Estimation of Forestry-Biomass using k-Nearest Neighbor(k-NN) method

Lee, Jung-Soo
Department of Forest management, College of Forest and Environmental Sciences, Kangwon National University

Yoshida, Shigejiro
Department of Forest and Forest Products Sciences, Faculty of Agriculture, Kyushu University

https://doi.org/10.5109/27365

出版情報：九州大学大学院農学研究院紀要. 58 (2), pp.339-349, 2013-09. Faculty of Agriculture, Kyushu University

J. Fac. Agr., Kyushu Univ., **58** (2), 339–349 (2013)

Estimation of Forestry–Biomass using *k–Nearest Neighbor(k–NN)* method

Jung–soo LEE[1] and Shigejiro YOSHIDA[2]

Laboratory of Forest management, Division of Forest sciences, Department of Agro–environmental Sciences,
Faculty of Agriculture, Kyushu University, Fukuoka 812–8581, Japan
(*Received April 26, 2013 and accepted May 9, 2013*)

The purpose of this study was to estimate of forestry–biomass using by *k–Nearest Neighbor (k–NN)* algorithm with the Landsat TM and field survey data in research forest of Kangwon national university. Optimum reference plots (k) were selected to estimate the forest biomass based on the minimum horizontal reference area (HRA) and spatial filtering using DN (Digital number), NDVI (Normalized difference vegetation index) and TC (Tasseled cap). The accuracy of RMSE was better in the order: DN, NDVI, and TC, respectively. In the DN value application, the RMSE of coniferous and broadleaved trees had the minimum value when k=11 of HRA 4 km and k=6 of HRA 10 km with 7by7 filtering. The bias of each was overestimated by 1.0 t/ha and 1.2 t/ha respectively. On the other hand, the minimum RMSE of *Pinus koraiensiss* had at k=8 and HRA of 4 km without filtering and the bias were underestimated by 1.6 t/ha. As a result, the estimated total forestry biomass was 802,000 t and 252 t/ha for *k–NN* methods. The results were higher than the plot data estimation by 16 t/ha. In this study, it is able to precise forest biomass at regional forest.

Key words: Kyoto protocol, GIS, remote sensing, k–Nearest Neighbor, forestry–biomass

INTRODUCTION

Forests are a renewable energy source that can substitute for fossil fuels and reduce greenhouse gas (GHG) emissions. Accordingly, forests are gaining in significance as a valuable and environmentally friendly green resource. International policies such as the Montreal Process have increased the importance of sustainable forest management, which includes practices such as forest resource monitoring and carbon stock assessment. South Korea has no current obligation to reduce GHG emissions during the commitment period because it was not included in Annex 1 of the Kyoto Protocol; however, South Korea is likely to be incorporated into Annex 1 after 2012. Therefore, the compilation of GHG absorption and emission statistics is an urgent necessity (Kwon *et al.*, 2005). The Korean Forest Service is building a GHG statistics system to cope with negotiations related to climate change mitigation, but there are limitations to the development of GHG–related databases using only statistical information and field surveys (Son *et al.*, 2007; Korea Forest Service, 2010).

Field surveys designed to assess forest carbon stocks using the estimated volume method are the most accurate, but are also labor–intensive and time–consuming. Moreover, it is difficult to conduct such surveys in inaccessible terrain, and these methods are not suitable for calculating the spatial distribution of forest biomass over large areas. South Korea currently estimates the national scale of its forest carbon stocks based on national forest resource survey data. According to Tier 3 of the Intergovernmental Panel on Climate Change Good Practice Guidance (IPCC GPG), it is necessary to clearly state the spatial distribution of forest biomass by utilizing remote sensing and Geographic Information Systems (GIS) in order to reliably estimate spatial GHG emission and absorption. Attaining statistics about GHGs became possible after the 16th Conference of the Parties, and it is claimed that the monitoring, reporting, and verification (MRV) system can be used to validate reductions. Therefore, in the development of statistically representative datasets, the utilization of GIS and remote sensing technology is of utmost importance.

Previous forest biomass estimations based on remote sensing estimated the forest stock, leaf area index, and age of a stand using regression analysis (Lee *et al.*, 2004; Scott *et al.*, 2010). However, more recent studies have utilized field surveys and remote sensing data to estimate the forest biomass in unsurveyed areas or to increase the accuracy of forest statistics (Tokola *et al.*, 1996; Tokola, 2000; Lee *et al.*, 2004; Holmgren *et al.*, 2000). In particular, studies estimating forest biomass using national forest inventory (NFI) data and forest type mapping, and studies estimating forest area biomass and developing thematic maps using NFI data and the *k*–Nearest Neighbor (*k–NN*) algorithm, are ongoing(Katila and Tomppo, 2001; 2002, Makela and Pekkarinen, 2004; Yim *et al.*, 2007; Jeong *et al.*, 2010). Fuchs *et al.* (2009) used high– and mid–resolution satellite images to create carbon maps based on the *k–NN* algorithm, and then employed linear regression analysis to compare and analyze the resulting maps. Studies have been conducted that employ the *k–NN* method in conjunction with both the basic band value of satellite images and with vegetation–sensitive ratio images in order to compare the carbon stock and storage of forest stands(Franco Lopez *et al.*, 2001; Yoo *et al.*, 2011). The *k–NN* technique has been applied not only to the Nordic region, which is a simple terrain area, but also to regions of complex terrain such as Switzerland and Italy. Most previous studies have

[1] Department of Forest management, College of Forest and Environmental Sciences, Kangwon National University, Korea
[2] Department of Forest and Forest Products Sciences, Faculty of Agriculture, Kyushu University, Fukuoka 812–8581, Japan
* Corresponding author(E–mail: jslee72@kangwon.ac.kr)

been conducted on a large scale (e.g., national and county levels). Therefore, research across smaller areas and at a finer scale is required in order to represent local (e.g., basin–level) variation and complexity in forests.

Consequently, this study utilized both the *k–NN* algorithm and remote sensing technology to estimate a regional forest biomass and to compare and analyze the accuracy of these estimates.

MATERIALS AND METHODS

Study area

The research forest of Kangwon National University is 3,144 ha in size and spans Chuncheon–si and Hongcheon–si in Gangwon–do, South Korea (Figure 1). The forest is composed of 27% coniferous trees, of which 89% are *Pinus koraiensis* and *Larix* spp., and 73% broadleaved trees, of which 98% are *Quercus* spp. The total volume of the forest is 755,700 m^3, 228,600 m^3 of which is made up of coniferous forest and 527,200 m^3 is broadleaved. The average volume of growing stock per ha is approximately 241 m^3; this is approximately 2.2 times higher than the national average. Distribution by age class was largely skewed to age class Ⅵ, which accounted for approximately 43% of the total area. In general, age classes V and higher accounted for 87% of the total area. Notably, 337 ha, or just over 10% of the study area, was classified as age class Ⅷ and was composed of *Larix* spp. and *Pinus* spp. (e.g., *P. koraiensis*) (Research Forest of Kangwon national university, 1999).

Conventional data

The GIS data used for this study included a forest type map, a forest compartment and sub–compartment map, a road network map for the research forest, and satellite images from Landsat 5 TM (row 115/path 34) obtained on 25 May 2009. Landsat 5 TM is a mid–resolution satellite that was launched in March of 1984. The satellite has a detection area of 185 km and an image collection cycle of 16 days. It consists of 7 bands, but this study used digital number (DN) values of only 6 bands and did not use the sixth band, which has a 120–m scale spatial resolution (Table 1). Sample plot survey field notes from Gangwon National University's research forest were used for the field survey data. The survey plot was designed based on the species and ratios of each age class. There were a total of 189 sample plots, and the survey was conducted between October and November of 2009. The species of trees, diameter at breast height, tree height, and number of stems >6 cm in size were measured in 20–m×20–m quadrangle sample plots. The data collected also included positional information (longitude and latitude) for the center of the sample plots measured with global positioning system devices (Table 2). ArcView 10.0, ENVI, and R–project were used to estimate the forest biomass.

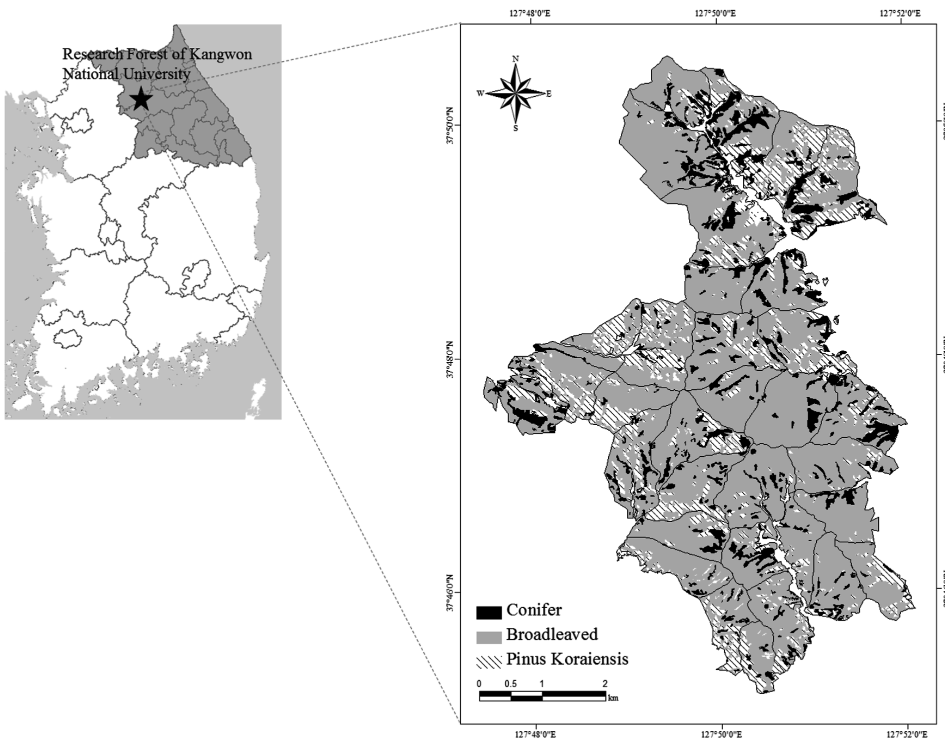


Fig. 1. The research forest of Kangwon national university in Korea.

[1] Stem volume (m^3/ha): forest volume per ha (m^3) (Korea Forest Service, 2009).
[2] Basic wood density (WD) (t/ m^3): dry weight to volume ratio
[3] Biomass expansion factor (BEF): ratio of the total aboveground biomass to the biomass of merchantable timber
[4] Root ratio: ratio of the total belowground biomass to the biomass of merchantable timber

Table 1. Characteristics of used data

Classification	Data	Organization
GIS data	Research Forest of Forest type map (2010.02)	Research Forest of Kangwon National University
	Plot location map (189point)	
Field data	Field survey (2009.10~11)	
Remote sensing data	Landsat TM–5(115/34)	Date: 2009/05/25

Table 2. The distribution of field plot by Forest type

Age class	Conifer Forest		Broadleaved Forest		*Pinus koraiensis*	
	n	Volume (m^3/ha)	n	Volume (m^3/ha)	n	Volume (m^3/ha)
Ⅰ	–	–	–	–	–	–
Ⅱ	–	–	1	104.8	–	–
Ⅲ	1	150.1	–	–	2	153.1
Ⅳ	40	198.8	–	–	47	209.2
Ⅴ	3	244.4	10	208.9	18	256.0
Ⅵ	2	288.1	19	235.7	2	295.8
Ⅶ	1	330.6	4	260.3	1	330.5
Ⅷ	24	376.9	–	–	14	361.4
Total	71	–	34	–	84	–

Processing

1. Preparation of the GIS Data and Sample Plots

GIS data related to the sample plot locations and forest tree surveys were established using sample plot forest tree survey field notes. These data were used to calculate the forest biomass based on the stem volume per ha, basic wood density, biomass expansion factor, and root ratio (equation 1). The stem volume per ha was calculated for each tree species by applying an alignment chart and the stand yield table of the Korean Forest Service. Furthermore, the basic wood density, biomass expansion, and root ratio were based on Korean forest carbon accounting standards (Korea Forest Service, 2010) (Table 3).

Forest Biomass (t)

$$=\text{stem volume per ha}^1 \times \text{basic wood density } (t/m^2)^2 \times \text{biomass expansion factor}^3 \times (1+\text{root ratio}) \quad (1)$$

2. Preprocessing of the Satellite Images

Because satellite images undergo spatial distortions for various reasons, including the relative movement of the Earth and the characteristics of the satellite or sensor, geometric correction is necessary(Lillesand *et al.*, 2008). Hence, for the Landsat 5 TM images, a geometric correction was applied to achieve a root mean square error (RMSE) of 25 m using a digital map, a forest compartment and sub–compartment map, and a forest road network map.

3. Forest Biomass Estimation using the *k–NN* Algorithm

The *k–NN* algorithm is a well–known non–parametric estimation method for classifying data from target sample plots (unobserved areas) based on the most analogical data value from researched reference plots by utilizing additional information (e.g., satellite images) (Tomppo, 1990). One advantage of applying the *k–NN* method to forestry is that practical estimations may be used instead of regression model estimations because *k–NN* refers to sample plot field survey data when estimating target sample plots(Tomppo, 1990; Yim, 2007).

3.1. Forest Type Classification and Establishment of a Reference Sample Plot

Scope setting is required to estimate forest data for unobserved target sample plots. Forests are affected by various factors, including the climate and topographic or stand–specific factors. To eliminate such effects, studies of how to set the scope of a reference sample plot based on various factors are ongoing (Tokola and Heikkilä, 1997; Katila and Tomppo, 2001). According to Yim *et al.* (2007), setting the reference scope using a classification

Table 3. Forest biomass Factor

Forest type	Basic wood density (t/m^3)	Biomass expansion factor	Root Ratio
Conifer Forest	0.44	1.44	0.27
Broadleaved Forest	0.61	1.36	0.35
Pinus koraiensis	0.41	1.85	0.32

method is effective in terms of minimizing the RMSE. Therefore, this study used a forest type map of the research forest to establish reference sample plots for mixed conifer (C), broadleaved (H), and dominantly *P. koraiensis* (PK) forest stands. Yim *et al*. (2009) stated that the establishment of reference sample plots should be based on the degree of similarity between the DN values for each satellite image band of a target sample plot (t) and a reference sample plot (r). Therefore, this study employed the Euclidian distance equation to determine the degree of similarity, where $d_{t,r}$ is the distance between the target sample plot (t) and reference sample plot (r); $x_{i,t}$ and $x_{i,r}$ are the DN values for each band of the target sample plot and reference sample plot at an image band i; and m is the number of satellite bands used (equation 2). This study used bands 1–5 and 7 from the Landsat 5 TM images.

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

Using the distance between the target and reference sample plots (i.e., the degree of similarity for the DN value in each band between two plots), higher weighted values were assigned (equation 3).

$$W_{t,r}=\frac{\frac{1}{d_{t,r}}}{\sum_{r=1}^{m}\frac{1}{d_{t,r}}} \quad (3)$$

As a result, an unobserved site estimation ($\hat{y}_t$) by the *k–NN* method is computed using the experimental value ($\hat{y}_r$) and weighted value for each reference sample plot ($w_{t,r}$) (equation 4).

$$\hat{y}_t=\sum_{i=1}^{k} w_{t,r}\times y_r \quad (4)$$

Furthermore, this study applied the vegetation index and DN values from Landsat 5 TM images to analyze the accuracy and to compare the values of the forest biomass estimates. Because satellite images render identical surfaces using different light intensities due to atmospheric or topographic influences, studies estimating a forest stock or biomass using image transformations (e.g., the vegetation index) in order to minimize these effects are ongoing(Franco–Lopez *et al*., 2001; Jensen, 2007; Yim *et al*., 2009; Yoo, 2011). This study used the normalized difference vegetation index (NDVI) and tassled cap (TC), which are the most popular tools for evaluating the vital degree of a forest for statistical verification of the biomass based on index alterations between images. Rouse *et al*. (1974) developed the NDVI, which is used to estimate the vital degree of vegetation or ground cover alteration (Lyon *et al*., 1998; Song *et al*., 2001). The TC transformation, which was developed by Crist and Cicone (1984), acquires the necessary data by amplifying Landsat TM images and transforming them into three categories: brightness, greenness, and wetness(Lee *et al*., 2008; Lee *et al*., 2004). This study used greenness because it greatly simplifies the identification of vegetated areas.

3.2. Horizontal Reference Area (HRA) Classification

In the forest resource survey, relationships between variables (forest biomass) and DN values of each image band diversified as the investigated area and area (number) of the referenced sample plots in the field surveys increased(Tokola and Heikkilä, 1997; Tokola, 2000; Yim *et al*., 2009; Jeong *et al*., 2010). This study used a range of HRAs (4, 7, and 10 km); of these, 10 km covered the entire target area.

3.3. Spatial Filtering

Image DN values are affected by the atmosphere, image sensor error, or noise generated in data transmission and reception because the research forest is located in rugged mountainous terrain(Tokola and Heikkila, 1997). Spatial filtering is required to reduce the influence of these factors. Spatial filtering refers to mathematically defined kernels of variation, representing rapid increases or decreases in spatially consecutive pixels. Normal kernels have odd numbers such as 3 by 3 or 5 by 5. The kernels move to the original images and they are assigned a central pixel value by computing a weighted value for each pixel. To remove Mixcell effects from the satellite images, we conducted statistical verification by varying the extent of filtering. With no filtering, the verification involved low–pass spatial filtering with three classifications: 3×3, 5×5, and 7×7. The central values were calculated as the average of neighboring values.

3.4 Statistical Verification

Cross–validation was employed to verify the estimation furnished by the *k–NN* method and set the optimum reference sample plot numbers (k) (Katila and Tomppo, 2001; Yim, 2009). The overall accuracy (OA) was computed with a fifth matrix(Franco–Lopez *et al*., 2001; Yim *et al*., 2007). Among cross–validations, the RMSE and a bias for estimation capacity evaluation were computed using equations 5 and 6.

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} \quad (5)$$

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

In these two equations, y_i is the forest biomass measurement, $\hat{y}_i$ is the forest biomass measurement using the *k–NN* method, and n is the number of reference sample plots. Meanwhile, measurements and estimations of the reference sample plots were separated into four classes and the overall accuracies were computed and compared with the biomass measurement furnished by the *k–NN* methods.

RESULT & DISCUSSION

1. Optimum HRA and Number Selection by Forest Type

1.1 RMSE Evaluation based on HRA Alteration (Pre–

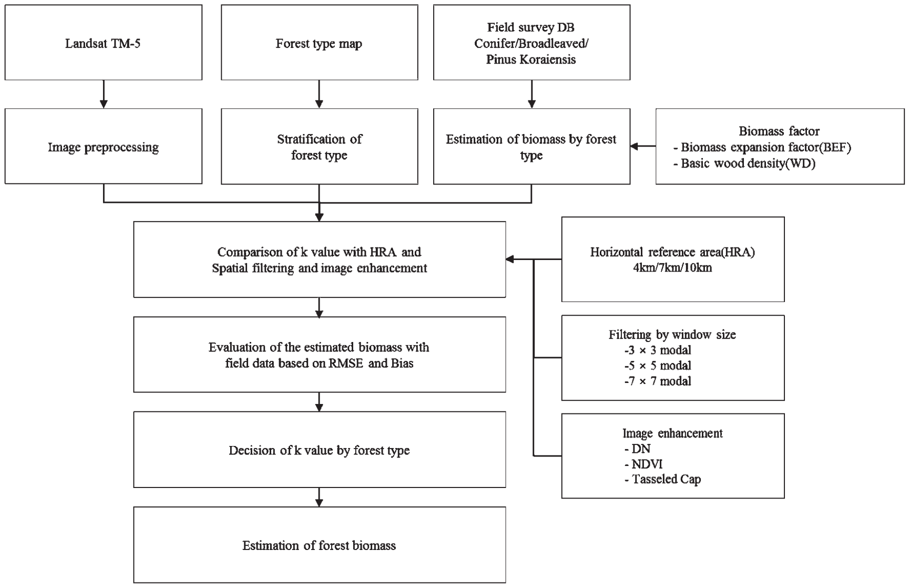
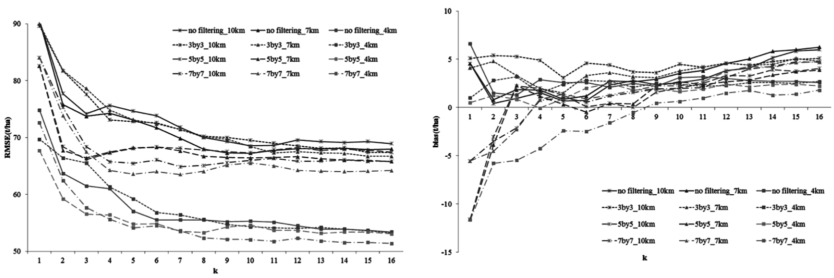
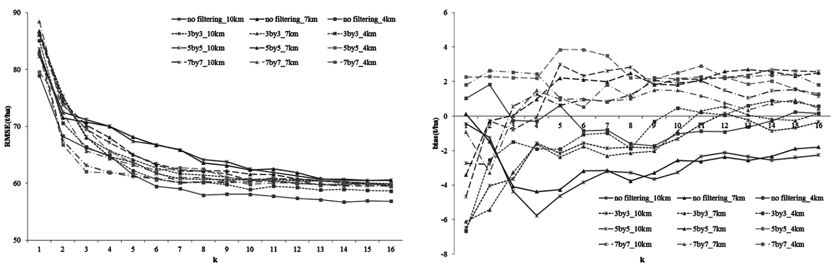


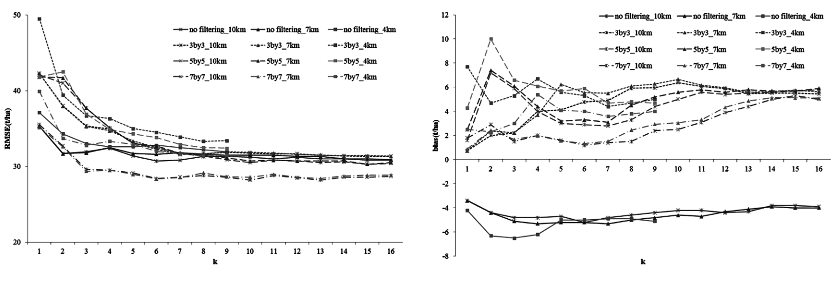
Fig. 2. Schematic methodology for estimating forest biomass using k-NN algorithmKorea.



a. Conifer



b. *Pinus koraiensis*



c. Broadleaved

Fig. 3. The RMSE and bias for different HRA and different number of neighbor plots by Original image (Left:RMSE, Right:Bias)

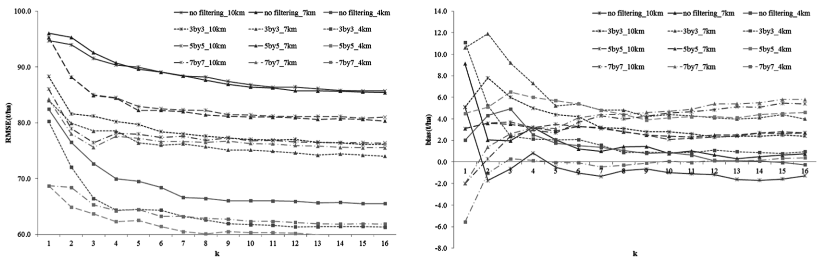
Filtering)

The RMSE of the forest biomass per ha using DN values were, from lowest to highest, 4 km>7 km>10 km for coniferous stands. For *P. koraiensis* stands, the values achieved were similar to those for mixed coniferous forests, and the RMSE was lowest for an HRA of 4 km; in addition, there was no significant difference between the RMSE at an HRA of 7 or 10 km. The average RMSE at an HRA of 4 km for mixed conifers and *P. koraiensis* was 57.0 and 60.0 t/ha, respectively (Figure 3a and b). Meanwhile, the order was 10 km>7 km>4 km for the broadleaved forest, and the average RMSE values were 31.3, 31.6, and 33.2 t/ha, respectively. Katila and Tomppo (2001) and Yim *et al*. (2007) stated that the RMSE tends to become smaller as the HRA increases (Figure 3c).

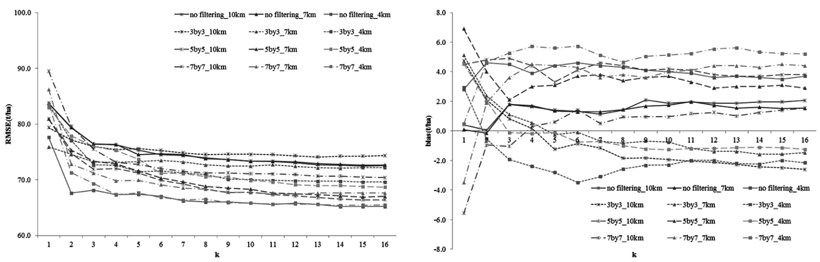
The RMSE for each HRA using the NDVI and TC were identical to the DN values. Conifers and *P. koraiensis* stands had the lowest RMSE at HRA=4 km, whereas broadleaved forest stands had the lowest RMSE at 10 km HRA. However, the RMSE for conifers and *P. koraiensis* increased more than the RMSE of the DN values regardless of the HRA. Conversely, the RMSE for broadleaved forests tended to decrease. Katila and Tomppo (2001) reported that the HRA of the minimum RMSE resulted from variations in forest structure, physiographic conditions, and sample plot plans. Previous studies chose regional levels (e.g., nation, city, or county) for their research, while this study was conducted at the basin level with a maximum length of 10 km to reduce the horizontal distance effect.

1.2 Error Evaluation based on HRA (Post–Filtering)

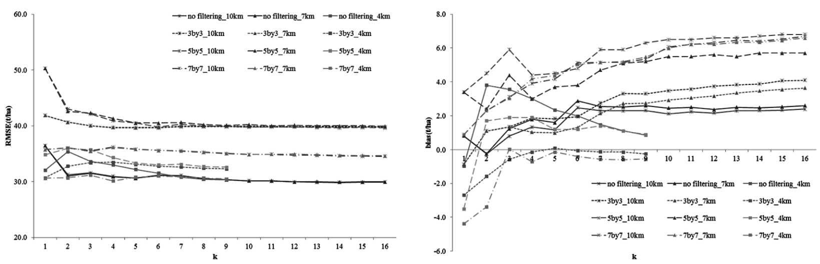
The forest biomass predicted using DN values had the lowest RMSE: an average of 54.5 t/ha at an HRA of 4 km with 7×7 filtering in the case of conifers. The RMSE was improved by 2.5 t/ha after filtering. Similar to coniferous forests, *P. koraiensis* had the lowest RMSE (61.7 t/ha at HRA=4 km). The RMSE was increased by 1.7 t/ha using 7×7 filtering. The RMSE of the biomass estimates for broadleaved forests tended to decrease as the window size increased; this was also true for coniferous forests. The RMSE of the biomass was the lowest at



a. Conifer



b. *Pinus koraiensis*



c. Broadleaved

Fig. 4. The RMSE and bias for different HRA and different number of neighbor plots by NDVI (Left:RMSE, Right:Bias)

HRA=10 km, and was improved by 2.0 t/ha with 7×7 filtering (Figure 3).

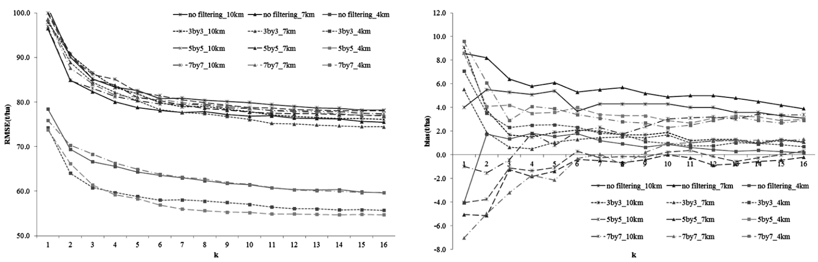
The RMSE of the forest biomass estimated using the NDVI was the lowest (61.3 t/ha for HRA=4 km) using 5×5 filtering for a coniferous forest. *Pinus koraiensis* had the lowest RMSE at HRA=4 km using 7×7 filtering, and broadleaved forests had the lowest RMSE at HRA=10 km using 7×7 filtering. However, none of these differences was statistically significant (Figure 4).

The RMSE of the forest biomass estimated using TC had the lowest RMSE at 57.6 and 60.7 t/ha for HRA=4 km using 3×3 filtering for coniferous stands and 7×7 filtering for *P. koraiensis* stands, respectively. Broadleaved forests had the lowest RMSE: 29.6 t/ha for HRA=10 km. Using 7×7 filtering improved the error by 1 t/ha.

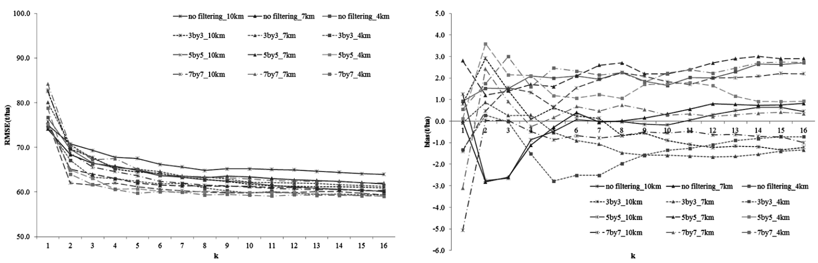
These results suggest that the optimum spatial filtering and HRA to use for RMSE minimization in the case of DNs were HRA=4 km with 7×7 filtering for coniferous forests, HRA=4 km without filtering for *P. koraiensis*, and HRA=10 km with 7×7 filtering for deciduous forest stands. In the case of NDVI, the optimal values were HRA=4 km with 5×5 filtering for conifers, HRA=4 km without filtering for *P. koraiensis*, and HRA=4 km with 7×7 filtering for deciduous forest stands. In the case of TC, the optimum values were HRA=4 km with 5×5 filtering for coniferous forests, HRA=4 km with 7×7 filtering for *P. koraiensis*, and HRA=10 km with 7×7 filtering for broadleaved forests (Figure 5).

1.3 Optimum Reference Sample Plot Number Selection and Error Evaluation

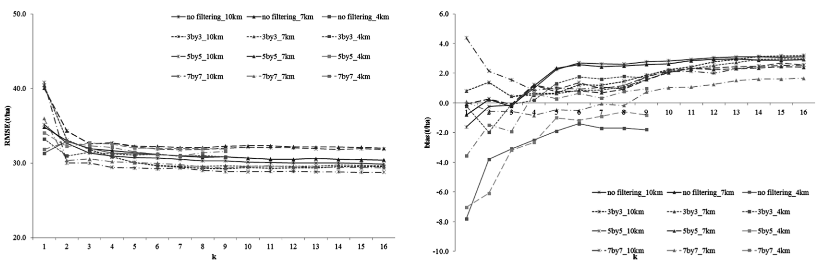
The minimum RMSE of the forest biomass based on reference plots (k) and DN values decreased rapidly when k was equal to 1–3 and HRA=4 km with 7×7 filtering for coniferous forests. Conversely, the RMSE increased when k was >11. Thus, k=11 was selected as the optimum reference plot size when RMSE was limited to 51.8 t/ha (Figure 3a). Furthermore, the trends for *P. koraiensis* (HRA=4 km with no filtering) were similar to those for conifers. The optimum reference plot size for conifers was k=8 when the RMSE was limited to 57.9 t/ha (Figure 3b). On the other hand, the RMSE of broadleaved stands (HRA=10 km with 7×7 filtering) decreased rapidly until k reached 1–2 and the RMSE gradually increased when



a. Conifer



b. *Pinus koraiensis*



c. Broadleaved

Fig. 5. The RMSE and bias for different HRA and different number of neighbor plots by Tasseled Cap (Left:RMSE, Right:Bias)

k was >6. Therefore, k=6 was selected as the optimum reference plot size when the RMSE was limited to 28.4 t/ha (Figure 3c).

In the case of NDVI estimates, the RMSE for coniferous forests (HRA=4 km and 5×5 filtering) declined sharply when k=1–2 and gradually decreased as k increased. The RMSE increased when k=8; thus, it was selected as the optimum reference plot when the RMSE was 60.1 t/ha (Figure 4a). Moreover, the RMSE of *P. koraiensis* (4 km without filtering) fluctuated when k was 1–4 and the error gradually decreased from k=5. The RMSE increased to 65.6 t/ha when k reached 11 (Figure 4b). In addition, the RMSE for broadleaved forests (HRA=4 km with 7×7 filtering) tended to increase slightly as k rose above 4. Therefore, k=4 was selected as the optimum reference plot size when RMSE was 30.1 t/ha (Figure 4c).

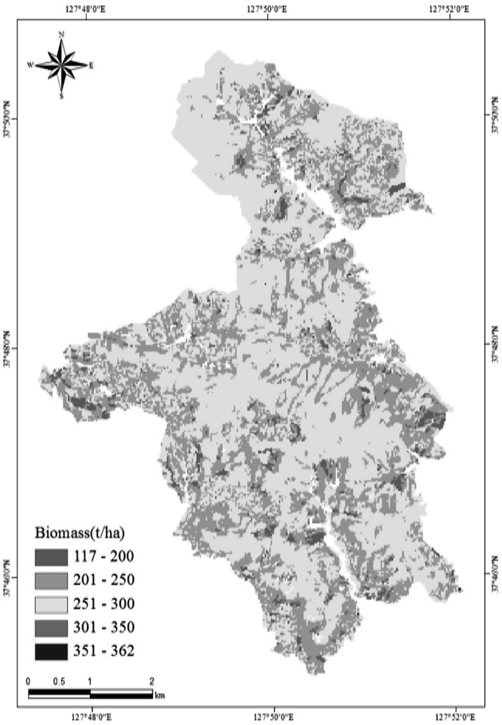
Lastly, in the case of TC, the RMSE of coniferous stand biomass estimates (HRA=4 km and 5×5 filtering) decreased rapidly when k was 1–3 and tended to decrease as k increased. The RMSE increased when k reached 10; hence, the optimum reference value was chosen as k=10 when the RMSE was 55.2 t/ha (Figure 5a). The RMSE of the *P. koraiensis* biomass estimate (HRA=4 km with 7×7 filtering) declined rapidly when k was 1–2 and gradually rose when k was >5. The optimum reference plot size was found to be k=5, at which the RMSE was 59.8 t/ha (Figure 5b). The RMSE of broadleaved stands (HRA= 10 km with 7×7 filtering) declined rapidly when k was 1–2 and increased as k exceeded 6. Therefore, the optimum reference plot was k=6 when the RMSE was 29.2 t/ha (Figure 5c). The reference plot k numbers in previous studies were distributed broadly based on the target area size, but mostly fell within the 5–10 range. Similarly, the results of this study were analogous with those of previous studies(Katila and Tomppo,2001; Franco–Lopez *et al*., 2001; Holmstrom, 2002; Jeong *et al*., 2010; Fuchs *et al*., 2009; Makela and Pekkarinen, 2004).

The variations in RMSE based on reference plot increases per forest type tended to decrease regardless of filtering. This finding is similar to those of Katila and Tomppo (2001) and Franco–Lopez *et al*. (2001). They demonstrated that RMSE values were inversely proportional to the number of reference plots. Yim *et al*. (2007) suggested an RMSE of 78.5–104.7 m^3/ha for tree stock volumes based on optimum reference plots; this corresponded to a forest biomass of 87.2–116 t/ha. This study achieved relatively low RMSE values (28.4–65.6 t/ha) because the number of reference plots was greater than for the NFI data. In other words, the accuracy was enhanced through a greater sample size.

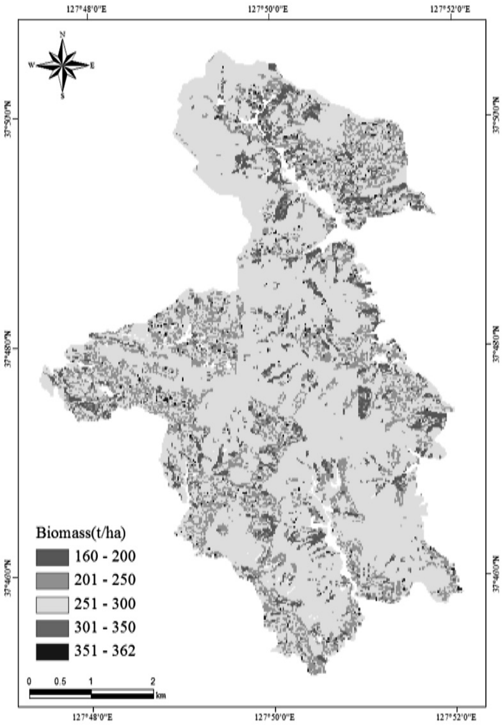
Based on the bias of coniferous forests using the optimum k, the estimations were too large regardless of the index transformation. The bias for DN and TC estimates tended to decrease as k increased, whereas the bias of the NDVI estimate tended to increase as k increased. The bias variation in NDVI had a small change of 3.9–6.5 t/ha. For broadleaved stands, the estimations of DN and TC were too large, with the exception of the NDVI. The bias for broadleaved stands was 2.9 times smaller than for conifers. Also, the bias decreased with increases in k in coniferous and deciduous stands when using the NDVI. The bias of DN and TC were increased when k was >6. Broadleaved stands had the smallest changes in bias with increases in k compared to mixed conifers and *P. koraiensis*. The biomass estimates for *P. koraiensis* stands were overestimated using the NDVI and TC, with

Table 4. Comparison of estimated forest biomass between field survey and *k–NN* method

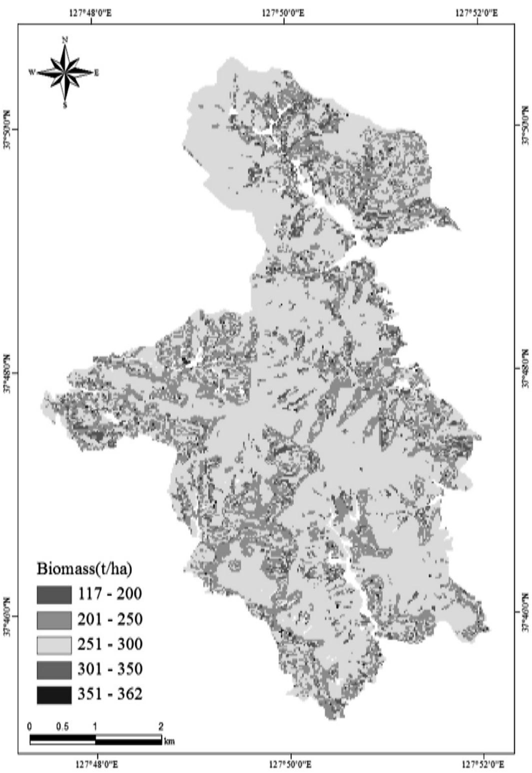
Forest type			Forest Biomass (t/ha)				Forest Biomass (t)
			Mean	Minimum	Maximum	Standard deviation	
Field survey (n=189)		Conifer	214	160	303	67	579,882
		Pinus koraiensis	248	209	362	57	126,728
		Broad–leaved	254	117	292	30	72,760
		Total	236	117	362	60	779,370
k–NN	DN	Conifer	211	160	303	23	76,331
		Pinus koraiensis	254	209	362	22	136,081
		Broadleaved	257	117	292	13	589,342
		Total	252	117	362	22	801,753
	NDVI	Conifer	210	160	303	50	75,852
		Pinus koraiensis	250	209	362	35	133,578
		Broad–leaved	262	195	292	13	601,309
		Total	254	160	362	30	810,740
	Tasseled Cap	Conifer	214	160	303	35	77,551
		Pinus koraiensis	248	209	362	35	132,448
		Broad–leaved	262	117	292	17	602,790
		Total	255	117	362	28	812,789



a. DN



b. NDVI



c. Tasseled Cap

Fig. 6. The distribution of forestry biomass by image enhancements.

a similar bias noted for conifers.

2. Forest Biomass Estimation based on Forest Type

The total study area biomass estimated by the *k–NN* method was 802,000 t for DN, 811,000 t for NDVI, and 813 t for TC; the average forest biomass per ha was 252, 254, and 255 t/ha, respectively. Compared to the estimations developed using reference plots, the average total biomass was overestimated by approximately 22–33 t and the average forest biomass per ha was overestimated by approximately 16–19 t/ha. This underestimation in conifer biomass is in accordance with that reported by Yim *et al*. (2007); however, broadleaved stands and *P. koraiensis* had conflicting results as they overestimated the results when compared with the reference plot estimates (Figure 6 and Table 4). The overall accuracy of the estimation based on *k–NN* was 0.21–0.26 t/ha, which is comparable to the value reported by Lim (2007) (0.23–0.28 t/ha). The accuracy was highest for NDVI and DN (0.38 t/ha at biomass estimates of 201–250 t/ha section). The TC method, however, had the highest accuracy (0.32 t/ha for biomass estimates of 251–300 t/ha). The accuracy of the biomass estimate for all index transformation images (301 t/ha) was 0.03, which is very low. This is comparable to the underestimation reported by Holmgren *et al*. (2000) (Table 5). According to the index of transformation comparison, in terms of OA the DN method using a basic band value yielded biomass estimates that were 3% lower than those produced using the NDVI; however, the RMSE and bias were higher by 5.9 and 1.74 t/ha, respectively, indicating that the index transformed images had no statistically significant impact. This result is similar to that of Yoo *et al*. (2011). Additionally, Franklin (1986) reported that there was no correlation between spectrum values and stands with closed crowns, which resulted in estimation errors.

CONCLUSION

This study established a forest biomass and distribution map for the research forest of Kangwon National University using the *k–NN* algorithm in combination with field survey reference plot data and band value data from Landsat 5 TM satellite images. The *k–NN* method determined the number of reference plots based on the RMSE and bias to select an optimum *k* value according to the HRA settings and filtering combination. Coniferous and *P. koraiensis* forests had smaller RMSEs as the HRA range decreased, whereas broadleaved forests had the opposite tendency. The RMSE tended to decrease as *k* increased, but the effect of filtering was insignificant.

Furthermore, according to the RMSE and bias analyses, the DN value application of the *k–NN* method was more effective than the ratio image value application with index transformation. Therefore, we present a methodology for estimating forest carbon absorption in a reliable spatial unit according to Tier 3 of the IPCC GPG.

In addition, the utilization of time series satellite image data can be useful in determining the spatial location and distribution of forest carbon stocks resulting from fragmentation at the regional or national level. Thus, we can use satellite imagery in an MRV system to measure the statistical relationships among GHGs, as well as

Table 5. Accuracy assessment of estimated biomss by *k–NN* method

Estimated data(t/ha) \ Reference data(t/ha)		<=200	201–250	251–300	301<=	total	Accuracy
DN	<=200	8	9	3	7	227	0.30
	201–250	18	23	27	19	87	0.26
	251–300	19	23	11	12	65	0.17
	301<=	–	6	3	1	10	0.10
	total	45	61	44	39	189	
	Accuracy	0.18	0.38	0.25	0.03		0.23
NDVI	<=200	12	11	8	11	42	0.29
	201–250	16	23	19	9	67	0.34
	251–300	17	21	13	18	69	0.19
	301<=	–	6	4	1	11	0.09
	total	45	61	44	39	189	
	Accuracy	0.27	0.38	0.30	0.03		0.26
Tasseled Cap	<=200	9	14	9	10	42	0.21
	201–250	18	15	16	16	65	0.23
	251–300	18	25	14	12	69	0.20
	301<=	–	7	5	1	13	0.08
	total	45	61	44	39	189	
	Accuracy	0.20	0.25	0.32	0.03		0.21

report and verify emission reductions in a variety of forest stand types.

ACKNOWLEDGEMENTS

This study was carried out with the support of 'Forest Science & Technology Projects (Project No. S211013L20130)' provided by Korea Forest.

REFERENCES

Crist, E. P. and Cicone, R. C. 1984 Comparisons of the dimensionality and features of simulated Landsat–4 MSS and TM data. *Remote Sensing of Environment*, **14**(1–3): 235–246

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**(3): 251–274

Franklin J. 1986 Thematic Mapper analysis of coniferous structure and composition. *Int. J. Remote Sensing*, **7**: 1287–1301

Fuchs, H., Magdon, P., Kleinn, C. and Flessa, H. 2009 Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. *Remote Sensing of Environment*, **113**: 518–531

Hampus Holmström. 2002 Estimation of single–tree characteristics using the *k–NN* method and plotwise aerial photograph interpretations. *Forest ecology and management*, **167**: 303–314

Holmgren, J., S. Joyce., M. Nilsson. and H. Olsson 2000 Estimating stem volume and basal area in forest compartments by combining satellite image data with field data. *Scandinavian Journal of Forest Research*, **15**: 103–111

Jensen, R. J. 2007 *Inroductory Digial Image Processing: A Remote Sensing Perspective. Sigmapress*, Vol3. pp. 338–352

Jo, Y. W., S. J. Kim and M. H. Jo 2009 An empirical study on discrimination of image algorithm for improving the accuracy of forest type classification. –Case of Gyeongju area using KOMPSAT–MSC image data– *Journal of the Korean society for geospatial information system*, **17**(2): 55–60

Jung, J. H., J. Heo., Y. S. Hong., K. M. Kim and J. B. Lee 2010 Estimation of Aboveground Biomass Carbon Stock in Danyang Area using *k–NN* Algorithm and Landsat TM Seasonal Satellite Images. *Journal of the Korean society for geospatial information system*, **18**(4): 119–129

Katila, M. and E. Tomppo 2001 Selecting estimation attributes for the Finnish multi source national forest inventory. *Remote Sensing of Environment*, **76**: 16–32

Katila, M. and E. Tomppo 2002 Stratification by ancillary data in multisource forest inventories employing *k–nearest neighbor* estimation. *Canadian Journal of forest research*, **32**(9): 1548–1561

Korea Forest Service 2009 Stem volume table and stand yield table. Korea Forest Service, Seoul (Korea) pp. 206–221

Korea Forest Service 2010 Study on the basis of forest carbon accounting in korea. Korea Forest Service, Seoul (Korea) pp. 109–120

Kwon, S. D., J. H. Seo., Y. M. Son and Y. K. Park 2005 Biomass carbon emissions according to conversion of forest land in korea Korea. *Forestry Energy Research Society*, **24**(2): 10–15

Lee, J. B., S.Jayakumar and J. Heo 2008 Basal area mapping using remote sensing and ecological data. *Korean Journal of Remote Sensing*, **24**(6): 621–629

Lee, J. Y., D. Y. Yang., J. Y. Kim and G. S. Chung 2004 Application of landsat ETM image indices to classify the wildfire area of Gangneung, Gangweon Province, Korea. *Journal of Korean Earth Science Society*, **25**(8): 754–763

Lee, S. H., C. M. Kim., H. K. Won., K. M. Kim and H. K. Cho 2004 Estimating the Growing Stock of Forest Resources Using Multi temporal Landsat TM. Proceeding of annual conference of Korea Forest Society, pp. 250–252

Lillesand, T. M., Kiefer R. W. and Jonathan W. Chipman 2008. *Remote Sensing and Image Interpretation*, pp. 486–490

Lyon, J. G., Yuan, D., Lunetta, R. S. and C. D. Elvidge 1998 A Change Detection experiment using vegetation indices. *Photogrammetric Engineering & Remote Sensing*, **64**(2): 143–150

Makela, H. and A. Pekkarinen 2004 Estimation of forest stand volumes by Landsat TM imagery and stand–level field–inventory data. *Forest Ecology and Management*, **196**: 245–255

Research forest of Kangwon national university 1999 The 6th forest management plan. Kangwon National University, Chuncheon. pp. 3–14

Rouse, j. w., Hass, R. H., Schell, J. A. and D. W. Deering 1974 Monitoring Vegetation System in the Great Plains with ERTS. Proceedings, 3rd Earth Resource Technology Satellite Symposium Vol 1. pp. 48–62

Scott, L. Powell, Warren B. Cohen, Sean, P. Healey, Robert E. Kennedy, Gretchen, G. Moisen, Kenneth, B. Pierce and J. L. Ohmann 2010 Qúantification of live above ground forest biomass dynamics with Landsat time–series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, **114**: 1053–1068

Son, Y. M., K. H. Lee and R. H. Kim 2007 Estimation of Forest Biomass in Korea. *Journal of Korean forest society*, **96**(4): 477–482

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P. and S. A. Macomber 2001 Classification and Change Detection Using Landsat TM Data : When and How to Correct Atmospheric Effects. *Remote Sensing of Environment*, **75**(2): 230–244

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

Tokola, T., Pitkanen, J., Partinen, S. 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**: 2333–2351

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

Tomppo, E. 1990 Satellite image–based national forest inventory of Finland. *International Archives of Photogrammetry and remote sensing*, **28**: 424

Yim, J. S., G. S. Kong., S. H. Kim and M. Y. Shin 2007 Forest thematic maps and forest statistics using the k–Nearest Neighbor for Pyeongchang–Gun, Gangwon–do. *Journal of Korean Forest Society*, **96**(3): 259–268

Yim, J. S., W. S. Han., J. H. Hwang., S. Y. Chung., H. K. Cho and M. Y. Shin 2009 Estimation of forest biomass based upon satellite data and national forest inventory data. *Korean Journal of Remote Sensing*, **25**(4): 311–320

Yoo, S. H., J. Heo., J. H. Jung., S. H. Han and K. M. Kim 2011 Estimation of Aboveground Biomass Carbon Stock Using Landsat TM and Ratio Images – *k–NN* algorithm and Regression Model Priority. *Journal of the Korean society for geospatial information system*, **19**(2): 39–48