mu_1 = \mu_2$). For illustration, the distribution of distances between pairs of sub-plots in the 25 pilot clusters of 10 points is shown in Figure 4.3, where the distances available range from 26m to 150m, with varying frequencies.

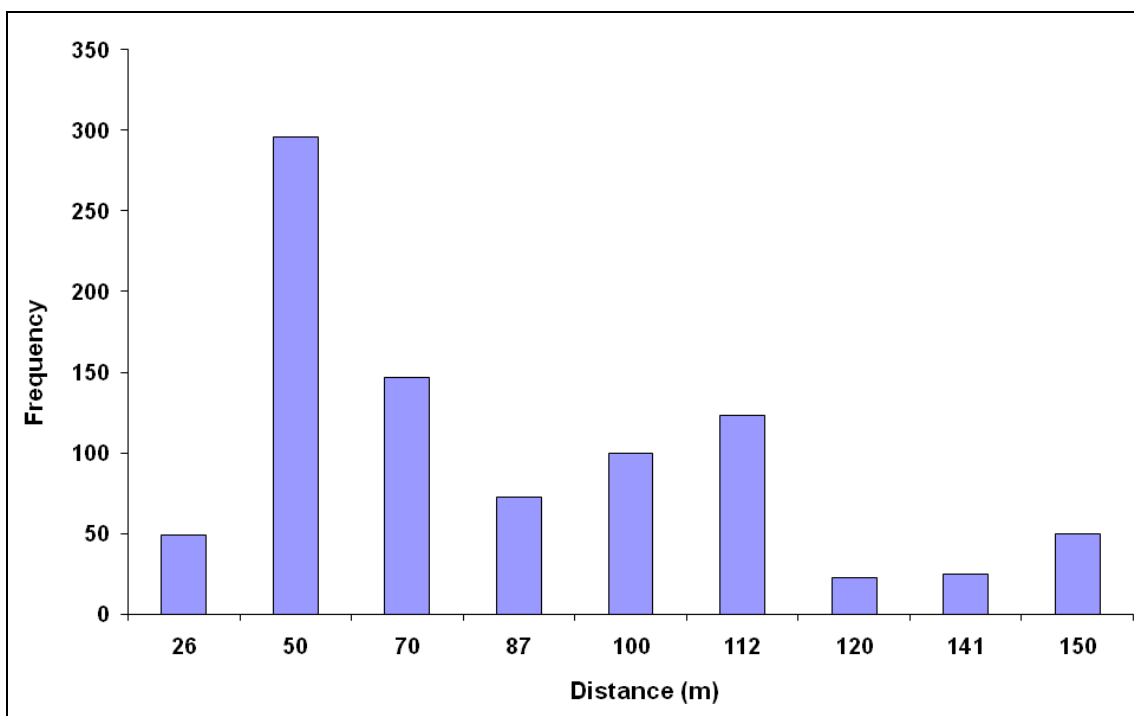


Figure 4. 3: Distance distribution between pairs of points for the 25 clusters in the pilot study.

Intra-cluster correlation

The intra-cluster correlation (ICC) is an important statistic used to determine the efficiency of a cluster plot (Tokola and Shrestha, 1999). The ICC measures the rate of homogeneity within a cluster. If sub-plots within a cluster are very similar to each other (high positive ICC), nothing is gained by making measurements on many sub-plots. If they are considerably different (negative ICC), the precision is high (even higher than that of simple random sampling). The latter case, however, rarely occurs in forest inventories. In cluster sampling, the total variance is a combination of two components: the variances between and within clusters. This can be expressed in the form shown in Table 4.1. The intra-cluster correlation coefficient (ρ), ranging from -1 to +1, can be defined as follows in terms of the variances (Cochran, 1977):

$$\rho = \frac{SS_b - SS_w / (m - 1)}{SS_t} \quad (4-4)$$

where SS_t : the total sum of squares between elements in the population,
 SS_b : the sum of squares between clusters, and
 SS_w : the sum of squares within clusters.

Table 4. 1: Variance table for cluster sampling (Scheuber and Köhl, 2003)

Source of Variation	Degrees of Freedom	Sum of Squares
Between clusters (s_b^2)	$n-1$	$m \sum_{i=1}^n (\bar{y}_i - \bar{y})^2$
Within cluster (s_w^2)	$n(m-1)$	$\sum_{i=1}^n \sum_{j=1}^m (y_{i,j} - \bar{y}_i)^2$
Total (s^2)	$nm-1$	$\sum_{i=1}^n \sum_{j=1}^m (y_{i,j} - \bar{y})^2$

where n = the number of clusters, m = the number of sub-plots per cluster,
 $y_{i,j}$ = the observation per sub-plot j in cluster i ,
 \bar{y} = the estimated mean per sub-plot, and
 \bar{y}_i = the estimated mean per sub-plot in cluster i .

Standard error

To determine an efficient cluster size, the standard error can be calculated as an indicator of precision, to evaluate the variability among the estimates for different cluster shapes and cluster sizes. This value is frequently denoted as the relative standard error (SE%):

$$SE\% = \frac{\frac{s}{\sqrt{n}}}{\bar{y}} \times 100 . \tag{4-5}$$

The total variance in Table 4.1 is approximately estimated because this variance estimator ignores the ICC. The total variance can be expressed as a function of the ICC (Equation 4-6), which is another way to estimate the variance for cluster sampling (Cochran, 1977), ignoring the sampling fraction:

$$s^2 = \frac{SS_t}{nm - 1} [1 + (m - 1)\rho] \tag{4-6}$$

where for $\rho = 0$, the variance estimator is identical to that of simple random sampling of size nm (Scheuber and Köhl, 2003).

4.2.4 Cost analysis

Cost is always an important limiting factor in forest inventory planning. In the cost function equation (4-2), the costs (c_2) depend on the cluster size (m) and total walking distance per cluster. Thus, this section addresses the total walking distances, which depend on the distances among sub-plots, and thus on the spatial arrangement of sub-plots in each cluster. The results of a pilot time study (KFS, 2004) were modified to calculate the required times per cluster.

First, the total walking distances within a cluster (D_w) were computed to assess the total times required by cluster configurations. These depend on the distances among sub-plots, cluster size, and spatial arrangement of sub-plots, as well as the walking travel route. In this study, two walking travel routes were considered: one (travel route

PLOT DESIGN OPTIMIZATION

1) chosen to minimize the walking distances, and the other (travel route 2) to reduce the location error of sub-plots, by surveying all sub-plots from a given starting sub-plot. When travel route 2 is used, the total walking distances computed include the distances from the starting sub-plot to the remaining sub-plots and back, while only the return distance from the last sub-plot to the starting sub-plot is included in travel route 1, as shown in Figure 4.4.

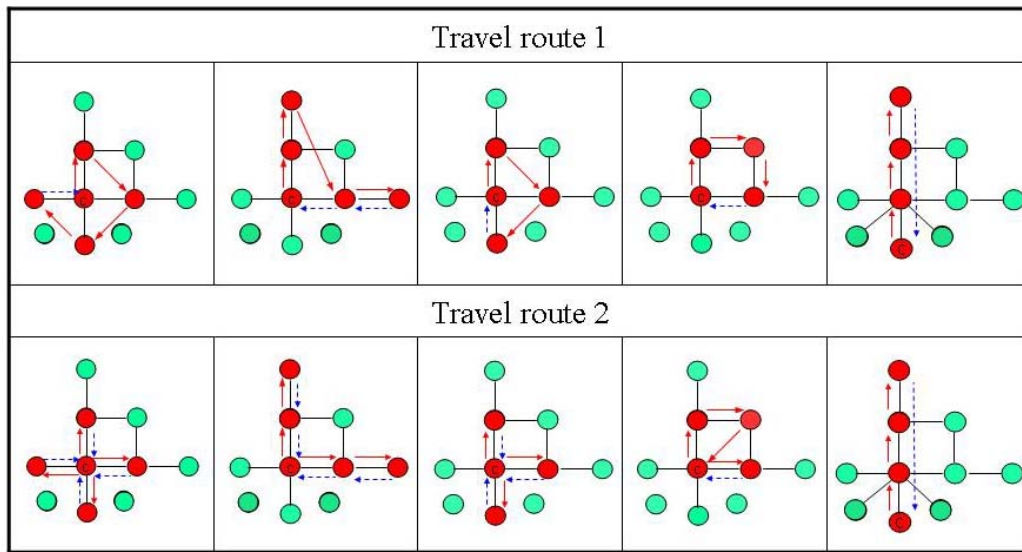


Figure 4. 4: Examples: walking distances for different travel routes and cluster configurations; the dotted lines are the return distances.

Thereafter, the total times required (T_c) for each cluster configuration is computed as follows:

$$T_c = T_t + T_s * m + (D_w / S) \quad (4-7)$$

where: T_c = total time for each cluster configuration,

T_t = total travel time from the office to a starting sub-plot within a cluster and back,

T_s = total time of establishment and inventory activities per sub-plot,

m = number of sub-plots within a cluster,

S = walking speed (in meters per minute), and

D_w = total walking distances (in meters) within a cluster.

4.3 RESULTS

4.3.1 Statistical characteristics

Covariance function

To determine the most efficient distance between pairs of sub-plots, distances from 26 m to 150 m were compared, as allowed by the pilot study. For the given attributes (growing stock, basal area and number of trees), the covariance functions were calculated for all pilot clusters. Figure 4.5 depicts the covariance functions for the key attributes, which oscillate more or less around a trend. They show the typical shape expected in natural resource assessment: the covariance functions increase with decreasing distance (d), and for $d=0$, the covariance equals the population variance of key attributes (Kleinn and Morales, 2001). They show a similar shape, which on average tends to descend from the maximum to close to zero with increasing distance when distance $d < 100$ m. At the smallest distance (26 m), the covariance for all key attributes was highest. The covariance functions approached absolute zero at distances of 87 m and 120 m.

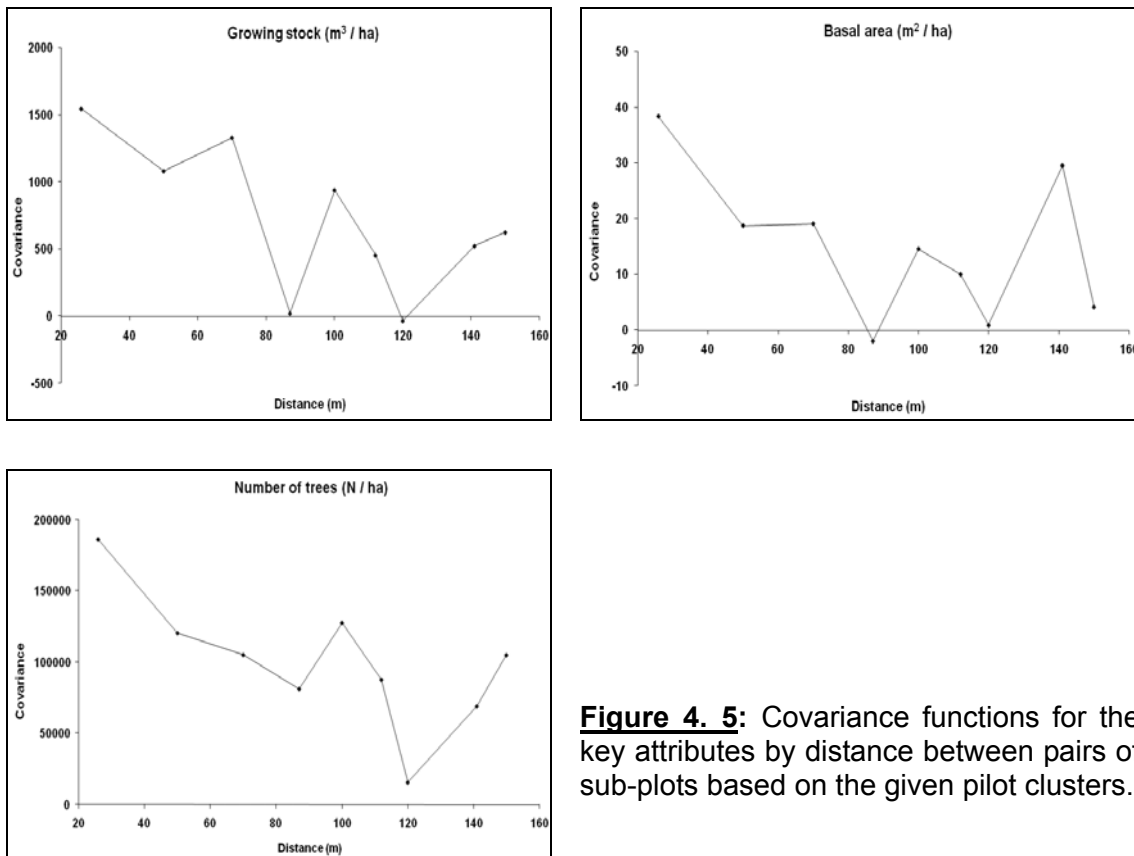


Figure 4. 5: Covariance functions for the key attributes by distance between pairs of sub-plots based on the given pilot clusters.

Intra-cluster correlation

The ICCs for clusters of 3-5 sub-plots were computed to evaluate the design efficiency of the cluster configurations for the key attributes, as shown in Figure 4.6. The ICC of the number of trees was much higher (0.46 ~ 0.74) than those of the other attributes regardless of cluster size. In the case of clusters of 5 sub-plots, the cross shaped cluster (cluster 1) was found to be more effective than the L-shaped cluster (cluster 2). For clusters of 4 sub-plots, the modified triangular cluster (cluster 4) gave better results than the other shapes on all the key attributes; the next best shape was the diamond shaped cluster (cluster 8).

When comparing the L-shaped clusters of different sizes, the ICC for cluster 9 ($m=3$) was found to be smaller than that for cluster 2 ($m=5$). This result was also observed in the line clusters (clusters 7 and 13, clusters 6 and 14). In the case of the L-shaped clusters of 3 sub-plots (clusters 9-12) which were designed with different directions of sub-plots, cluster 10 (with sub-plots to the east and south) generated better results than the others, except for the number of trees. In the smallest cluster size ($m=3$), the triangular cluster with a distance of 87m among sub-plots (cluster 16) had the lowest ICCs for all key attributes: -0.01, -0.04, and 0.46 for the volume, basal area, and number of trees, respectively.

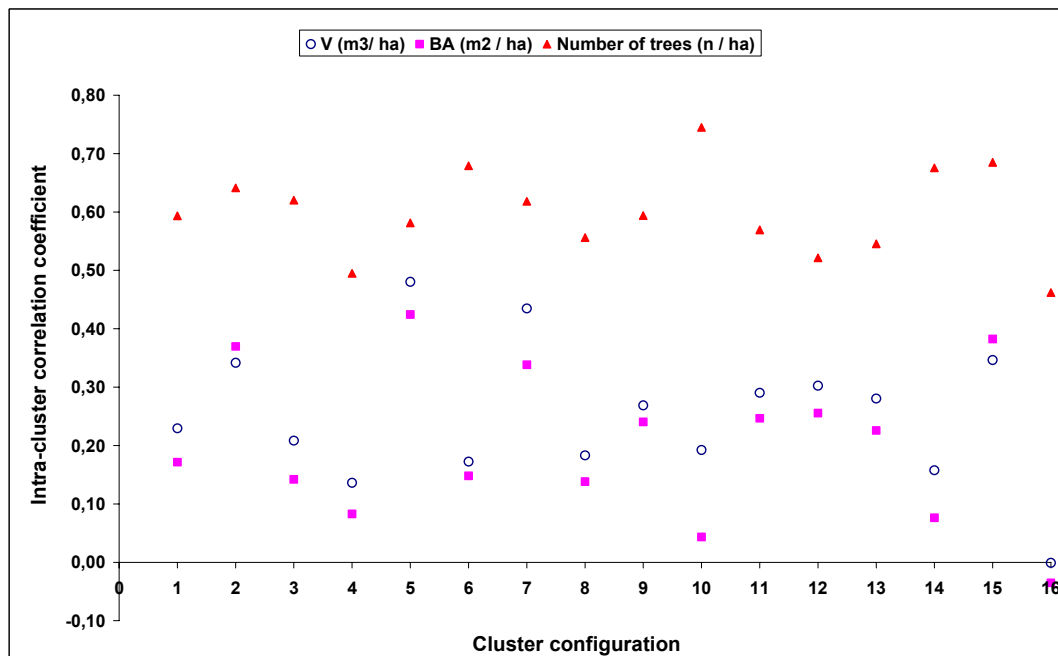


Figure 4. 6: Intra-cluster correlation coefficients for different key attributes and different cluster configurations (see Figure 4.2).

Standard error

The relative standard errors (SE%) of the estimation were computed from the 25 clusters of the pilot study. The SE% for each key attribute show similar trends, as shown in Figure 4.7. The variations for basal area and volume were small, whereas the number of trees had a relatively large variability. For clusters of 3 sub-plots (cluster 9-16), the SE% observed had higher variability than those of the other cluster sizes. The differences among cluster sizes, however, were not notable; this means that with smaller field efforts (smaller numbers of sub-plots) the precision is not significantly affected.

When different spatial arrangements for clusters of 4 sub-plots (clusters 3-8) were compared, the variability according to their shapes varied with the forest attributes. For the basal area, the modified cross shape (cluster 3) used in the 4th NFI gave the smallest SE% (2.4%), while for the volume (cluster 6) and number of trees (cluster 7), line clusters resulted in the smallest SE%. A comparison of the clusters having identical shapes and cluster sizes but different orientations of sub-plots (like clusters 6 and 7, clusters 9-12 and clusters 13 and 14) was especially interesting. Significant differences were observed between different sub-plots orientations. When comparing the results of the ICC, the clusters with smaller ICCs gave more precise results, except for number of trees (cluster 6, cluster 10, and cluster 14).

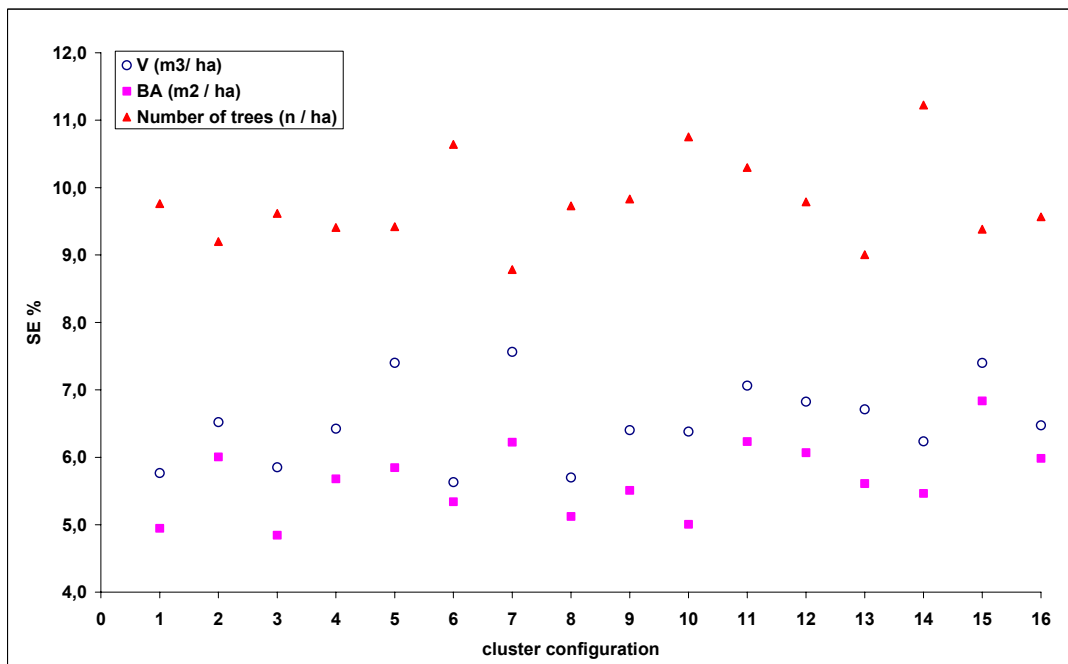


Figure 4. 7: Relative standard errors for each attribute by cluster configuration.

4.3.2 Cost analysis

Table 4.2 summarizes the travel and inventory times based on the pilot research carried out by a three-person field crew (KFS, 2004). On average, the travel time from the office to a starting sub-plot per cluster and back was about 134 minutes. The walking speed in the forest was very slow, averaging 8.3 m/minute. The times for inventory activities and establishment of each sub-plot were about 64 minutes and 10 minutes, respectively. In this study, the inventory activities included not only timber-production variables such as DBH, height, and tree species for standing trees (DBH \geq 6cm), but also multi-resource variables such as regeneration, dead trees, understory vegetation, soil attributes, *etc.*

Table 4. 2: Summary of the time study for the pilot research (modified from KFS, 2004)

Description	Average time required (in minutes)
From the office to a closed road of a cluster	30
From a closed road of a cluster to a center sub-plot within a cluster	37
To establish a sub-plot within a cluster	10
Inventory activities per sub-plot	64
From a centre sub-plot to a sub-plot (distance : 50m)	6
Walking distance per minute	8.3m

The total walking distances per cluster for the two travel routes and different cluster configurations were simulated (see Figure 4.4). In practice, closed cluster shapes (square and triangle) are more efficient than open cluster shapes (line-shape) for equal cluster sizes. This was observed in our case: for clusters of 4 sub-plots (clusters 3-8), the total walking distance along travel route 1 increased from the square (cluster 5) to the modified cross (cluster 3), the modified triangular (cluster 4), the diamond (cluster 8), and the line (cluster 6 and 7), as shown in Figure 4.8.

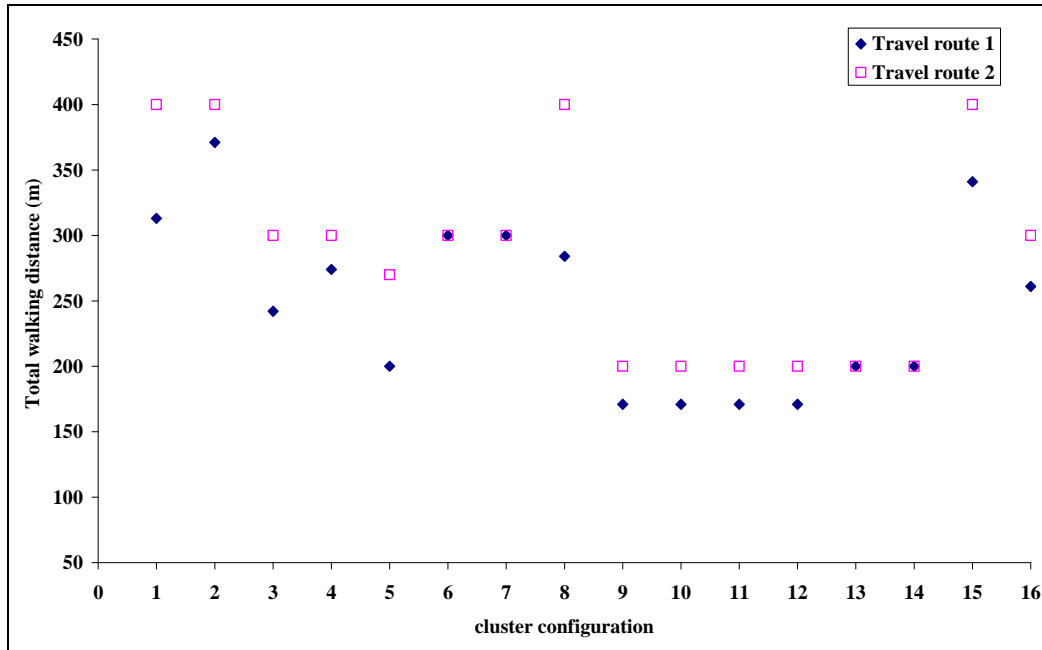


Figure 4. 8: Total walking distances for the two travel routes and different cluster configurations.

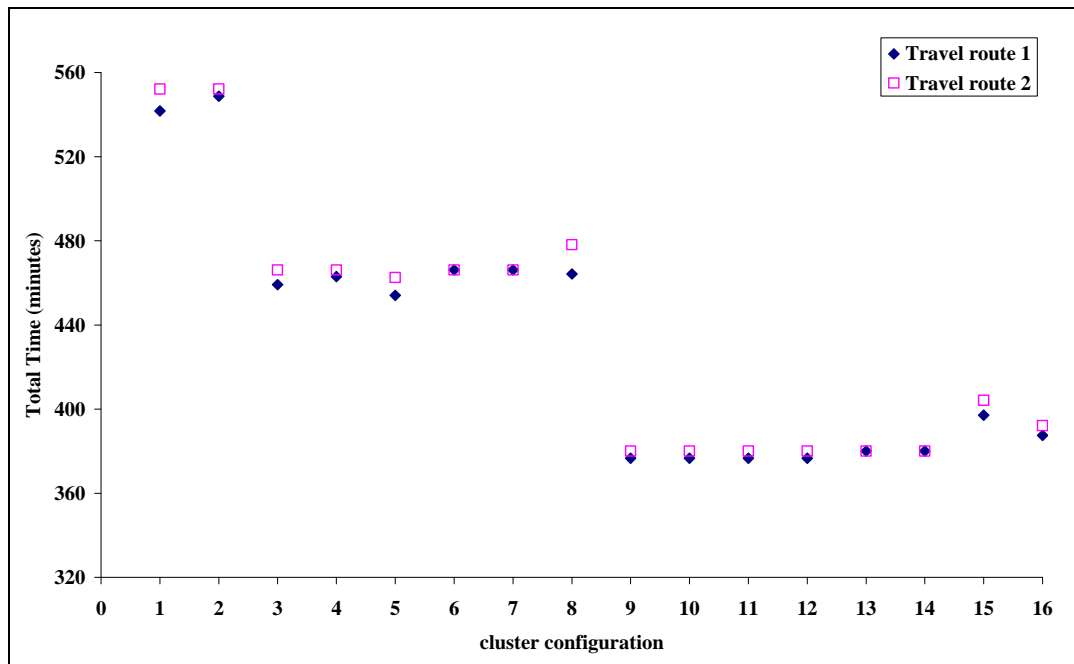


Figure 4. 9: Total times for the two travel routes and different cluster configurations.

PLOT DESIGN OPTIMIZATION

Travel route 2 required greater walking distances than travel route 1, except for the line clusters. The closed shapes (square and diamond) were particularly sensitive to the influence of the travel route chosen. In the case of cluster 4, the difference in total walking distance between the two routes was only about 26 m, whereas for cluster 5 (square shape) the difference amounted to about 142 m.

The costs were investigated here in terms of units of time, assuming that other costs (such as equipment, *etc.*) are constant for all cluster configurations. In the total time function (Equation 4-7), the cluster size (m) played an important role for determining the most cost-efficient cluster plot.

Figure 4.9 depicts the total times (in minutes) for the various cluster configurations. If the working day is defined as 8 hours (480 minutes), clusters of 5 sub-plots (clusters 1 and 2) cannot be carried out in a single day, while clusters of 3 sub-plots (clusters 9-16) have a residual time of about 90 minutes. As a result, with clusters of 4 sub-plots (clusters 3-8) the inventory activity can be fully accomplished in one working day. With the square cluster (cluster 5), the travel distance was minimized, but travel route 1 was not easy to implement in difficult field conditions. In practice, the modified cross (cluster 3) and triangular (cluster 4) clusters were found to be more efficient shapes for the given forest conditions. This is because the inventory activities could be fully achieved in one working day, and these two cluster shapes also reduced the errors in location of the sub-plots.

4.4 DISCUSSION AND CONCLUSION

Cluster plots are very useful as sampling units in natural resource assessment over large areas, such as regional and national forest inventories. This chapter was conducted to determine the optimum cluster plot for Korean forests, which must consider both statistical soundness and cost-effectiveness. In this chapter, following three factors in cluster plot planning were analyzed: distance between pairs of sub-plots, geometrical arrangement of sub-plots, and number of sub-plots.

The pilot clusters ($n=25$) used in this study were randomly selected within a municipality unit. Although the spacing between the clusters is also an important factor that should influence precision, it was not addressed in this chapter.

What is the optimal shape of the cluster ?

Spatial auto-correlation is very relevant for the cluster plot optimization. It can be described by the covariance function of the sub-plots at a distance d apart from each other (Kleinn, 1994). From a statistical point of view, the covariance function is expected to decrease with increasing distance for typical forest settings. This means that a longer distance between pairs of sub-plots leads to a smaller covariance between observed values, and then can provide more additional information per cluster. However, a very long distance, such that the expected value of the covariance function approaches to zero ($cov(d) \rightarrow 0$), is obviously difficult to achieve in natural populations.

In this study, the covariance functions, however, did not decrease much more with longer distances beyond 150 m; this might be related to forest conditions such as highly fragmented forests, as shown in Figure 4.10. Under the forest conditions of the test area, the distance of 87m between pairs of sub-plots was found to be an efficient minimum distance. The covariance at this distance was lower than those at the other distances (50m, 71m, and 100m; found in cluster 3) as used in the 4th NFI system.

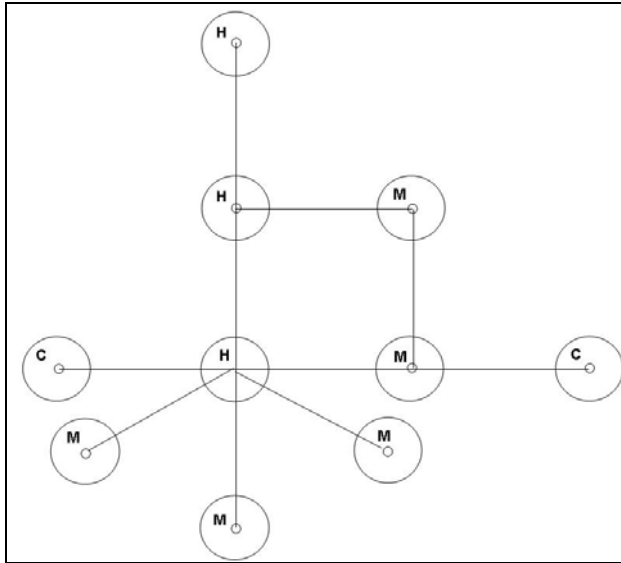


Figure 4. 10: An example: forest strata per sub-plot within a cluster (C; coniferous forest, H; deciduous forest, and M; mixed forest)

What is the optimal sample plot configuration ?

The ICCs for the given cluster configurations were computed to evaluate the homogeneity among sub-plots per cluster configuration. From a purely statistical point of view, open shapes like the line and L-shape are superior because of their larger spatial arrangement and the larger average distances between pairs of sub-plots. Additionally, they are usually better when the intra-cluster correlation is lower (Kleinn, 1994; 1996). In this sense, given a constant cluster size, the line-shaped clusters (clusters 6 and 7) are the preferable shape. However, in practice, it is better to use compact closed shapes such as a square (cluster 5) and a triangle (cluster 16), which reduce the walking distances per cluster. The cross-shaped cluster (cluster 1) is not preferable due to the relative proximity of the sub-plots.

The ICC for the cross shape (cluster 1), however, was smaller than that for the open cluster shape (L-shape, cluster 2) at the same cluster size ($m=5$). This result was also observed in the clusters of 4 sub-plots: the ICC for the line cluster (cluster 7) was higher than those for the modified triangular clusters (clusters 3 and 4). This means that in the natural environment, the correlation between sub-plots must be more sensitive to the forest structure and landscape conditions than to their shape and size. In the given pilot cluster, despite having the smallest cluster size ($m=3$), the triangle shaped cluster (cluster 16) had a smaller ICC for all key attributes. It is concluded that this cluster plot can provide more information than the others.

However, Kleinn (1996) pointed out that it is not possible to identify one cluster shape as generally superior, as the cluster shape also relies on the cost function to ensure that it will be practical in the field. In some simulation studies on cluster plot optimization (Kleinn, 1994; Tokola and Shrestha, 1999), the differences in precision between different cluster shapes were minor. In this study, however, the SE% was affected more cluster shape than by cluster size (Figure 4.7). This result may be relevant to the Korean forest conditions such as the complexity of landscape and forest structure.

What is the optimal cluster size ?

In this study, the cluster sizes were limited to the range of 3 to 5 sub-plots per cluster, and then analyzed by relative standard errors for the cluster sizes and cluster shapes (Figure 4.7). The difference in SE% between the cluster sizes of 4 and 5 was merely about 1%. In addition to the optimal cluster size, the cost of field operations then becomes a major factor when determining the cluster size and allocation of a sample. According to the pilot time survey for the multi-resource inventory (Table 4.2), the inventory time per sub-plot (about 60 minutes) increases with increasing number of the forest variables of interest, compared to the 4th NFI system. Assuming that this inventory time per sub-plot is constant, this time is an important factor in determining the optimum cluster size. As a result, clusters of 5 sub-plots could not be surveyed in one working day, as shown in Figure 4.9.

In South Korea, most of the working time is spent on reaching the target sample point due to the low forest road density (below 3 m/ha) (KFS, 2004a). In the pilot study, the traveling times accounted for about 30% of the total working time (Table 4.2). Moreover, it is difficult to reach the correct field location due to the complexity of the landscapes and forest structure. To address this concern, two travel routes were compared to minimize walking distance and to reduce location errors. Although the total walking distances for the closed shapes responded differently to the travel routes, the difference in total working time was small. Particularly in multi-resource forest inventories, the total working time is affected more by the time for inventory activities per sub-plot than by the traveling time.

PLOT DESIGN OPTIMIZATION

Consequently, the search for the most efficient cluster plot in difficult forest conditions is more driven by practical restrictions of accessibility, forest conditions, and user objectives than by statistical characteristics.

Although numerous forest attributes have been collected in the field, for this study only timber-oriented attributes were analyzed. In addition, the pilot field data were collected in a specific small area. There are more research needs to develop an efficient cluster plot for the Korean NFI: (i) field data to cover the forest conditions over the whole country are required; (ii) key attributes of interest in the field inventory must be determined before plot design planning; and (iii) the plot design cannot be performed independently but must be dependent on a given sampling design, i.e., spatial distribution of samples.

5. SAMPLING DESIGN OPTIMIZATION

5.1 INTRODUCTION

In South Korea, the process of National Forest Inventory began in 1971 and a stratified systematic sampling with cluster plots was applied. Forest cover types were identified and delineated from 1:15,000 black and white aerial photos and then used as stratification criteria for field sampling. Despite the increasing need for information on forest resources and technological development, the inventory design remained the same until the 4th NFI (1996-2005). The main goal of the NFI was to provide information for the reforestation plan over the destroyed forest areas. Thus, the NFI addressed estimates of the total growing stock for the entire country, as well as for different stratifications such as forest cover types, dominant tree species, age classes, etc. for forest conditions, ownerships, and administrative units.

At the planning stage of the NFI, a total sample size for the entire country was determined to fulfill a specified precision requirement of the total growing stock. For the NFI, a rotation system by provinces was applied and therefore the total sample size was divided into ten sub-sample sizes by provinces. This means that the NFI was carried out in one province per year over a 10-year period. Thus, the NFI could not provide reliable information for the entire country at the same time (KFS, 2002).

Currently, the NFI is in its fifth cycle (2006-2010) and has been reorganized and expanded to provide information for sustainable forest management. The inventory design has changed to a systematic cluster sampling and the NFI has been carried out in about 20% of the total sample size over the entire country per year (KFRI, 2006). However, to search the most efficient sampling design for forest resource assessment, more statistically based foundational research is required. Most of the research on sampling design for forest resources inventory was conducted on small study areas in the 1960s by Kim (1965, 1966, and 1973). Since then, there have been few studies on this topic (KFS, 2005; Shin and Han, 2006).

Many different sampling techniques for forest inventories have been suggested. Two of the basic sampling designs are random sampling and systematic sampling. In large area forest inventories, systematic sampling has been widely applied because systematic samples are well-spread across the population and give several advantages in practice. Another design is stratified sampling, which helps to reduce

the error variance in many cases. Stratification can take two forms, depending on whether the ancillary information is used before or after the sample selection: pre-stratification (mostly, stratified sampling) and post-stratification. Stratified sampling is feasible when the entire population can be divided into different sub-populations (e.g. forest cover types) since samples are independently selected in each stratum. This can be performed on maps, aerial photographs, or satellite imagery. Post-stratification can be combined with different sampling designs, for example, systematic sampling with post-stratification (Saborowski and Cancino, 2007). In this approach, samples are taken under a given sampling design and are then stratified into strata; that means in the samples a categorical or indicator variable is recorded as stratification criterion. These variables indicate which sample belongs to which stratum. Furthermore, the three major sampling designs may be used not only individually, but also in combination with different sampling designs such as in two or multi-stage sampling and multi-phase sampling (Lanz, 2000).

Alternative sampling designs for forest inventories can be evaluated and compared by several means. One way is to carry out actual inventories in a target forest (Kim, 1973; Saborowski and Cancino, 2007). Under normal circumstances, however, the necessary sample sizes in the evaluation of each sampling design cannot be realized over a large area. Computer simulations offer a cheaper and more flexible possibility. The main advantages of simulations are that they allow controlled experimentation and sensitivity analyses. Scott and Köhl (1993) developed an interactive computer program for an extensive forest inventory, called SIZE, which is able to simulate alternative sampling designs using combinations of several characteristics such as sampling designs, sample sizes, cost function, and precision levels.

In recent years, artificial populations for relatively large areas have been computed and generated by combining information from digital satellite data and ground truth data. Notwithstanding the drawbacks of artificial populations, which include deviations from reality, they also allow for the simulations of various sampling designs such as, for example, estimation of land use for systematic sampling from a land use classified image (Dunn and Harrison, 1993), cluster plot optimization (Tokola and Shrestha, 1999), and sampling simulation for small area statistics using outputs of a large area forest inventory (Katila and Tomppo, 2006).

The objective of this chapter is to optimize sampling design for the Korean forest conditions. To achieve this objective, different sampling designs and sample grid sizes are simulated and compared on the basis of an artificially generated forest population for a municipality area.

5.2 MATERIALS AND METHODS

5.2.1 Artificial forest population

In order to simulate various sampling designs, an artificial forest population was derived from the results elaborated in chapter 3. This procedure consists of the following three main steps:

- Modeling of a forest attribute of interest (growing stock map);
- Production of a forest cover map for stratification;
- Generation of an artificial forest population by combining the two thematic maps.

Using the *k*-NN technique, thematic maps of growing stock, forest/non-forest classes and forest cover types over the test area (Yang-Pyeong County) were generated with the characteristics listed in Table 3.8.

5.2.2 Simulation of sampling designs

The following four sampling designs that have been most commonly applied to large area forest inventories were compared:

- stratified systematic sampling,
- systematic sampling,
- systematic sampling with post-stratification, and
- systematic cluster sampling.

The characteristics of different sampling designs are summarized in Table 5.1.

SAMPLING DESIGN OPTIMIZATION

Table 5. 1: Summary of characteristics for different sampling designs (Cochran, 1977; Johnson, 2000; Köhl *et al.*, 2006)

Sampling design	Advantages	Disadvantages
Stratified random Sampling (STR)	<ul style="list-style-type: none"> - more precise than SRS - estimations for each stratum 	<ul style="list-style-type: none"> - need for ancillary information for stratification - time-consuming & expensive - complex to organize samples
Systematic sampling (SYS)	<ul style="list-style-type: none"> - more precise than SRS - simple to implement - simple to explain & control - well distributed samples - easy to combine with other sampling designs 	<ul style="list-style-type: none"> - lack of randomization of samples - no general unbiased estimator - "auto-correlation" between samples
Stratified systematic Sampling (sys+pre)	<ul style="list-style-type: none"> - more precise than STR - well distributed samples - simple to organize samples than STR 	<ul style="list-style-type: none"> - need for ancillary information for stratification - time-consuming & expensive
Systematic sampling with post-stratification (sys+post)	<ul style="list-style-type: none"> - estimations for each stratum - meaningful sub-division - quicker & less expensive than STR 	<ul style="list-style-type: none"> - unknown sample sizes for each stratum - laborious & dubious for stratification - need for definitions of strata at a field observation level
Systematic cluster Sampling (sys+clu)	<ul style="list-style-type: none"> - more information - more precise for an equal sample size - cost-effective 	<ul style="list-style-type: none"> - need for the optimum cluster plot as the sampling unit

Stratified sampling

In the Korean NFI, there are several possible types of stratification such as forest cover type, age class, DBH class, *etc.* for forest conditions, ownership, and political units (KFRI, 1996). The stratification by forest conditions is performed with aerial photographs. This study used the generated thematic map of forest cover types as the basis for stratification. Since the stratum sizes in the artificial population could be easily calculated from the thematic map, the proportional allocation method was adopted. The mean of estimates was computed as follows:

$$\text{Mean} \quad \bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h = \frac{1}{N} \sum_{h=1}^L N_h \bar{y}_h \quad (5-1)$$

where \bar{y}_h : estimated mean per stratum h ,

N_h : stratum size, and

W_h : stratum weight.

Systematic sampling

Systematic sampling comprises a large group of sampling designs that have one important characteristic from a statistical point of view: the lack of randomization of the sample to be selected. In other words, only the selection of a starting point is at random and from this random starting position, the remaining sample points of the sampling procedure follow a pre-defined pattern. Despite the problems with variance estimation, systematic sampling has a series of advantages with respect to practical and statistical considerations (Table 5.1). For these reasons, systematic sampling is popularly used for large area forest inventories in many countries. Applying the SRS estimators to systematic sampling produces a conservative estimation of standard error; that is, the true standard error is commonly over-estimated.

When systematic sampling is applied, the spatial spacing between samples is an important factor that should influence precision. The spatial spacing applied for large area forest inventories vary with countries and forest conditions. In this study, four sample grid sizes (1km, 1.5km, 2km, and 4km) were compared.

Stratified systematic sampling

There are cases in which the systematic sampling can be combined with stratified sampling. Stratified systematic sampling is similar to STR except that the sample points are systematically selected within each stratum, where starting points are independently determined for each stratum. This design was applied in the last NFI with a square grid size of 1km (KFRI, 1996).

Through actual sample sizes and stratum weights for each stratum are variable because the starting points are randomly selected for each simulation and therefore it is not possible to exactly determine the sample size before the sample is selected (Kleinn, 2007). However, since this design uses pre-defined strata and their stratum weights, the estimators for stratified sampling are taken into account (Table 5.2).

Table 5. 2: Mean estimators for the simulated systematic sampling designs (Cochran, 1977; Johnson, 2000)

Sampling designs	Mean estimator
Stratified systematic sampling	$\bar{y}_{sts} = \sum_{h=1}^L \frac{N_h}{N} \bar{y}_h$
Systematic sampling with post-stratification	$\bar{y}_{post} = \sum_{h=1}^L W'_h \bar{y}_h$
Systematic cluster sampling	$\bar{y}_{clu} = \frac{\sum y_i}{\sum m_i}$

where n : total sample size, L : number of strata,
 n_h : sample size for each stratum h , \bar{y}_h : estimated mean per stratum h ,
 W'_h : stratum weight per stratum h for post-stratification,
 y_i : total per cluster i , and
 m_i : cluster size per cluster i .

Systematic sampling with post-stratification

In systematic sampling with post-stratification (sys+post), samples are systematically selected across an entire population and the selected systematic samples are post-stratified into strata by their indicator variable. In this technique, the stratum proportions are estimated from the proportion of samples in each stratum. Thus, the optimal sample sizes for each stratum are unknown. The stratum weight (W'_h) for each stratum (h) may then be estimated as $W'_h = n'_h / n'$. In this study, selected samples were post-stratified based on the thematic map of forest cover types such as pre-stratification.

Systematic cluster sampling

This design uses a cluster plot consisting of various sub-plots instead of a single plot. Under systematic sampling, clusters are distributed across the whole population using the pre-defined systematic intervals. This study employed the cluster plot design ($m=4$) developed in Chapter 4 as a sampling unit.

In most applications of clusters for natural resources assessment, clusters contain different numbers of sub-plots. In this case, the ratio estimator can be applied (Table 5.2), with the cluster size serving as an ancillary variable.

5.2.3 Comparison**Sampling error**

Because the samples are only a subset of the total population, each estimate for any given sampling design will contain an error, so called "sampling error". The results of repeated data collections using the same sampling design have a certain variance, which is due to the randomness of the sample. This variance can be estimated through a Monte Carlo simulation with 1000 repetitions, in which samples for each simulation are independently and identically distributed (Engeman *et al.*, 1994). The variance is most easily calculated as follows:

$$\hat{var}(\tilde{y}) = \frac{\sum_{i=1}^k (\bar{y}_i - \tilde{y})^2}{k-1} \quad (5-2)$$

where k : the number of simulations,

\bar{y}_i : the estimated mean of the i -th simulation, and

\tilde{y} : the mean of the estimated means by repeated simulations.

In this estimator, the denominator $k-1$ was used to estimate an unbiased variance in sampling with replacement. The square root of this variance is the standard error.

Relative efficiency

The relative efficiency is given by two unbiased estimators z_1 and z_2 of a variable z with variances $Var(z_1)$ and $Var(z_2)$, respectively. The efficiency of z_1 relative to z_2 is given by

$$RE \left(\frac{z_1}{z_2} \right) = \frac{Var(z_2)}{Var(z_1)}. \quad (5-3)$$

In this study, the variance $Var(z_1)$ is calculated from SRS as a baseline and the variance $Var(z_2)$ is estimated for the simulated systematic sampling designs. If the efficiency is greater than 1, SRS is preferable; conversely, if its value for a candidate design is less than 1, this design is more precise than SRS.

5.3 RESULTS

5.3.1 Artificial forest population

In this study, the forest area in an artificial model forest consists of cells (or pixels) of 25 m, and the set of cells within the forest forms the population, as shown in Figure 5.1. In the NNC classified image, the forest area is 73,189 ha. About 38% (28,094 ha) of the total forest area is covered by coniferous forests, 36% by deciduous forests, and 25% by mixed forests (Table 5.3).

The parametric mean of growing stock per pixel and for each stratum is summarized in Table 5.3. The parametric mean value are calculated to be 7.16 m³, 7.01 m³, 6.95 m³ per unit for coniferous, broadleaved, and mixed forests, respectively. In the case of the mixed forest, its variance is highest, which means that the variability among elements is larger than the other forest cover types. The difference in variance between strata, however, is very small.

Table 5. 3: Summary statistics of growing stock (m³) per pixel unit (0.0625 ha) for each stratum for the given artificial forest population

Classification		Area (ha)	Mean (m ³)	Min. (m ³)	Max. (m ³)	Variance (m ³) ²
Sub-Population	Coniferous	28,094	7.16	1.44	12.64	3.16
	Deciduous	26,522	7.01	1.19	12.44	3.00
	Mixed	18,573	6.95	0.94	12.75	3.19
Population		73,189	7.06	0.94	12.75	3.12

Variance: the variance per element of the population and sub-populations

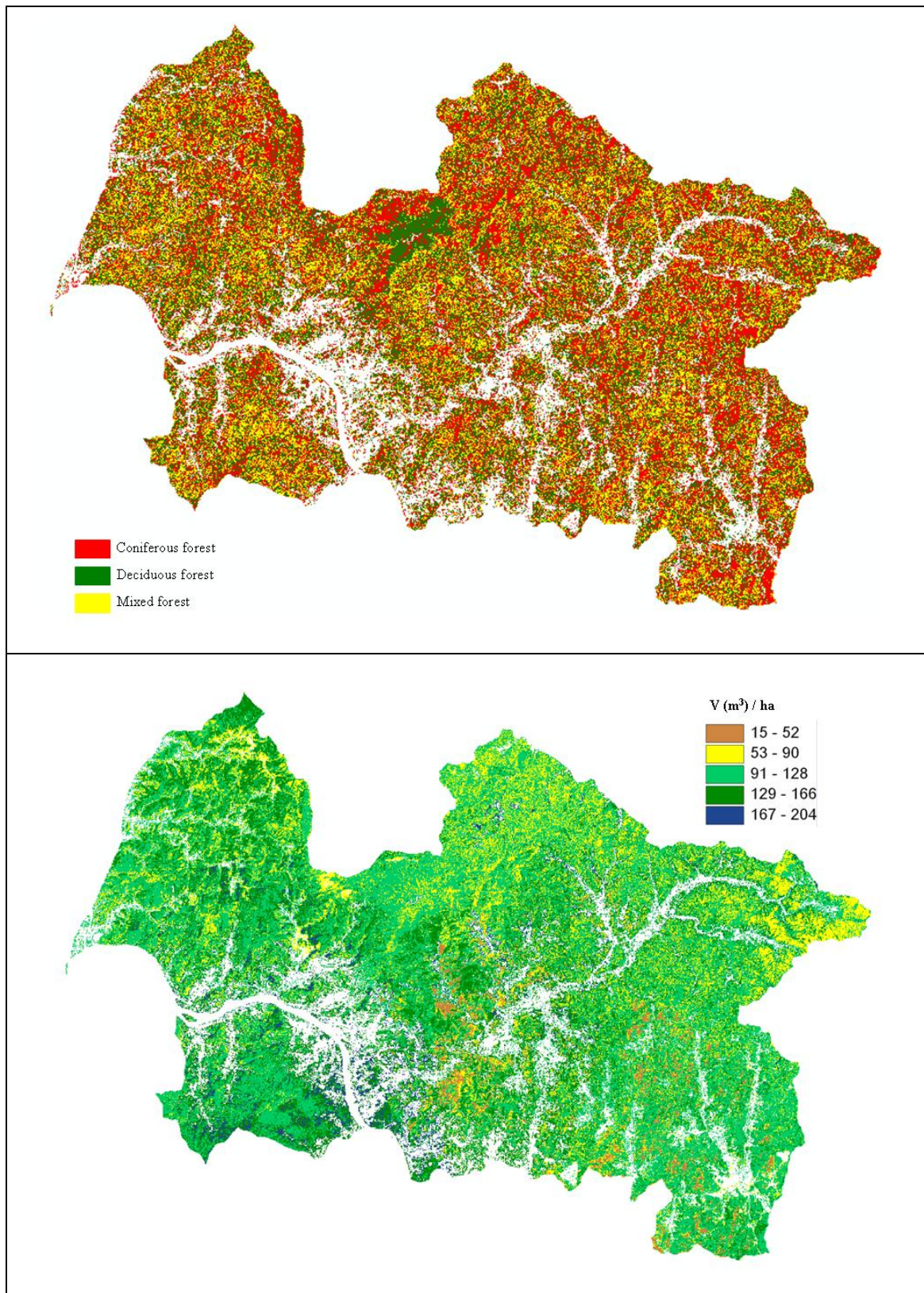


Figure 5. 1: Thematic maps used as an artificial forest population in this study: forest cover types (top) and growing stock per hectare within the forest (bottom).

5.3.2 Sampling simulation

Sampling error

Each sampling design was simulated with different sample sizes without replacement and was repeated 1,000 times, respectively. When systematic sampling is applied, sample size is a random variable that varies with sample grid size (1km, 1.5km, 2km and 4km) and starting point for each simulation. Thus, mean sample size was computed for each grid size. Simple random sampling (SRS) as a baseline was compared to the simulated systematic sampling designs.

For evaluating the benefit of stratification (forest cover types), *stratified random sampling* was compared to SRS (Table 5.4). For the smallest sample size ($n=50$), the sampling error of STR was slightly less than that of SRS, while for the other sample sizes, the difference in sampling error was similar.

Table 5. 4: Summary of estimations by simulation for different sample sizes under stratified random sampling and simple random sampling

	Sample size	3,000	2,000	1,000	500	100	50
Stratified random sampling	Mean (m ³)	7.053	7.054	7.053	7.052	7.047	7.055
	Error variance	0.0011	0.0016	0.0031	0.0066	0.0318	0.0587
	SE	0.03	0.04	0.06	0.08	0.18	0.24
	SE%	0.48	0.56	0.79	1.15	2.53	3.43
Simple random sampling	Mean (m ³)	7.055	7.054	7.055	7.056	7.061	7.056
	Error variance	0.0011	0.0015	0.0033	0.0059	0.0316	0.0619
	SE	0.03	0.04	0.06	0.08	0.18	0.25
	SE%	0.46	0.55	0.81	1.09	2.52	3.53

SE : standard error

SE%: standard error of the estimated mean

SAMPLING DESIGN OPTIMIZATION

The SE% for *systematic sampling* (SYS), ranging from 0.81% to 3.54%, was smaller than that for SRS, except for the grid size of 4km. The SE% for the largest sample size was about four times smaller than that for the smallest sample size. This result was also observed for the other simulated systematic sampling designs, as presented in Table 5.5. The sampling error for *stratified systematic sampling* was slightly higher than that for SYS due to pre-defined stratum weightings.

For reducing costs and obtaining improved precision through stratification, *systematic sampling with post-stratification* was applied. The SE% for this design ranged from 0.80% to 3.60% and was slightly less than that for *stratified systematic sampling*. Furthermore, this design yielded slightly more precise results than SYS when $n \geq 183$.

Systematic cluster sampling was superior to the other sampling designs, as presented in Table 5.5. Additionally, as the distance between grids gets smaller, i.e., as sample size increases, the precision becomes increasingly higher. In contrast to other sampling designs that include only one element, the cluster plot consists of four elements per cluster.

With respect to the distances between samples, a greater improvement in sampling error was observed between the sample grid sizes of 4km and 2km for all simulated systematic sampling designs. The decrease in sampling error was relatively small compared to the increase in the number of samples from 183 to 732.

Table 5. 5: Summary of estimations by sample size for different sampling designs

Simple random sampling	Sample size	732	325	183	46
	Mean (m ³)	7.055	7.057	7.055	7.066
	Error variance	0.0039	0.0095	0.0174	0.0580
	SE	0.06	0.10	0.13	0.24
	SE%	0.89	1.38	1.87	3.41
Grid size		1km*1km	1.5km*1.5km	2km*2km	4km*4km
Systematic sampling	Sample size*	732	325	183	46
	Mean (m ³)	7.058	7.054	7.063	7.065
	Error variance	0.0033	0.0083	0.0162	0.0641
	SE	0.06	0.09	0.13	0.25
	SE%	0.81	1.29	1.81	3.54
Stratified systematic sampling	Sample size*	734	326	185	48
	Mean (m ³)	7.056	7.055	7.054	7.073
	Error variance	0.0038	0.0090	0.0165	0.0685
	SE	0.06	0.09	0.13	0.26
	SE%	0.87	1.34	1.82	3.70
Systematic sampling with post-stratification	Sample size*	732	326	183	46
	Mean (m ³)	7.056	7.056	7.057	7.067
	Error variance	0.0032	0.0079	0.0158	0.0644
	SE	0.06	0.09	0.13	0.25
	SE%	0.80	1.26	1.78	3.60
Systematic cluster Sampling	Sample size*	730	325	184	46
	Mean (m ³)	7.011	7.004	7.008	7.01
	Error variance	0.0012	0.0045	0.0071	0.031
	SE	0.03	0.07	0.08	0.18
	SE%	0.49	0.95	1.20	2.51

Sample size* : mean sample size according to systematic selection

Relative efficiency

The error variance for SRS was used as a baseline to calculate relative efficiency. Figure 5.2 depicts the relative efficiency as a function of sample size. The efficiency of the sampling designs using the pre-stratification (STR and sys+pre) was small. For the simulated systematic sampling designs, the efficiency was higher than SRS, which increased with increasing sample size. When the same sampling effort was used, on average systematic sampling with post-stratification (sys+post) had the highest efficiency. The efficiency of systematic cluster sampling for all sample sizes was about two times higher than that of SRS.

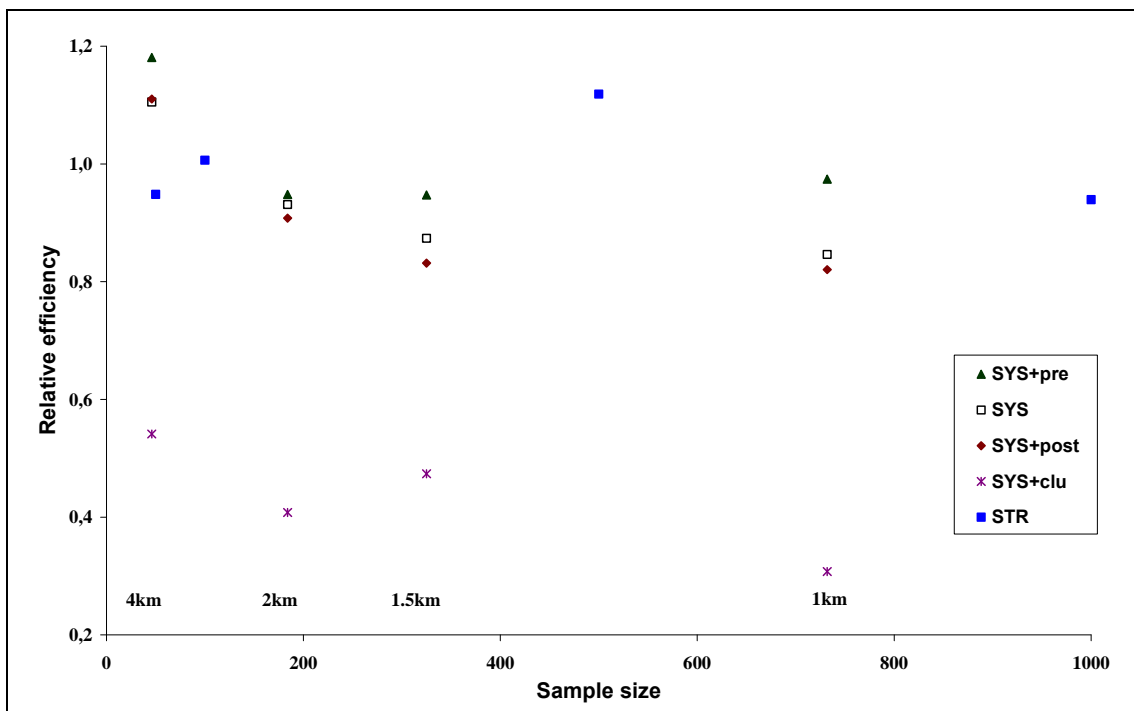


Figure 5. 2: Relative efficiency for sample sizes; sampling designs: SRS (simple random sampling), STR (stratified random sampling), SYS (systematic sampling), sys+pre (stratified systematic sampling), sys+post (systematic sampling with post-stratification), and sys+clu (systematic cluster sampling).

5.4 DISCUSSION AND CONCLUSION

The objective of this chapter was to develop the optimum sampling design in Korean forest conditions. In order to simulate various sampling designs, an artificial forest population was built from the results of Chapter 3.

According to Scott and Köhl (1993), the most efficient sampling design is seldom simple and easy to understand. Hence, they suggested that despite a loss in statistical efficiency, a simple design can be more practical for large area forest inventories. In this study, five simple sampling designs were employed as candidates for the most efficient sampling: stratified random sampling, stratified systematic sampling, systematic sampling, systematic sampling with post-stratification, and systematic cluster sampling.

It is a known fact that the efficiency of stratified sampling depends upon the stratification. In the last NFI, the characteristics of stratification (e.g., stratum weight) for the forest conditions, however, were not used for the allocation of the samples. These stratifications were only used to select samples within each stratum and to estimate total growing stocks (KFRI, 1996). Aerial photos were used to identify the information about stratification criteria for forest conditions. However, photograph-based stratification process was laborious and expensive (Kim *et al.*, 1989). Since aerial photographs were mainly used to estimate areas for different stratifications by forest conditions, the characteristics of the forest conditions would not contribute to the precision of estimates (Shin *et al.*, 2002).

In the given artificial forest population, the differences in mean and variance of growing stock between strata are similar (Table 5.3), because the forest population was derived from a small number of samples. This means that there is not much benefit to stratifying the given population by forest cover types. If each forest cover type is sub-stratified into age classes which are closely correlated to growing stock, then the benefit of stratification can be realized.

In pre-stratification procedure, there are several factors utilized for the selection of samples such as stratum weight and stratum size, and sample size under stratified systematic sampling. These factors are sensitive to conditions of the strata at a point in time. If the stratification criteria change over time, the factors must change according to the given strata at different times. The stratifications can be divided into two types; political units and ownerships are time-invariant, whereas stratification

criteria for forest conditions are time-variant. In the last NFI, the samples were selected depending upon forest conditions so that they may change at a future time. Considering the objective of the current NFI that provides reliable information about forest resources and monitor their change over time as well, pre-stratification by time-variant variables (e.g. forest types) is inappropriate to the task of monitoring forest resources for changes (Scott, 1998).

Categorical information about such forest conditions can be obtained through post-stratification. When compared to pre-stratification, the variance of the post-stratification estimator is usually higher because strata sizes are estimated. In this study, post-stratified systematic sampling, however, gave more precise estimations than stratified systematic sampling (Table 5.5). It is concluded that post-stratification is to be a very useful and cost-effective approach for large area forest inventories. In order to apply this approach to the Korean NFI, it is necessary to have clear definitions of forest strata per field observation unit (e.g., forest cover types per subplot). If an automated process cannot be used to stratify, or if the strata are not clearly defined, not only can it be a difficult and time-consuming task for the field crew, but it may also cause classification errors depending on the interpreter's decision.

The current NFI has adopted systematic sampling. In this NFI, a sample grid size of 4 km, which is driven by the pre-defined total sample size in the last NFI, has been applied (KFRI, 2006). Although the spatial spacing between samples is an important factor that affects the precision, it was not considered in the allocation of samples, because the total sample size was affected more by the specified precision requirement and budget available. In terms of forest proportion by region over the country in South Korea, the southwestern region is low flat land with a small forest proportion on average, whereas the northeastern region (in particular Kyung-Buk and Kangwon provinces) includes highly mountainous areas with a large forest proportion (KFS, 2004a). In this context, there is a need for more research on an appropriate spatial spacing based on the forest characteristics of the population and provinces: for example, the sample grid sizes for the German NFI varied by states (BMVEL, 2001). The application of varying sample grid sizes via forest proportions is expected to improve the precision at the national and provincial levels.

In South Korea, most of the time within one working day has been spent to reach the samples due to the limited accessibility and difficult terrain. According to a pilot time study (KFS, 2004a), the average travel time from an office to a sample is at least two hours. In addition, the forest variables of interest in the field increase to provide multi-sources information and then the inventory time per sample is required more than one hour. Consequently, it is hardly possible to measure more than two samples within one working day. In order to reduce the traveling time and obtain more additional information at each sample, a cluster plot as the sampling unit was applied. When comparing different plot designs (one element and four elements per cluster), for 184 samples, the one element design requires at least 92 days, whereas the cluster plot design requires only 46 days. Nevertheless, the difference in sampling error between the two plot designs was small (Table 5.5).

The scope of the Korean NFI is expanded to support sustainable forest management planning. In this context, the forest variables of interest in the field are increasing and therefore the optimum sampling design for field data collection is needed. The NFI also provides data and information for the entire country, as well as for different stratifications. In order to obtain stratifications by forest conditions, either ancillary information is required (pre-stratification), or the collected samples have to be post-stratified. While the former may be expensive and time-consuming, the latter might be laborious and the stratification procedure may be indefinite. If the objective is to obtain estimates for the entire country, systematic sampling is considered to be the most cost-efficient and practical sampling design. It not only achieves the objective, but the systematic sample is fixed and therefore allows the monitoring of net changes in forest resources over time. Since using a cluster plot reduces traveling costs, it can provide more information at lower cost. Moreover, if strata per field observation unit in systematic cluster sampling are clearly defined, estimates for different stratifications can also be provided and the precision can be improved by means of a post-stratification procedure.

6. OVERALL CONCLUSION

In large area forest inventories, methodological research is focused heavily on remote sensing and sampling issues and deals with the question of how to efficiently combine the two. In this context, this thesis addresses the integration of satellite imagery with forest inventory data from the Korean NFI and sampling design issues.

With the development of remote sensing, GIS, and GPS techniques, the use of satellite data increases, particularly natural resources assessment over large areas. In this thesis, their possibility into the Korean NFI was evaluated. Even though there are a variety of errors in the application of satellite imagery, digital satellite data can represent reasonably useful and cost-effective data source for forest cover classification and forest attribute estimation. In the current NFI cycle, GPS-based field data have been collected annually over the entire country and then can easily combine with digital satellite data. With respect to the definition of forest, the difference in minimum area per observation unit may affect classification accuracy. In order to successfully integrate digital satellite data into the Korean NFI, there are needs for more research on the selection of a suitable satellite imagery regarding plot design and user objectives, and post-classification processing to correspond to the definition of the forest in the Korean NFI.

What is the optimum sampling design for the Korean NFI? This thesis was carried out to assist in arriving at the answer to this question. For the NFI, sampling design would be determined to fulfill a specified precision and to achieve its objective for a limited budget. When considering difficult forest conditions in Korea, systematic sampling is the more efficient sampling design and allows the monitoring of net changes in forest resources over time. With respect to sample allocation, there is a need for more research relevant to spatial and forest proportions by regions over the entire country. The most optimal cluster plot design is driven more by practical restrictions and user objectives rather than statistical characteristics.

Consequently, in order to efficiently implement a forest inventory, it requires clear definitions per observation unit. In the both issues, there are many advantages of post-stratified samples with respect to statistical and economical grounds. In order to fully achieve post-stratification procedures to the Korean NFI, it is necessary to have clear definitions of strata of interest per field observation unit.

7. SUMMARY

The required information about forest resources has been gradually increased not only for timber-oriented attributes but also non-timber goods and services, including biodiversity. Frequently, however, the budget available for its implementation is not sufficient to fulfill all these demands. Efficient inventory techniques that help guarantee reliable information for a specific situation are in demand.

Recently, the Korean NFI has been reorganized and expanded to support sustainable forest management also on a more small area level. Despite several advantages of digital satellite data, these data are still not integrated into the NFI. Sampling and plot designs for field surveys need more attention to research for optimization, considering both statistical soundness and cost-effectiveness.

One of the major points in optimizing the Korean NFI and therefore in many other inventories as well is to attempt to reduce field work efforts and provide more detailed information about forest resources. This should be done not only at a national level but also at a small area level, by efficiently integrating digital satellite data and forest inventory data (Chapters 2 and 3) and by optimizing sampling and plot designs (Chapters 4 and 5).

Forest cover maps, which are used as an important baseline for forest managers and other policy makers, are a key product of the Korean NFI. Since 1970s, aerial photographs have been used to produce these maps. However, the use of aerial photographs may not be suitable for forest cover classification over large areas due to the ability to repeatedly obtain photographs, the uncertainty of interpretation process, cost-effectiveness, *etc.* With the development of remote sensing and GIS technologies, the application of digital satellite data is possible to enhance natural resources assessment. Despite several advantages of them, these data were not fully used for forest cover classification in the Korean NFI yet. In Chapter 2, the possibility of satellite remotely sensed data for forest cover classification in Korean forests is evaluated. In order to reduce radiometric distortions in mountainous regions, Landsat TM imagery was topographically normalized using the Minnaert constant method. In the image classification process, field data per sub-plot from the NFI were served as training data. Classified images by two pixel-based classifiers, Maximum likelihood (MLC, parametric approach) and Nearest Neighbor (NNC, non-parametric approach), were produced and compared with exiting digital forest maps from interpretation of

SUMMARY

aerial photographs. The accuracy for NNC was superior to MLC and the digital forest maps. Moreover, a NNC-classified image was close to a classification result from field plot data. When comparing the NNC-classified image with the digital forest maps, it appears that there are different definitions of observation unit; a minimum area of the forest defined for aerial photograph interpretation is much larger than that for satellite data classification which is dependent upon the spatial resolution of satellite data to be used. There is a need for more research on post-classification processing to improve the quality of classifications and to correspond to the definition in the NFI.

Field plot data from a large-area forest inventory may not provide reliable information for a small area at a municipal level, since a number of available samples is small. The relationship between spectral features on satellite data and forest characteristics is observed and then imputation procedures have been applied to estimate forest attributes across an entire area of interest, which is covered by remotely sensed data to be used. As an imputation procedure, the *k*-Nearest Neighbor (*k*-NN) estimator that is one of non-parametric regressions has been widely applied. In this procedure, the similarity between observations and un-observations is usually determined based on distances in feature spaces. In Chapter 3, the potentiality of the *k*-NN technique in Korean forests is analyzed, including different satellite sources, distance-weightings for neighbors and for spectral features, image enhancement, and stratification for the selection of reference samples. The forest of a test area has a relatively small variation in growing stocks since it is relatively young, mostly less than 50 years old. Despite a small sample size (191 sub-plots) that served as training data, the precision of estimates obtained for growing stock was relatively small, ranging from 40 to 60% in relative RMSE. When comparing two satellite images, the precision using ASTER was slightly better than that obtained using Landsat ETM+. The use of ASTER imagery for large area inventories, however, may be expensive and this represents a trade-off between precision and budget. The application of stratification for the selection of reference samples at a target point, particularly based on forest conditions, can improve the precision and preserve the variability of observations as well. In this study, the effectiveness of stratification, however, is not achieved because a number of available reference samples for each stratum is different and relatively small. Despite the inherent sources of errors using the *k*-NN technique, the *k*-NN map that is generated allows for supporting forest management plans and for reporting forest statistics for large- and small-area units.

Consequently, in order to fully integrate digital satellite data into the Korean NFI, the following factors must be taken into consideration: (i) selection of a suitable satellite sensor regarding the plot design and user objectives, (ii) accurate information of field plot locations for spatial matching them to digital maps and satellite data, and (iii) clear definitions of strata of interest per field observation unit.

The last Korean NFI cycle (1996-2005) used a stratified systematic sampling with clusters. Since the inventory was implemented with a rotation system by province over a 10-year period, it is hardly possible to provide reliable information about the forest resources for the entire country at the same time. In regard to plot design, a cluster plot was designed completely on empirical findings rather than scientific grounds. Moreover, as the required information in the field is increasing, it becomes necessary to develop the most efficient sampling strategy for the NFI, for which three basic design elements must be considered: sampling design, plot design, and estimation design.

In order to develop the most efficient cluster plot as a plot design, a pilot study is carried out in Chapter 4. When optimizing a cluster plot, several questions are relevant: distance between sub-plots per cluster; spatial arrangement of sub-plots; and cluster size with respect to statistical efficiency. To answer these questions, three statistical criteria are taken into account; covariance function, Intra-cluster correlation, and standard error as a criterion of precision. In addition, a cost function as an economic criterion also has to be developed. To address these criteria, a pilot cluster plot consisting of 10 sub-plots ($n=25$) was empirically designed and collected in a small area. In this study, 16 cluster configurations that are able to implement within a single working day were employed as candidates for the most efficient cluster plot. From covariance functions for key attributes (growing stock, basal area, and tree density), a distance of 87 m between pairs of sub-plots was found to be an efficient minimum distance. The precision was affected more by cluster shape than by cluster size. When clusters of 5 sub-plots were applied, the cluster size was not accomplished within one working day because much more inventory times per sub-plot are needed in a multi-resource forest inventory. Accordingly, from the results of both statistical and practical efficiencies, a modified triangular-shaped cluster of 4 sub-plots was found to be the most efficient cluster plot under the conditions of the test area. Although numerous forest attributes are collected in the field, for this study only timber-oriented attributes (growing stock, basal area, and tree density) were analyzed. Moreover, the pilot field data collected cover in a specific small area. There

SUMMARY

are more research needs to develop an efficient cluster plot for the Korean NFI: (i) field data to cover the forest conditions over the whole country are required; (ii) key attributes of interest in the field inventory must be determined before plot design planning; and (iii) the plot design cannot be performed independently but must be dependent on a given sampling design, i.e., spatial distribution of samples.

With regard to sampling design optimization, an artificial forest population for growing stock and forest cover types as stratification criterion was generated from the results in Chapter 3. Several sampling designs that have been mostly applied for large area forest inventories were simulated, including systematic sampling, stratified systematic sampling, systematic sampling with post-stratification, and systematic cluster sampling. When systematic sampling is applied, the spatial spacing between samples is an important factor that should influence precision. In this study, four sample grid sizes (1km, 1.5km, 2km, and 4km) were compared. Error variances for different sampling designs and different sample sizes were estimated through a Monte Carlo simulation with 1000 repetitions. In addition, relative efficiency was also compared, in which the error variance for simple random sampling used as a baseline. In general, pre-stratified sampling produces higher precision as the difference in parametric variance between strata is larger. In the given population, since the difference between strata is small, the benefit of stratification could not be realized. In addition to pre-stratified sampling, this procedure requires ancillary information for obtaining stratification criteria such as forest cover types and therefore is expensive. Accordingly, systematic sampling was to be the most cost-efficient and practical sampling design. This procedure also allows the monitoring of net changes in the forest resources over time. When comparing same sampling efforts, systematic sampling with post-stratification was superior to the others. It is concluded that post-stratification is a very useful and cost-effective approach because it not only achieves the objective, but can improve the precision. In order to apply this approach to the Korean NFI, it is necessary to have clear definitions of forest strata per field observation unit. In the simulated systematic sampling design, a greater improvement in sampling error was observed between grid sizes of 2km and 4km. However, the decrease in sampling error was relatively small compared to the decrease in sample grid size from 2km to 1km. Due to difficult forest conditions in Korea most of the time within one working day is spent on reaching the samples. Since using a cluster plot as plot design reduces traveling costs, it allows the collection of additional information at lower cost.

8. REFERENCES

- Abrams, M.** 2000. ASTER: data products for the high spatial resolution imager on NASA's EOS-AM1 platform. *International Journal of Remote Sensing* 21:847-861.
- Ahern, F.J., Erdle, T., Maclean, D.A., and Knepeck, I.D.** 1991. A quantitative relationship between forest growth rates and Thematic Mapper reflectance measurements. *International Journal of Remote Sensing* 12: 387-400.
- Altman, N.S.** 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46(3): 175-184.
- Arvanitis, L.G. and O'Regan, W.G.** 1972. Cluster or satellite sampling in forestry: a Monte Carlo computer simulation study. pp 191–205. Proceedings of the 3rd conference of the IUFRO Advisory Group of Forest Statisticians. Volume 72, Number 3. Institut National de la Recherche Agronomique, Jouy-en-Josas, France.
- BMVEL.** 2001. Survey instructions for Federal Forest Inventory II. 106p.
- Chen, D. and Stow, D.A.** 2002. The effect of training strategies on supervised classification at different spatial resolution. *Photogrammetric Engineering and Remote Sensing* 68: 1155-1162.
- Cho, H.K., Akça, A., and Lee, W.K.** 1999. Correction of topographic effects on Landsat TM imagery for forest type classification in mountainous area. *Korean Journal of Forest Measurements* 2(2): 50-59.
- Cho, H.K.** 2002. Untersuchungen über die Erfassung von Waldflächen und deren Veränderungen mit Hilfe der Satellitenfernerkundung und segmentbasierter Klassifikation. Dissertation der Forstwissenschaftlichen Fakultät der Georg-August-Universität Göttingen. 120p.
- Chung, D.J.** 1996. Konkurrenzverhältnisse und Struktur eines natürlichen *Pinus densiflora-Quercus variabilis* Mischwaldes in Korea. Dissertation der Forstwissenschaftlichen Fakultät der Georg-August-Universität Göttingen. 187p.
- Chung, K.H., Lee, W.K., Lee, J.H., Kwon, K.H., and Lee, S.H.** 2001. Classification of forest type using high resolution imagery of satellite IKONOS. *Korean Journal of Remote Sensing* 17(3): 275-283.
- Cochran, W.G.** 1977. Sampling techniques. 3rd edition. John Wiley & Sons. 428p.
- Colby, J.D.** 1991. Topographic normalization in rugged terrain. *Photogrammetric Engineering and Remote Sensing* 57: 531-537.
- Congalton, R.G.** 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37: 35-46.

REFERENCES

- Davies, E.R.** 1988. Training sets and a priori probabilities with the nearest neighbor method of pattern recognition. *Pattern Recognition Letters* 8:11-13.
- Drăgut, L. and Blaschke, T.** 2006. Automated classification of landform elements using object-based image analysis. *Geomorphology* 81:330-344.
- Dunn, R. and Harrison, R.A.** 1993. Two-dimensional systematic sampling of land use. *Applied Statistics* 42(4): 585-601.
- Engeman, R.M., Sugihara, R.T., Pank, L.F., and Dusenberry, W.E.** 1994. A comparison of plotless density estimators using Monte Carlo simulation. *Ecology* 75(6):1769-1779.
- European Communities.** 1997. Study on European Forestry Information and Communication System. Reports on forestry inventory and survey systems. - Volume 1 (673p) and Volume II (655p.).
- FAO.** 1997. Working paper No: APFSOS/WP/06 - In-depth country study in the Republic of Korea -. 42p.
- FAO.** 2004. FRA 2005 terms and definition. www.fao.org/forestry/site/fra2005-terms/en
- Finley, A., McRoberts, R.E., and Ek, A.R.** 2006. Applying an efficient neighbor search to forest attribute imputation. *Forest science* 52(2): 130-135.
- Foody, G.M.** 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80: 185-201.
- Franco-Lopez, H, Ek, A.R., and Bauer, M.E.** 2001. Estimation and mapping of forest stand density, volume and cover type using the k-nearest neighbors method. *Remote Sensing of Environment* 77: 251-274.
- Franklin, J., Phinn, S.R., Woodcock, C.E., and Rogan, J.** 2003. Rationale and conceptual framework for classification approaches to assess forest resources and properties. pp279-300. In Wulder, M.A. and Franklin, S. E. (eds.). *Remote sensing of forest environments - Concepts and case studies*. Kluwer.
- Haapanen, R., Ek, A.R., Bauer, M.E., and Finley, A.** 2004. Delineation of forest/non-forest land use classes using nearest neighbor methods. *Remote Sensing of Environment* 89: 265–271.
- Halme, M. and Tomppo, E.** 2001. Improving the accuracy of multi-source forest inventory estimates by reducing plot location error - a multi-criteria approach. *Remote Sensing of Environment* 78:321-327.
- Hansen, M.H., Hurwitz, W.N., and Madow, W.G.** 1953. Sample Survey methods and theory. John Wiley & Sons. NY. 638p.
- Hardin, P.J.** 1994. Parametric and Nearest neighbor methods for hybrid classification: a comparison of pixel assignment accuracy. *Photo-grammetric Engineering and Remote Sensing* 60(12): 1439-1448.

- Hardin, P.J. and Thomson, C.N.** 1992. Fast nearest neighbor classification methods for multi-spectral imagery. *Professional Geographer* 44(2): 191-201.
- Holmgren, P. and Thuresson, T.** 1998. Satellite remote sensing for forestry planning - a review. *Scandinavian Journal of Forest Research* 13: 90-110.
- Ince, F.** 1987. Maximum likelihood classification: Optimal or problematic? A comparison with nearest neighbor classification. *International Journal of Remote Sensing* 8(12): 1829-1838.
- Itten, K.I., Meyer, P., Kellenberger, T., Leu, R., Bitter, P., and Seidel, K.** 1992: Correction of the impact of topography and atmosphere on Landsat-TM forest mapping of Alpine regions. Remote sensing series vol. 18, 47p.
- Jensen, J.** 1996: Introductory digital processing: a remote sensing perspective. Englewood Cliffs, NJ: Prentice-Hall. 318p.
- Johnson, E.W.** 2000. Forest sampling desk reference. CRC Press LLC. 985p.
- Kangas, A.** 1993. Estimating the attributes of systematic cluster sampling by model based inference. *Scandinavian Journal of Forest Research* 8: 571-582.
- Katila, M.** 2004. Controlling the estimation errors in the Finnish multisource National Forest Inventory. Academic dissertation. Finnish Forest Research Institute, Research Papers 910. 36p.
- Katila, M.** 2006. Empirical errors of small area estimates from the multisource national forest inventory in eastern Finland. *Silva Fennica* 40(4): 729-742.
- Katila, M. and Tomppo, E.** 2001. Selecting estimation attributes for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment* 76: 16-32
- Katila, M. and Tomppo, E.** 2002. Stratification by ancillary data in multisource forest inventories employing k-nearest-neighbor estimation. *Canadian Journal of Forest Research* 32(9): 1548-1561.
- Katila, M. and Tomppo, E.** 2006. Sampling simulation on multi-source output forest maps – an application for small areas. pp 613-624. In Caetano, M and Painho, M. (eds.), processing of the 7th International symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. July 5-7, 2006. Lisboa, Instituto Geográfico Português.
- Khunrattansiri, W.** 2005. Development of forest inventory techniques with remote sensing for forest resources assessment. Dissertation der Forstwissenschaftlichen Fakultät der Albert-Ludwigs-Universität Freiburg. 129p.
- Kim, C. M.** 1991. Studies on analysis of forest distribution and characteristics in Gumo-san provincial park using Landsat-TM data and DTM. Dissertation, Seoul National University, Korea.
- Kim, C.M.** 2004. Forest geographic information systems in Korea. Proceedings of the 4th conference of the international society for ecological informatics, Oct. 24-28 2004, Pusan,

REFERENCES

- South Korea.
- Kim, J.B., Cho, M.H., Kim, I.H., and Kim, Y.G.** 2003. A study on extraction of damaged area by Pine wood Nematode Using high resolution satellite images and GPS. *Journal of Korean Forest Society* 92(4):362-366.
- Kim, K.D.** 1965. Studies on sampling unit. *Journal of Korean Forest Society* 4:26-29.
- Kim, K.D.** 1966. A study on the forest inventory work. *Journal of Korean Forest Society* 5:10-15.
- Kim, K.D.** 1973. On study of forest sampling methods in natural deciduous forest. *Journal of Korean Forest Society* 17: 35-42.
- Kim, K.D.** 1987. Studies on the application of remote sensing technique to forestry. *Journal of Korean Forest Society* 76(1): 41-50.
- Kim, K.D., Lee, S.H., and Kim, C.M.** 1989. Classification and mapping of forest type using Landsat TM and B & W infrared photographs. *Journal of Korean Forest society* 78(3):263-273.
- Kleinn, C.** 1994. Comparison of the performance of line sampling to other forms of cluster sampling. *Forest Ecology and Management* 68:365-373.
- Kleinn, C.** 1996. Ein Vergleich der Effizienz von verschiedenen Clusterformen in forstlichen Grossrauminventuren. *Forstw. Cbl.* 115: 378-390.
- Kleinn, C.** 2002. Review and discussion of new technological and methodological options for national forest inventories. Background paper. Kotka IV Expert Consultation on Forest Resources Assessment.
- Kleinn, C. and Jost, A.** 1994. Überlegungen zur Planung und Auswertung systematischer Stichproben. *Schweizerische Zeitschrift für Forstwesen* 145(9): 703-719.
- Kleinn, C. and Morales, D.** 2001. Tree resources outside the forest project. WP4 Report: Development of sampling strategies. CATIE, Costa Rica. 53p.
- Kleinn, C., Corrales, L., and Morales, D.** 2002. Large area forest cover estimates in the tropics- the case of Costa Rica. *Environmental Monitoring and Assessment* 73(1): 17-40.
- Köhl, M.** 1986. Effektivität von Gruppenstichproben. Mitteilungen der Abteilung für Forstliche Biometrie und der Abteilung für Luftbildmessung and Fernerkundung der Universität Freiburg, 86-1. 165p.
- Köhl, M., Magnussen, S., and Marchetti, M.** 2006. Sampling methods, Remote Sensing and GIS Multi-resource Forest Inventory. Springer. 373p.
- KFRI.** 1996. Manual of the fourth National Forest Inventory in South-Korea. 46p.
- KFRI.** 2004. Forest sites of Korea: Forest Soil. 621p.
- KFRI.** 2006. Manual of the fifth National Forest Inventory in South-Korea. 34p.
- KFS.** 2002. The reorganization of National Forest Inventory System by a change of demands

- of society and international process for forest resource I. 273p.
- KFS.** 2003. The reorganization of National Forest Inventory System by a change of demands of society and international process for forest resource II. 271p.
- KFS.** 2004a. Statistical yearbook of Forestry. 434p.
- KFS.** 2004b. The reorganization of National Forest Inventory System by a change of demands of society and international process for forest resource III. 284p.
- KFS.** 2005. The reorganization of National Forest Inventory System by a change of demands of society and international process for forest resource IV. 290p.
- Koukal, T.** 2004. Nonparametric assessment of forest attributes by combination of field data of the Austrian forest inventory and remote sensing data. Dissertation zur Erlangung des Doktorgrades an der Universität für Bodenkultur Wien. 105p.
- Lanz, A.** 2000. Optimal sample design for extensive forest inventory. Dissertation of Swiss federal institute of technology Zurich. 250p.
- Lee, J.S. and Han, S.S.** 1986. Studies on Research Plot Size for Some Characters (II) - Application to the 6 - year - old Korean Pine Plantation -. *Journal of Korean Forest Society* 65: 96-99.
- Lee, K.S.** 1994. Vegetation cover type mapping over the Korean peninsula using multi-temporal AVHRR data. *Journal of Korean Forest Society* 83(4): 441- 449.
- Lee, K.S. and Kim, J.H.** 2000. Change analysis of forest area and canopy conditions in Kaesung, North Korea using Landsat, SPOT and KOMPSAT data. *Korean Journal of Remote Sensing* 16(4): 327-338.
- Lee, K.S. and Yoon, J.S.** 1997: Radiometric correction of terrain effects for SPOT and Landsat Thematic Mapper imagery in mountainous forest area. *Korean Journal of Remote Sensing* 13(3): 277-292.
- Lee, S. H.** 1991. Studies on the detection of forest cover change using Landsat MSS and TM image data. Dissertation, Seoul national university, Korea. 104p.
- Lee, S.H., Chung, S.H., and Song, J.H.** 1998. Forest Resources inventory of North Korea using satellite remote sensing data. *KFRI, Journal of Forest Science* 58(1): 1-13.
- Lee, S.H., Kim, C.M., and Kim, K.H.** 1994. Analysis of stand structure and growth characteristics by remote sensing and topographic information. Research reports of Forest Research Institute 50: 80-95.
- Lee, S.H., Kim, C.M., Won, H.K., Kim, K.M., and Cho, H.K.** 2004a. Estimating the Growing stock of Forest resources using multi-temporal Landsat TM. pp250-252. Proceedings of the 2004 summer meeting of Korean Forest Society.
- Lee, W.K., Von Gadow, K., Chung, D.J., Lee, J.L., and Shin, M.Y.** 2004b. DBH growth model for *Pinus densiflora* and *Quercus variabilis* mixed forests in central Korea.

REFERENCES

- Ecological Modelling* 176: 187-200.
- Lillesand, T. M., Kiefer, R. W., and Chipman, J. W.** 2004. Remote sensing and image interpretation. 5th ed. John Wiley & Sons, Inc. 763p.
- Lund, H. G.** (coord.). 2007. Definitions of Forest, Deforestation, Afforestation, and Reforestation. Available from: <http://home.comcast.net/~gyde/DEFpaper.htm>. Misc. pagination. (last visited 11.September 2007).
- Magnussen, S., Boudewyn, P., Wulder, M., and Seemann, D.** 2000. Predictions of forest inventory cover type proportions using Landsat TM. *Silva Fennica* 34(4): 351-370.
- Maselli F., Chirici, G., Bottai, L., Corona, P., and Marchetti, M.** 2005. Estimation of mediterranean forest attributes by the application of k-NN procedures to multi-temporal Landsat ETM+ images. *International Journal of Remote Sensing* 26(17): 3781-3796.
- McRoberts, R.E., Wendt, D.G., Nelson, M.D., and Hansen, M.H.** 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment* 81: 36-44.
- McRoberts, R.E. and Tomppo, E.** 2007. Remote sensing support for national forest inventories. *Remote Sensing of Environment* 110: 412-419.
- Minnaert, J.L.** 1941. The reciprocity principle in lunar photometry. *Astrophysics Journal* 93: 403-410.
- Mitchell, T.M.** 1997. Machine learning. The McGraw-Hill Companies, INC. 414p.
- Nelder, J. A. and Mead, R.** 1965. A simplex method for function minimization. *Computer Journal* 7: 308-313.
- Nilsson, M.** 1997. Estimation of forest variables using satellite image data and airborne lidar. Doctoral thesis. Acta Univ. Agric. Suec. 17.
- Päivinen, R.** 1987. Metsän inventoinnin suunnittelumalli. Summary: A planning model for forest inventory. University of Joensuu, Publication in Science 11(30). 168p.
- Paola, J.D. and Schowengerdt, R.A.** 1995. A review and analysis of back propagation neural networks for classification of remotely sensed multi-spectral imagery. *International Journal of Remote Sensing* 16: 3033-3058.
- Park, K.H.** 1986. Stand composition of the deciduous forest in Korea by use of aerial photographs. *Journal of Korean Forest Society* 74: 67-81.
- Park, S.M., Im, J.H., and Sagong, H.S.** 2001. Land Cover Classification of a Wide Area through Multi-Scene Landsat Processing. *Korean Journal of Remote Sensing* 17(3): 189-197.
- Reese, H., Nilsson, M., Granqvist Pahlén, T., Hagner, O., Joyce, S., Tingelöf, U., Egberth, M., and Olsson, H.** 2003. Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. *Ambio* 32: 542-548.

- Rho, K.M. and Lee, S.I.** 1995. The analysis of change of forest cover and vegetation index used Landsat MSS data and TM data. Proceedings of the 1995 annual meeting of Korean Forest Society. pp99-100.
- Saborowski, J.** 1992. Ein Diskussionsbeitrag zum Thema: Systematische Stichproben in der Waldinventur. *AFJZ* 163 (6): 107-110.
- Saborowski, J. and Cancino, J.** 2007. About the benefits of post-stratification in forest inventories. *Journal of Forest Science* 53(4): 139-148.
- Scheuber, M and Köhl, M.** 2003. Assessment of non-wood-goods and services by cluster sampling. pp 157-171. In Corona, P., M. Köhl and M. Marchetti. (Editor). *Advanced in forest inventory for sustainable forest management and biodiversity monitoring*. Kluwer Academic Publishers, Netherlands.
- Scott, C.T.** 1981. Design of optimal two-stage multi-resource surveys. Dissertation. University of Minnesota, St. Paul, Minnesota, USA. 138p.
- Scott, C.T.** 1991. Optimal design of a plot cluster for monitoring. pp233-242. Proceedings of an IUFRO S4.11 Conference held on 10-14 Sep. 1991 at the University of Greenwich, London, UK.
- Scott, C.T.** 1998. Sampling Methods for Estimating Change in Forest Resources. *Ecological Applications* 2: 228-233.
- Scott, C.T. and Köhl, M.** 1993. A method of comparing sampling design alternatives for extensive inventories. *Mitteilungen der Eidgenössischen Forschungsanstalt für Wald, Schnee und Landschaft* 68 (1): 3-62.
- Shin, M.Y., Yim, J.S., and Lee, D.K.** 2001. A study on stand structure and competition status by site types in natural deciduous forest of Pyeongchang, Kangwon-do. *Journal of Korean Forest society* 90(3): 295-305.
- Shin, M.Y., Yim, J.S., and Kong, G.S.** 2002. A comparative study of national forest inventory system by nation. *Korean Journal of Forest Measurements* 5(2): 1-9.
- Shin, M.Y. and Han, W.S.** 2006. Development of a forest inventory system for the sustainable forest management. *Journal of Korean Forest Society* 95(3): 370-377.
- Smith, J. A., Lin, T. L., and Ranson, K.J.** 1980. The Lambertian Assumption and Landsat Data, *Photogrammetric Engineering and Remote Sensing* 46(9): 1183-1189.
- Stoney, W.E.** 2006. ASPRS Guide to land imaging satellites. 17p. <http://www.asprs.org/news/satellites/index.html> (06/2006)
- Tokola, T.** 2000. The influence of field sample data location on growing stock volume estimation in Landsat TM-based forest inventory in eastern Finland. *Remote Sensing of Environment* 74(3): 422-431.

REFERENCES

- Tokola, T., Pitkänen, J., Partinen, S., and Muinonen, E.** 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. *International Journal of Remote Sensing* 17(12): 2333-2351.
- Tokola, T. and Heikkilä, J.** 1997. Improving satellite image based forest inventory by using a priori site quality information. *Silva Fennica* 31(1): 67-78.
- Tokola, T. and Shrestha, S.M.** 1999. Comparison of cluster-sampling techniques for forest inventory in southern Nepal. *Forest Ecology and Management* 116: 219-231.
- Tokola, T., Sarkeala, J., and van der Linden, M.** 2001. Use of topographic correction in Landsat TM-based forest interpretation in Nepal. *International Journal of Remote Sensing* 22(4): 551–563.
- Tomppo, E.** 1991. Designing a satellite image-aided national forest survey in Finland. In *The Usability of Remote Sensing for Forest Inventory and Planning. SNS/IUFRO Workshop, 26–28 February 1990, Report 4* (Umeå : Swedish University of Agricultural Sciences, Remote Sensing Laboratory), pp. 43–47.
- Tomppo, E.** 1996. Multi-source national forest inventory of Finland. In *New Thrusts in Forest Inventory. Proceedings of the subject Group S4.03-00 "Forest Resource Inventory and Monitoring" and Subject Group S4.12-00"Remote sensing Technology". Vol.1. IUFRO XX World congress, 6-12 Aug. 1995, Tampere, Finland.* Edited by Päivinen, R., Vanclay, J., and Miina, S. European Forest Institute, Joensuu, Finland. pp 27-41.
- Tomppo, E. and Halme, M.** 2004. Using coarse scale forest variables as ancillary information and weighting of variables in *k*-NN estimation: a genetic algorithm approach. *Remote Sensing of Environment* 92: 1-20.
- Tuominen, S., Holopainen, M., and Poso, S.** 2006. Multiphase sampling. pp235-252. In Kangas, A. and Maltamo, M. (eds.). *Forest inventory - Methodology and Application.* Springer. 326p.
- USDA.** 2005. Forest Inventory and Analysis National core field guide version 3.0. 203p.
- Wilkinson, G.G.** 1996. A review of current issues in the integration of GIS and remote sensing data. *International Journal of Geographical Information Systems* 10(1): 85-101.
- Yim, J.S. and Shin, M. Y.** 2006. Comparison of plot sizes for forest inventory in a natural deciduous forest. *Journal of Korean Forestry Society* 95(5): 595-600.
- Zhang, H., Berg, A.C., Maire, M., and Malik, J.** 2006. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. Proceeding of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 17-22, 2006, New York, NY, USA.

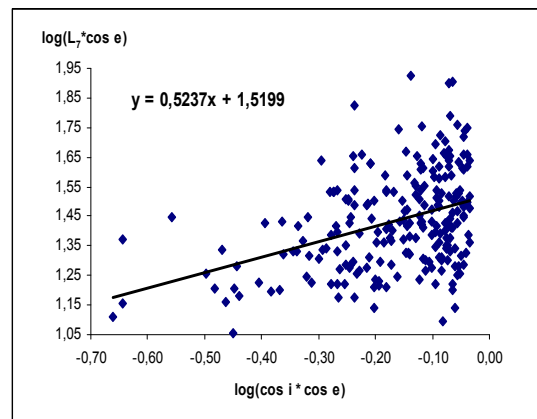
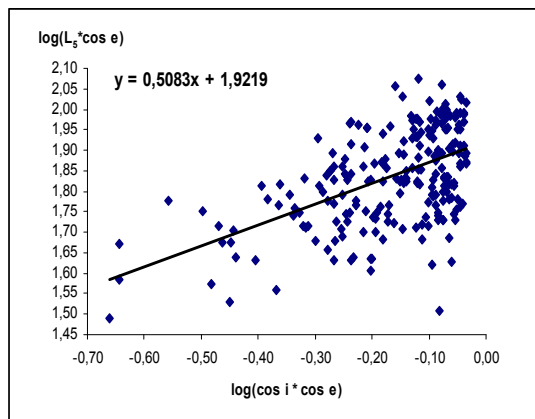
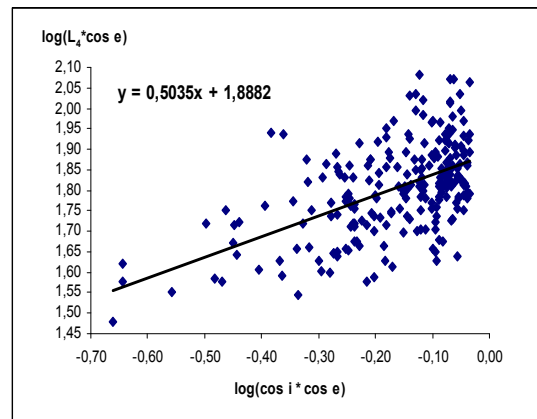
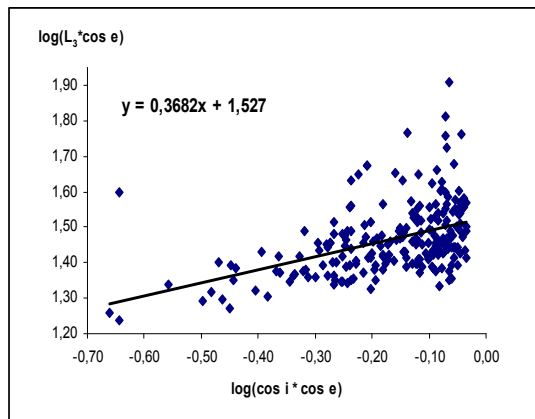
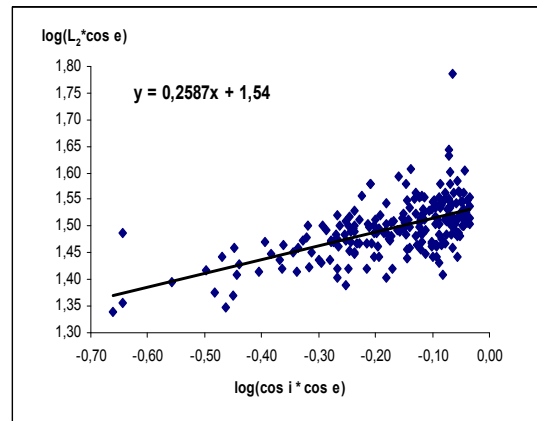
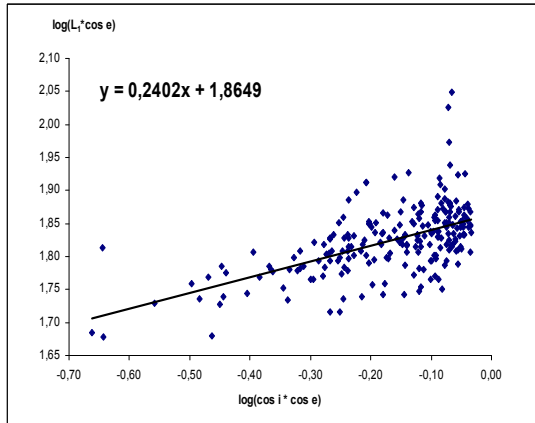
9. ANNEXES

Annex. 1: General forest strata and their description in the Korean NFI (KFRI, 1996).

Classification		Label	Description	
Stocked Land	Tree Species	<i>Pinus densiflora</i>	D	
		Planted <i>Pinus densiflora</i>	PD	
		Planted <i>Pinus koraiensis</i>	PK	
		Planted <i>Larix leptolepis</i>	PL	
		Planted <i>Pinus rigida</i>	PR	
		<i>Quercus</i>	Q	Crown closure or number of trees of a specified tree species $\geq 75\%$
		Planted <i>Quercus</i>	PQ	
		<i>Populus</i>	Po	
	<i>Castanea crenata</i>	Ca		
	<i>Cryptomeria japonica</i>	Cr		
	<i>Chamaecyparis obtuse</i>	Co		
	Stand cover type	Coniferous forest	C	
		Deciduous forest	H	Crown closure or number of tree of deciduous species $\geq 75\%$
Mixed forest		M	Crown closure or number of tree of coniferous species or deciduous species 25-74%	
Un- stocked land	Cutover area	F	The land that temporarily lost standing trees and bamboos	
	Treeless area	O	The land with $50\% \leq$ crown closure $\leq 75\%$	
	Denuded area	E	The land with crown closure $\leq 50\%$	
	Grass area	LP	Grass and pasture	
	Cultivated area	L	Agriculture and orchard area	
Non-forest area		R	Road, rock, graveyard, military area, etc.	
		W	Water, stream, marshland, etc.	
Bare Land		○	The land with crown closure $\leq 30\%$ within stocked area	

ANNEXES

Annex. 2: Regression lines for each band to estimate the Minnaert constants (k) using Landsat TM.



Annex. 3: Volume functions by tree species and DBH classes (KFRI, 1996).

$$V = a_1 * DBH^{a_2} * H^{a_3}$$

Tree species	DBH Class (cm)	Coefficients		
		a_1	a_2	a_3
PD	≤ 10	0.00005611	1.6416	0.9137
	12 ≤ DBH ≤ 20	0.00005944	1.7927	1.0422
	22 ≤ DBH ≤ 30	0.00012246	1.5616	1.0531
	32 ≤ DBH ≤ 40	0.00002324	2.0479	1.0536
	≥ 40	0.00026321	1.6155	0.7746
PK	≤ 10	0.00006730	1.8523	0.9128
	12 ≤ DBH ≤ 20	0.00004947	1.9594	0.9682
	≥ 22	0.00010498	1.7566	0.9050
PR	≤ 10	0.00087862	1.9427	0.6909
	12 ≤ DBH ≤ 20	0.00004079	2.0354	0.9056
	≥ 22	0.00004263	2.1605	0.8176
PL	≤ 10	0.00005303	1.8553	0.9995
	12 ≤ DBH ≤ 20	0.00003623	1.8252	1.1871
	≥ 22	0.00006748	1.7988	1.0025
Q	≤ 10	0.00005595	1.8062	1.0084
	12 ≤ DBH ≤ 20	0.00005464	1.7676	1.0602
	22 ≤ DBH ≤ 30	0.00005139	1.8254	1.0103
	≥ 32	0.00003147	1.8110	1.1957

PD : *Pinus densiflora*PK : *Pinus koriensis*PL : *Larix leptolepis*PR : *Pinus rigida*Q : *Quercus* and other deciduous tree species

ANNEXES

Annex. 4: Forest strata per sub-plot unit for the 25 pilot clusters.

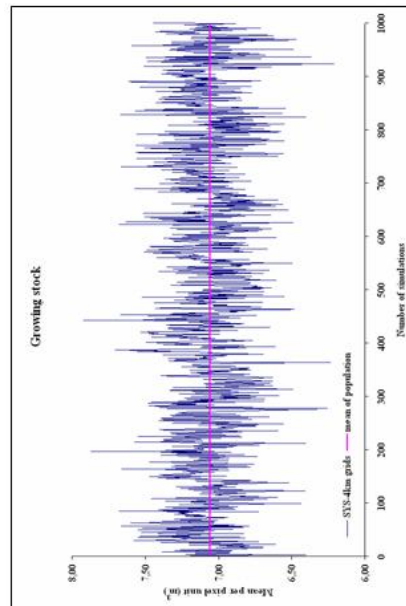
Subplot Cluster	1	2	3	4	5	6	7	8	9	10
1	H	H	H	H	H	H	H	H	H	H
2	PK	M	PK	PK	PL	PK	M	PK	PK	D
3	M	PL	M	H	M	M	M	M	M	H
4	M	PR	H	PR	PR	PR	PR	PR	PR	PR
5	PL	PK	PL	PK	PL	PK	PK	PK	PK	PK
6	PL	M	M	H	PL	M	M	M	M	M
7	PK	PL	PK	PK	PK	PL	PL	Non	PK	PK
8	H	PL	PL	H	M	PL	M	M	M	PK
9	H	H	Non	H	H	H	H	H	H	H
10	H	H	H	H	H	M	H	PL	H	PL
11	M	M	M	M	PL	PK	M	M	H	M
12	PK	M	PK	PK	PK	PK	PK	PK	H	M
13	PR	PR	PR	PR	PR	PR	PR	PR	PR	PR
14	PK	PK	PK	PK	PK	M	PK	PK	PK	PK
15	PR	M	PR	M	H	H	H	M	H	PR
16	PK	PK	PK	PK	M	M	PR	M	PK	PK
17	H	H	M	H	H	H	H	H	H	H
18	H	H	H	M	PK	M	M	H	PK	M
19	PK	PK	PK	PR	PR	PR	PK	PK	PK	PR
20	H	M	H	H	M	PR	PR	M	H	PR
21	M	PR	PR	PR	PR	PR	PL	PL	PL	PL
22	PK	PK	PK	M	PK	PK	PK	PK	PK	PR
23	H	H	H	H	H	H	H	H	H	M
24	M	PK	PK	PK	PK	PK	PK	PK	PK	PK
25	M	M	PR	PR	M	PR	PR	PR	M	M

Non : non-forest, C : coniferous forest (D, PK, PL, PR),

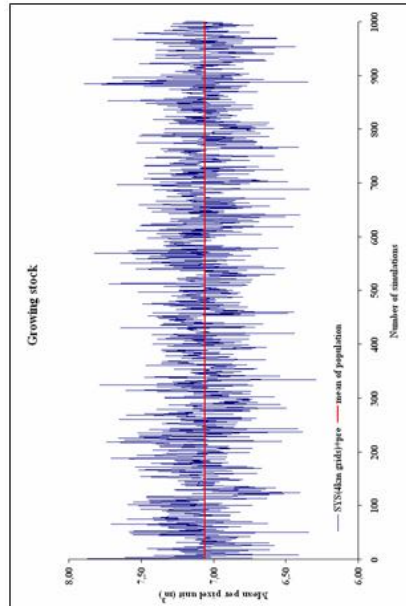
H : deciduous forest, M : mixed forest.

*Forest cover types per sub-plot are defined as the proportion of stems by dominant tree species (see Annex. 1).

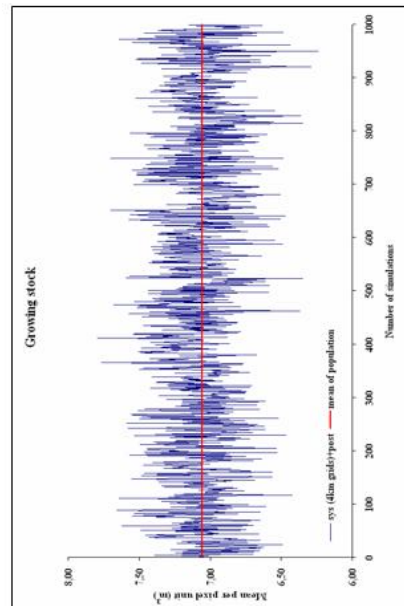
Annex. 5: Examples: distribution of estimated means ($k=1000$) for different systematic sampling designs with the grid size of 4km, where the line is the parametric mean value.



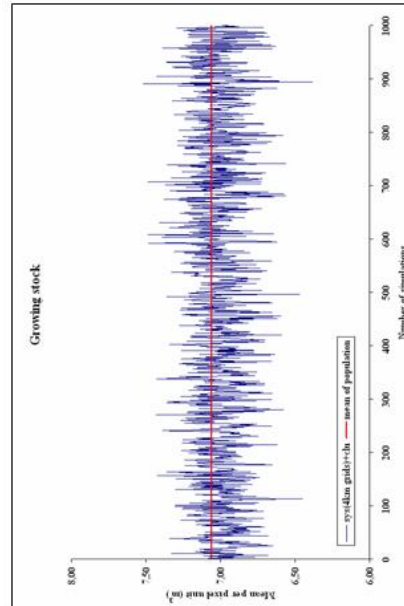
systematic sampling



stratified systematic sampling



systematic sampling with post-stratification



systematic cluster sampling

CURRICULUMS VITAE

Personal Data

Name : Jong-Su Yim
Nationality : Republic of Korea
Date of birth : 04. September 1974
Place of birth : Hwa-Seong, Republic of Korea
Family status : Married to Song-Hui Lee (since 2004), one daughter (Suh-Yoon)

Education

2003 - 2008 PhD candidate, Department of Forest Inventory and Remote Sensing,
Faculty of Forest Sciences and Forest Ecology,
Georg-August-University, Göttingen, Germany
2000 - 2002 Master course, Graduate School of College of Forest Science,
Kook-Min University, Seoul, Republic of Korea
Degree : Master of Science in Agriculture
1993 - 2000 Department of Forestry,
Kyung-Hee University, Suwon, Republic of Korea
Degree : Bachelor in Agriculture

Practical Experience

2005 - 2007 DFG-KOSEF Cooperative Program (Researcher at Göttingen university)
Topic: *Research on an efficient forest assessment design as planning tool
for sustainable forest management in South-Korea*
2002 - 2003 Assistant Researcher at Kook-Min University
Topic: *Research on a new national forest inventory system in South-Korea*

Military Service

1994 - 1997 Republic of Korea Air Force