

## kNN 기법을 이용한 강원도 평창군의 산림 주제도 작성과 산림통계량 추정

任鍾洙<sup>1</sup> · 孔知洙<sup>2</sup> · 金性鎬<sup>2</sup> · 申萬鏞<sup>3\*</sup>

<sup>1</sup>독일 괴팅겐 대학 임학과, <sup>2</sup>국립산림과학원 산림자원정보과, <sup>3</sup>국민대학교 산림자원학과

## Forest Thematic Maps and Forest Statistics Using the *k*-Nearest Neighbor Technique for Pyeongchang-Gun, Gangwon-Do

Jong-Su Yim<sup>1</sup>, Gee Su Kong<sup>2</sup>, Sung Ho Kim<sup>2</sup> and Man Yong Shin<sup>3\*</sup>

<sup>1</sup>Institute of Forest Management, Büsingenweg 5, D-37077 Göttingen University, Germany

<sup>2</sup>Division of Forest Resource Information, Korea Forest Research Institute, Seoul 130-712, Korea

<sup>3</sup>Department of Forst Resources, Kookmin University, Seoul 136-702, Korea

**요약:** 본 연구는 야외조사 자료와 원격탐사 자료를 연계하여 야외조사가 이루어지지 않은 미 관측지점의 산림정보를 추정하고 산림 주제도를 작성할 수 있는 kNN 기법을 이용하여 강원도 평창군을 대상으로 산림정보별 주제도를 작성하고, 산림통계량을 산출하였다. 수평참조범위 반경을 20, 40, 60 km로 구분한 후, 각 반경별 추정치의 오차를 비교하였다. 반경 60 km일 때, 최소 오차를 갖는 것으로 분석되었지만, 반경 40 km와 비교하면 차이가 없는 것으로 파악되어, 반경이 작은 반경 40 km를 효율적인 참조범위로 선정하였다. 선정된 수평참조범위에서 최적의 참조 표본점의 개수를 선정하기 위하여 오차행렬을 분석한 결과, k=5가 최적의 참조 표본점개수로 관측되었다. 따라서 최소 수평반경 40 km와 k=5의 참조표본점수를 이용하여 평창군 산림의 ha당 제적, 흉고단면적, 그리고 본수에 대한 주제도를 작성하였다. 작성된 주제도에 의해 추정된 산림통계량은 야외조사에 의한 추정치보다 과소추정치를 나타내었다.

**Abstract:** This study was conducted to produce forest thematic maps and estimate forest statistics for Pyeongchang Gun using the *k*NN technique, which has been applied to produce thematic maps of variables of interest including unobserved plots by combining field plot data, remotely sensed data and other digital map data in forest inventories. The estimation errors for three horizontal reference areas (HRAs), whose radii are 20, 40 and 60 km respectively, were compared. Although the precision for the 40 km radius was lower compared to that for the 60 km radius, the 40 km radius was found to be an efficient HRA because their difference in precision was modest. At a value of *k*=5 nearest neighbors for the selected HRA, the overall accuracy was high. As a result, using the *k*=5 neighbors within the HRA of 40 km radius, thematic maps of number of trees, basal area, and growing stock per hectare were generated. As compared to the forest statistics based on field sample plots, the estimated means of each parameter from the produced maps were underestimated.

**Key words:** *k*-nearest neighbor technique, forest thematic map, satellite imagery, national forest inventory

### 서론

전국 산림자원조사는 산림자원 및 관련 분야에 대한 국가 단위의 의사결정과 모니터링을 주목적으로 대부분의 나라에서 수행되고 있다. 또한 산림자원조사는 지역별 또는 행정구역별 자료를 제공하며, 국제적인 산림 평가 및 산림자원의 지속 가능한 경영을 위한 보고서 작성에 이용된다(산림청, 2004b). 이러한 산림자원조사는 효율적인 표

본 설계에 따른 야외 표본 조사와 대면적의 정보를 획득할 수 있는 원격탐사 기법을 이용하고 있으며, 또한 두 방법이 효율적으로 연계되어 수행되기도 한다. 산림분야와 관련하여 위성영상을 이용한 원격탐사 기법은 주로 산림경영계획, 산림자원조사, 그리고 산림피해 등의 모니터링을 위해 이용되고 있다. 특히 산림자원조사 분야에서는 임상도의 제작, 산림 정보의 추정, 산림의 단편화(Fragmentation) 구획, 경관 생태 분석, 그리고 전국 산림자원조사와 같은 대면적 산림자원조사에 활용되고 있다(Holmgren and Thuresson, 1998).

\*Corresponding author  
E-mail: yong@kookmin.ac.kr

우리나라의 경우 산림자원조사와 관련하여 원격탐사 기법은 항공사진 판독 및 위성영상을 이용한 임상도 및 토지이용 구분도의 제작이 주를 이루고 있다. 전국 산림자원조사에서 수집된 야외 표본점 자료가 항공사진 판독에 작성된 임상도의 정확성 검정을 위한 자료로 이용되기도 하였지만(임업연구원, 1996), 위성영상과 같은 다른 부가자료와 연계한 연구는 미흡한 편이다(이승호 등, 2004). 위성영상과 야외 표본점 자료를 통합하기 위해서는 우선 야외표본점의 정확한 위치 정보의 수집이 요구된다(Tokola *et al.*, 1996; Halme and Tomppo, 2001). 하지만 지난 산림자원조사에서 수집된 야외 표본점의 위치정보는 다른 장비의 이용없이 1:50,000 지형도 도면상에 표본점의 위치를 표시하였기 때문에, 표본점 자료와 다른 자료와의 연계가 이루어 지지 못하였다. 그러나 최근 위치 추적을 위한 다양한 GPS 장비의 도입과 지리 정보시스템의 발달로 표본점의 위치에 대한 오차를 감소시킬 수 있으며(Kleinn, 2002), 특히 제5차 산림자원조사에서는 야외 표본점의 위치 정보를 추적하기 위하여 GPS(Global Positioning System)가 이용되는 만큼 야외 표본점의 자료와 위성영상에서 획득된 자료의 연계가 가능하다(산림청, 2005). 또한 산림지리정보시스템(FGIS)에서 제작된 산림관련 주제도를 활용하여 보다 정도(精度)가 높은 산림 정보의 추정이 이루어 질수 있을 것으로 기대된다.

k-Nearest Neighbor 기법은 훈련 자료(training data)를 이용하여 새로운 자료의 등급을 구분하거나 추정하는 알고리즘의 하나이다. 이 기법은 산림자원조사 분야에 있어서, 야외조사 자료와 위성영상 등의 부가자료와 연계하여 야외조사가 이루어지지 않은 지점에 대한 산림정보를 추정하고, 추정된 정보의 연속적인 수치 레이어(digital layer), 즉 주제도(thematic map)를 작성하는 데 이용되고 있다(Tomppo, 1990; Katila and Tomppo, 2001). 핀란드에서는 1990년도 이후 국가 산림자원조사 자료와 연계하는 다목적 산림자원조사에서 kNN 기법을 적극적으로 활용하고 있으며(Tomppo, 1990), 스웨덴의 국가 산림자원조사(Reese *et al.*, 2002; 2005)뿐만 아니라 미국에서도 주별 산림 정보를 추정하기 위하여 적용되고 있다(McRoberts *et al.*, 2002). 최근에는 kNN 기법에 의한 추정치의 오차를 감소시키기 위하여, 참조 표본점의 범위 선정에 관한 연구(Tokola and Heikkilä, 1997; Katila and Tomppo, 2001), 다양한 거리 산출 방법의 적용(Franco-Lopez *et al.*, 2001), 밴드별 가중치에 관한 연구(Tokola *et al.*, 1996; Tomppo and Halme, 2004), 그리고 야외 표본점의 위치정보 보정을 위한 영상 처리 기법의 적용(Franco-Lopez *et al.*, 2001) 등의 다양한 연구가 진행되고 있다.

본 연구는 지형적 변이가 심한 우리나라 산림을 대상으로 kNN 기법을 적용하여 야외조사가 이루어지지 않은 지

점을 포함한 연구 대상지의 산림을 대상으로 산림변수별 주제도를 작성하고, 산림 통계량을 추정함으로써 우리나라 전국산림자원조사에서 kNN 기법의 활용방안과 적용 가능성을 검토하기 위해 수행하였다.

## 재료 및 방법

### 1. 연구 대상지

본 연구의 대상지는 비교적 험준한 산악지역에 위치한 강원도 평창군을 대상으로 하였다. 평창군의 중심 경위도는 북위 37.50와 동경 128.42이며, 총 면적은 약 1,463 km<sup>2</sup>인데 그 중에서 약 84%가 산림이고 10%는 농지, 그리고 기타 용지가 6%로 구성되어 있다(산림청, 2004a). 평창군의 해발고도는 약 210m~1,570m의 범위에 있으며, 평균 해발고도는 약 600m이다. 이 지역의 주요 수종으로는 소나무(*Pinus densiflora*), 잣나무(*Pinus koraiensis*), 낙엽송(*Larix leptolepis*) 그리고 신갈나무(*Quercus mongolica*) 및 기타 활엽수들이 분포하고 있다.

### 2. k-Nearest Neighbor(kNN) 기법

kNN 기법은 비모수(non-parameter) 그리고 다변량(multivariate) 회귀 추정식의 하나로 새로운 또는 알려지지 않은 목표 표본점의 정보를 얻기 위하여 위성 영상과 같은 부가정보를 활용한다. 이를 위해 목표 표본점과 야외 조사가 실시된 표본점의 부가정보의 유사함을 산출하고, 목표 표본점의 정보를 조사된 측정치와 가장 유사한 부가정보를 갖는 표본점의 측정치로 분류 및 추정하는 기법이다(Tomppo, 1990). 여기에서 유사성(similarity)은 훈련 데이터 셋(training data set) 내의 참조 표본점과 목표 표본점 사이의 밴드들의 거리에 의해 측정되어 진다.

산림자원조사에 있어서, kNN 기법은 야외 조사에서 수집된 자료를 이용하여 알려지지 않은, 즉 조사가 이루어지지 않은 표본점에 대하여 산림 정보를 추정하는데 이용되고 있다. 이 기법에 있어서 야외조사 자료는 훈련 표본점이 되며, 또한 추정오차를 산출하는 데 있어서 목표 표본점으로도 이용되어 진다. Figure 1은 kNN 기법의 적용 순서를 나타낸 그림이다. 이 기법을 적용하기 위해서는 우선 위치정보를 포함한 야외 표본점의 자료, 위성영상 또는 수치 항공사진 자료, 그리고 산림과 비산림이 구분되어 있는 주제도가 요구된다.

#### 1) 참조 표본점의 선정

조사가 이루어지지 않은 목표 표본점의 산림 정보를 추정하기 위해서는 우선 훈련 표본점내에서 목표 표본점의 정보를 추정하기 위한 참조 표본점을 선정하여야 한다. 참조 표본점의 선정은 목표 표본점( $t$ )과 훈련 데이터셋 내에

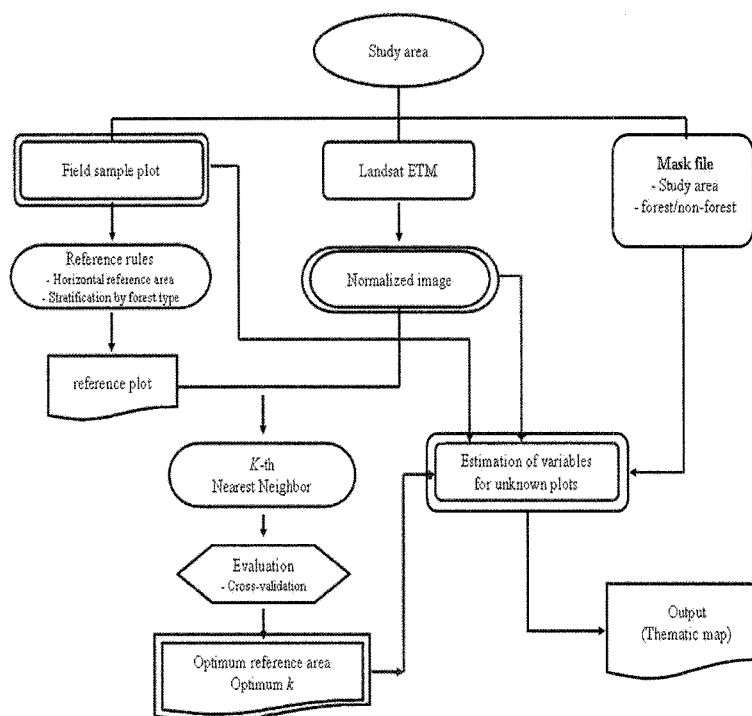


Figure 1. Flow chart of the kNN technique application.

있는 참조 표본점( $r$ )에서 추출된 각 픽셀에서의 밴드별 값(digital number)들의 거리에 의해 산출되며, 본 연구에서는 아래의 유클리디안 거리식을 적용하였다. 여기서,  $d_{t,r}$ 은 목표 표본점( $t$ )과 훈련 데이터셋 내의 참조 표본점( $r$ )간의 거리를 나타내며,  $x_{i,t}$ 와  $x_{i,r}$ 은 각각 영상 밴드  $i$ 에서의 목표 표본점과 참조 표본점의 밴드별 값, 그리고  $m$ 은 위성영상내의 밴드 개수를 나타낸다.

$$d_{t,r} = \sqrt{\sum_{i=1}^m (x_{i,t} - x_{i,r})^2}$$

2) 참조 표본점별 거리 가중치

목표 표본점과 참조 표본점들 간의 밴드 값에 의한 거리를 산출한 후, 목표 표본점과 참조 표본점이 거리가 가까울수록, 즉 두 표본점의 밴드들의 값들이 유사할수록 보다 큰 가중치를 부여하는 방법 중에서, 본 연구에서는 가장 일반적으로 이용되는 방법인 거리 역산에 의해 가중치를 부여하는 아래의 식을 이용하여 참조 표본점별 가중치를 산출하였다.

$$w_{t,r} = \frac{1}{d_{t,r}} \sum_{r=1}^k \frac{1}{d_{t,r}}$$

3) 목표 표본점의 추정치

kNN 기법에 있어서,  $k$ 번째 이웃 참조 표본점의 선정은 목표 표본점과 참조 표본점에서의 밴드들의 값에 의한 거

리에 의해 결정되어지며, 목표 표본점( $t$ )에서의 산림 정보의 추정치( $\hat{v}_t$ )는 앞에서 선정된 참조표본점의 산림 정보의 측정치( $v_r$ )와 각 참조 표본점별 산출된 가중치( $w_{t,r}$ )를 이용하여 다음과 같은 수식에 의해 추정되어 진다.

$$\hat{v}_t = \sum_{r=1}^k w_{t,r} \times v_r$$

3. 연구 방법

1) 야외 표본점 자료의 수집

야외 조사 표본점의 위치를 선정하기 위해서 본 연구에서는 연구 대상지인 강원도 평창군 전체에 걸쳐 고루 표본점이 배치될 수 있도록 계통적 추출방법에 의해 표본점을 배치하였다. 즉, 연구 대상지를 4 km×4 km 격자로 나누는 후 각 교점을 표본점으로 하였다(Figure 2). 전체 표본점은 91개인데, 제3차 산림자원조사에서 제작된 임상도와 중첩한 결과 산림지역에 포함된 표본점은 75개 그리고 비산림지역에는 16개로 분류되었다. 하지만 산림지역에 포함된 75개 표본점 중에서 군사지역 및 토지이용 용도가 변경된 지역, 그리고 접근이 곤란한 표본점들을 제외한 61개 표본점에 대해서만 2005년 여름에 야외 조사를 실시하였다.

표본점은 0.05ha 크기의 원형 표본점이며, 각 표본점에서 흉고직경 6 cm 이상의 모든 입목을 대상으로 수종명을 기록한 후 흉고직경과 수고를 측정하였으며, 또한 소형 GPS 장비(GARMIN V DELUXE)를 이용하여 각 표본점

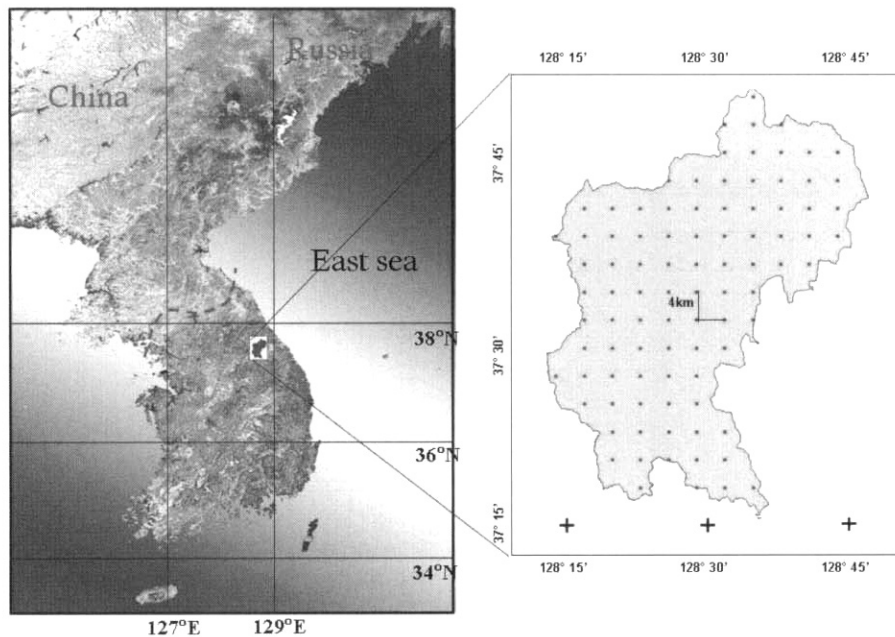


Figure 2. Location of study area and distribution of field sample plots.

Table 1. Summary statistics of the field sample plots.

Variables	No. of Sampling Points	Mean	Minimum	Maximum	Standard Deviation
N/ha	61	1068	240	2200	431.9
BA/ha(m <sup>2</sup> )	61	32.5	1.6	70.3	13.1
V/ha(m <sup>3</sup> )	61	177.8	5.3	430.3	82.7

중심의 해발고도와 경위도의 위치정보를 수집하였다. 각 표본점의 경위도 좌표는 국립지리원에서 배포하는 좌표변환 프로그램을 이용하여 우리나라의 좌표계인 TM으로 변환되었다. 야외조사에서 수집된 자료를 이용하여, 각 표본점의 ha당 흉고단면적과 본수를 산출하고, 전국 산림자원조사요령의 각 수종 및 지역별 재적 추정식을 이용하여 ha당 재적을 산출하였다(임업연구원, 1996). Table 1은 야외조사 표본점, 즉 참조 표본점으로 이용될 자료의 요약이다.

2) 위성영상 자료의 이용

본 연구에서 이용된 위성영상은 2002년 4월 28일에 촬영된 Landsat-7 ETM 영상(row 115/path 34)이다. Landsat-7 ETM 영상은 7개의 밴드로 구성되어 있으며, 본 연구에서는 공간해상도 60m를 나타내는 열적외선 밴드인 6번 밴드를 제외한 6개 밴드를 이용하였다. 이 영상의 대한 태양의 고도각과 방위는 각각 59.4°와 134.6°이다. 또한 이 영상은 구름이 없으며, 수치 지형도를 활용하여 픽셀당 전체 RMSE(Root Mean Square Error)가 25m 미만으로 기하보정 되었다.

3) 주제도의 이용

연구 대상지인 평창군은 험준한 산악지역에 위치하고

있으므로, 태양의 고도와 위치에 따른 지형오차가 나타나므로 이러한 오차를 감소시키기 위해 위성영상에 대하여 지형보정 기법중의 하나인 Minnaert 상수에 의한 지형보정을 수행하였다(Lee and Yoon, 1997; Cho, 2002). 밴드별 Minnaert 상수를 추정하기 위하여 1:25,000 수치지형도를 이용하여 위성영상의 픽셀크기와 동일한 25m×25m 픽셀 크기를 갖는 Digital Elevation Model(DEM) 자료를 생성하였으며, 이 자료와 위성영상 촬영시의 태양의 고도각과 방위정보를 이용하여 영상에 나타난 지형에 따른 오차를 보정하였다.

또한, kNN 기법은 산림뿐만 아니라 연구 대상지내 모든 픽셀에 대하여 추정식을 실행하므로, 관심 대상인 연구대상지내 산림지역의 추정치를 산출하기 위해서는 산림지역과 비산림지역에 대한 구분이 된 주제도가 요구된다. 하지만 조사 시점에서의 산림과 비산림으로 구분된 주제도를 획득하는데 한계가 있어 본 연구에서는 제3차 전국 산림자원조사에서 제작된 수치임상도를 이용하여 연구 대상지의 산림과 비산림을 구분하였다.

4) 참조 표본점의 범위 결정

산림은 지형, 토양 그리고 기후 조건에 따라 임분 구조가 상이하기 때문에, 참조 표본점의 범위를 동일한 특성

을 갖는 등급으로 구분한 후 참조표본점의 범위를 목표 표본점과 동일한 등급 내에서 선정하는 것이 보다 효율적인 추정치를 산출할 수 있다. 이러한 참조 표본점의 구분을 위하여 수치토양도(Tokola and Heikkilä, 1997)와 수치 토지이용구분도(Katila and Tomppo, 2002) 등의 다양한 주제도들이 이용되고 있으며, 또한 수평참조범위와 해발고도 등의 지형특성을 이용한 방법들도 적용되고 있다(Tokola, 2000; Katila and Tomppo, 2001). 본 연구에서는 참조 표본점의 범위를 수평참조범위와 임상을 이용한 방법을 적용하였다.

(1) 수평참조범위(Horizontal Reference Area, HRA)에 의한 방법

산림자원조사에서 야의 조사를 실시한 면적이 증가할 수록 산림 변수(재적 등)와 영상의 밴드 값들과의 관계는 매우 다양하게 나타난다. 따라서 수평참조범위가 증가하면 참조 표본점의 개수가 증가하게 되며 추정치는 편차가 발생될 수 있으며, 한편 수평참조범위가 작을 경우 참조 표본점의 개수가 감소하기 때문에 이 경우에는 표본의 모든 변량을 포함하는 최소 수평참조범위의 선정이 요구된다. 본 연구에서는 수평참조범위에 의한 범위를 각 목표 표본점을 중심으로 반경 20 km, 40 km, 그리고 60 km의 3가지로 구분하였다. 반경 60 km의 경우에는 연구 대상지의 모든 참조 표본점을 포함하게 된다. 수평참조범위에 의한 참조 표본점의 선정은 연구 대상지의 모든 픽셀과 참조 표본점의 xy 좌표에 의한 유클리디안 거리를 산출하여 적용하였다. 하지만 수평참조범위에 의한 방법을 이용할 경우 야외조사가 연구대상지에서만 수행되었기 때문에 경계 부분에 위치한 목표 표본점들은 상대적으로 적은 참조 표본점의 개수를 포함하게 된다.

(2) 임상별 층화에 의한 방법

산림의 특성은 기후, 지형조건, 토양, 수종 그리고 토지 이용구분 등에 의해 차이가 있기 때문에 다양한 부가 자료를 이용하여 참조 표본점의 범위를 구분한 후 참조 표본점을 선정하는 방법에 관한 연구가 진행되고 있다(Tokola and Heikkilä, 1997; Katila and Tomppo, 2001). 본 연구에서는 산림의 특성을 임상에 따라 구분하여 참조 표본점을 선정하는 방법을 적용하였다. 즉, 목표 표본점이 임상도 상에 침엽수림 산림에 위치할 경우, 참조 표본점의 범위를 혼련 표본점내에서 동일한 침엽수림 내에 있는 표본점으로 제한하는 방법이다. 야의 표본점의 자료를 임상별로 구분할 경우, 침엽수림 32개 표본점, 활엽수림 21개 표본점 그리고 혼효림 8개 표본점에 불과하여 최대 참조 표본점의 개수를 10개까지만 이용하였다.

5) 통계적 검증

kNN 기법에 의한 추정치의 검증과 최적의 참조 표본점

의 개수(k)를 선정하기 위하여 본 연구에서는 교차 대조법(cross-validation)에 의한 RMSE와 편차(bias), 그리고 측정치와 추정치의 등급별 오차행렬(error matrix)을 이용한 전체 정확도(Overall Accuracy, OA)를 산출하였다.

(1) 교차 대조법

kNN 기법의 추정 능력을 평가하기 위하여 참조 표본점으로 이용된 야의 표본점에서 실측된 재적의 측정치와 kNN 기법에 의한 재적 추정치를 이용하는 교차 대조법을 이용하였다. 본 연구에서 사용한 교차대조법에 의한 추정치의 오차는 다음의 수식에 의해 산출하였다.

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}$$

여기서  $y_i$ 는 참조 표본점  $i$ 에서의 재적 측정치,  $\hat{y}_i$ 는 kNN 기법에 의한 재적 추정치, 그리고  $n$ 은 참조 표본점의 개수이다. 또한, 편차(bias)는 아래 식에 의하여 산출하여 비교하였다.

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

(2) 오차 행렬

kNN 기법에 의해 얻어진 재적 추정치를 평가하기 위하여 분류기법에서 이용되는 오차행렬 방법을 사용하였다(Franco-Lopez *et al.*, 2001). 본 연구에서 야의 표본점에서 재적의 추정치와 측정치를 각각 5개의 등급(0~100, 100~150, 150~200, 200~250, 그리고 250 m<sup>3</sup>/ha 이상)으로 구분하였으며, 전체 정확도(OA)를 각 참조 범위 선정방법과 참조 표본점의 개수별로 산출하여 비교하였다.

6) 산림주제도 작성 및 산림통계량 산출

산림 정보별 주제도를 작성하기 위하여, 위성영상에서 연구 대상지내에 포함된 모든 픽셀에 대하여 중심 위치 정보(xy좌표)와 밴드 값을 추출하였다. 하지만 산림주제도의 작성에 있어서, 수집된 임상도의 임상은 제3차 산림자원조사에서 작성되었기 때문에 야외조사와 위성영상 자료가 수집된 시점과는 시간의 경과에 따른 산림의 변화가 있을 것으로 판단되어, 참조 표본점의 선정방법은 수평참조범위만을 적용하였으며, 선정된 최적의 참조 표본점의 개수를 이용하여 평창군의 모든 픽셀에 대하여 ha당 흉고단면적, 임목분수, 그리고 재적을 추정하였다. 마지막으로, 관심대상이 산림이므로 작성된 산림과 비산림의 주제도를 이용하여 산림지역내의 정보를 추출하고, 각 변수별 주제도를 작성한 후 산림통계량을 산출하였다.

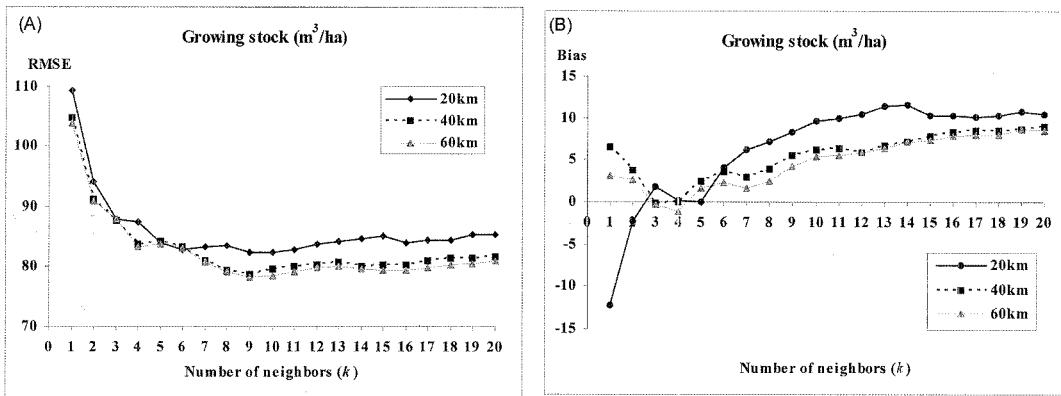


Figure 3. The root mean square error(A) and bias(B) for different horizontal reference areas and different number of neighbor plots( $k$ ).

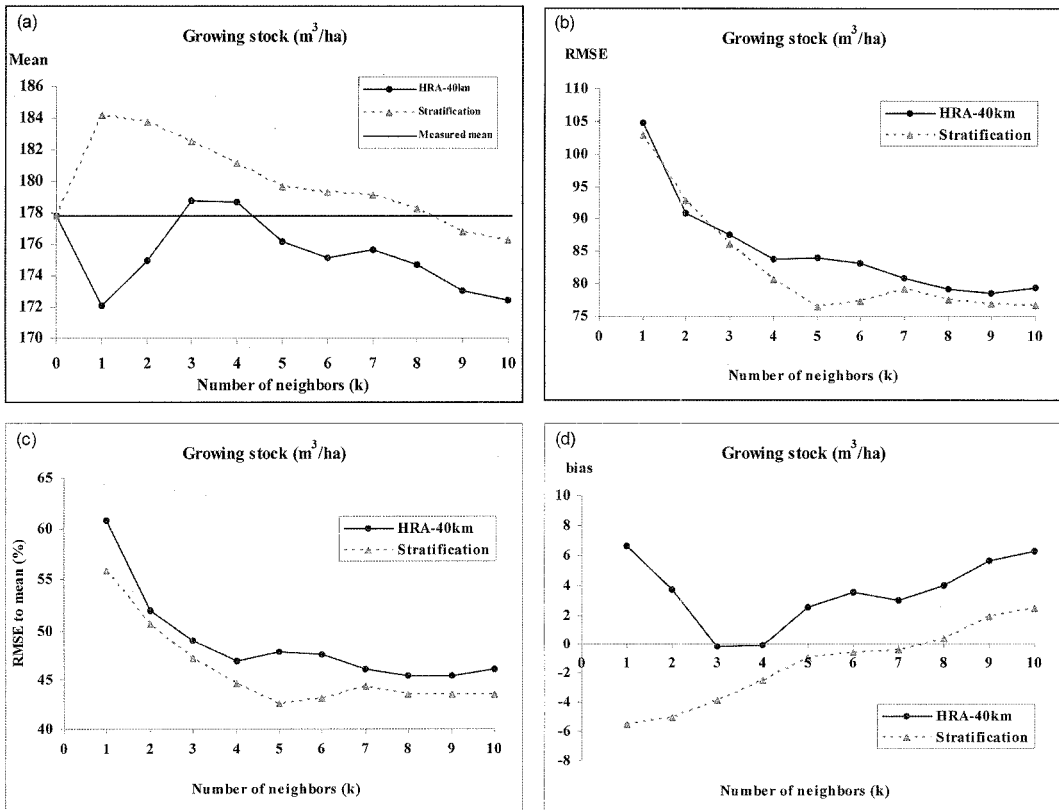


Figure 4. Mean(a), root mean square error(b), RMSE to mean(c) and bias(d) for different selecting rules and different number of neighbors( $k$ ).

### 결과 및 고찰

#### 1. 참조 범위의 선정

##### 1) 수평참조 범위(Horizontal Reference Area, HRA)

Figure 3은 본 연구 대상지에서 3가지 수평참조범위 반경(20 km, 40 km, 60 km)과 참조 표본점의 개수에 따른 ha당 재적의 RMSE와 편차를 산출한 결과이다. RMSE는 참조 범위의 반경과 참조 표본점의 개수가 증가할수록 감소하였다. 이러한 결과는 전형적인 kNN 기법의 결과와 유사하다(Katila and Tomppo, 2001). 즉 수평참조범위가 증가할수록 참조 표본점으로 이용가능한 표본점의 수가 증

가하여, 보다 다양한 변량이 추정을 위해 이용되기 때문에 추정 오차가 감소하게 된다.

참조 표본점의 개수가 1에서 4개로 증가할 때 RMSE는 급격히 감소하였으며, 특히 수평참조범위 20 km 반경에서 5개 이상 그리고 반경 40 km와 60 km에서는 참조 표본점 10개 이상부터는 참조 표본점의 개수가 증가하여도 차이가 없는 것으로 관측되었다. RMSE는 수평참조범위 반경이 60 km이고 참조 표본점의 개수가 9개일 때 최소값을 나타냈으며, 반경 40 km일 경우에도 반경 60 km의 RMSE와 그 재적의 차이는 1  $m^3/ha$  미만에 불과하였다.

Altman(1992)은 kNN 기법에 의한 추정치는 참조 표본

점의 개수가 증가할수록 편차가 증가되는 경향을 나타낸다고 하였는데, 본 연구에서도 Figure 3에서 볼 수 있듯이 편차는 참조 표본점의 개수가 1개에서 4개까지는 감소하였으며, 그 이상부터는 참조 표본점의 개수가 커질수록 증가하였다. 수평참조범위별로 살펴보면, RMSE와 편차는 수평참조범위가 증가할수록 감소하였으나, 반경 40 km와 60 km에서는 큰 차이가 없는 것으로 나타났다. 즉, 반경 40 km 보다 넓은 수평참조범위를 적용하는 것은 시간 소모적인 것으로 판단되며, 따라서 본 연구에서는 반경 40 km를 갖는 수평참조범위가 효율적인 참조 표본점을 갖는 최소 수평참조범위로 판명되었다. 한편, 최소 수평참조범위와 관련하여, Katila and Tomppo(2001)은 최소 수평참조범위는 연구 대상지의 산림구조와 지형조건 및 야외 표본점의 표본설계에 따라 차이가 발생한다고 발표하였다.

2) 참조표본점 선정방법의 비교

참조표본점의 선정방법과 참조 표본점의 개수에 따른 추정치의 평균, RMSE, 상대 RMSE 그리고 편차를 비교하였다(Figure 4). Figure 4-a는 참조범위 선정방법과 참조 표본점의 개수에 따른 추정치의 평균을 비교한 것으로, 참조 표본점의 개수 0( $k=0$ )은 야외 표본점에서 측정된 ha당 재적의 평균을 의미한다. 수평참조범위에 의한 선정방법에서는 참조 표본점의 개수가 3과 4개일 때 과대추정치를 나타내었으며, 반면 층화에 의한 방법은 참조 표본점의 개수가 9와 10개일 때만 과소추정 하였으며 9개보다 적은 참조 표본점의 개수를 이용할 경우에는 과대 추정하는 것으로 나타났다.

한편 RMSE는 참조 표본점의 개수가 증가할수록 감소하였으며, 참조 표본점의 수가 3개 이상에서는 층화에 의한 추정치가 수평거리에 의한 추정치보다 정도(精度)가 높은 것으로 관측되었는데, 추정치의 평균을 고려한 상대 RMSE(%)를 보면 보다 뚜렷한 경향을 알 수 있다(Figure 4-c). 또한 편차의 경우에도, 참조표본점의 개수가 3과 4개일 때는 수평참조범위에 의한 편차가 작았으며, 다른 참조 표본점의 개수에서는 층화에 의한 편차가 작은 것으로 관측되었다. 또한 오차행렬에 있어서도, 참조표본점의 개수가 2개일 때를 제외하고는 임상 구분에 의한 층화에 의한 방법이 높은 정확도로 추정되었다(Figure 5). 이러한 결과는 연구 대상지의 면적이 작기 때문에, 수평참조범위에 의한 지형적 특성보다는 산림의 특성에 고려하여 참조 표본점을 선정하는 것이 추정치의 오차를 감소시킬 수 있는 것으로 판단된다.

2. 최적의 참조 표본점 개수의 선정

1) 교차 대조법

일반적으로, kNN 기법에서 추정치의 오차를 나타내는

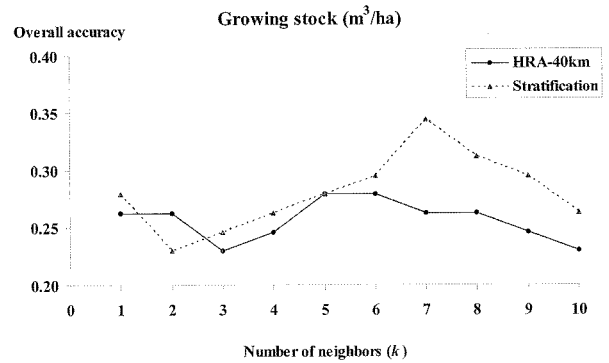


Figure 5. Overall accuracy for different selecting rules and different number of neighbors(k).

RMSE는 참조 표본점의 개수가 증가할수록 감소하는 경향을 나타낸다(Franco-Lopez *et al.*, 2001). 본 연구에서도 참조범위를 선정하는 두 방법 모두에서 RMSE는 참조 표본점의 개수가 증가할수록 감소하였다(Figure 3). 하지만, 추정된 RMSE의 값은 핀란드(Katila and Tomppo, 2001)와 미국(Franco-Lopez *et al.*, 2001)에서 수행된 연구결과와 비교하면, 최소 수평참조범위에 의한 추정치의 오차는 비교적 큰 78.5~104.7 m<sup>3</sup>/ha의 범위로 관측되었으며, 최소 RMSE는 참조 표본점의 개수가 9개에서 관측되었다. 이러한 결과는 본 연구에서 이용된 참조 표본점의 개수가 충분하지 않기 때문에, 즉 한정된 변량에 의한 원인으로 판단된다. 수평참조범위 반경 40 km에서 RMSE가 안정화되는 참조 표본점의 4개와 최소의 RMSE 값의 차이는 대략 7%인 것으로 나타났다. 편차의 경우에도 참조 표본점의 개수가 4개일 때, 최소인 -0.07 m<sup>3</sup>/ha로 관측되었다.

2) 오차 행렬

참조 표본점의 개수에 따른 전체 정확도(OA)를 분석한 결과는 Figure 5와 같다. 본 연구에서는 전체 정확도는 참조 표본점의 개수에 따라 증가하였다가 다시 감소하는 경향을 보이고 있다. 수평참조범위 반경 40 km에서 0.23~0.28의 범위로 추정되었으며, 이는 다른 연구결과에 비해 낮은 정확도를 보이고 있다(Franco-Lopez *et al.*, 2001). 이러한 결과는 본 연구에서 수집된 표본점의 개수가 상대적으로 적기 때문으로 여겨진다. 최대 전체정확도는 참조 표본점의 개수가 5개일 때 0.28로 분석되었다.

3. 산림 주제도 작성과 산림통계량 산출

수평참조범위에 의한 선정방법 중에서 반경 40 km가 효율적인 최소 수평참조범위로 선정되었으며, 교차대조법과 전체정확도의 결과를 바탕으로 선정된 참조 표본점의 개수( $k=5$ )를 적용하여 평창군 전체의 변수별 주제도를 작성한 후, 산림과 비산림으로 구분된 주제도를 이용하여, 산림지역내의 추정치를 추출하였다(Figure 6).

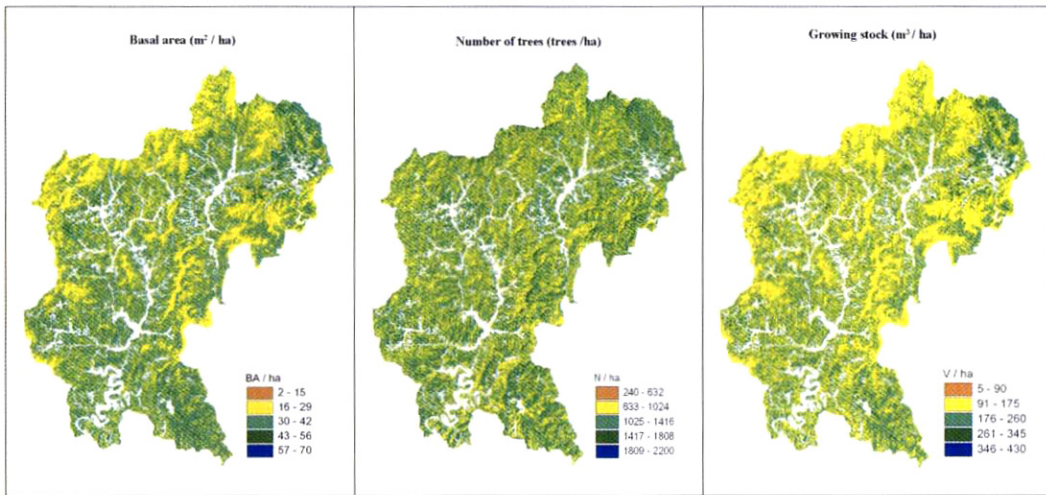


Figure 6. Thematic maps for each forest parameter from the  $k$ NN technique within a horizontal reference area of 40 km radius( $k=5$ ).

주제도의 작성과 관련하여, Franco-Lopez *et al.*(2001)은  $k$ NN 기법의 적용에 있어서, 주제도 작성을 목적으로 할 경우에는 비록 추정치의 오차는 가장 크지만, 최근린(the nearest neighbor,  $k=1$ ) 표본점을 선택하는 것이 표본의 다변량(多變量) 범위를 갖으며, 비편향의 추정치가 산출되기 때문에 최적이라고 추천하였다. Figure 6에 제시한 산림 주제도 역시 산림의 구분은 수치 임상도를 이용하였기 때문에, 현재의 산림현황과는 차이가 있을 것으로 판단된다. 하지만,  $k$ NN 기법을 이용하여 임상별 그리고 수종별 주제도의 작성도 가능하기 때문에(Franco-Lopez *et al.*, 2001), 동일한 시기의 위성영상과 야외조사 자료를 이용하여 작성된 임상도와 항공사진 판독에 의한 임상도의 비교를 통하여 위성영상을 이용한 임상도의 제작 가능성을 고려해 볼 필요가 있다.

$k$ NN 기법의 최대 장점은, 야외조사가 이루어지지 않은 미 관측지점에 대하여, 야외에서 수집된 대부분의 산림 변수(임령, 재적, 임분밀도, 바이오매스 등)에 대하여 위성영상 및 항공사진 등의 원격탐사 자료를 이용하여 각 픽셀별 추정치를 산출할 수 있다는 것이다. Table 2는 연구 대상지인 강원도 평창군에 대한 ha당 분수, 흉고단면적, 재적을  $k$ NN 기법에 의해 추정된 결과이다. 이 추정치는 연구 대상지의 미 관측 지점을 포함한 평창군 전체 산림을

전수(全數)조사한 결과와 같은 산림통계량으로, 앞에서 작성된 산림 주제도로부터 도출된 결과이다. 즉,  $k$ NN 기법에 의해 작성된 주제도는 25m×25m의 크기를 갖는 픽셀들의 집약이며, 각 픽셀은 밴드 값 대신 산림변수별 추정치를 포함하고 있는 것이다. 여기서 픽셀의 크기는 위성영상의 공간 해상도와 관련이 있으며, 산림 주제도의 변수별 추정치들은 야외에서 수집된 정보 즉, 참조 표본점의 측정치와 참조 표본점의 개수에 의존하게 된다. 결과적으로 Table 1의 추정치는 61개 야외 조사 표본점에 근거한 산림통계량인 반면, Table 2의 추정치는 훨씬 많은 픽셀들의 요약이기 때문에 표준편차가 작은 것으로 나타났다. 따라서 본 연구에서 얻은 산림통계 추정치는 비록 현실의 산림분포와는 다소 차이가 있을 것으로 예상되지만, 연구 대상지의 산림전체를 모집단으로 산출된 결과로 표본에 의한 추정치보다 좀 더 현실에 가까운 추정치가 될 것으로 평가된다.

## 결론

본 연구는 위성영상과 야외조사 자료를 연계하여 다양한 산림주제도를 작성하고, 야외조사가 이루어지지 않은 미 관측 지점의 정보를 추정하여 산림통계를 산출할 수 있는  $k$ NN 기법의 적용 가능성과 활용방안을 검토하고자 수행하였다.

참조범위를 선정하는 방법에 있어서 임분의 특성을 이용하여 참조 표본점을 선정하는 층화 방법이 수평거리를 이용한 방법보다 RMSE와 편차가 작은 것으로 추정되었으며, 또한 전체 정확도도 높아 보다 효율적인 선정방법으로 나타났다. 하지만, 본 연구에서 이용된 임상도는 위성영상과 야외자료가 수집된 시점과 상당한 차이가 있어,

Table 2. Summary statistics of forest thematic maps from the  $k$ NN technique within a horizontal reference area of 40 km radius in the study area( $k=5$ ).

Variables	Mean	Minimum	Maximum	Standard Deviation
N/ha	1065	240	2200	195.3
BA/ha(m <sup>2</sup> )	29.9	2.0	70.0	5.8
V/ha(m <sup>3</sup> )	162.3	5.0	430.0	36.4



최적의 참조 표본점의 선정과 산림 주제도의 작성에서는 이용되지 않았다. 결과적으로 본 연구에서는 수평참조범위 반경 60 km를 적용하였을 경우보다 추정치의 오차는 크지만, 반경 40 km와 비교하여 차이가 없는 것으로 파악되어, 반경이 작은 40 km가 효율적인 최소 반경으로 선정되었다. 추정치의 오차를 나타내는 RMSE는 참조 표본점의 개수가 증가할수록 감소하는 경향을 나타내기 때문에, 최소 RMSE를 나타내는 참조 표본점의 개수를 찾는 것은 시간 소모적인 것이라 판단되며, 따라서 RMSE가 안정화되는 시점과 편차가 최소가 되는 시점에서의 참조 표본점 개수( $k=4$ )를 선정하였다. 또한 참조 표본점의 개수에 따른 전체 정확도를 도출하여, 두 조건을 만족시킬 수 있는 참조 표본점의 개수를 최적의 참조 표본점의 개수로 선정하였다( $k=5$ ). 최적 참조 표본점의 개수는 연구 대상지의 산림현황, 야외 표본점의 개수, 표본 설계 방법, 그리고 위성영상의 공간적 해상도 등에 따라 차이가 있으므로, 우리나라 산림환경에 적합한 참조 표본점의 개수를 선정하기 위해서는 기준을 마련하기 위해 kNN 기법을 적용하기 전에 사전연구가 요구된다. 선정된 최소 수평참조범위 반경과 참조표본점의 개수를 적용하여 평창군의 산림주제도를 작성하고, 산림통계량을 산출하였다. 하지만, 본 연구에서 수집된 야외 표본점의 자료가 충분하지 않기 때문에, 추정된 산림통계량의 정도(精度)는 높지 않을 것으로 판단된다. 또한 야외 조사 시점에서의 산림과 비산림의 주제도 대신 제3차 산림자원조사에서 생산된 임상도를 이용하여 산림지역을 추출하였기 때문에 현실의 산림 면적과는 차이가 있으며, 이에 따라 작성된 재적 및 흉고단면적의 분포와는 차이가 있을 것으로 판단된다.

하지만 제5차 전국산림자원조사에서는 집락으로 표본점이 배치되고, 연구 대상지 주변 지역에 대한 표본 자료들이 수집되면 참조표본점의 개수는 증가하게 되어, 결과적으로 추정치의 오차를 감소시킬 수 있을 것으로 기대된다. 앞으로 전국 산림자원조사에서 수집된 자료와 연계하여 kNN 기법을 활용하면 전국단위의 산림통계뿐만 아니라 기본계획구별 그리고 시군구 단위별 산림통계 및 주제도의 작성이 가능할 것이다. 산림주제도는 문서로 작성되는 임업통계요람과는 다르게 GIS와 연계하여 시각적으로 산림 현황 파악이 쉽게 이루어질 수 있다. 또한, 전국 산림자원조사에서 배치된 야외 표본점의 수가 적어 산림통계를 산출하는데 어려움이 있을 것으로 예상되는 일부 시군구의 경우에는 kNN 기법을 통해 통계연보 작성에 필요한 산림통계를 산출할 수 있을 것이다.

kNN 기법을 통하여 야외 조사에서 산출된 정보를 이용하여, 조사가 이루어지지 않은 지역을 포함한 연구 대상지 전체 산림을 대상으로 산림주제도를 작성함으로써 산림정보의 공간적 분포를 확인할 수 있었다. 하지만 야외

표본점의 위치 정보의 정확성, 한정된 야외 자료에 의한 편향된 추정치 그리고 이웃하는 표본점들에서 밴드 값 간의 자기 상관관계 등의 문제점이 지적되고 있다. 따라서 보다 정확한 추정치를 얻기 위해서는 영상과 산림 정보들과의 상관관계에 의한 밴드별 가중치의 적용과 관심있는 변수와 상관관계가 높은 부가정보의 활용이 요구된다. 이 기법을 우리나라에서 적용하기 위해서는 야외 조사 표본점의 정확한 위치정보의 수집과 지형 및 임분 특성 등의 다양한 조건에 따른 충분한 자료의 수집이 요구된다.

## 감사의 글

이 논문은 한국과학재단의 지원으로 수행된 국제공동연구 “우리나라의 지속가능한 산림경영에 필요한 효율적인 표본조사 방법의 개발”(과제번호 : F01-2004-000-10105-0)에서 얻어진 연구결과의 일부입니다. 연구비 지원에 감사드립니다. 또한 논문의 질을 높일 수 있도록 좋은 의견을 주신 두 분의 심사위원께도 감사를 드립니다.

## 인용문헌

1. 산림청. 2004a. 임업통계요람. 434p.
2. 산림청. 2004b. 국내외 여건변화에 따른 산림자원조사체계 개편연구(III). 284p.
3. 산림청. 2005. 국내외 여건변화에 따른 산림자원조사체계 개편연구(IV). 290p.
4. 이승호, 김철민, 원현규, 김경민, 조현국. 2004. Landsat TM 위성영상을 위한 산림자원량 산정. 한국임학회 정기학술발표 논문집, pp 250-252.
5. 임업연구원. 1996. 전국산림자원조사요령. 49p.
6. Altman, N.S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46(3): 175-184.
7. Cho, H.K. 2002. Untersuchungen ber die Erfassung von Waldflächen und deren Vernderungen mit Hilfe der Satellitenfernerkundung und segmentbasierter Klassifikation. Dissertation zur Georg-August-Universitt Gttingen. 120p.
8. Franco-Lopez, H., A.R. Ek, and M.E. Bauer. 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.
9. Halme, M. and E. Tomppo. 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.
10. Holmgren, P. and T. Thuresson. 1998. Satellite remote sensing for forestry planning a review. *Scandinavian Journal of Forest Research* 13:90-110.
11. Katila, M. and E. Tomppo. 2001. Selecting estimation parameters for the Finnish multisource national forest

- inventory. *Remote Sensing of Environment* 76: 16-32.
12. Katila, M. and E. Tomppo. 2002. Stratification by ancillary data in multisource forest inventories employing  $k$ -nearest-neighbour estimation. *Canadian Journal of Forest Research* 32(9): 1548-1561.
  13. Kleinn, C. 2002. New technologies and methodologies for national forest inventories. *Unasylva* 210(53): 10-18.
  14. Lee, K.S. and J.S. Yoon. 1997. Radiometric correction of terrain effects for SPOT and Landsat Thematic mapper imagery in mountainous forest area. *Journal of the Korean society of Remote sensing* 13(3): 277-292.
  15. McRoberts, R.E., M.D. Nelson, and D.G. Wndet. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the  $k$ -nearest neighbors technique. *Remote sensing of Environment* 82: 457-468.
  16. Reese, H., M. Nilsson, P. Sandstrm, and H. Olsson. 2002. Applications using estimates of forest parameters derived from satellite and forest inventory data. *Computers and Electronics in Agriculture* 37(1): 37-55.
  17. Reese, H., T. Granqvist-Pahlen, M. Egberth, M. Nilsson, and H. Olsson, 2005. Automated estimation of forest parameters for Sweden using Landsat data and  $k$ NN algorithm. The 31<sup>st</sup> International symposium on Remote sensing Environment June 20-24 2005, Saint Petersburg, Russia.
  18. Tokola, T., J. Pitknen, S. Partinen, and E. Muinonen. 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.
  19. Tokola, T. and J. Heikkilä. 1997. Improving satellite image based forest inventory by using a priori site quality information. *Silva Fennica* 31(1): 67-78.
  20. 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.
  21. Tomppo, E. 1990. Satellite image-based national forest inventory of Finland. *International Archives of Photogrammetry and Remote Sensing* 28: 419-424. Proceedings of the Symposium on Global and Environmental Monitoring, Techniques and Impacts, 17-21 Sept. 1990. Victoria, British Columbia, Canada.
  22. Tomppo, E., and M. Halme. 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.
- 
- (2007년 2월 6일 접수; 2007년 4월 19일 채택)